



Semantics at the Crossroads

From Theoretical Explorations to Implementations

Edited by

Aikaterini-Lida Kalouli

Tracy Holloway King

Stephen Pulman

Annie Zaenen

Semantics at the Crossroads

From Theoretical Explorations to Implementations

Edited by

Aikaterini-Lida Kalouli

Tracy Holloway King

Stephen Pulman

Annie Zaenen

Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.). 2025. *Semantics at the Crossroads: From Theoretical Explorations to Implementations*. Konstanz: PubliKon.

Bibliographic information published by the Deutschen Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at <https://portal.dnb.de>.

Published by PubliKon,
Universität Konstanz, Kommunikations-, Informations-, Medienzentrum (KIM),
2025.
Universitätsstraße 10 78464 Konstanz Germany

The online version of this publication is permanently available for free under an open access license on PubliKon, the e-book platform of the KIM Konstanz, <https://www.kim.uni-konstanz.de/publikon/>

© 2025, the authors

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):
<http://creativecommons.org/licenses/by/4.0/>

This license applies only to the original material. All marked third party content (e.g., illustrations, photos, quotations, etc.) is excluded from the CC license, and it may be necessary to obtain further permission from the copyright holder for its reuse.

ISBN: 978-3-89318-101-8 (Digital)
978-3-89318-100-1 (Hardcover)

Cover and concept of design: Ulrike Harbort; Cover image created with Generative AI by Aikaterini-Lida Kalouli
Fonts: Libertinus, Arimo, DejaVu Sans Mono
Typesetting software: Xe₂La_TE_X

Contents

	Preface	iii
1	Introduction Annie Zaenen	1
I	Formal Computation	
2	Conservation as a Principle of Grammar Ronald M. Kaplan	9
3	Computational Challenges from Theoretical Semantic Modeling in LFG Mark-Matthias Zymła	25
4	Morphological Exponence and Morphosemantics Ash Asudeh	55
5	Glue Semantics for Hopf-Algebraic Minimalism Avery Andrews	79
II	Reasoning and Discourse	
6	The Linguistic Discourse Model at 40 Livia Polanyi & Martin H. van den Berg	111
7	From Network Theory to Linguistic Insights Alexandros Tantos & Kosmas Kosmidis	137
8	Modal and Higher Order Logical Reasoning with SUMO Adam Pease	165

Contents

9 Variation in Natural Reasoning	
Robin Cooper	187
III Lexicon	
10 <i>Out</i>-Applicatives are Concealed Comparatives	
Cleo Condoravdi & Paul Kiparsky	211
11 Revisiting the Light Verb Jungle: How Good Is GPT at Hacking Away?	
Annette Hautli-Janisz, Erisa Bytyqi, Pranshu Gupta & Miriam Butt	237
12 Poking at GPT’s NL Generalizations	
Annie Zaenen, Alexis Cooper, Felix Zheng, Susan W. Brown & Martha Palmer	265
IV Applications	
13 Semantic Representation in Biomedical Text Mining	
David Milward	301
14 An Information Theoretic Evaluation of AI Cooperation	
Marisa Ferrara Boston	323
15 Multi-modal Search: To Embed or not to Embed	
Tracy Holloway King	333
Index	355
Bibliography	377

Preface

We would like to thank our external reviewers for their insightful comments on the chapters: Tafseer Ahmed, David Ahn, Ash Asudeh, Emily Bender, Christoph Benzmüller, Stergios Chatzikyriakidis, Robin Cooper, Mary Dalrymple, Annette Hautli-Janisz, Ronald M. Kaplan, Beth Levin, Martha Palmer, Chirag Shah, Martin van den Berg, Leonie Weissweiler, Mark-Matthias Zymla,

We would like to thank Miriam Butt and the KIM at the Universität Konstanz for publication of this volume via PubliKon.

The book cover was created via generative AI (www.canva.com/ai-image-generator/) by Aikaterini-Lida Kalouli. The book \LaTeX is based on the templates at LangSci Press.

Aikaterini-Lida Kalouli (Bundesdruckerei-Gruppe)

Tracy Holloway King (Adobe Inc.)

Stephen Pulman (University of Oxford)

Annie Zaenen (Stanford University & Universität Konstanz)

Chapter 1

Introduction

Annie Zaenen^a

^aStanford University & Universität Konstanz

The articles to this volume reflect the variety of Dick Crouch’s interests and contributions to linguistics and NLP. Throughout his career he has had a constant interest in the interaction between semantics and reasoning, while preferring insights embodied in implementations. Dick’s PhD at Cambridge University was on Tense, Modals and Conditionals in English. Upon joining SRI Cambridge, he worked on natural language and specifically the interface with planning as well as general purpose language processing systems. His theoretical work focused on using the resource sensitivity of linear logic to model context update in the interpretation of utterances, and on constraints on logical derivations to model semantic underspecification. Moving to PARC he became more involved in implementation problems such as adding a semantic component for the Xerox Linguistic Environment (XLE), developing a broad coverage semantic lexicon, and implementing algorithms for efficiently finding all derivations (millions of derivations per second) and other issues connected with packing and ambiguity management. After leaving PARC, he went to Powerset and later Microsoft with PARC colleagues to integrate their semantic search technology into search engines. Later he worked on medical question answering, passage retrieval, textual inference, and virtual assistants for customer care at Nuance, product search at Amazon A9, and at Chegg to help students learn more in less time and at a lower cost. In 2022, he rejoined PARC and in 2024, SRI’s California lab.

The contributions in this volume, which comprises a Festschrift for Dick, reflect the range of his interests and the impact he has had on the field and the researchers he collaborated with.

Four papers discuss the formal computational angle. Ron Kaplan, *Conservation as a Principle of Grammar* revisits the principle of conservation that “pos-

tulates that the grammatical mapping between sentences and underlying representations of their meanings must not add or subtract unbounded amounts of information.” He proposes a formalization of this principle in LFG as the “proper anchoring condition”, elaborated in collaboration with the late Jürgen Wedekind.

Another paper that discusses the formal computational angle is that of Mark-Matthias Zymla, *Computational Challenges from Theoretical Semantic Modeling in LFG*. It discusses the tension between LFG as a theoretical framework and its implementations that should be faithful to the theory. While the mutually constraining correspondence between f- and c-structures reflects that idea, Zymla argues that the non-monotonicity of Glue semantics combined with commutativity leads to problems of unwanted ambiguity and that there is no mutual constraining. Hence the LFG correspondence architecture plays no role. This not only leads to parsing inefficiency but also diminishes the way the theory can provide linguistic insights.

Glue semantics plays a role in two other contributions. Ash Asudeh, *Morphological Exponence and Morphosemantics*, discusses a version of L_{RF} , a framework that combines traditional LFG with Distributed Morphology. The combination gives up the strong lexicalism of LFG: terminal nodes in the c-structure are labelled with a category and are paired with a f(unctional)-description but contain no information about form. Instead, this is captured in L_{RF} by exponence (in v-structure) and realization. A single morpheme can be the exponent of multiple nodes at the same time. The paper discusses how to capture morphosemantics in this framework, allowing a direct morphology-semantics interface so that mutual constraints are possible.

Avery Andrews, *Glue Semantics for Hopf-Algebraic Minimalism*, is an exploratory paper that combines propositional Glue-semantics with Minimalist syntax using Hopf algebras. A Hopf algebra of unordered trees provides the basic machinery for External and Internal Merge as proposed in Minimalism since 2013. It provides a concept of workspace as essentially a set or multiset of trees together with a coproduct that provides the materials for Internal Merge. An operation of disjoint union for workspaces and a B^+ operator handle syntactic Merge. However, these accounts do not provide an account of scope and various other semantic problems that glue semantics treats well. Hence this attempt to combine them.

Four papers relate to Dick’s interest and work on discourse and reasoning. Livia Polanyi and Martin van den Berg, *The Linguistic Discourse Model at 40*, present the Linguistic Discourse Model (LDM), a discourse theory that proposes a level of syntactic structure above the sentence, namely discourse parsing. LDM models how discourse units are related to each other in a tree structure through

subordination, coordination or an n-ary node. The paper illustrates the importance of the Right-Edge Constraint in discourse parsing, but also discusses how the framework handles changing contexts and quotations.

Alexandros Tantos and Kosmas Kosmidis, *From Network Theory to Linguistic Insights: Exploring Motifs and Antimotifs in the Discourse GraphBank Corpus*, use network analysis to study discourse properties. More specifically they argue that one can find motifs (sub-graphs that repeat themselves in a specific network or even among various networks) and anti-motifs in discourse structures that resemble mechanisms of biological signaling and that those structures give insights into discourse coherence and dynamics. They apply their methods to the GraphBank corpus.

Adam Pease, *Modal and Higher Order Logical Reasoning with SUMO*, discusses an implementation of a quantified multi-modal logic using an automated theorem prover. The work uses SUMO (Suggested Upper Merged Ontology) and shows how different modalities can coexist in the same framework, presenting epistemic, deontic and temporal operators supported by formulas that allow multi-modal logic to be tested in common sense reasoning. The paper describes how SUMO and its set of multi-modal axioms can be translated into classical higher order logic in the THF language and integrated with the Vampire theorem prover. The approach is illustrated by automated reasoning results on test cases.

Robin Cooper, *Variation in Natural Reasoning*, addresses the relation between an intuitive notion of natural reasoning and the kind of inference that has been discussed in computational linguistics as many of the inferences that people draw depend on “real world knowledge” in addition to what is present in the text. This means that different people with different amounts of knowledge and assumptions can draw different inferences from the same text. He tackles this problem by invoking the Aristotelian notion of topos and a notion of witness types. Topoi are recurrent plausible argumentative patterns. Part of the characterization of a persona is a distribution over topoi the agent is likely to use, the way you might expect a person to argue. Different kinds of people argue in different ways and will argue differently in different contexts depending in part on their assumptions about the people they are talking to. The paper illustrates this idea with how dogwhistles are interpreted.

Three papers focus on lexical semantic concerns in quite different ways. Cleo Condoravdi and Paul Kiparsky, *Out-Applicatives are Concealed Comparatives*, discuss various types of *out*-applicatives and argue that they all have a comparative meaning, derivable from a base with its complete argument structure and Theta-grid. This is contrasted with the analysis of event *out*-applicatives made in the

framework of Distributed Morphology by Ahn (2022) in which the meaning is claimed to come from *out* and not from the base. The paper discusses the historical development of the construction that seems to be unique to English.

Annette Hautli-Janisz, Erisa Bytyqi, Pranshu Gupta and Miriam Butt, *Revisiting the Light Verb Jungle: How Good Is GPT at Hacking Away?*, focus on V-V complex predicates in Urdu/Hindi and investigate how well GPT-3.5 and GPT-4 can identify and generate these predicates. In their investigation they vary prompts between English and Urdu or Hindi, and linguistic tasks like grammaticality judgments, part-of-speech recognition, and word order permutations, as well as lexical semantics aspects such as translations and minimal pair generation. They find that with zero-shot prompting and one-shot prompting, the performance of both GPT-3.5 and GPT-4 is very variable but that GPT-4 performs better than GPT-3.5. The results for syntactic aspects are variable depending on the settings. Translation gives better results for Hindi than for Urdu but again with much variation. For minimal pairs, the results for Hindi are quite good. Overall the results for lexical semantics properties were better than those for grammatical patterns.

Annie Zaenen, Alexis Cooper, Feliz Zheng, Susan Brown, and Martha Palmer, *Poking at GPT's NL Generalizations*, examine whether GPT-4 uses subcategorization information the same way humans would. They focus on inferences, specifically those triggered by the Caused Motion Construction (CMC) as a sophisticated case. They argue that while GPT-4 seems to have a good grasp of what are grammatical transitive and intransitive verbs, there is no evidence that it recognizes anything that corresponds to the CMC and that their experimental evidence suggests that GPT-4's linguistic integration is rather low level. They point out that there are no studies about how humans understand the CMC so that a more full-fledged comparison would require a study of human subjects too.

A final set of papers focuses on applications. David Milward, *Semantic Representation in Biomedical Text Mining*, looks at the mapping from natural language text to abstract semantic representations that can be used for inference in the field of biomedical text mining. Example inferences are represented in the style of the FraCaS test suite (a test suite that Dick was one of the authors of) and compared with FraCaS examples. The paper discusses the semantic representation features of several formats used to represent biological texts: tabular formats as used in spreadsheets; semantic triples allowing integration with open-source semantic web data, including ontological information; document-oriented NoSQL databases which allow for more natural semantic representations while providing the benefits of a database for access control and capabilities for data visualization and querying; the Biological Expression Language format which represents complex relationships of interest in systems biology in a more human readable

format which is useful for curation. The paper shows that as biomedical representations have a strong focus on the semantics of entities, relatively simple representations for relationships or properties still provide strong capabilities for sorting, filtering or joining which would not be available when using free text directly. Even though some representations do not provide a rich context, there are contextual filtering methods possible within text mining strategies, and ontology filtering can often be used as a substitute for ontology inference.

Marisa Boston, *An Information Theoretic Evaluation of AI Cooperation*, aims to advance the development of AI systems capable of effective human collaboration. She uses LLMs for a new way to evaluate cooperative behavior by calculating the unpredictability of unique words (entropy) in a corpus of messages and illustrates this through a study of the Enron corpus. This approach shows a sharp decline in entropy and of vocabulary size that correlate with the Enron crisis of 2001. The decline in entropy occurs at the moment of the crisis, the change in vocabulary a bit later. It is hypothesized that these changes correlate with changes in cooperative behavior.

Tracy Holloway King, *Multi-modal Search: To Embed or not to Embed*, discusses an issue that at first blush might have little to do with work in linguistics but that impacts all of us every day: multimodal search and more specifically search involving images and text. The paper gives an overview of the problems and approaches used at Adobe and concentrates on the advantages and disadvantages of using embedding (vector-based) techniques compared to classic (keyword) search, illustrating the broader use of techniques used in NLP and pointing out problems in search that cannot be solved (yet) with these techniques.

The collection gives an idea of the range of Dick's interests in and contributions to the field of language and inference. The enthusiasm of those who have contributed to this volume testifies to Dick's personal qualities of openness, generosity, supportive and cheerful collaboration, and general good companionship.

Part I

Formal Computation

Chapter 2

Conservation as a Principle of Grammar

Ronald M. Kaplan^a

^aIndependent Scholar

The basic architecture of Lexical Functional Grammar has remained relatively stable since its original formulation by Kaplan & Bresnan (1982). It was shown at the beginning that the recognition and parsing problems are undecidable for arbitrary LFG grammars, and a prohibition against nonbranching dominance (NBD) chains in constituent structure was proposed to suppress this undesirable property. It is now known that many other questions are undecidable even with the NBD restriction (Kaplan & Wedekind 2023) and also that the NBD restriction is incompatible with natural accounts of certain constructions. This paper discusses “proper anchoring”, an alternative and linguistically appealing way of constraining the excessive power of c-structure derivations. Proper anchoring conforms to the Principle of Conservation, the pretheoretic claim that the grammatical mapping between utterances and their meanings does not add or subtract arbitrary amounts of information.

1 Introduction

The basic architecture of Lexical Functional Grammar (LFG) defines a relation between the strings that make up the sentences of a language, a hierarchical constituent structure that represents the grouping of words into phrases, and a functional structure that characterizes the grammatical functions and features carried along by those words and phrases (Kaplan & Bresnan 1982). This relation was specified by setting up a straightforward correspondence between the structures of two mathematically simple frameworks, context-free grammars and the quantifier-free theory of equality. It was shown at the beginning that this simple arrangement is overly expressive in that the recognition problem (the question whether a given sequence of words belongs to the language of an arbitrary LFG

Ronald M. Kaplan. 2025. Conservation as a Principle of Grammar. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 9–24. Konstanz: PubliKon.

grammar) is undecidable (Kaplan & Bresnan 1982). It was later shown that the emptiness problem (whether there are any word-strings at all in the language of an arbitrary grammar) and the realization problem (whether there is at least one string that corresponds to a given f-structure) are also undecidable (Roach 1983, Nishino 1991, Wedekind 2014).

Kaplan & Bresnan (1982) introduced a particular restriction on the derivations of the context-free c-structure grammar to guarantee the decidability of the recognition and parsing problems. Since a specific segment of an input string carries a limited amount of linguistically significant information, they argued pretheoretically that c-structures beyond a certain size provide only redundant encodings of that information. Such large derivations therefore can be excluded from the candidate set of c-structures whose f-structure requirements must be evaluated. This intuition was formalized as the nonbranching dominance condition (NBD) that disallows c-structures containing two nodes that cover the same part of the input and are labeled with the same category. This restriction means that the linguistically relevant c-structures for a given string are finite in number and bounded by the length of the string.¹ Each of those can be examined to determine whether it is assigned a well-formed f-structure according to the functional annotations of its nodes.

It was observed subsequently that the NBD restriction is too strong. It disallows recursive nonbranching dominance chains in every situation, even when an errant subderivation is a component of a discontinuous constituent and its depth correlates intuitively with elements elsewhere in the string. Johnson (1986) first pointed out that NBD proscribes the straightforward analysis of the Dutch double infinitive construction as provided by the grammar of Bresnan et al. (1982). Haug (2017) used the example (1) to illustrate the same point with respect to Latin discontinuous constituents.

- (1) Maximilianus boni trusit Frederici filium.
Maximilian.NOM good.GEN pushed Frederick.GEN SON.ACC
'Maximilian pushed good Frederick's son.'

The adjective *boni* is understood as a modifier of *Frederici* but the two are separated by the intervening verb *trusit*. The natural relationship of *Frederici* to the object head *filium* follows from the assignments of the OBJ and POSS functions in the branching subtree sketched in Figure 1. The dashed recursive subtree on the

¹The NBD constraint in LFG is an early example of a family of what have become known as *Off-line Parsability* conditions. Related constraints have been proposed for other grammatical frameworks.

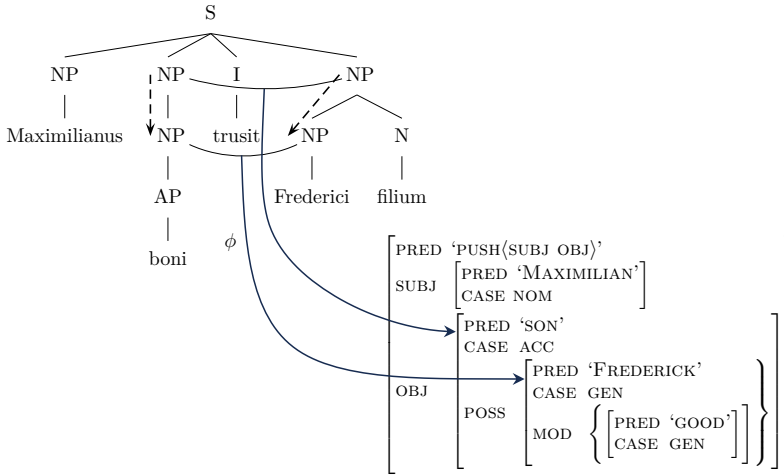


Figure 1: ϕ correspondences for the recursive subderivations (dashed) of sentence (1).

left naturally makes the parallel assignment for *boni*, but this recursive subtree does not branch and hence (1) would be excluded by the NBD restriction.

On the other hand, it has been noted that the NBD restriction is also too weak. Kaplan & Wedekind (2023) investigate in some detail the formal properties of LFG grammars. They identify the problems listed in Table 1 as undecidable for arbitrary LFG grammars. The NBD restriction reduces the complexity only of parsing/recognition and the other problems tagged with “NBD”. What is common to these problems is that they take a single string as input. The realization/generation problems are also of computational interest, but their input is a single f-structure, not a string of words or morphemes. They remain undecidable whether or not the NBD restriction is imposed on the c-structures of an arbitrary LFG grammar.

The exclusion of nonbranching dominance chains is thus not an appropriate restriction, linguistically or formally, on the generative power of LFG grammars. Kaplan & Wedekind (2023) propose as an alternative a condition that guarantees the decidability of problems with either string or f-structure inputs and also admits a broader range of linguistically desirable analyses. Their “proper anchoring” condition conforms to a general Principle of Conservation, the pretheoretic postulate that the grammatical mapping between sentences and underlying representations of their meanings must not add or subtract unbounded amounts of information. This paper summarizes the motivation for their proposal and some of its key features.

Table 1: Decision problems for arbitrary and NBD-restricted LFG's

NBD Recognition/parsing: Are there any f-structures for string s ? What are they?
NBD Satisfiability of constraining equalities and existentials = $c \neq (\dots) \neg(\dots)$
NBD Completeness/Coherence
NBD Infinite ambiguity: Is there any string with infinitely many f-structures?
Realization/generation: Are there any strings for f-structure f ? What are they?
Infinite expressiveness: Are any f-structures expressed by infinitely many strings?
Economy: Is it possible to find the smallest c-structure for f-structure f ?
OT-LFG ranking: Is it possible to find the derivation of f with fewest violations?
Emptiness: Is any string assigned an f-structure?
Ambiguity-preserving generation: Does any string express more than one f-structure?
Generation from underspecified f-structures: Is there a string for some extension of a given f ?
Cyclic f-structure: Is a cyclic f-structure assigned to any string?

2 Conservation of information

Kaplan & Wedekind (2023) attribute the complexity of parsing and generation for arbitrary LFG grammars to the fact that there is no finite number of annotated c-structures whose inspection is sufficient to determine whether there is a valid derivation for a given input string (in parsing) or f-structure (in generation). As a consequence, there is no systematic relationship between the length of a string and the size of a corresponding f-structure (the number of its attribute-value pairs) or the size of an f-structure and the length of any generated string. Moreover, a grammar can assign infinitely many f-structures to a single string and infinitely many strings to a single f-structure. Kaplan and Wedekind note that these properties seem implausible for language as a medium of communication. A related idea is expressed in Asudeh's (2012) Resource Sensitivity Hypothesis.²

²This intuition is also reminiscent of early technical proposals that were introduced to restrict the power of transformational grammars, such as Recoverability of Deletions (Peters & Ritchie 1973) and Subsistence (Wasow 1978).

This intuition can be formulated more explicitly in terms of the general “grammatical mapping problem”, the problem of characterizing in an explanatory and computable way the relation Γ between the utterances of a language and representations of their meanings (presumably logical formulas that can be interpreted in a non-linguistic universe). The pretheoretic claim is that Γ is an information-conserving relation in the following sense.

(2) a. *Conservation of information*

A relation R is *conservative* iff there are bounding functions \vec{b}_R and \tilde{b}_R such that $|y| \leq \vec{b}_R(|x|)$ and $|x| \leq \tilde{b}_R(|y|)$ for every $(x, y) \in R$.

b. *Conservation Principle*

The grammatical mapping Γ is conservative.

If u is an utterance of a language and m represents one of its meanings (that is, $(u, m) \in \Gamma$), then Γ is conservative if the sizes of u and m are mutually constrained. The size $|u|$ of an utterance is some measure of the amount of information it contains, often its length, and the size $|m|$ of a meaning representation can be defined in any reasonable way. The crucial property is that the derivational machinery that maps between u and m does not add or subtract, in either direction, arbitrary amounts of information.

Within the LFG Correspondence Architecture the syntactic component is one of a constellation of modules that together determine how utterances relate to their meanings (Kaplan 1987, 1989). Representations in the different modules are typically of formally different types, but they are tied together by piecewise correspondences that permit structures at one level to be described in terms of the properties of structures defined in other modules. Each module is equipped with its own type-specific derivational machinery to calculate the set of structures that satisfy those descriptions. No matter how the cross-module descriptions are specified (by co-description, by analysis, or by inverse correspondence, see for example Kaplan (1989)), the (u, m) mappings of Γ belong to the composition of the satisfying derivations of the individual modules. It follows from (3) that Γ is conservative if each of its components defines a conservative relation.

(3) The composition of conservative relations is conservative.

This is because the end-to-end relation is bounded by compositions of the per-component bounding functions (2a).³

There are still open questions about the inventory of modules and how they interact with each other and with syntax. For present purposes it suffices to take a high-level view and consider the modules (morphology, phonology, prosody...) that together mediate between utterances u and syntactic word-strings s as defining a monolithic “Pre-syntax” relation of (u, s) pairs. Similarly, a single “Post-syntax” relation can incorporate the modules (semantics, anaphora, discourse, ...) that together determine the relation between f-structures f and meanings m . The relation of syntactic derivations (s, f) for an LFG grammar G is itself the traditional composition of string-to-c-structure and c-structure-to-f-structure relations:

- (4) a. The set of syntactic derivations Δ_G for an LFG grammar G is the relation $C_G \circ F_G$, where
- b. $(s, c) \in C_G$ iff c is an annotated c-structure that satisfies the context-free rules of G , and s is its terminal string.
- c. $(c, f) \in F_G$ iff f is the minimal model for a satisfiable f-description instantiated from c .

At this high level of abstraction the grammatical mapping Γ is defined as in (5).

$$(5) \Gamma = \text{Pre-syntax} \circ \Delta_G \circ \text{Post-syntax} \\ = \text{Pre-syntax} \circ C_G \circ F_G \circ \text{Post-syntax}$$

Γ is conservative if there are bi-directional bounds for each of its component relations.

There have not been detailed technical analyses of all the theoretical variants that might make up the Pre- and Post-syntax relations, but it appears that most if not all past and current proposals are naturally conservative. With respect to the Pre-syntax modules, Kaplan & Kay (1994) showed that the input-output relations of certain phonological rule systems are regular and can be implemented

³The simple view for this discussion is that composition is based on explicit, shared interface elements (strings and f-structures). The Correspondence Architecture encourages a more sophisticated and perhaps theoretically more attractive approach wherein composition is interpreted as the conjunction of cross-module correspondence-based descriptions. The structures of each module would be formally distinct, but they would be conditioned by their connections to structures in other modules. Under that interpretation the conservation bounding-functions would constrain the size of the descriptions, not the size of the described structures.

by finite-state transducers. The computational approaches outlined by Koskeniemi (1983), Karttunen et al. (1992), and Beesley & Karttunen (2003) make use of regular relations for morphology; Asudeh & Siddiqi (2023) suggest that the current morphological theories with the most relevance to LFG are also regular in their expressive power. As noted by Kaplan (2005), information at the interface between morphology and LFG syntax is measured not in terms of the length of individual strings but in terms of the number of distinct suffix string-sets defined by such a regular mapping, and this convention has been incorporated into computational implementations such as XLE (Crouch 2006). Under this interpretation the composition with morphology is conservative even for regular relations that are not linear bounded.⁴

Bögel (2023) surveys various LFG-related proposals for integrating prosodic information into the grammatical mapping. Some of them are explicit in their use of finite-state technology to account for string-level discrepancies between prosody and syntax (e.g. Bögel et al. 2009, 2010). Others exploit the Correspondence Architecture in much deeper ways to formalize and clarify how prosody interacts with syntax and potentially with other modules. The proposals of Butt & King (1998) and Dalrymple & Mycock (2011) are bounded by the non-recursive Prosodic Hierarchy, while Bögel’s (2015) approach is described as carrying a fixed amount of information between modules.

The semantic component is the most developed of the Post-syntax modules. Dalrymple et al. (1993) first introduced the idea of controlling the mapping between *f*-structure and meaning by means of resource-sensitive deductions in linear logic, and Glue semantics is now the standard theoretical approach to semantic interpretation in LFG (Dalrymple 1999, Asudeh 2023). The number of Glue meaning constructors for a given *f*-structure *f* is proportional to its size, and the number of meaning constructors determines the length of the normalized linear logic proof that governs how the meaning representation is assembled. This establishes an upper bound on the size of the meaning representation, and a lower bound follows from the meaning terms that have been deployed in linguistically motivated constructors.⁵ For practical applications, semantic interpretation has typically been implemented by the rewriting rules of XLE’s transfer component

⁴Even reduplication, an apparent exception to the regularity claim, is regular if it is defined on a separately specified vocabulary of morphological segments. The constraints of other morphological processes may involve formal machinery that is somewhat stronger but still conservative: linear grammars may be more appropriate for derivational morphology while compounding may be modeled by a somewhat less restricted subclass of the context free grammars.

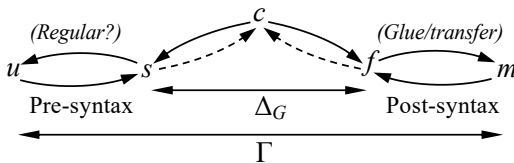
⁵Explicit restrictions on the form or interpretation of Glue premises may still be needed to control deductions that pathologically discard all linguistic content. For example, if meaning constructors such as $\lambda x.\lambda y.TRUE : (\uparrow OBJ)_\sigma \multimap (\uparrow SUBJ)_\sigma \multimap \uparrow_\sigma$ are allowed, the (f, m) relation

(Crouch 2005, Crouch & King 2006, Crouch et al. 2011). The number of transfer rule applications and their overall output are bounded by the fact that the rules apply in an ordered sequence without repetition and with limited recursion.

Returning to the syntactic components C_G and F_G , for an arbitrary LFG grammar the terminal string s of a c -structure c is obviously bounded by the number of nodes in c , but there is no limit on the size of c -structures over a given string. Likewise, the size of the functional description instantiated from a c -structure's node annotations depends on the number of nodes, and this limits the size of the f -description's minimal model. However the opposite does not hold: the description for a given f -structure can be matched to the annotations of arbitrarily many nodes.

The solid lines in (6) represent the natural bounding functions of the Pre- and Post-syntax modules and the syntactic bounding functions that require no further stipulation.

(6) Bounding functions



The dashed lines stand for the bounding functions that are not defined for arbitrary LFG grammars.

The nonbranching dominance restriction can be seen as a first attempt to augment this picture by supplying a bound for $|c|$ as a function of $|s|$. If nonbranching dominance chains are excluded, the maximum number of nodes of a c -structure over strings of a given length can be calculated from a grammar's rules and lexical entries. The decidability of the parsing problem follows from the simple (but clearly inefficient) strategy of enumerating all c -structures of size up to the string-dependent bound and evaluating the (bounded) f -descriptions for each of them. Δ_G is not conservative and generation remains undecidable because NBD does not bear on the dashed f to c dependency. The alternative proper anchoring condition proposed by Kaplan & Wedekind (2023) supports a linguistically more appropriate s -to- c bound and completes this picture by also providing a bound on $|c|$ as a function of $|f|$. It delineates a conservative subclass of syntactic derivations without proscribing the natural analysis of discontinuous constituents.

might not be conservative despite the resource sensitivity of all Glue derivations (Mary Dalrymple, p.c.).

3 Properly anchored syntactic derivations

The proper anchoring condition constrains the number and distribution of recursive subderivations in an admissible c-structure c . A recursive subderivation is a subtree of c headed by a node with category A that dominates another A node but does not include the daughters of that lower A node. Proper anchoring focuses on recursive subderivations because c-structures of unlimited size can be assigned to particular strings or f-structures only if recursive subderivations are allowed to stack arbitrarily, one on top of another.

Under proper anchoring, recursive subderivations are controlled by first requiring each of them to be anchored by particular words and f-structure units and then insisting that the anchors must be distinct for recursive subderivations in a stack. The anchors for recursive subderivations are specified in a linguistically sensible way, in terms of the traditional correspondence ϕ of c-structure nodes to f-structure units and its inverse ϕ^{-1} . The technical details of proper anchoring are laid out in the following definitions, from Kaplan & Wedekind (2023).

Let c be an annotated c-structure and let n and n' be two distinct nodes in c with n dominating n' . Then

- (7) a. The *subderivation from n to n'* , denoted by $c_{n'}^n$, is the derivation obtained from c by removing from the subderivation rooted by n the subtree under n' .
- b. A subderivation $c_{n'}^n$ is *recursive* if n and n' are both labeled with the same annotated category.
- c. Two subderivations $c_{n'}^n$ and $c_{n''}^{n''}$ are *stacked* if the bottom node of one dominates the top node of the other.

The admissibility of recursive subderivations is then defined in terms of f-structure and string anchors.

- (8) Let c be an annotated c-structure with terminal string s and f-structure f . A recursive subderivation $c_{n'}^n$ is
 - a. *f(unctionally)-anchored in f_k* iff there is a node \bar{n} of $c_{n'}^n$ such that $\phi(\bar{n}) = f_k$ and
 - b. *s(tring)-anchored in s_j* iff there is a node \bar{n} of $c_{n'}^n$ such that \bar{n} or another node in $\phi^{-1} \circ \phi(\bar{n})$ dominates s_j .

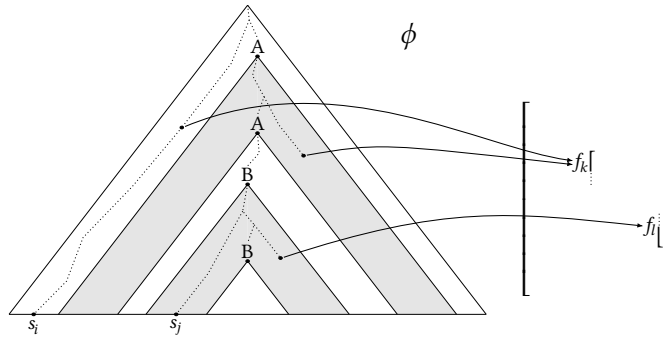


Figure 2: A c-structure with two stacked recursive subderivations, highlighted in gray. The subderivations are f- and s-anchored at f_k , s_i and f_l , s_j , respectively.

The functional anchor of a recursive derivation is defined directly in terms of the ϕ correspondence for at least one of its nodes. The string-anchor definition incorporates the extended-head composition $\phi^{-1} \circ \phi$ that is also used in formalizing the f-precedence and functional category (CAT) predicates.

The effects of these definitions are shown in Figure 2, again based on Kaplan & Wedekind (2023). C-structure dominance relations are represented by dotted lines, and solid arrows show the mapping of nodes to f-structure units according to the ϕ correspondence. The recursive subderivations highlighted in gray are in a stacked configuration because the lower A-labeled node dominates the upper B node. The lower subderivation includes all the nodes under its B-labeled root except for the nodes under the lower B node. This subderivation is f-anchored in f_l because one of its nodes maps to that f-structure unit (other f-anchors might arise from the ϕ mappings of other internal nodes). It also has a string anchor s_j which in this case happens to lie within the subderivation. The higher A-rooted recursive subderivation illustrates the more complicated case of a discontinuous string anchor. The string element s_j is not dominated by any node in the subderivation, but it can still serve as an s-anchor. That is because at least one of its ancestor stands as an extended head for one of the nodes internal to the subderivation—they both map to the f-structure unit f_k .

The subclass of properly anchored derivations is then defined as follows.

- (9) a. A syntactic derivation $(s, f) \in \Delta_G$ is *properly anchored* iff for every annotated c-structure c with terminal string s and f-structure f
 - b. every recursive subderivation c_n^l of c is f- and s-anchored and

- c. the *f*-anchors of any two recursive subderivations in a stack are distinct, and so are their *s*-anchors.

The qualifier “in a stack” in (9c) is crucial to this formulation. A *c*-structure is properly anchored even if the same words or *f*-structure units serve as anchors for recursive subderivations that are not vertically related. However the maximum depth of a *c*-structure and hence its size is determined by the length of the root-to-frontier path for the stack containing the most recursive subderivations, and that is what is limited by the requirement of distinctness.

Specifically, if N is the set of annotated nonterminal categories for a grammar G , any subderivation c_n^n with a path-length equal to or greater than $|N|$ must be recursive. Given distinctness, the annotated *c*-structures of properly anchored derivations are thus bounded in depth by the respective sizes of their corresponding strings and *f*-structures, as stated in the following lemma.

- (10) The depth of every *c*-structure c for all properly anchored derivations $(s, f) \in \Delta_G$ is bounded by $|N|(|s| + 1)$ and $|N|(|f| + 1)$, respectively, for a string of length $|s|$ and an *f*-structure of $|f|$ units.

Lemma (10) provides both of the bounding functions needed to ensure that the properly anchored derivations for an arbitrary LFG grammar are conservative. In accordance with the Conservation Principle (2b) only these derivations of the grammar are linguistically relevant. It follows for this subrelation of Δ_G that recognition/parsing and other problems with single-string inputs are decidable, as with the NBD restriction, but now generation/realization and problems with *f*-structure inputs are also decidable. This is because all of these problems can be solved by inspecting a finite number of annotated *c*-structures up to a depth determined by the sizes of the given *s* or *f* inputs. The problem list (1) is updated in Table 2 with “PA” now added to mark the generation-related problems that are decidable if only properly anchored derivations are considered.⁶

As noted above, the recursive subderivation in the analysis of Haug’s (2017) discontinuous Latin example (1) is nonbranching and thus runs afoul of the NBD condition. However the recursive *c*-structure needed to account for such a discontinuous constituent is not excluded by the proper anchoring condition. Proper anchoring respects the fact that, as in this case, functionally-related but separated subtrees are naturally correlated in depth.

⁶However, emptiness and other problems that require consideration of string or *f*-structure inputs of all possible sizes remain undecidable even for properly anchored derivations. Those problems are decidable under the restrictions that characterize the k -bounded subclass of LFG grammars (Kaplan & Wedekind 2023).

Table 2: Decision problems for properly anchored derivations

NBD **Recognition/parsing:** Are there any f-structures for string s ? What are they?

NBD **Satisfiability of constraining equalities and existentials** $=c \neq (\dots) \neg(\dots)$

NBD **Completeness/Coherence**

NBD **Infinite ambiguity:** Is there any string with infinitely many f-structures?

PA **Realization/generation:** Are there any strings for f-structure f ? What are they?

PA **Infinite expressiveness:** Are any f-structures expressed by infinitely many strings?

PA **Economy:** Is it possible to find the smallest c-structure for f-structure f ?

PA **OT-LFG ranking:** Is it possible to find the derivation of f with fewest violations?

Emptiness: Is any string assigned an f-structure?

Ambiguity-preserving generation: Does any string express more than one f-structure?

Generation from underspecified f-structures: Is there a string for some extension of a given f ?

Cyclic f-structure: Is a cyclic f-structure assigned to any string?

Thus, Figure 1 shows that ϕ maps the top NP nodes of both recursive subderivations to the same OBJ f-structure unit. That unit is therefore an f-anchor for both subderivations. Also, by virtue of that f-structure correspondence and the extended-head provision in definition (8b), either or both of the subderivations can be string-anchored by either of the words *boni* or *filium*. The rule that generates the nonbranching subderivation over *boni* can be reapplied to produce a next larger c-structure, but then *boni* would be its only possible string anchor. Given the distinctness condition (9c), *boni* would not be available as an anchor for any subderivations stacked after that, and for this sentence those would be inadmissible. More of those subderivations would be properly anchored, however, for sentences with more elaborate branching structures on the right (e.g. if sentence (1) is extended with additional possessive levels on the right).

4 Summary

The proper anchoring condition (9a) establishes a manageable relationship between strings and f-structures by controlling the size of permissible c-structures in linguistically relevant syntactic derivations. This relationship is information-conserving in the sense of (2a) and guarantees the decidability of both parsing and generation. It crucially depends on the ϕ correspondence and the linguistically motivated notion of extended heads to limit the depth of c-structure re-

cursion by the sizes of strings and f-structures, as indicated by Lemma (10). The configuration of modules in (6) is thus complete for a syntactic component that supplies only properly anchored derivations. With syntax restricted in this way, the overall grammatical mapping between utterances and meanings adheres to the pretheoretic Conservation Principle for communicative language.

Acknowledgments

Dick Crouch and I have worked together for almost 25 years, first at PARC, then Powerset, then Microsoft, then Nuance, and finally Amazon. I have gained so much from his deep technical insights and the clarity of his thinking as much as from his warmth and friendship. It is with great pleasure that I contribute this paper in his honor.

The ideas presented here have evolved over many years in a close and fruitful collaboration with Jürgen Wedekind. This paper would have been cleaner and clearer if he were still with us.

References

- Asudeh, Ash. 2012. *The logic of pronominal resumption* (Oxford Studies in Theoretical Linguistics). Oxford University Press. DOI: 10.1093/acprof:oso/9780199206421.001.0001. <https://doi.org/10.1093/acprof:oso/9780199206421.001.0001>.
- Asudeh, Ash. 2023. Glue semantics. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 651–697. Language Science Press. DOI: 10.5281/zenodo.10185964. <https://doi.org/10.5281/zenodo.10185964>.
- Asudeh, Ash & Daniel Siddiqi. 2023. Morphology in LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 855–901. Language Science Press. DOI: 10.5281/zenodo.10185976. <https://doi.org/10.5281/zenodo.10185976>.
- Beesley, Kenneth & Lauri Karttunen. 2003. *Finite-state morphology*. CSLI Publications.
- Bögel, Tina. 2015. *The syntax-prosody interface in Lexical Functional Grammar*. University of Konstanz. (Doctoral dissertation).
- Bögel, Tina. 2023. Prosody and its interfaces. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 779–821. Language Science Press. DOI: 10.5281/zenodo.10185970. <https://doi.org/10.5281/zenodo.10185970>.

- Bögel, Tina, Miriam Butt, Ronald M. Kaplan, Tracy Holloway King & John T. Maxwell III. 2009. Prosodic phonology in LFG: A new proposal. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG09 conference*, 146–166. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2009/lfg09.html>.
- Bögel, Tina, Miriam Butt, Ronald M. Kaplan, Tracy Holloway King & John T. Maxwell III. 2010. Second position and the prosody-syntax interface. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG10 conference*, 106–126. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2010/lfg10.html>.
- Bresnan, Joan, Ronald M. Kaplan, Stanley Peters & Annie Zaenen. 1982. Cross-serial dependencies in Dutch. *Linguistic Inquiry* 13(4). 613–635. <https://www.jstor.org/stable/4178298>. Reprinted in Savitch, Bach, Marsh & Safran-Naveh (1987: 286-319).
- Butt, Miriam & Tracy Holloway King. 1998. Interfacing phonology with LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG98 conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG1998/index.html>.
- Crouch, Dick, Mary Dalrymple, Ronald Kaplan, Tracy Holloway King, John T. Maxwell III & Paula Newman. 2011. *XLE Documentation*. Palo Alto Research Center (PARC). <https://ling.sprachwiss.uni-konstanz.de/pages/xle/index.html>.
- Crouch, Richard. 2005. Packed rewriting for mapping semantics to KR. In *Proceedings of the sixth international workshop on computational semantics (IWCS-6)*, 103–114.
- Crouch, Richard. 2006. Packed rewriting for mapping text to semantics and KR. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: variations on a theme by Ronald M. Kaplan*, 389–416.
- Crouch, Richard & Tracy Holloway King. 2006. Semantics via f-structure rewriting. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG06 conference*, 145–165. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2006/lfg06.html>.
- Dalrymple, Mary (ed.). 1999. *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*. The MIT Press.
- Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell III & Annie Zaenen (eds.). 1995. *Formal issues in Lexical-Functional Grammar*. CSLI Publications.
- Dalrymple, Mary, John Lamping & Vijay Saraswat. 1993. LFG semantics via constraints. In *Proceedings of the sixth conference of the European chapter of the Association for Computational Linguistics (EACL) (EACL '93)*, 97–105. Associa-

- tion for Computational Linguistics. DOI: 10.3115/976744.976757. <https://doi.org/10.3115/976744.976757>.
- Dalrymple, Mary & Louise Mycock. 2011. The prosody-semantics interface. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 conference*, 173–193. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2011/lfg11.html>.
- Haug, Dag Trygve Truslew. 2017. Syntactic discontinuities in Latin. In Victoria Rosén & Koenraad De Smedt (eds.), *The very model of a modern linguist: In honor of Helge Dyvik*, 75–96. Bergen Language & Linguistic Studies.
- Johnson, Mark. 1986. The LFG treatment of discontinuity and the double infinitive construction in Dutch. In *Proceedings of the fifth west coast conference on formal linguistics*, 102–118. CSLI Publications.
- Kaplan, Ronald M. 1987. Three seductions of computational psycholinguistics. In Peter Whitelock, Mary McGee Wood, Harold L. Somers, Rod Johnson & Paul Bennett (eds.), *Linguistic theory and computer applications*, 149–188. Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 337–367). Academic Press.
- Kaplan, Ronald M. 1989. The formal architecture of Lexical-Functional Grammar. *Journal of Information Science and Engineering* 5(4). Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 1–27), 305–322.
- Kaplan, Ronald M. 2005. A method for tokenizing text. In Antti Arppe, Lauri Carlson, Krister Lindén, Jussi Piitulainen, Mickael Suominen and Martti Vainio, Hanna Westerlund & Anssi Yli-Jyrä (eds.), *Inquiries into words, constraints and contexts: Festschrift for Kimmo Koskenniemi on his 60th birthday*. 57–66. CSLI Publications. <https://web.stanford.edu/group/cslipublications/cslipublications/koskenniemi-festschrift/>.
- Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 173–281. The MIT Press. Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 29–130).
- Kaplan, Ronald M. & Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics* 20. 331–478.
- Kaplan, Ronald M. & Jürgen Wedekind. 2023. Formal and computational properties of LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 1035–1082. Language Science Press. DOI: 10.5281/zenodo.10185982. <https://doi.org/10.5281/zenodo.10185982>.
- Karttunen, Lauri, Ronald M. Kaplan & Annie Zaenen. 1992. Two level morphology with composition. In *Proceedings of COLING 92*, vol. I, 141–148.

- Koskenniemi, Kimmo. 1983. *Two-level morphology: A general computational model for word-form recognition and production*. Tech. rep. 11. Department of General Linguistics, University of Helsinki.
- Nishino, Tetsuro. 1991. Mathematical analysis of Lexical-Functional Grammars—complexity, parsability, and learnability. *Language Research* 27(1). 867–915.
- Peters, Stanley & R. W. Ritchie. 1973. On the generative power of transformational grammars. *Information Sciences* 6. 49–83. DOI: 10.1016/0020-0255(73)90027-3. [https://doi.org/10.1016/0020-0255\(73\)90027-3](https://doi.org/10.1016/0020-0255(73)90027-3).
- Roach, Kelly. 1983. LFG languages over a one-letter alphabet. Unpublished manuscript, Xerox Palo Alto Research Center.
- Savitch, Walter J., Emmon W. Bach, William Marsh & Gila Safran-Naveh (eds.). 1987. *The formal complexity of natural language* (Studies in Linguistics and Philosophy). Springer. DOI: 10.1007/978-94-009-3401-6. <https://doi.org/10.1007/978-94-009-3401-6>.
- Wasow, Thomas. 1978. On constraining the class of transformational languages. *Synthese* 39(1). 81–104. <http://www.jstor.org/stable/20115307> (17 September, 2024).
- Wedekind, Jürgen. 2014. On the universal generation problem for unification grammars. *Computational Linguistics* 40(3). 533–538.

Chapter 3

Computational Challenges from Theoretical Semantic Modeling in LFG

Mark-Matthias Zymla

This paper reflects on one of the great strengths of Lexical Functional Grammar (LFG), namely, its tight relation between theoretical and computational modeling of the framework. Moving from syntactic to semantic considerations, the paper aims to highlight some inconsistencies in elements of the theory that have not been computationally tested or still pose open issues. Generally, the paper shows that there is a trend in which LFG analyses tend to overgenerate. In theoretical LFG, this overgeneration can be handled by accommodating certain presuppositions, but computational implementations do not have this possibility. Thus, new learners of computational LFG often hit similar pitfalls when trying to translate theoretical analyses into computational ones. However, the real caveat of this paper is a view on this topic from the point of the re-invigorated interest in computational Glue semantics. Concretely, I discuss commutativity as a problem and a feature of Glue semantics and its repercussions for the syntax-semantics interface, clarifying the price Glue semantics pays for its highly flexible combinatory system.

1 Introduction

The goal of this paper is to reminisce about the roots of Lexical Functional Grammar (LFG), in particular, the way in which it closely relates formally rigorous theoretical work with precise computational implementations. Lexical Functional Grammar (e.g. Kaplan & Bresnan 1982, or, more concisely, Börjars 2020) is a formalism that is well-known for its formal rigor which allowed it a fairly faithful computational implementation in the form of the Xerox Linguistics Environment (XLE; Crouch et al. 2011). In fact, a large chunk of the early LFG literature illustrates how close theoretical and computational efforts in the theory were in its

Mark-Matthias Zymla. 2025. Computational Challenges from Theoretical Semantic Modeling in LFG. in Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 25–54. Konstanz: PubliKon.

early years. This overlap was perpetuated by efforts like the large international ParGram project, which had the goal of developing cross-linguistically parallel computational LFG grammars (Butt et al. 2002). However, over time, computational interest has waned. Thus, theoretical LFG has started to develop at its own pace. Consequently, the two strains of LFG research, theoretical and computational, are in danger of becoming estranged. As interest in computational semantics based on LFG has been on the rise recently, this paper takes the chance to observe this issue from that perspective.

To explore this topic, I discuss challenges for new grammar writers hinting at different experiences with theoretical and computational LFG. One of the core topics of this paper is *optionality* which requires careful consideration from a computational perspective as it easily introduces unwanted ambiguities.

The discussion of LFG’s semantic framework of choice, Glue semantics (Dalrymple 1999, Asudeh 2022), tackles some more substantial challenges of the divide between theoretical and computational LFG, expanding on the issue of optionality, which is in danger of being used too lavishly in semantics. I argue that appropriate use of *optionality* is dependent on various considerations concerning the module (c-structure, f-structure, and semantics) it is used in.

Another issue of Glue semantics are spurious ambiguities due to commutativity. On the one hand, commutativity is a feature of Glue semantics required for the successful analyses of, e.g. scope interactions (Dalrymple et al. 1997); on the other hand, the non-monotonicity of Glue semantics combined with commutativity leads to severe issues from a computational perspective.

Exploring the different causes of ambiguity leads to some more general questions about the philosophy behind formal and computational linguistics, including the question of whether theoretical LFG runs the same risk as other linguistic formalisms of becoming a *meta language* rather than a *formal language*, thus diverging from its roots in the sense that formal rigor is reduced in favor of descriptively sufficient analyses that may or may not hold up when embedded in a fully formalized version of LFG – as is required for computational purposes. More generally, I want to point out that linguistic theory “suffers” from presuppositions, i.e. assumptions about how certain things work, without concretely testing them.¹

The next section introduces the core concepts of LFG, focusing on the syntax, semantics, and their interface. In section 3, I first discuss the distinction

¹Of course, usually the caveats are well understood but ignored or accommodated because it does not deter from the overall linguistic intuitions that researchers want to sell. However, this arguably diminishes the role of the formal system, questioning its use in the first place.

between defining and constraining equations to illustrate monotonic and non-monotonic aspects of LFG, then, optionality at the level of f-structure as a cause of unwanted ambiguities. Taking the previous discussion into consideration, section 4 discusses the commutative property of Glue semantics as its main cause of spurious ambiguities and its potential to cause bugs in semantic analyses. Section 5 presents some conceptual challenges pertaining to the syntax-semantics interface from a computational perspective, including the non-monotonicity of Glue semantics and the previously discussed topics of commutativity and optionality. Section 6 concludes.

2 An introduction to LFG

The topic of this paper requires a fairly thorough understanding of LFG's formalism. In this section, I outline the main properties of LFG, focusing on syntactic and semantic formalization. Conceptually, LFG has coined the term parallel projection architecture wherein different levels of linguistic analysis are performed by dedicated formal systems. This has been most prominently explored in LFG's c(onstituent)-structure and f(unctional)-structure, the syntactic module.

2.1 Syntax in LFG

The c-structure describes the surface structure of a linguistic expression, i.e. the linear order and constituency information. The f-structure describes syntactic properties that abstract away from surface order, e.g. distinguishing between the grammatical functions of *subject*, *object*, etc. It also includes functional features such as tense and aspect, case marking, gender, and so on.

Figure 1 presents a very simple LFG grammar. It consists of a lexicon on the right and phrase structure rules on the left. Phrase structure rules define context-free grammars and provide the c-structure, i.e. phrase structure trees. The context-free grammar and its lexicon are constrained by functional annotations, i.e. statements in the quantifier-free logic of equality (Kaplan 1989). Among these statements are equations (e.g. the VP of the S rule) and functional applications (e.g. on the NP in the S rule). In this conception, f-structures serve as functions. Their domain consists of grammatical functions and functional attributes, and their co-domain consists of f-structures and values of functional attributes.²

²Strictly speaking, function application can be applied to all indices used in the creation of a phrase structure tree, i.e. we can specify attributes and grammatical functions for c-structure nodes or nodes of other projections.

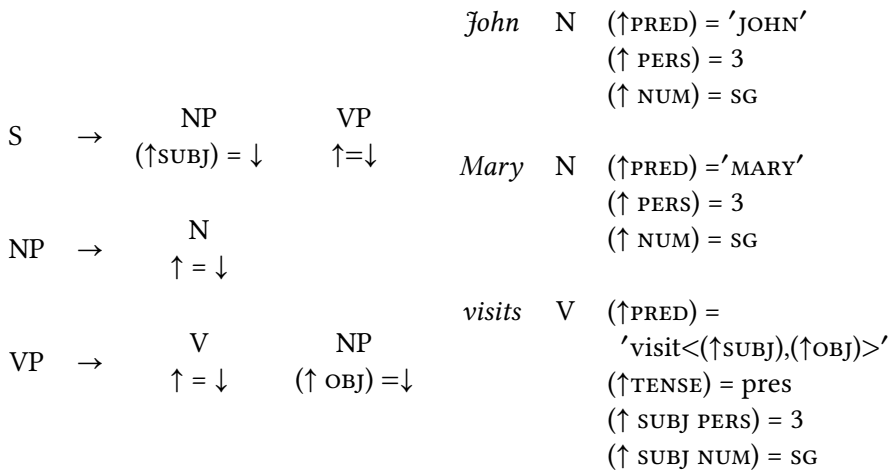


Figure 1: A simple grammar and lexicon

As Figure 1 suggests, f-structures are identified via the meta-variables \uparrow and \downarrow . These meta-variables point to indices in the c-structure as shown in Figure 2. By equating c-structure indices, LFG produces a more condensed structure describing the relations between sets of c-structure nodes: the f-structure. This mapping is highlighted in color in Figure 2 and is called a correspondence function. As multiple c-structure nodes are mapped onto a single f-structure node, this function is a many-to-one mapping. Conversely, there exists an inverse mapping onto the power set of the domain of a given correspondence function. For c- and f-structure, each c-structure node maps onto an f-structure index (the function ϕ), and an f-structure node may map onto a set of c-structure indices (the inverse correspondence ϕ^{-1} ; see chapter 7.2, footnote 10 in Bresnan et al. 2016).

Linguistic dependencies like agreement are handled in LFG by ascertaining consistency in functional annotations. In the current example, consistency needs to be checked between the SUBJ NP *John* and the verb *visits*, as both of them specify constraints on the grammatical function (GF) SUBJ. If verb and subject were incongruent regarding number and person, the f-structure formation would fail due to conflicting functional annotations. To put it another way, if a feature does not behave as a function within an f-structure, f-structure formation fails (e.g. in *I visits Mary*, where $(\uparrow \text{PERS}) = 1$ and $(\uparrow \text{PERS}) = 3$ would be provided for the SUBJ). Thus, the mechanism of resolving equations and functional annotations is used to define the interaction between the different linguistic elements specified in an f-structure.

3 Computational Challenges from Theoretical Semantic Modeling in LFG

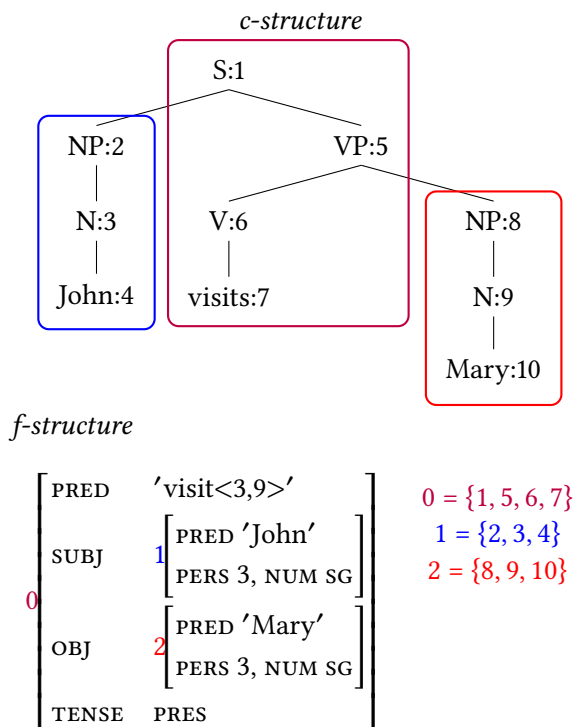


Figure 2: c- and f-structure: *John visits Mary*

Finally, f-structures must follow certain felicity conditions to be valid: *completeness* and *coherence*. These conditions are tied to the PRED value which provides a *semantic form*. The semantic form of verbs is special in that it provides subcategorization information. Completeness and coherence intuitively state that all and only the arguments which a predicate subcategorizes for may occur in its f-structure. Thus, the f-structure in Figure 2 is valid because it specifies the SUBJ and OBJ of the verb, no more, no less. Consequently, completeness and coherence provide essential linguistic constraints on f-structure formation.

This outline describes the key features of LFG’s syntactic component that play a role in this paper but leaves out some nuances of the theory. For more in-depth explanations of LFG, there exist various textbooks, e.g. Dalrymple (2001), Börjars et al. (2019), Falk (2011), or the more recent overview article by Börjars (2020). There also exists a new comprehensive handbook on LFG: Dalrymple 2024. In the next section, I provide a concise introduction to LFG’s semantic component.

2.2 Semantics in LFG

In LFG, the semantics is handled via Glue semantics, a resource-conscious approach to assembling meanings.³ More concretely, linear logic (Girard 1995) is used to guide compositional processes. Thus, glue derivations are essentially proofs based on the premises provided by the grammar for a given sentence. Before diving into the specifics of how meaning representations are tied to the overall grammar, I first introduce the anatomy of a *meaning constructor*, LFG’s semantic representations. A meaning constructor (MC) consists of two components: a meaning representation and a linear logic representation, which provides instructions for how to combine a set of premises (meaning constructors).

Linear logic is a resource-conscious logic as it does away with the rules of weakening and contraction (Girard 1995). Effectively, this means that each premise can be used only once in a linear logic proof. In a Glue semantics proof, then, premises roughly correspond to words, and a semantic derivation succeeds if all premises take part in a derivation.⁴ This is illustrated in (1). There, *John* and *Mary* provide atomic premises, and the verb *visits* provides a complex premise, indicated by the use of *linear implication* (\multimap).

- (1) **John** $j : g$
Mary $m : h$
visits $\lambda x.\lambda y.visit(x, y) : g \multimap (h \multimap f)$

The linear implication indicates that the corresponding premise wants to consume a meaning of a certain kind. In (1), first, the verb wants to consume the subject and then the object of the sentence. As shown in Figure 3 on the right, the process of consuming a premise corresponds to function application on the meaning side. (This process is reminiscent of modus ponens in classical logic. However, resource-sensitivity prohibits re-use of already used premises.) Applying this knowledge to the meaning constructors in (1), we can derive the proof in Figure 4, here shown as a proof tree.

³More comprehensive introductions are provided in, e.g. Dalrymple (2001, 1999), or the overview articles by Asudeh (2022, 2023). More recent varieties of Glue semantics differ substantially from their original conception described in, e.g. Dalrymple et al. (1993, 1997). In this paper, I subscribe to (a version of) the now standardly used *new Glue* (Dalrymple et al. 1999).

⁴It is possible that a word introduces more than one meaning constructor or that meaning constructors are introduced by non-lexical elements.

$$\frac{\begin{array}{c} [x : A]^i \\ \vdots \\ f(x) : B \end{array}}{\lambda x. f(x) : A \multimap B} \multimap_{I,i} \qquad \frac{f : A \multimap B \quad a : A}{f(a) : B} \multimap_E$$

Figure 3: Implication introduction and elimination

$$\frac{\frac{\lambda x. \lambda y. \text{visit}(x, y) : g \multimap (h \multimap f) \quad j : g}{\lambda y. \text{visit}(y, m) : h \multimap f} \quad m : h}{\text{visit}(j, m) : f}$$

Figure 4: Derivation of *John visits Mary*.

2.3 Accounting for ambiguities

As Figure 3 on the left indicates, Glue semantics not only provides a mechanism for function application but also for lambda abstraction, thus completing the correspondence between linear logic and lambda calculus: the Curry-Howard isomorphism (Curry et al. 1958, Howard 1980). This allows linear logic to account for, e.g. scopal ambiguities by making assumptions that are discharged later in a derivation. This is illustrated by virtue of the classic quantifier ambiguity in Figures 5 and 6. The meaning constructors for the sentence *everyone sees someone* are given in (2). None of the resources are immediately combinable. This is where assumptions come into play.

- (2) **see** : $\lambda x. \lambda y. \text{see}(x, y) : m_\sigma \multimap (b_\sigma \multimap f_\sigma)$
everyone : $\lambda Q. \forall y [\text{person}(y) \rightarrow Q(y)] : (b_\sigma \multimap f_\sigma) \multimap f_\sigma$
someone : $\lambda P. \exists x [\text{person}(x) \wedge P(x)] : (m_\sigma \multimap f_\sigma) \multimap f_\sigma$

Figure 5 begins with making an assumption m that combines with the lexical entry of the verb. The verb, one slot saturated by the assumption, is now compatible with the meaning constructor for *every person*. Next, the assumption is *discharged*, meaning reintroduced as shown in Figure 3 on the left. This is accompanied by abstraction over the variable on the meaning side of the original assumption, X . Thus, the resulting element is now compatible with the meaning constructor for *someone*, yielding the final denotation corresponding to the surface scope of the sentence. The inverse scope interpretation works by making two assumptions and then discharging them in the order they were made, as shown in Figure 6.

$$\frac{
 \frac{
 [X : m_\sigma]^1 \quad \lambda x.\lambda y.\text{see}(x, y) : m_\sigma \multimap (b_\sigma \multimap f_\sigma)
 }{
 \lambda y.\text{see}(c, y) : b_\sigma \multimap f_\sigma \quad \lambda Q.\forall y[\text{person}(y) \rightarrow Q(y)] : (b_\sigma \multimap f_\sigma) \multimap f_\sigma
 }{\multimap_E}
 }{
 \frac{
 \frac{
 \lambda P.\exists x[\text{person}(x) \wedge P(x)] : (m_\sigma \multimap f_\sigma) \multimap f_\sigma \quad \lambda x.\forall y[\text{person}(y) \rightarrow \text{see}(x, y)] : m_\sigma \multimap f_\sigma
 }{\multimap_{I,1}}
 }{
 \exists x[\text{person}(x) \rightarrow \forall y[\text{person}(y) \rightarrow \text{see}(x, y)]] : f_\sigma
 }{\multimap_E}
 }{\multimap_E}
 }$$

 Figure 5: Glue proof: *Someone sees everyone* surface scope

$$\frac{
 \frac{
 \frac{
 [X : m_\sigma]^1 \quad \lambda x.\lambda y.\text{see}(x, y) : m_\sigma \multimap (b_\sigma \multimap f_\sigma)
 }{
 \lambda y.\text{see}(X, y) : b_\sigma \multimap f_\sigma \quad [Y : b_\sigma]^2
 }{\multimap_E}
 }{
 \text{like}(X, Y) : f_\sigma \quad \lambda x.\text{see}(x, Y) : m_\sigma \multimap f_\sigma
 }{\multimap_E}
 }{
 \lambda P.\exists x[\text{person}(x) \wedge P(x)] : (m_\sigma \multimap f_\sigma) \multimap f_\sigma \quad \lambda x.\text{see}(x, Y) : m_\sigma \multimap f_\sigma
 }{\multimap_{I,1}}
 }{
 \frac{
 \frac{
 \frac{
 \exists x[\text{person}(x) \wedge \text{see}(x, Y)] : f_\sigma
 }{
 \lambda y.\exists x[\text{person}(x) \wedge \text{see}(x, y)] : b_\sigma \multimap f_\sigma \quad \lambda Q.\forall y[\text{person}(y) \rightarrow Q(y)] : (b_\sigma \multimap f_\sigma) \multimap f_\sigma
 }{\multimap_{I,2}}
 }{
 \forall y[\text{person}(y) \rightarrow \exists x[\text{person}(x) \wedge \text{see}(x, y)]] : f_\sigma
 }{\multimap_E}
 }{\multimap_E}
 }$$

 Figure 6: Glue proof: *Someone sees everyone* inverse scope

Due to the Curry-Howard isomorphism, Glue semantics is a system that captures semantic ambiguity via type-logical considerations (Gotham 2018), automatically inferring the order of computation.⁵ This is a very flexible system that can be combined with various meaning representations and syntactic formalisms.⁶ In the next section, I show how the syntax-semantics interface of Glue semantics works in tandem with LFG’s syntax.

2.4 The syntax-semantics interface in LFG

The syntax-semantics interface is one of the main points of contention in Glue semantics. Approaches can be broadly divided between *co-descriptive* approaches and *description-by-analysis* approaches. Co-description goes back to Dalrymple et al.’s (1993) original conception of Glue semantics.⁷ This approach is charac-

⁵Section 4 discusses the issue of ambiguity management in more detail.

⁶Regarding syntax, e.g. dependencies (Gotham & Haug 2018, Garrette & Klein 2009), HPSG (Asudeh & Crouch 2002), L-TAG (Frank & van Genabith 2001), and minimalist syntax (Gotham 2015). Regarding semantics, e.g. DRT (Garrette & Klein 2009) and its variants (van Genabith & Crouch 1997, Kokkonidis 2006), first-order lambda calculus (Dalrymple 1999).

⁷The debate goes back further, see, e.g. Halvorsen & Kaplan (1988) for co-description and Kaplan & Wedekind (1993) for description-by-analysis. However, it is noteworthy that the discussion was not necessarily about choosing between description-by-analysis vs. co-description but rather to what extent the two approaches should be employed. Wedekind & Kaplan (1993) illustrates this concretely. A caveat of this earlier literature is that it does not rely on the resource-sensitive approach of Glue semantics.

terized by evaluating semantic descriptions in parallel to c- and f-descriptions. Consequently, semantic information is annotated in parallel in the corresponding parts of the grammar. In (3), this is exemplified via the lexical entry for the verb *visits*. There, in addition to the functional annotations, a meaning constructor is included. This MC is added to the premise set whenever the verb is used in the analysis of a sentence. As the linear logic side of the MC shows, f-structure indices play an integral role in determining Glue resources.⁸ That is, semantic resources are co-notated with f-structure indices. The variable *%stem* is taken from XLE and refers to the lemma of the lexical entry (e.g. *visit* for *visits*). The use here makes concrete the relation between the semantic form and the semantics encoded in the meaning constructor.

(3) **Co-description:**

$$\begin{aligned} \text{visits } V \quad (\uparrow \text{PRED}) &= '\%stem <(\uparrow \text{SUBJ}),(\uparrow \text{OBJ})>' \\ (\uparrow \text{TENSE}) &= \text{pres} \\ (\uparrow \text{SUBJ PERS}) &= 3 \\ (\uparrow \text{SUBJ NUM}) &= \text{sg} \\ \lambda x.\lambda y.\%stem(x, y) &: (\uparrow \text{SUBJ})\multimap((\uparrow \text{OBJ})\multimap \uparrow) \end{aligned}$$

(4) **Description-by-analysis:**

$$\begin{aligned} \#f \text{ PRED } \%stem \ \& \ \text{SUBJ } \#g \ \& \ \#f \ \text{OBJ } \#h \\ \implies \#f \ \text{GLUE } \lambda x.\lambda y.\%stem(x, y) &: g\multimap(h\multimap f) \end{aligned}$$

The description-by-analysis approach exemplified in (4) also makes use of f-structure variables to determine meaning constructors. Here, they are not built up in parallel but rather take the f-structure as input to a set of rules that derives the premises for a given f-structure. The rule in (4) identifies f-structure nodes by virtue of variables of the form # [A-B]+ and feature value-variables take the form % [A-B]+. Thus, this rule for transitive verbs applies to any f-structure that has a SUBJ and an OBJ, introducing a fitting meaning constructor.⁹

⁸In this example, resources are identified by f-structure nodes. However, to make the system more modular, one could use a semantic projection (typically called the σ -projection) of f-structure nodes. As this example shows, this additional projection step is not strictly necessary.

⁹Description-by-analysis adds the (potential) complication of dealing with the question of how to choose the rules that apply to a given input f-structure, and the answer depends on the exact approach. For example, XLE's transfer system can be used for a version of description-by-analysis (Crouch 2005, Crouch & King 2006). There, rules are applied sequentially but also in a destructive manner (giving them a resource-sensitive character). For the sake of this paper, we do not presuppose any one special mechanism but only provide those rules that are relevant to the current examples.

Both of these approaches have been explored both theoretically and computationally. Nonetheless, there is a clear trend for co-descriptive approaches to be more popular in theoretical Glue semantics and description-by-analysis to be more popular in computational Glue semantics. We come back to this divide in section 5. Now that we have sketched an overview of LFG’s architecture (at least the syntactic and semantic components), we can start our expedition into the the gulf between theory and practice.

3 Lexical Functional Syntax between theory and practice

As indicated in the introduction, LFG has a long-standing history where theoretical assumptions and computational implementations go hand-in-hand. This approach has led to innovations in the theory and also unveiled some challenges of formal language modeling (see King (2016) for some interesting examples). In the next section, I present one such instance from LFG’s early days.

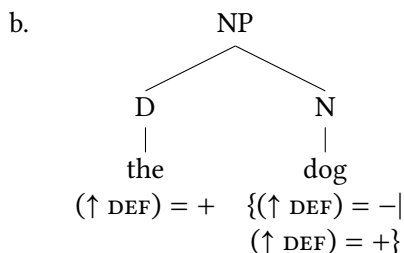
3.1 Defining and constraining equations

To ease ourselves into the topic, let’s start with a common error among beginner grammar writers: the distinction between defining and constraining equations (Kaplan & Bresnan 1982).¹⁰ Although this distinction is an important part of LFG’s theory, it is often overlooked. Intuitively, defining equations are the cornerstone of the underlying logic of LFG’s f-structure (the quantifier-free logic of equality). Constraining equations are, in principle, optional in the sense that they can be mimicked by virtue of using a more complex space of features (Kaplan & Bresnan 1982).¹¹ I assume that this comes down to the distinction between structures and descriptions (Kaplan 1989). As Kaplan points out, descriptions can be satisfied by increasingly complex structures. Conversely, it is sometimes easy to overlook how many structures a description allows for. Consider the naive example in (5b) for implementing the intuition that unmarked nouns are interpreted as indefinite, whereas definiteness must be explicitly marked (found, e.g. in Welsh).

- (5) a. Y ci (Welsh)
DEF dog
the dog

¹⁰As these are challenges observed in beginners in LFG and XLE, they should not be taken as serious problems but rather as small hurdles to keep in mind when teaching these concepts.

¹¹However, they have other useful applications, see Kaplan & Wedekind (2023).



The mistake here lies in the fact that the disjunction introduces two possibilities, neither of which is contradicted by the absence of a definiteness marker. Thus, the functional descriptions in (5b) correctly capture the fact that NPs marked for *definiteness* are, in fact, definite descriptions; however, for descriptions that are supposed to be indefinite, the analysis leads to an unwanted ambiguity because it also allows a definite interpretation for unmarked nouns.

(6) **Definite NP:**

$$\begin{bmatrix} \text{PRED} & \text{'dog'} \\ \text{DEF} & + \end{bmatrix}$$

Indefinite NP:

$$\begin{bmatrix} \text{PRED} & \text{'dog'} \\ \text{DEF} & \text{A1: -, A2: +} \end{bmatrix}$$

The underlying idea is that $(\uparrow \text{DEF}) = -$ is a kind of *elsewhere*-condition that applies in the absence of other possibilities.¹² As mentioned above, I suspect the culprit to be the distinction between description and structure: the description describes the desired structure but not only the desired structure. One way to remedy this issue is to replace the defining $(\uparrow \text{DEF}) = +$ constraint with a constraining equation $(\uparrow \text{DEF}) =_c +$. Then, the noun can be definite, but only if this is enforced elsewhere, i.e. in the determiner. However, now we have the “problem” that the structure does not reflect the description fully, at least when we do not infer it by contrasting definite structures with indefinite structures. That is, the f-structure makes no distinction between defining and constraining equations.¹³

Another such instance discussed by Kaplan & Bresnan (1982) is that of managing interactions between auxiliary verbs and participle forms. Concretely, the English auxiliary *be* is compatible with the present participle, whereas *have* is compatible with the past participle (in non-passive sentences). If the auxiliary

¹²The *elsewhere*-condition intuitively states that an analysis is used in the absence of a more specific one. It is traditionally attributed to Anderson (1969) and Kiparsky (1973). In LFG, it has been used, e.g. in morphological analyses (Andrews 1990).

¹³As a reviewer points out, this problem exists more clearly in grammar development (with the XLE), where the process of collecting defining equations and filtering via constraining equations is handled by the algorithm and not clearly presented to the user.

verb defines an equation along the lines of (\uparrow PARTICIPLE) = *present* and a verb has no participle feature whatsoever, then the auxiliary can impose upon the verb the status of a participle, even if it is not one, e.g.:

- (7) * A girl is hands the baby a toy. (Kaplan & Bresnan 1982: p. 33)

The problem becomes exactly the same as the previous one if we discuss past participles that are congruent with their simple past form, e.g. *walked*, *hugged*, *slept*. The verb introduces a disjunction between a participle and a past feature, but a defining equation on the auxiliary is compatible with both disjuncts, leading to mistaken analyses.

As previously discussed, such problems can be solved without resorting to the distinction between defining and constraining equations. For example, this can be achieved by assuming another feature value PARTICIPLE = *none*, but such features that express the absence of something (let's call them null features) can quickly make f-structures more unwieldy (Kaplan & Wedekind 2023). Thus providing an alternative source for introducing errors.

We can understand the introduction of defining and constraining equations as an update to the original declarative constraint satisfaction idea behind LFG that allows the removal of unnecessary or uninformative clutter at the level of f-structure. However, as Saraswat (1999) explains, the distinction makes the LFG architecture procedural: first, defining equations are used to build up possible f-structures, then constraining equations are applied to weed out undesired structures.¹⁴ However, LFG's ideal architecture is declarative in nature: multiple mutually constraining parallel projections. This is also reflected in the fact that f-structures give no hints on which features are aligned based on constraining equations and which features are aligned based on defining equations. This is more of a challenge for grammar development than LFG in general (see footnote 13), as XLE does not encourage inspecting f-descriptions explicitly but puts more emphasis on comparing structures.¹⁵

As pointed out at the beginning of this section, the errors pertaining to the use of defining and constraining equations seem to be particularly prevalent among

¹⁴This is reminiscent of unification-based grammars more generally as opposed to purely context-free grammars (CFGs). CFGs have a lot of expressive power, but writing them in such a way that they account for all sorts of agreement and other interactions makes the grammars immensely complex and ugly (Jurafsky & Martin 2024).

¹⁵The ParGram project (Butt et al. 2002) used f-structure comparisons to find cross-linguistically comparable analyses. Such an approach can be challenging for new learners of computational LFG since not all functional annotations can be easily inferred from the f-structure.

people new to grammar development. This is true both for people with and without a background in LFG.¹⁶ I think that this is because we often focus on the structures that we want to build as opposed to those that we want to avoid and the fact that constraining equations are only a part of descriptions but not of structures. This indicates how easily small slips in conceptualizing linguistic analyses can lead to computational issues. However, the possibility to debug such issues granted by the XLE also grants a deeper understanding of LFG grammars and even allows us to detect issues in the first place (King 2016).

3.2 Optionality

I want to expand on this point a bit more by talking about *optionality*.¹⁷ In a sense, the definiteness example above models an optional default value. In LFG, optionality is traditionally encoded at the level of c-structure, e.g. by making determiners optional. As shown above, this optionality interacts with the definiteness feature on the noun, requiring us to introduce a constraining equation. Example (8) illustrates how optionality is generally marked by wrapping an element in parentheses.

(8) NP → (D) AP* N.

Optionality can be found in functional descriptions, too. The challenge of optionality in the functional domain is that it is not related to the presence or absence of words, i.e. optionality at the c-domain means *add iff present*. However, in the functional domain, optionality means *always add, where compatible*. By this, I mean that functional equations always have to be resolved with the optional item in place and without it. This can lead to issues very similar to the ones presented above with defining equations. If some annotation is optional, it is always there if it does not conflict with anything. This leads to unwanted ambiguities by, potentially unnecessarily, adding an optional feature.

A typical use of optionality of f-structure descriptions are *pro-drop* languages, e.g. Warlpiri (Dalrymple 2001: p. 128) or Serbo-Croatian (Asudeh 2005: p. 495).

¹⁶Based on anecdotal evidence.

¹⁷As a reviewer points out, the actual culprit here is that of disjunction. (Furthermore, other extensions of LFG, e.g. optimality theory, a version of which is also implemented in XLE, are even more opaque and challenging.) Thus, other aspects of LFG provide challenges for trainees in grammar development. Optionality is taken as an example here because it has the potential to result in unwanted ambiguity, even for simple analyses, i.e. it is an early candidate for teaching newcomers to LFG and grammar development the importance of carefully handling disjunctions. Furthermore, this section highlights some other aspects of *optionality* that play a role across LFG's different projections.

We use Dalrymple’s example to show that optionality must be constrained by other mechanisms, making it non-transparent in f-structure descriptions:

$$\begin{aligned}
 (9) \quad \textit{panti-rni} \quad \vee \quad & (\uparrow \text{PRED}) = \text{'SPEAR<SUBJ,OBJ>'} \\
 & ((\uparrow \text{SUBJ PRED}) = \text{'PRO'}) \\
 & (\uparrow \text{SUBJ CASE}) = \text{ERG} \\
 & ((\uparrow \text{OBJ PRED}) = \text{'PRO'}) \\
 & (\uparrow \text{OBJ CASE}) = \text{ABS}
 \end{aligned}$$

This example uses optionality markers to indicate that both subject and object can be optional in this language. However, the optionality is constrained by the f-structure’s well-formedness conditions (completeness and coherence).

The formal implementation of LFG in XLE implements optionality as a disjunction between some f-description and TRUE.¹⁸ This is illustrated in (10), based on the example in (9).

$$\begin{aligned}
 (10) \quad \textit{panti-rni} \quad \vee \quad & (\uparrow \text{PRED}) = \text{'SPEAR<SUBJ,OBJ>'} \\
 & \{ (\uparrow \text{SUBJ PRED}) \mid \\
 & \quad (\uparrow \text{SUBJ PRED}) = \text{'PRO'} \} \\
 & (\uparrow \text{SUBJ CASE}) = \text{ERG} \\
 & \{ (\uparrow \text{OBJ PRED}) \mid \\
 & \quad (\uparrow \text{OBJ PRED}) = \text{'PRO'} \} \\
 & (\uparrow \text{OBJ CASE}) = \text{ABS}
 \end{aligned}$$

The gist of (10) is that we use the logic of a conditional on encoding optionality: ($p \rightarrow q \equiv (\neg p \vee q)$), where p is the f-structure without the optional condition and q is the f-structure with the optional condition in place. As the equivalence suggests, optionality in the present example is in fact conditionality, i.e. if the SUBJ has no PRED-value, then add it. In this sense, the optionality marking presented in (9) leaves more information implicit, making the use of the operator less transparent. This approach replaces optionality by the use of constraining equations, which add an inelegance of a different kind to the LFG machinery.

This excursion begs the question of whether all instances of optionality are, in fact, instances of conditionality. If we assume that syntactic structure strictly expresses the structural information encoded in the input string, we might want to

¹⁸Another possibility can also be envisioned: one could impose optional constraints by putting them on an optional empty node that is a sister to the node that it constrains. This approach has the same drawbacks as that of optionality without a condition. Even more so, because it obfuscates the c-structure (see also <https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/notations.html#N2>).

believe that optionality should not have a place in syntactic descriptions. However, we might be convinced to add optionality at the level of semantics, e.g. for manager resources (Asudeh 2004) or for optional arguments (Asudeh & Giorgolo 2012). The use of manager resources can be tied to conditions on various levels of analysis, e.g. information structure or prosody (Haug 2008). There, they are used to introduce ambiguity that would not be there otherwise. However, our goal is to unveil challenges pertaining to unwanted ambiguities. The case of optional arguments is more critical. Why is that? Consider the passive lexical entry from Asudeh & Giorgolo (2012):

$$\begin{aligned}
 (11) \quad & \textit{eaten} \quad V \quad (\uparrow \text{PRED}) = \textit{'eat'} \\
 & ((\uparrow \text{VOICE}) = \text{PASSIVE}) \\
 & \lambda x. \lambda y. \lambda e. \textit{eat}(e) \wedge \textit{eat}(e) \wedge \textit{agent}(e) = x \wedge \textit{patient}(e) = y : \\
 & (\hat{*}_{\alpha} \text{AGENT})_{\theta \rightarrow} (\uparrow \text{SUBJ}_{\sigma \rightarrow} (\uparrow \text{EVENT})_{\sigma \rightarrow} \uparrow_{\sigma} \\
 & (\lambda P. \exists x [P(x)] : ((\hat{*}_{\alpha} \text{AGENT})_{\theta \rightarrow} \uparrow_{\sigma}) \rightarrow \uparrow_{\sigma})
 \end{aligned}$$

This lexical entry states that a transitive verb like *eat* can optionally introduce a resource saturating the agent position of the described action. This agent is not further semantically specified: it is represented by an existential quantifier. This resource is only necessary if there is no oblique agent introduced via a *by*-phrase, thus making it optional. This is illustrated in the following example.

- (12) a. The bone was eaten.
 b. The bone was eaten by the dog.

Saturating the agent optionally is meant to express that it is needed in case it is not specified explicitly, i.e. a condition on the surface string in parsing. However, the kind of optionality presented in (11) makes the condition semantic: add the optional element just in case the proof fails. For parsing, this means that each of the sentences in (12) induces two different proofs: one with the optional agent and one without it. This ambiguity is not necessary if semantics listens to the syntax. A corresponding lexical entry in XLE could look like (13). There, we use the presence of an oblique agent as a condition for the optionality. This approach is reminiscent of the null subject approach discussed before in the sense that a condition is related to the presence or absence of a grammatical function.¹⁹

¹⁹As discussed in section 5, this approach depends on the fact that completeness and coherence are dealt with at the level of syntax and not semantics (as argued for by Andrews 2008, Asudeh 2004, Kuhn 1998 i.a.).

$$\begin{aligned}
 (13) \quad & \textit{eaten} \quad \text{V} \quad (\uparrow \text{PRED}) = \textit{'eat'} \\
 & \dots \\
 & \{ (\uparrow \text{PERF}) = + \mid \\
 & \quad (\uparrow \text{VOICE}) = \text{PASSIVE} \\
 & \quad \{ (\uparrow \text{OBL-AG}) \mid \\
 & \quad \quad \lambda P. \exists x [P(x)] : ((\hat{*}_\alpha \text{AGENT})_{\theta \rightarrow} \uparrow_\sigma) \rightarrow \uparrow_\sigma \} \}
 \end{aligned}$$

Overall, optional constraints, either syntactic or semantic, are a known cause for unwanted ambiguity. Optional semantic elements, in particular, invite a lot of unnecessary computation in a fully modular architecture. This is discussed in more detail in section 5.

In this section, we have seen the pitfalls of theoretical analyses. They make assumptions or, in linguistic terms, presuppositions. We presuppose that there is a condition on optionality.²⁰ We have also seen that some presuppositions may eventually be asserted in the analysis. The distinction between defining and constraining equations is an early development to constrain unwanted feature propagation. In summary, “inconsequential” assumptions are not stated explicitly and require (presupposition) accommodation. However, this leads to a discrepancy between formal modeling and computational modeling. In the next section, we discuss more such issues from the domain of semantics.

4 The commutativity conspiracy

When new concepts are introduced into a theory, issues are bound to emerge. Some of these issues only become apparent under the scrutiny of a computational implementation. Thus, theoretical tools that have been explored more recently (e.g. optionality at the level of semantics) that have not been accompanied by computational exploration require careful attention, as their formal properties may not have been explored with the rigor necessary for computational modeling. Formal rigor is accommodated for by implicit knowledge, i.e. knowledge you (seemingly) do not need to worry about when thinking about linguistic problems theoretically. An exemplary instance of this is Glue semantics’s commutativity.

4.1 The worst case

Asudeh & Giorgolo (2020) write that semanticists have the bad habit of generalizing to the worst case. Arguably, that is also true of Glue semantics. Concretely,

²⁰For an example of this in the recent literature, see Asudeh & Rad (2023), which produces some fairly complex templates using optional elements. However, the resolution of these optional elements is mediated by an unspecified pragmatic component.

3 Computational Challenges from Theoretical Semantic Modeling in LFG

the greatest feature of Glue semantics is also its most significant source of problems. The issues discussed in section 3.2 have reasonable fixes that are easy to implement. However, commutativity, which plays a crucial role in modeling ambiguities, leads to considerable computational inefficiency or sometimes even to plain overgeneration. This issue has received considerable attention in the literature without a satisfying resolution (e.g. Findlay & Haug 2022, Gotham 2019, Lev 2007, Crouch & van Genabith 1999).

To understand this conundrum, take the noun phrase *big black dog* (f-structure and meaning constructors in Figure 7). The typical treatment of adjectives uses *modifier* premises. Modifier premises, intuitively, are premises that, given some input X of type τ , return the same object, i.e. $X_{\tau} \multimap X_{\tau}$.²¹ Here, commutativity comes into play. The property of linear logic proofs allows us to apply modifier premises of the same type in any order, as in Figures 8 and 9.

$$\begin{array}{l}
 \left[\begin{array}{l} \text{PRED 'dog'} \\ \text{ADJUNCT } \left\{ \begin{array}{l} \left[\text{PRED 'big'} \right], \\ \left[\text{PRED 'black'} \right] \end{array} \right\} \end{array} \right] \\
 \mathbf{1}
 \end{array}
 \quad
 \begin{array}{l}
 (\mathbf{1}_e \multimap \mathbf{1}_t) \multimap (\mathbf{1}_e \multimap \mathbf{1}_t) : \\
 \lambda P. \mathit{big}(P) \\
 \\
 (\mathbf{1}_e \multimap \mathbf{1}_t) \multimap (\mathbf{1}_e \multimap \mathbf{1}_t) : \\
 \lambda P. \mathit{black}(P) \\
 \\
 (\mathbf{1}_e \multimap \mathbf{1}_t) : \mathit{dog}
 \end{array}
 \end{array}$$

Figure 7: F-structure and meaning constructors for *big black dog*

$$\frac{\frac{(\mathbf{1}_e \multimap \mathbf{1}_t) \multimap (\mathbf{1}_e \multimap \mathbf{1}_t) : \mathit{black} \quad (\mathbf{1}_e \multimap \mathbf{1}_t) : \mathit{dog}}{(\mathbf{1}_e \multimap \mathbf{1}_t) : \mathit{black}(\mathit{dog})} \quad (\mathbf{1}_e \multimap \mathbf{1}_t) \multimap (\mathbf{1}_e \multimap \mathbf{1}_t) : \mathit{big}}{(\mathbf{1}_e \multimap \mathbf{1}_t) : \mathit{big}(\mathit{black}(\mathit{dog}))}$$

Figure 8: Derivation of *big black dog*

$$\frac{\frac{(\mathbf{1}_e \multimap \mathbf{1}_t) \multimap (\mathbf{1}_e \multimap \mathbf{1}_t) : \mathit{big} \quad (\mathbf{1}_e \multimap \mathbf{1}_t) : \mathit{dog}}{(\mathbf{1}_e \multimap \mathbf{1}_t) : \mathit{big}(\mathit{dog})} \quad (\mathbf{1}_e \multimap \mathbf{1}_t) \multimap (\mathbf{1}_e \multimap \mathbf{1}_t) : \mathit{black}}{(\mathbf{1}_e \multimap \mathbf{1}_t) : \mathit{black}(\mathit{big}(\mathit{dog}))}$$

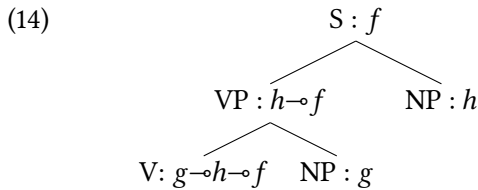
Figure 9: Spurious ambiguity in *big black dog*

²¹More complex modifier types are possible. The quantifiers shown in section 2.2 are ultimately modifier premises (Lev 2007). So are premises of type $X_{\tau} \multimap y_{\sigma} \multimap X_{\tau}$.

On the surface, this example does not seem all that problematic, but it causes a snowball effect. Combined with the quantifier ambiguity presented in section 2.2, this ambiguity would be propagated, leading to four readings for *every big black dog chased a cat*. Formally and more generally, for any number of unconstrained modifiers n , there are $n!$ possibilities of applying these modifiers, quickly leading to thousands of solutions even for relatively simple sentences (see also Bunt 2008). Consequently, constraining (compositional) ambiguity means constraining modifier premises in Glue semantics.

Opposed to modifier premises are skeleton premises (Gupta & Lamping 1998), which usually do not introduce additional ambiguity.²² I bring them up as they allow for an interesting analogy.

Skeleton modifiers correspond roughly to binary-branching trees, whereas modifier premises correspond to multiple-branching trees. Consider the first example in section 2.2. There, there is a clear correspondence between a syntactic tree and the meaning representations forming a functional application tree.²³



For multiple-branching trees, the relation is a one-to-many mapping, as shown in Figure 10. There, the two possible solutions correspond to the two possible orderings of the terminal nodes of the function-application tree (while adhering to the condition that the functor is always on the left and the argument always to the right, eliminating further equivalent trees).

²²In linear logic, the following equality holds:

(i) $a \multimap b \multimap c \equiv b \multimap a \multimap c$

This is due to the mechanism of introducing assumptions. Generally, any proof involving linear implication can have infinitely many solutions (Lev 2007: p. 137f). However, in this paper, I only assume normal form proofs, intuitively, proofs that make no unnecessary assumptions.

²³For these trees, I assume that the functor is always on the left and the argument is always on the right to annoy syntacticians and semanticists all at once. Due to commutativity, the tree could be re-ordered to match English word order or any other word order for that matter.

3 Computational Challenges from Theoretical Semantic Modeling in LFG

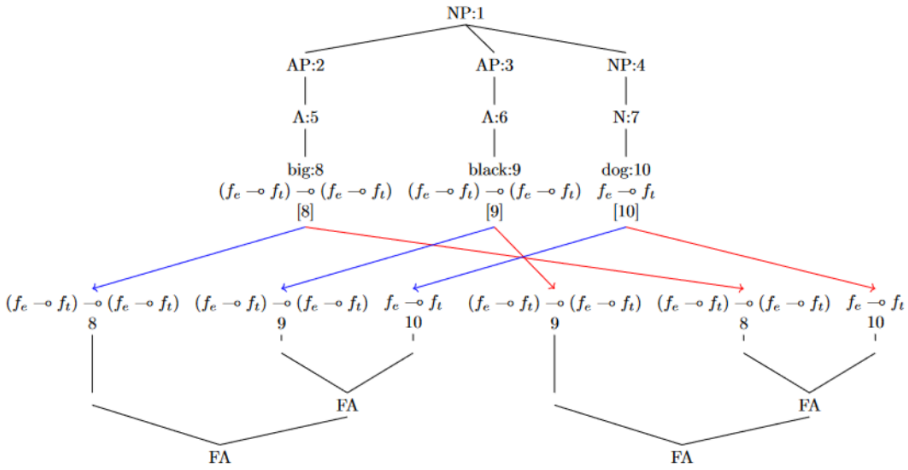


Figure 10: Semantic derivation as ordering: *big black dog*

Consequently, linear logic provides a partial ordering on the terminal nodes of a function-application tree. Skeleton premises provide a total ordering, whereas modifier premises make the order partial (in normal form proofs).²⁴

It seems obvious that skeleton premises avoid spurious ambiguities. Nonetheless, a recent trend is to represent the arguments of a verb in terms of modifiers. The merits of such an approach are, generally speaking, more flexibility in dealing with the selectional restrictions of predicates. Concretely, approaches that treat arguments as modifiers are usually instances of implementing a (variant of) Neo-Davidsonian event semantics (e.g. Asudeh et al. 2014, Findlay & Haug 2022, Przepiórkowski & Patejuk 2023). Thus, the role of *modifier* is shared between the glue and meaning side (Parsons 1990) of an MC. Consequently, their order in semantic combination is irrelevant. A simplified representation is given in (15b). There, each argument introduces two meaning constructors: its denotation and its argument role.

²⁴There are more potential sources of partial ordering. Consider the set $a, a, a \multimap a \multimap b$. There, either a can combine with the consuming formula first, providing two possible proofs of b , though such combinations are not generally expected in Glue semantics proofs. However, some instances of Glue semantics beyond the implicational fragment used here use multiplicative conjunction \otimes , which can introduce similar spurious ambiguities.

(15) a. Jordan visits Mary.

b. $j : g$
 $m : h$
 $\lambda P.\lambda x.P \wedge ag = x : f \multimap g \multimap f$
 $\lambda P.\lambda x.P \wedge th = x : f \multimap h \multimap f$
 $f : \text{visit.}$

When comparing (1) and (15b), it is clear that the latter is more complex even in its simplified form. The modifier treatment of verb arguments causes spurious ambiguities.²⁵ However, what is interesting here is that the f-structure, the linear logic type, and the semantic representation are all in unison. In comparison, (14) and (1) effectively show an approach that reintroduces information from c-structure (i.e. constituency information, but not necessarily linear ordering).

4.2 Commutativity’s fundamental flaw

This means that by moving from skeleton arguments to modifier arguments, redundant structural information is eliminated from semantic derivations. The spurious ambiguity that accompanies this step would not be an issue, even computationally, if the commutativity of Glue semantics was not fake. The problem is that modifier premises form a heterogeneous group: it is not possible to infer from the linear logic types alone whether a set of meaning constructors only allows one distinct interpretation or multiple. Recall Figures 5 and 6. Although these examples obfuscate the fact, quantifiers are modifier premises with the minor difference that they are anchored in the derivation, constraining their scope.

Approaches to resolving Glue derivations based on compilation (Hepple 1996, Lev 2007) make this intuition clear. They rely on the fact that normal-form proofs can be found without making any actual assumptions.²⁶ Rather, in a given set of Glue premises, we can first polarize premises and then match all occurrences of atomic elements such that each positive element has a negative counterpart. This is exemplified in (16). There, all atomic elements are matched except for the result that has been specified for the proof, namely, t , which is of positive charge.

(16) $(e \multimap t) \multimap t, e \multimap t \vdash t$ e.g. *a dog barks*

²⁵This is not the only approach to handle Neo-Davidsonian event semantics in LFG, see, e.g. Asudeh et al. (2008). Their approach circumvents the spurious ambiguity.

²⁶These intuitions are also the foundation of proof nets (de Groote 1999).

3 Computational Challenges from Theoretical Semantic Modeling in LFG

$$(e^- \multimap t^+)^+ \quad ((e^+ \multimap t^-)^- \multimap t^+)^+$$

This knowledge can be used to simplify linear logic formulas with complex antecedents such as those found in quantifiers. This is shown in (17a). As illustrated in (17b), decomposed elements can be freely used in a Glue derivation. However, they have to ultimately reconnect with their host, as indicated by the subscript i . A thorough explanation of this process is given in Lev (2007: p. 141ff). The important fact here is that quantifiers are ultimately modifiers of type $t \multimap t$.

(17) a. $(e \multimap t) \multimap t \rightarrow_{comp} [e]^i, t_i \multimap t$

$$\frac{[e]^i \quad e \multimap t}{t^i \quad t_i \multimap t} \quad t$$

b.

Thus, compositional ambiguity can be reduced to the existence of modifier premises. However, in the case of quantifiers, the ambiguity is desired, and in the case of other modifiers, the ambiguity is often spurious.

In sum, it may be fair to say that Glue semantics suffers from an architectural issue. This seems to stem from the fact that the communication between the linear logic side and the meaning side of meaning constructors is imperfect.²⁷ In the next section, I discuss some potential repercussions of these flaws for the overall grammar architecture of LFG.

5 Ambiguities in grammar architecture

Let's recapitulate. In the previous two sections, we discussed two types of ambiguities: unwanted ambiguities introduced, e.g. by optionality, and spurious ambiguities introduced by modifier premises. Unwanted ambiguities are primarily (but not exclusively as we will see) a challenge for grammar writers, whereas spurious ambiguities are challenges for computational linguists interested in implementing Glue semantics in a way that is efficient and practical. However, ultimately, both points have computational repercussions and highlight potential flaws in LFG's syntax-semantics interface.

²⁷A related point is made by Findlay (2021). Findlay argues that there is no real connection between the semantics of a linguistic expression and the remaining correspondence structure, i.e. in the conception of Glue presented in this paper, the semantics lacks the mutually constraining quality of LFG's various projections.

5.1 Unwanted ambiguities and optionality

The issue of unwanted ambiguities, as introduced by semantic optionality, seems computational in nature. Recall that the issue here was that the semantics did not listen to the syntax when specifying optional arguments. Thus, when relying on an analysis where completeness and coherence are handled in the domain of semantics (Andrews 2008, Kuhn 1998, Asudeh 2004), completely unwanted derivations emerge (the unwanted derivations fail, but it is questionable whether they should occur in the first place).

In comparison, at f-structure, completeness and coherence can be ascertained by checking a small number of features without compromising the structures provided by the defining equations of a given analysis. In essence, f-descriptions naturally describe a set of f-structures, whereas meaning constructors describe multiple possible derivations, which all need to be checked when they contain optional elements without a separate condition, e.g. in the syntax, that constrains them. Thus, constraining optional arguments via completeness and coherence at f-structure is computationally more efficient. However, the fact that one even has to attempt to compute an obviously wrong derivation when putting completeness and coherence in the semantics suggests that putting those conditions in the f-structure is more elegant formally, and certainly computationally.

More generally, conditions on optional elements are a challenge from a computational perspective. I briefly discuss three points that are highlighted by the example we have explored in this paper: First, in LFG's architecture, conditions on optionality can create a kind of long-distance dependency. Conceptually, the projection architecture is parallel, but computationally and from the perspective of speech production and perception one would expect some procedural processing (see Jackendoff 2010). In this case, the repercussions of such conditions may become increasingly non-transparent as discussed in section 3.2.

A second related point is that it seems wise to resolve optionalities early as the underlying ambiguities could propagate through the form-to-meaning mapping, potentially inviting further unwanted or spurious ambiguity and redundant computations. This is illustrated by the example above. By resolving completeness and coherence early in parsing, unnecessary computations are avoided. This also shows that a shorter distance is not always better. Rather, the issue of relating optional elements to conditions is nuanced. This point deserves further research as it is also dependent on the directionality of parsing and generation. Here, we have only considered parsing.

Thirdly, certain formal tools allow us to deal with optionality more naturally and efficiently. In particular, monotonic or truly commutative representations

fare more easily with respect to optionality, whereas the non-monotonicity and fake-commutativity in Glue semantics may cause computational challenges.²⁸

5.2 Spurious ambiguity and commutativity

Spurious ambiguity seems to be a more fundamental issue due to the heterogeneous nature of modifier premises. In theoretical Glue semantics, such spurious ambiguities can be filtered out, and this is often assumed in papers on theoretical Glue (see, e.g. the discussion in Findlay & Haug 2022 on Asudeh et al.'s (2014) event semantics). This filtering must be implemented computationally to make Glue derivations computationally feasible; however, this filtering is not trivial, which highlights the fact that there is a tension between linear logic and meaning representation in Glue semantics.

As mentioned in the previous section, spurious ambiguities are also bound to propagate through the syntax-semantics interface. More generally, in computational modeling, inefficiency in one place causes issues for computational modeling in another place. Consequently, all modules have to be efficient for the overall system to be efficient. Thus, making it a worthwhile goal to work on the issue of commutativity in Glue semantics.

Let me leave you with a final somewhat unsatisfying note on the topic: Gotham (2021) explains that categorical grammar uses a non-commutative logic because it also needs to capture word-order phenomena in its representations (see also Moot & Retoré 2012). As there exist languages with varying word-orders and many exhibit some degree of word-order variation, commutativity may be a property that needs to be distributed across different dimensions of linguistic analysis. For example, the research of Gotham (2019, 2021) shows that quantification seems to be more restricted in a language with more free word order like German. However, the nature of this relation between word order and restrictions on quantification is left for future work.

All in all, the kinds of ambiguities discussed in this paper highlight some nuances of formal semantic research and linguistic research more generally that are challenging, especially when considered from a computational perspective. Thus, they require due attention in the (computational) semantic research to come, hopefully, revealing fruitful insights for LFG and linguistics as a whole.

²⁸F-structures are not build fully monotonically. However, the building of f-structures is efficient and the resulting structures are easy to filter. In Glue semantics, non-monotonicity is more deeply intertwined making dealing with optionality harder (but maybe not impossible).

6 Summary

In a sense, this paper recapitulates what was already said by Ronald Kaplan:

- (18) Linguists don't really design formalisms, or at least, they don't seem to design very good ones. It is the business of computational linguists to do this, and this is what we're skilled at. But we've got to design formalisms that linguists will use, to make sure that they don't come up and start using formalisms that are unimplementable. (Kaplan 1987).

What is the perceived wisdom here? Gradual modifications of a framework always invite new issues. Thus, it is always helpful to test one's ideas computationally. A particularly noteworthy process is that of *regression testing* which encourages us to make new innovations without compromising existing insights (Chatzichrisafis et al. 2007).

In this paper, I have shown some discrepancies between the formal and computational modeling of LFG and related them to the terms *issues* and *features*. The caveat of the whole ordeal is that both formal models and computational models tend to move complexity around or, in other words, outsource it. However, the crucial differences between formal and computational models is that formal models can outsource certain constraints to the cognitive apparatus of humans, whereas a computational analysis needs to deal with the complexity in one place or another. In other words, computational analyses do not allow us to presuppose information as easily. If information is omitted in a computational analysis, this hole cannot be filled by our knowledge of the subject matter.

In my eyes, this has long been a strength of LFG: a close communication between computational and formal needs. However, this paper serves to nudge people to contemplate these roots and take them seriously. If LFG devolves into a formalism of assumed mechanisms, then it runs the danger of becoming a meta language of linguistic research that is not transparent with respect to its formal underpinnings. Although such approaches allow us to further our understanding of language conceptually, they make discussions about the suitability of linguistic frameworks obsolete. In this case, theoretical LFG is seriously threatened by the sheer size of other frameworks, which establish many of the same presuppositions that are shared by a much wider array of researchers. Similarly, computational modeling will lose its appeal of being very close to its theoretical counterpart. Thus, keeping the close relationship between computational and theoretical research alive should be a worthwhile goal for researchers in LFG.

Acknowledgments

The origins of this paper can be traced to a visit to California in the fall of 2017, when Moritz Meßmer, a colleague and friend of mine, and I traveled to California to work on a first implementation of a prover for Glue semantics under the supervision of Dick Crouch. Back then, we managed to develop a prover that dealt with some classic examples often discussed in Glue semantics, e.g. quantifier ambiguities and a treatment of adverbs and adjectives. For us, it was an achievement, and we will be forever grateful to Dick and Valeria, who agreed to host us and take care of us for those couple of weeks. Since then, the Glue Semantics Workbench, as we called it, has undergone considerable revisions and has gathered some attention, particularly in the LFG community. It also has cemented itself as one of my primary research interests. Thus, this paper is written as a thanks to Dick for setting me on this path and supporting us back then. I am also thankful to Miriam Butt who made this collaboration possible in the first place.

I am grateful to the reviewers who gave valuable feedback and helped smooth some rough edges. Furthermore, I want to extend this thanks to the editors of this book, particularly Tracy Holloway King, who also lives on the edge of theoretical and computational linguistics. This paper has also benefited from discussions with various individuals dear to me (in no particular order): Tina Bögel, Frances Dowle, Songfalo Lacina Silué, Jamie Findlay, and Ash Asudeh. Finally, I am grateful to the Glue semantics community for showing continued interest in my research.

References

- Anderson, Stephen Robert. 1969. *West Scandinavian vowel systems and the ordering of phonological rules*. Massachusetts Institute of Technology. (Doctoral dissertation).
- Andrews, Avery D. 1990. Unification and morphological blocking. *Natural Language & Linguistic Theory* 8(4). 507–557.
- Andrews, Avery D. 2008. The role of PRED in LFG+glue. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG08 conference*, 46–76. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2008/lfg08.html>.
- Asudeh, Ash. 2004. *Resumption as resource management*. Stanford University. (Doctoral dissertation).
- Asudeh, Ash. 2005. Control and semantic resource sensitivity. *Journal of Linguistics* 41(3). 465–511.

- Asudeh, Ash. 2022. Glue semantics. *Annual Review of Linguistics* 8. 321–341.
- Asudeh, Ash. 2023. Glue semantics. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 651–697. Language Science Press. DOI: 10.5281/zenodo.10185964. <https://doi.org/10.5281/zenodo.10185964>.
- Asudeh, Ash & Richard Crouch. 2002. Glue semantics for HPSG. In *Proceedings of the 8th international HPSG conference*, 1–19. CSLI Publications.
- Asudeh, Ash, Mary Dalrymple & Ida Toivonen. 2008. Constructions with lexical integrity: Templates as the lexicon-syntax interface. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG08 conference*, 68–88. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2008/lfg08.html>.
- Asudeh, Ash & Gianluca Giorgolo. 2012. Flexible composition for optional and derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG12 conference*, 64–84. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2012/lfg12.html>.
- Asudeh, Ash & Gianluca Giorgolo. 2020. *Enriched meanings: Natural language semantics with category theory*, vol. 13. Oxford University Press.
- Asudeh, Ash, Gianluca Giorgolo & Ida Toivonen. 2014. Meaning and valency. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG14 conference*, 68–88. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2014/lfg14.html>.
- Asudeh, Ash & Siavash Rafiee Rad. 2023. Persian perception verbs. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG23 conference*, 45–65. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/2>.
- Börjars, Kersti. 2020. Lexical-Functional Grammar: An overview. *Annual Review of Linguistics* 6. 155–172.
- Börjars, Kersti, Rachel Nordlinger & Louisa Sadler. 2019. *Lexical-Functional Grammar: An introduction*. Cambridge University Press.
- Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-functional syntax*. 2nd edn. Wiley-Blackwell.
- Bunt, Harry. 2008. Semantic underspecification: Which technique for what purpose? In *Computing meaning*, 55–85. Springer.
- Butt, Miriam, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi & Christian Rohrer. 2002. The parallel grammar project. In *Proceedings of the 2002 workshop on grammar engineering and evaluation*, vol. 15, 1–7. Association for Computational Linguistics.

3 Computational Challenges from Theoretical Semantic Modeling in LFG

- Chatzichrisafis, Nikos, Dick Crouch, Tracy Holloway King, Rowan Nairn, Manny Rayner & Marianne Santaholma. 2007. Regression testing for grammar-based systems. In *Proceedings of the grammar engineering across frameworks (GEAF07) workshop*, 128–143. CSLI Publications.
- Crouch, Dick, Mary Dalrymple, Ronald Kaplan, Tracy Holloway King, John T. Maxwell III & Paula Newman. 2011. *XLE Documentation*. Palo Alto Research Center (PARC). <https://ling.sprachwiss.uni-konstanz.de/pages/xle/index.html>.
- Crouch, Richard. 2005. Packed rewriting for mapping semantics to KR. In *Proceedings of the sixth international workshop on computational semantics (IWCS-6)*, 103–114.
- Crouch, Richard & Tracy Holloway King. 2006. Semantics via f-structure rewriting. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG06 conference*, 145–165. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2006/lfg06.html>.
- Crouch, Richard & Josef van Genabith. 1999. Context change, underspecification, and the structure of glue language derivations. In Mary Dalrymple (ed.), *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*, 117–189. The MIT Press.
- Curry, Haskell Brooks, Robert Feys, William Craig, J. Roger Hindley & Jonathan P. Seldin. 1958. *Combinatory logic*, vol. 1. North-Holland Amsterdam.
- Dalrymple, Mary (ed.). 1999. *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*. The MIT Press.
- Dalrymple, Mary. 2001. *Lexical Functional Grammar*, vol. 34 (Syntax and Semantics). Academic Press.
- Dalrymple, Mary (ed.). 2024. *Handbook of Lexical Functional Grammar*. Language Science Press.
- Dalrymple, Mary, Vineet Gupta, John Lamping & Vijay Saraswat. 1999. Relating resource-based semantics to categorial semantics. In Mary Dalrymple (ed.), *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*, 261–280. The MIT Press.
- Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell III & Annie Zaenen (eds.). 1995. *Formal issues in Lexical-Functional Grammar*. CSLI Publications.
- Dalrymple, Mary, John Lamping, Fernando Pereira & Vijay Saraswat. 1997. Quantification, anaphora and intensionality. *Logic, Language and Information* 6. Appears with some modifications in Dalrymple (1999), pp. 39–90, 219–273.
- Dalrymple, Mary, John Lamping & Vijay Saraswat. 1993. LFG semantics via constraints. In *Proceedings of the sixth conference of the European chapter of the Association for Computational Linguistics (EACL) (EACL '93)*, 97–105. Associa-

- tion for Computational Linguistics. DOI: 10.3115/976744.976757. <https://doi.org/10.3115/976744.976757>.
- de Groote, Philippe. 1999. An algebraic correctness criterion for intuitionistic multiplicative proof-nets. *Theoretical Computer Science* 224(1-2). 115–134.
- Falk, Yehuda. 2011. *Lexical-Functional Grammar*. Oxford University Press.
- Findlay, Jamie Y. 2021. Meaning in LFG. In I Wayan Arka, Ash Asudeh & Tracy Holloway King (eds.), *Modular design of grammar: Linguistics on the edge*, 340–374. Oxford University Press.
- Findlay, Jamie Y. & Dag Trygve Truslew Haug. 2022. Managing scope ambiguities in glue via multistage proving. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG22 conference*, 144–163. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/1>.
- Frank, Anette & Josef van Genabith. 2001. Linear logic-based semantics construction for LTAG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG01 conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2001/lfg01.html>.
- Garrette, Dan & Ewan Klein. 2009. An extensible toolkit for computational semantics. In *Proceedings of the eighth international conference on computational semantics*, 116–127.
- Girard, Jean-Yves. 1995. Linear logic: Its syntax and semantics. *London Mathematical Society Lecture Note Series*. 1–42.
- Gotham, Matthew. 2015. Towards glue semantics for Minimalist syntax. *Cambridge Occasional Papers in Linguistics* 8. 56–83.
- Gotham, Matthew. 2018. Making logical form type-logical: Glue semantics for Minimalist syntax. *Linguistics and Philosophy* 41(5). 511–556.
- Gotham, Matthew. 2019. Constraining scope ambiguity in LFG+Glue. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG19 conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2019/index.shtml>.
- Gotham, Matthew. 2021. Approaches to scope islands in LFG+glue. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG21 conference*, 146–166. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2021/lfg21.html>.
- Gotham, Matthew & Dag Trygve Truslew Haug. 2018. Glue semantics for universal dependencies. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG18 conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2018/index.shtml>.

3 Computational Challenges from Theoretical Semantic Modeling in LFG

- Gupta, Vineet & John Lamping. 1998. Efficient linear logic meaning assembly. In *Proceedings of the 17th international conference on computational linguistics – volume 1*, 464–470.
- Halvorsen, Per-Kristian & Ronald M. Kaplan. 1988. Projections and semantic description in Lexical-Functional Grammar. In *Proceedings of the international conference on fifth generation computer systems*, 1116–1122. Also published in Dalrymple, Kaplan, Maxwell & Zaenen (1995), pp. 279–292. Institute for New Generation Computer Technology.
- Haug, Dag Trygve Truslew. 2008. Tense and aspect for glue semantics: The case of participial XADJs. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG08 conference*, 291–311. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2008/lfg08.html>.
- Hepple, Mark. 1996. A compilation-chart method for linear categorial deduction. In *Proceedings of the 16th conference on computational linguistics – volume 1*, 537–542.
- Howard, William A. 1980. The formulae-as-types notion of construction. In J.P. Seldin & J.R. Hindley (eds.), *To H. B. Curry: Essays on combinatory logic, lambda calculus, and formalism*, vol. 44, 479–490.
- Jackendoff, Ray. 2010. The parallel architecture and its place in cognitive science. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 583–605. Oxford University Press.
- Jurafsky, Daniel & James H. Martin. 2024. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models*. 3rd edn. Online manuscript released August 20, 2024; print 2nd edition 2009. <https://web.stanford.edu/~jurafsky/slp3/>.
- Kaplan, Ronald M. 1987. Three seductions of computational psycholinguistics. In Peter Whitelock, Mary McGee Wood, Harold L. Somers, Rod Johnson & Paul Bennett (eds.), *Linguistic theory and computer applications*, 149–188. Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 337–367). Academic Press.
- Kaplan, Ronald M. 1989. The formal architecture of Lexical-Functional Grammar. *Journal of Information Science and Engineering* 5(4). Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 1–27), 305–322.
- Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 173–281. The MIT Press. Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 29–130).
- Kaplan, Ronald M. & Jürgen Wedekind. 1993. Restriction and correspondence-based translation. In *Proceedings of the sixth conference of the European chapter of the Association for Computational Linguistics (EACL)*, 193–202.

- Kaplan, Ronald M. & Jürgen Wedekind. 2023. Formal and computational properties of LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 1035–1082. Language Science Press. DOI: 10.5281/zenodo.10185982. <https://doi.org/10.5281/zenodo.10185982>.
- King, Tracy Holloway. 2016. Theoretical linguistics and grammar engineering as mutually constraining disciplines. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the joint 2016 conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, 339–359. CSLI Publications. DOI: 10.21248/hpsg.2016.18.
- Kiparsky, Paul. 1973. “Elsewhere” in *phonology*. Indiana University Linguistics Club.
- Kokkonidis, Miltiadis. 2006. A glue/ λ -DRT treatment of resumptive pronouns. In *Proceedings of the eleventh ESSLLI student session*, 51–63.
- Kuhn, Jonas. 1998. *Resource sensitivity in the syntax-semantics interface and the German split NP construction*. Universität Stuttgart, Fakultät Philosophie.
- Lev, Iddo. 2007. *Packed computation of exact meaning representations*. Stanford University. (Doctoral dissertation).
- Moot, Richard & Christian Retoré. 2012. *The logic of categorial grammars: A deductive account of natural language syntax and semantics* (Lecture Notes in Computer Science volume 6850). Springer.
- Parsons, Terence. 1990. *Events in the semantics of English*. MIT Press.
- Przepiórkowski, Adam & Agnieszka Patejuk. 2023. Filling gaps with glue. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG23 conference*, 223–240. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/2>.
- Saraswat, Vijay. 1999. LFG as concurrent constraint programming. In Mary Dalrymple (ed.), *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*, 281–318. The MIT Press.
- van Genabith, Josef & Richard Crouch. 1997. How to glue a donkey to an f-structure or porting a dynamic meaning representation language into LFG’s linear logic based glue language semantics. In *International workshop for computational semantics*, 52–65.
- Wedekind, Jürgen & Ronald M. Kaplan. 1993. Type-driven semantic interpretation of f-structures. In *EACL ’93: Proceedings of the sixth conference of the European chapter of the Association for Computational Linguistics*, 404–411. DOI: 10.3115/976744.976791.

Chapter 4

Morphological Exponence and Morphosemantics

Ash Asudeh^a

^aUniversity of Rochester

This paper briefly introduces the relatively recent constraint-based theory of morphosyntax, *Lexical-Realizational Functional Grammar*. L_RFG is a marriage of a constraint-based version of Distributed Morphology (DM), a prominent approach to morphosyntax, with a familiar, constraint-based LFG syntax. Unlike traditional DM, which is bolted onto a derivational, Minimalist syntax, L_RFG is purely constraint-based. The paper also introduces a refinement and revision of the L_RFG theory of exponenda. A feature of the framework is its ability to directly model the morphology-semantics interface, such that semantics can condition exponence, too. The main contribution of the paper is a revision of the mechanism for morphosemantic blocking in L_RFG .

1 Introduction

This paper has three main aims:

1. It briefly introduces the relatively recent constraint-based theory of morphosyntax, *Lexical-Realizational Functional Grammar* (L_RFG), which is a marriage of a constraint-based version of Distributed Morphology (DM), a prominent approach to morphosyntax, with a familiar, constraint-based LFG syntax.¹ Unlike traditional DM, which is bolted onto a derivational, Minimalist syntax, L_RFG is ‘constraints all the way down.’

¹DM is a *lexical-realizational* framework per the classification of Stump (2001). Although we could not resist naming the framework using this terminology, the term *lexical* is potentially confusing, since it has various orthogonal usages. Therefore, I will prefer the term *morphemic* below instead of lexical.

2. It next introduces a novel refinement of the L_R FG theory of exponenda and Vocabulary lookup.
3. It ends with a refinement of the mechanisms for morphosemantic blocking in L_R FG.

The structure of the paper is simple: each item above is dealt with in turn, followed by a conclusion.

2 Background: Lexical-Realizational Functional Grammar

L_R FG combines Distributed Morphology (DM; Halle & Marantz 1993)—a morphemic, realizational approach to morphology—with Lexical-Functional Grammar (LFG; Kaplan & Bresnan 1982)—a constraint-based theory of syntax. Standard DM is traditionally combined with a derivational approach to syntax, the framework known as the Minimalist Program (Chomsky 1995). In contrast to standard DM, then, L_R FG is a version of DM with ‘constraints all the way down’ (Asudeh, Melchin, et al. 2024).

L_R FG is based on a version of LFG’s *Correspondence Architecture* (Kaplan 1987, 1989, Asudeh 2006, 2012). The L_R FG correspondence architecture is shown in Figure 1. C(onstituent)-structure is mapped to f(unctional)-structure by the standard ϕ correspondence function from LFG. Similarly, f-structure is mapped to s(emantic)-structure by the standard σ correspondence feature familiar from work in LFG+Glue (among others, Dalrymple et al. 1993, 2019, Dalrymple 1999, Asudeh 2012) and s-structure is mapped by the ι correspondence function to i(nformation)-structure (Dalrymple & Nikolaeva 2011). This lower path through the architecture is called the *Structure-Function-Meaning Path* (*SFM path* for short). C-structure is also mapped to v(ocabulary)-structure via a new correspondence function, ν (meant to be reminiscent of a lower case ν). V-structure is the site of morphosyntactic *exponence*; the ν function maps from bundles of information (categories and features) in c-structure terminal nodes to representations of the information’s morphophonological exponent. V-structure is in turn mapped by the ρ correspondence function to p(rosodic)-structure (Bögel 2015) and p-structure is mapped by the o correspondence function to the p(honological)-string (o is meant to be reminiscent of *output*). P-structure and the p-string are represented in p-diagrams (Bögel 2015). This upper path through the architecture is called the *Morphology-Prosody-Phonology Path* (*MPP path* for short).

An L_R FG analysis has three main components (Melchin et al. 2020, Asudeh & Siddiqi 2023, Asudeh et al. 2023):

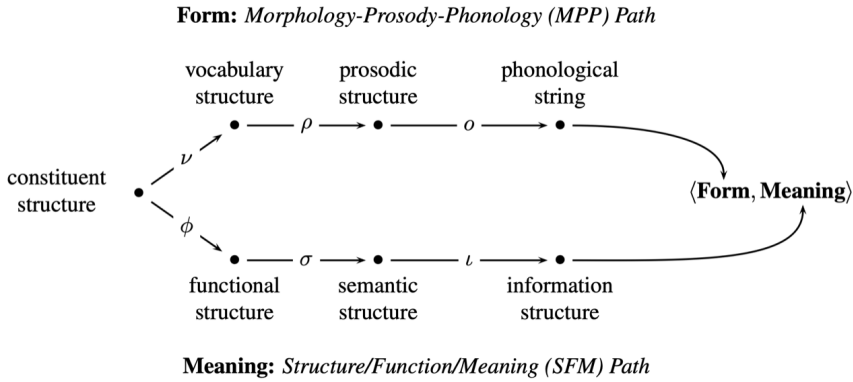


Figure 1: The L_RFG correspondence architecture (Asudeh et al. 2023)

1. A set of annotated c-structure rules, as in standard LFG.
2. A set of *vocabulary items* (VIs), each of which is a mapping from an exp-
 onendum to an exponent. The exponent is mapped to the realization. Thus,
 L_RFG separates exponence and realization as follows (Asudeh et al. 2023):

$$(1) \text{exponendum} \xrightarrow{\nu} \text{exponent} \xrightarrow{o \circ \rho} \text{realization}$$

The set of vocabulary items is called the *Vocabulary*.

3. A set of constraints for choosing the best available VI.
 - a) A family of **MostInformative** constraints which respectively prefer
 a VI based on its c-structural information (**MostInformative_c**), its f-
 structural information (**MostInformative_f**), and its semantic infor-
 mation (**MostInformative_s**)
 - b) A constraint, **MostSpecific**, which prefers a VI based on information
 in its exponent (i.e. its ν -structure)

The main distinction between L_RFG and standard LFG is that, in order to adopt the morphemic and realizational assumptions of DM, L_RFG gives up LFG's Strong Lexicalism (Chomsky 1970, Lapointe 1980, Bresnan et al. 2016). In particular:

1. The terminal nodes of L_RFG c-structure are labelled with a category (as
 in standard LFG) and are paired with a f(unctional)-description (as in stan-
 dard LFG), but contain no information about *form*. This is instead captured
 in L_RFG by *exponence* (in ν -structure) and *realization* (the rest of the MPP

path). In other words, the terminal nodes contain information about the exponendum which must be matched against exponenda in the left-hand side of VIs in the Vocabulary for exponence.

2. The terminal nodes of L_RFG c-structures map to *morphemes* (more precisely, *allomorphs*), not *words*. L_RFG assumes an operation known in the DM literature as *spanning* (Ramchand 2008, Merchant 2015, Svenonius 2016, Haugen & Siddiqi 2016), such that a single morpheme may simultaneously be the exponent of multiple nodes; in other words, the mapping from c-structure to exponence is many-to-one.

Despite this deviation, much of L_RFG 's syntax is similar to LFG's. In particular, the division of syntactic information into a structural component (c-structure) and a functional component (f-structure) is important in L_RFG .

Let's look at a simple example from English:

- (2) Every dog likes most snacks.

The L_RFG c-structure and f-structure for (2) are shown in Figure 2. The exponent v-structures are abbreviated with the English string in a morphemic breakdown. Which c-structure nodes map to which v-structures are shown by arrows marked *v*. Note that L_RFG makes the assumption that there is prosodic inversion of the agreement inflection and the verb stem here. This allows maintenance of a head-initial word order for English, with I preceding VP, but also recognition of the fact that the prosodic word in question is *likes*, not *slike*. See Asudeh et al. (2023) for further details.

There is no equivalent of LFG's lexicon in DM (Marantz 1997), and this is also true of L_RFG , as a variety of DM. If there is no lexicon, the question arises: What populates the c-structure terminals? L_RFG adapts the standard LFG answer: annotated phrase-structure rules license the c-structure terminals. Here are the rules for the c-structure (and its correspondence to f-structure) in Figure 2:

- (3) a. $IP \rightarrow DP \quad I'$
 $(\uparrow \text{SUBJ}) = \downarrow \quad \uparrow = \downarrow$
- b. $I' \rightarrow I \quad VP$
 $\uparrow = \downarrow \quad \uparrow = \downarrow$
 $@AGR!$

- c. VP → vP DP
 $\uparrow=\downarrow$ (\uparrow OBJ) = \downarrow
- d. vP → $\sqrt{\quad}$ v
 $\uparrow=\downarrow$ $\uparrow=\downarrow$
 @ROOT($___$)
- e. DP → D NP
 $\uparrow=\downarrow$ $\uparrow=\downarrow$
 @DET!
- f. NP → #P
 $\uparrow=\downarrow$
- g. #P → nP (#)
 $\uparrow=\downarrow$ $\uparrow=\downarrow$
 @NUM!
- h. nP → $\sqrt{\quad}$ n
 $\uparrow=\downarrow$ $\uparrow=\downarrow$
 @ROOT($___$)

As it happens, these rules all license binary branching trees, but L_RFG is not committed to binary branching. The standard adherence to binary branching in DM is not a feature of that framework per se either, but rather derives from DM’s usual assumption of a Minimalist syntax, which in turn assumes a binary Merge operation (Chomsky 1995). Thus, L_RFG c-structures can in principle be n-ary branching (Everdell et al. 2021), just like in LFG. Also as in LFG, L_RFG allows c-structure nodes to be optional if they are not necessary for other aspects of well-formedness in the architecture.

L_RFG rules make heavy use of LFG’s templates (Dalrymple et al. 2004, Asudeh et al. 2013), which we call *macros* instead (*template* has a different usage in the DM literature, which could potentially be confusing). There are three major kinds of macros in the rules in (3). The *ROOT* macro individuates roots by their *PRED* values, which are uniquely instantiated *semantic forms*, as in LFG. Root individuation is a big topic in DM (see Harley 2014 and replies), because roots are generally assumed to be category-less, such that the root itself does not determine its category (Marantz 1997; also see Borer 2005a,b,c for proposals in *Exo-Skeletal Syntax* which have also been influential in DM). DM assumes that the categorial status of a root is determined by *categorizers*, such as *v* (‘little v’) and *n* (‘little n’) in Figure 2. L_RFG’s solution to root individuation is that *roots are individuated by their PREDs* (Asudeh & Siddiqi, forthcoming). The argument of the *ROOT* macro is underspecified, which means that any of the language’s possible *PRED* values can

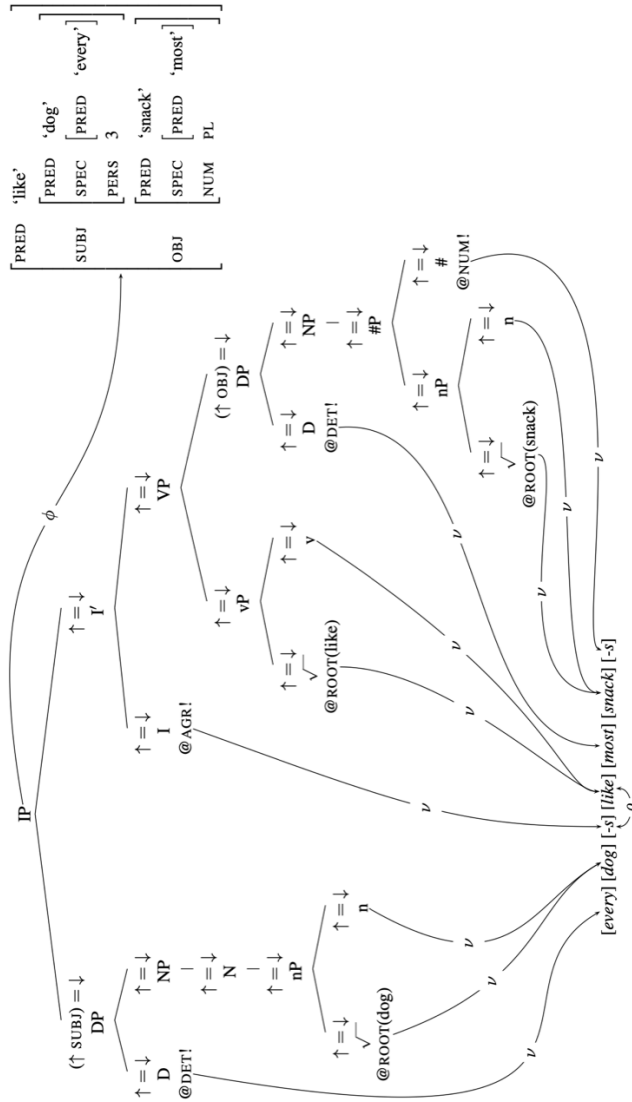


Figure 2: C-structure and f-structure for English *Every dog likes most snacks*

satisfy it. A rule containing this kind of underspecification is in effect a *metarule* (Gazdar et al. 1985) over the possible instantiated versions of the rule. The possible instantiations of PRED values are listed using the LIST macro, which is also how the theory handles arbitrary morphological facts, such as declension and

conjugation classes in languages that feature these. I do not show the LIST macro here, but further details can be found in Asudeh, Bjorkman, et al. (2024).

The second kind of macros featured above are L_{RFG} 's *bang macros*, which enumerate choices. For example, the DET! (pronounced *det-bang*) macro in (3b) lists the possible basic quantificational determiners in English.² The other bang macros are not particularly interesting, because of English's relative morphological impoverishment and its consequent lack of many options for number (NUM!) and agreement (AGR!), but they do present an opportunity for talking about a couple of other aspects of the L_{RFG} framework. The theory makes use of (Paninian) *blocking* and uses underspecification to capture the Elsewhere Condition (see, e.g. Kiparsky 1973). Thus, specification of plural by NUM! as the only option is legitimate because singular number is assumed to be the underspecified, elsewhere case (at least for English). Singular is the *lack* of plural marking and is unmarked. Notice that L_{RFG} uses the terms *marked* and *unmarked* in the traditional morphological way, meaning to be or to not be marked (i.e. have morphophonological content). We do not use 'marked' to mean 'odd,' 'typologically rare,' 'infrequent' or in any of the other ways that the term has been stretched to cover. Similarly, the English AGR macro can be defined to be about just one case, third singular present tense subject agreement, since it is the only case of (regular) agreement marking in the language.^{3,4} The agreement macro in turns calls the 3SG macro. This last macro is the third kind of macro. It does not represent choice (unlike bang macros) or individuate roots, but is rather just a simple packaging of recurrent information. This is the most basic kind of macro. It also provides an opportunity to observe that macros can call other macros.

The macros used in (3) are these:

- (4) a. $ROOT(x) := (\uparrow PRED) = 'x'$
 b. $DET! := (\uparrow SPEC PRED) = \{ 'every' \mid 'some' \mid 'most' \mid \dots \}$
 c. $NUM! := (\uparrow NUM) = PL$

²The complex determiners would be built up around these compositionally. I have specified the DET! macro here as a disjunctive list of choices of PRED values for a SPEC. As mentioned above, this would now be taken care of by a call to the LIST macro.

³Standard DM would likely propose a separate T node for tense, but there is little evidence of this in English, where agreement seems to be more fusional. Nevertheless, the standard DM approach could be accommodated by making the equation for TENSE a *constraining* equation ($=_c$) which requires the value to be specified by something else such that agreement only requires its presence but does not itself *define* the f-structure as having present tense.

⁴There are the agreement patterns of the auxiliaries *have* and *be* to account for, too. I leave these aside here, but the theory is fully capable of accommodating these.

d. $\text{AGR!} := @3\text{SG}$

e. $3\text{SG} := (\uparrow \text{SUBJ PERS}) = 3$
 $(\uparrow \text{SUBJ NUM}) = \text{SG}$

The semantics of macro invocation is very simple and identical to how it works in LFG (Dalrymple et al. 2004): each @ call of a macro substitutes the macro's definition (on the right-hand side of the := symbols above) at the invocation point.

The last element that we need to complete this illustrative analysis is the relevant subset of English's Vocabulary:

- (5) a. $\langle [\text{D}], (\uparrow \text{PRED}) = \text{'every'} \rangle \xrightarrow{v} [\text{every}]$
 $(\text{SPEC } \uparrow)$
 $\lambda R \lambda S. \text{every}(\text{R})(\text{S}) :$
 $[[(\text{SPEC } \uparrow)_\sigma \text{VAR}] \multimap ((\text{SPEC } \uparrow)_\sigma \text{RESTR})] \multimap$
 $\forall X. [(\text{SPEC } \uparrow)_\sigma \multimap X] \multimap X$
- b. $\langle [\sqrt{\quad}], (\uparrow \text{PRED}) = \text{'dog'} \rangle \xrightarrow{v} [\text{dog}]$
 $\text{dog} : (\uparrow_\sigma \text{VAR}) \multimap (\uparrow_\sigma \text{RESTR})$
- c. $\langle [\sqrt{\quad}], (\uparrow \text{PRED}) = \text{'like'} \rangle \xrightarrow{v} [\text{like}]$
 $\text{like} : (\uparrow \text{OBJ})_\sigma \multimap (\uparrow \text{SUBJ})_\sigma \multimap \uparrow_\sigma$
- d. $\langle [\text{Agr}], (\uparrow \text{SUBJ PERS}) = 3 \rangle \xrightarrow{v} [-s]$
 $(\uparrow \text{SUBJ NUM}) = \text{SG}$
- e. $\langle [\text{D}], (\uparrow \text{PRED}) = \text{'most'} \rangle \xrightarrow{v} [\text{most}]$
 $(\text{SPEC } \uparrow)$
 $\lambda R \lambda S. \text{most}(\text{R})(\text{S}) :$
 $[[(\text{SPEC } \uparrow)_\sigma \text{VAR}] \multimap ((\text{SPEC } \uparrow)_\sigma \text{RESTR})] \multimap$
 $\forall X. [(\text{SPEC } \uparrow)_\sigma \multimap X] \multimap X$
- f. $\langle [\sqrt{\quad}], (\uparrow \text{PRED}) = \text{'snack'} \rangle \xrightarrow{v} [\text{snack}]$
 $\text{snack} : (\uparrow_\sigma \text{VAR}) \multimap (\uparrow_\sigma \text{RESTR})$
- g. $\langle [\#], (\uparrow \text{NUM}) = \text{PL} \rangle \xrightarrow{v} [-s]$
 $\text{PI} : [(\uparrow_\sigma \text{VAR}) \multimap (\uparrow_\sigma \text{RESTR})] \multimap [(\uparrow_\sigma \text{VAR}) \multimap$
 $(\uparrow_\sigma \text{RESTR})]$

The general form of an L_RFG vocabulary item (VI) is as follows (Asudeh et al. 2023):

$$(6) \left\langle \begin{array}{l} [C_1, \dots, C_n] \\ \text{distribution} \end{array}, \begin{array}{l} F \cup G \cup I \\ \text{function/meaning} \end{array} \right\rangle \xrightarrow{v} \left[\quad \right]_{v\text{-structure}}$$

A vocabulary item is a mapping from an exponendum, the left-hand side, to an exponent, the right-hand side (again, these are abbreviated here). The exponendum consists of two parts:

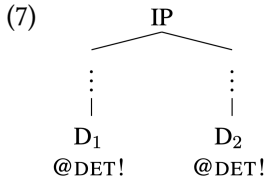
1. A *list* of categories, minimally (and typically) of length one. However, an item may be listed in the Vocabulary as *spanning* more than one category, so the list is allowed to be longer in principle.
2. A set *F* of f-structural information (an f-description), a set *G* of meaning information, modelled as a set of Glue meaning constructors, and a set *I* of i-structural information. These are presented in a union of sets which we call a *fugui* (pronounced “foo-gooley”), based on the resemblance of the set-theoretic union term, $F \cup G \cup I$, to this nonce word of English.

The observant reader may have noticed a couple of things by this point. First, VIs can have content that is not in the c-structure terminal nodes. For example, all the VIs in (5) have Glue meaning constructors (except the agreement morpheme, since I am leaving tense aside in the semantics, and agreement itself is not contentful).⁵ Second, something slightly mysterious seems to be going on, with the *v*-mapping being used to indicate the mapping from c-structure to exponents and the mapping in the Vocabulary from exponenda to exponents. The first point allows direct modelling of *morphosemantics*; i.e. the interface between morphology and semantics (Asudeh & Siddiqi 2022). Clarifying the second point is not only important in its own right, it also forces a slight reconsideration of how morphosemantics works in the framework. These are the topics of the next sections. I turn first to the issue of mismatch between the targets of exponence in the c-structure and the exponenda in the Vocabulary.

3 Exponenda and Vocabulary lookup

Let us look at the nodes labelled D in Figure 2. The content of these nodes is repeated here:

⁵One could just add a **Pres** tense operator to (5d), if one were to want to add tense to the semantics, given the fusional nature of the morpheme.



I have arbitrarily identified these as nodes 1 and 2 in the c-structure.

As part of the contentful grammatical vocabulary of the language, determiners have PREDs, but they are not roots, since there is no evidence that determiners are labile in the way roots are. Thus, they are not associated with the @ROOT macro by the rule (3e) but rather with the @DET! macro. This macro was defined as follows above, repeated here:

$$(4b) \text{ DET!} := (\uparrow \text{ SPEC PRED}) = \{ 'every' \mid 'some' \mid 'most' \mid \dots \}$$

The vocabulary items for *every* and *most* are repeated here:

$$\begin{array}{l}
 (5) \quad \text{a.} \quad \langle [D], \quad (\uparrow \text{ PRED}) = 'every' \quad \rangle \xrightarrow{v} [every] \\
 \quad \quad \quad (\text{SPEC } \uparrow) \\
 \quad \quad \quad \lambda R \lambda S.\text{every}(R)(S) : \\
 \quad \quad \quad [((\text{SPEC } \uparrow)_\sigma \text{ VAR}) \rightarrow ((\text{SPEC } \uparrow)_\sigma \text{ RESTR})] \rightarrow \\
 \quad \quad \quad \forall X. [(\text{SPEC } \uparrow)_\sigma \rightarrow X] \rightarrow X \\
 \quad \quad \quad \text{e.} \quad \langle [D], \quad (\uparrow \text{ PRED}) = 'most' \quad \rangle \xrightarrow{v} [most] \\
 \quad \quad \quad (\text{SPEC } \uparrow) \\
 \quad \quad \quad \lambda R \lambda S.\text{most}(R)(S) : \\
 \quad \quad \quad [((\text{SPEC } \uparrow)_\sigma \text{ VAR}) \rightarrow ((\text{SPEC } \uparrow)_\sigma \text{ RESTR})] \rightarrow \\
 \quad \quad \quad \forall X. [(\text{SPEC } \uparrow)_\sigma \rightarrow X] \rightarrow X
 \end{array}$$

The question is, how do we map the information in the nodes to the correct instance of a determiner VI in each case?

In earlier versions of L_RFG (e.g. Melchin et al. 2020, Asudeh & Siddiqi 2022, 2023), we postulated a *bridging function*, Φ ('big phi') between v-structure and f-structure so that v-structure could be sensitive to f-structural content. However, in recent work we have dispensed with this usage (Asudeh et al. 2023). Here I propose to bring big phi back as part of Vocabulary lookup, i.e. how information in c-structural terminal nodes is matched against potential vocabulary items.

Let us propose a matching function, μ , which we can define as follows:

$$(8) \text{ For all nodes } n \in T, \text{ where } T \text{ is the set of terminal nodes in a c-structure,}$$

$$\mu(n) = \{ \langle x, y \rangle \mid x = \lambda(n) \wedge \phi(n) = y \}$$

In other words, μ returns a pair consisting of a category—the label of μ 's argument, n , per LFG's usual labelling function, λ (Kaplan 1989)—and an f-structure constructed by applying LFG's standard c-structure/f-structure correspondence function, Φ , to n . This f-structure is further populated and constrained by the f-structural information associated with n . In this case, this is a disjunction defined by DET!. However, this is the same disjunction in both instances of D in Figure 2, as emphasized in (7) above. Thus, in both cases, the same set is delivered by μ :

$$(9) \quad \mu(1) = \mu(2) = \left\{ \begin{array}{l} \left\langle D, \left[\text{SPEC} \left[\text{PRED} \text{ 'every'} \right] \right] \right\rangle \\ \left\langle D, \left[\text{SPEC} \left[\text{PRED} \text{ 'some'} \right] \right] \right\rangle \\ \left\langle D, \left[\text{SPEC} \left[\text{PRED} \text{ 'most'} \right] \right] \right\rangle \\ \dots \end{array} \right\}$$

Note that μ is in essence just defining how *lexical insertion* works (but where the Vocabulary is an inert 'listicon' that does not contain any operations of its own, such as lexical rules). There is no substantive difference here between using just the category for choosing suitable lexical items, as in standard DM or LFG. Standard LFG treats the terminal node as the lexical item for *every* or *most*, which includes a matching category for lexical insertion (given the c-structure rule), an f-description, a meaning constructor, and a form. L_RFG treats the terminal node as (the same) bundle of information *except form*, such that the information is mapped to an exponent that specifies further constraints on the realization of the form.

In order to match the f-structure in the second coordinates of (9), we now use big phi, Φ . We define Φ as follows:

$$(10) \quad \text{Given some vocabulary item, } v_i, \text{ such that } v_i = \langle \langle \dots, F \cup G \cup I \rangle, \dots \rangle, \\ \Phi(v_i) \text{ is the minimal f-structure that satisfies } F.$$

Given the ν -mapping in (6), the Vocabulary is a set of vocabulary items such that each member of the set is a pair consisting of an exponendum, the left-hand side of a vocabulary item, and an exponent, the right-hand side of the vocabulary item; i.e. when we treat VIs set-theoretically, they are pairs consisting of an exponendum, which is itself a pair, and an exponent. In short, Φ does the same thing as ϕ , but whereas ϕ takes a c-structure node as an argument and returns the minimal f-structure satisfying its associated f-description, Φ takes a vocabulary

item as an argument and returns the minimal f-structure satisfying the F in the vocabulary item's fugui.⁶

The output of Φ applied to each of the nodes labelled D in (7), i.e. node 1 and node 2, is such that they can respectively correspond to the following f-structures, arbitrarily labelled as f and g :

- (11) a. $f \left[\text{SPEC} \left[\text{PRED} \text{ 'every'} \right] \right]$
 b. $g \left[\text{SPEC} \left[\text{PRED} \text{ 'most'} \right] \right]$

The **MostInformative_f** constraint (**MostInformative** in f-structural terms) can now be defined as follows:

(12) **MostInformative_f**

Given a c-structure node, n , that is the target of exponence and two Vocabulary Items, α and β ,

$$\text{MostInformative}_f(\alpha, \beta) =$$

$$\begin{cases} \alpha \text{ if } \exists t \exists f \forall g. f = \Phi(\alpha) \wedge g = \Phi(\beta) \wedge t \in \phi(n) \wedge f, g \sqsubseteq t \wedge g \rightarrow f \\ \beta \text{ if } \exists t \exists f \forall g. f = \Phi(\beta) \wedge g = \Phi(\alpha) \wedge t \in \phi(n) \wedge f, g \sqsubseteq t \wedge g \rightarrow f \\ \perp \text{ otherwise} \end{cases}$$

The constraint returns one of three cases. In the first two cases, we check whether there are two VIs that each subsume the f-structure of the target node (f-structure t). This allows the exponenda in the VIs to be more general than the one they are trying to match (capturing DM's Subset Principle) but does not allow them to be inconsistent with the target's f-structure; this is modelled by subsumption on f-structures (\sqsubseteq ; Bresnan et al. 2016). The last term in each of the first two cases checks which of the two VI's f-descriptions creates an f-structure that contains strictly more information than the other; this is modelled by proper subsumption on f-structures. In other words, **MostInformative_f** checks whether there is a vocabulary item whose corresponding f-structure subsumes the target f-structure and which is most informative with respect to other candidate VIs for the target. It returns whichever VI is most informative with respect to f-structure. If there are no such candidates or if more than one candidate fits the bill, the constraint instead 'borks' and returns \perp : there is no *most* informative VI.

⁶Recall that *fugui* is L_RFG 's nonce term for the second coordinate of the left-hand side of a vocabulary item, $F \cup G \cup I$; see (6) above.

In the case of the f-structures in (11), neither subsumes (never mind *properly* subsumes) the other, because they have distinct PREDs. Therefore, **MostInformative_f** chooses neither over the other and borks. It is important to note that borking means the constraint is indecisive, not violated. Thus, one available instantiation of the relevant c-structure is the one in Figure 2; thus the relevant sentence gets a successful parse and is deemed grammatical.

In contrast to our previous approach, though, the Glue meaning constructors do not populate the terminal node, but rather just come along for the ride with the VIs that match the c-structural information (category) and f-structural information (the minimal f-structure satisfying the f-description) on the target c-structure node. In effect, we have stretched the function cascade in (1) as follows:

$$(1') \quad \underbrace{\text{target of exponence}}_{c\text{-structure}} \xrightarrow{\mu} \underbrace{\text{exponendum}}_{\text{Vocabulary}} \xrightarrow{\nu} \underbrace{\text{exponent}}_{\text{form}} \xrightarrow{\sigma \circ \rho} \underbrace{\text{realization}}_{\text{form}}$$

This in turn means that the ν -annotated arrows in L_RFG diagrams like Figure 2 should really be labelled $\nu \circ \mu$. However, to avoid clutter, we can take the μ mapping as read and continue to label the arrows with just ν .

As I mentioned above, the Glue meaning constructors in this approach do not populate terminal nodes in c-structure, the targets of exponence, but are rather associated directly with exponenda in the Vocabulary. The meaning constructors, once instantiated to the arbitrary mnemonic labels added to the f-structure in Figure 2, as in (13), can construct the usual two possible Glue proofs (one corresponding to surface scope, *every* > *most*, and the other to inverse scope, *most* > *every*); the two possibilities are shown in Figure 3.

$$(13) \quad \left[\begin{array}{l} \text{PRED} \quad \text{'like'} \\ \\ \text{SUBJ} \quad d \left[\begin{array}{l} \text{PRED} \quad \text{'dog'} \\ \text{SPEC} \quad \left[\begin{array}{l} \text{PRED} \quad \text{'every'} \end{array} \right] \\ \text{PERS} \quad 3 \end{array} \right] \\ \\ \text{OBJ} \quad s \left[\begin{array}{l} \text{PRED} \quad \text{'snack'} \\ \text{SPEC} \quad \left[\begin{array}{l} \text{PRED} \quad \text{'most'} \end{array} \right] \\ \text{NUM} \quad \text{PL} \end{array} \right] \end{array} \right]$$

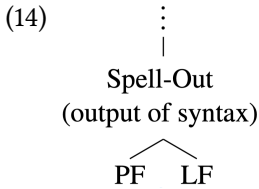
Moreover, as in standard Glue semantics, this is achieved *without* an otherwise unmotivated syntactic ambiguity to derive the distinct interpretations (Asudeh

$$\begin{array}{c}
\text{like} : \\
\frac{s_\sigma \rightarrow d_\sigma \rightarrow l_\sigma}{\lambda y_1 \lambda x_1 \text{like}(y_1)(x_1) : \Rightarrow \eta} \\
\frac{s_\sigma \rightarrow d_\sigma \rightarrow l_\sigma}{\lambda x_1 \text{like}(y)(x_1) : d_\sigma \rightarrow l_\sigma} \frac{[y : s_\sigma]^1}{\text{like}(y)(x) : l_\sigma} \frac{[x : d_\sigma]^2}{\rightarrow \varepsilon, \Rightarrow \beta} \\
\text{snack} : \quad \text{PI} : \\
\frac{(v_2 \rightarrow r_2) \rightarrow \forall X.(s_\sigma \rightarrow X) \rightarrow X}{\lambda R \lambda S \text{most}(R)(S) :} \quad \frac{(v_2 \rightarrow r_2) \rightarrow \forall X.(d_\sigma \rightarrow X) \rightarrow X}{\lambda Y \text{PI}(\text{snack}) :} \\
\frac{\lambda S \text{most}(\text{PI}(\text{snack}))(S) :}{\forall X.(s_\sigma \rightarrow X) \rightarrow X} \quad \frac{\lambda Y \text{PI}(\text{snack}))(Y) : l_\sigma}{\lambda Y \text{like}(y)(x) :} \frac{\rightarrow_{I.1}}{s_\sigma \rightarrow l_\sigma} \frac{\rightarrow_{\varepsilon, \forall \mathcal{E}[l_\sigma/X], \Rightarrow \beta}}{\text{most}(\text{PI}(\text{snack}))(\lambda y \text{like}(y)(x)) : l_\sigma} \\
\frac{\lambda R \lambda S \text{every}(R)(S) :}{(v_1 \rightarrow r_1) \rightarrow \forall X.(d_\sigma \rightarrow X) \rightarrow X} \quad \frac{\lambda X \text{every}(\text{dog})(S) :}{\forall X.(d_\sigma \rightarrow X) \rightarrow X} \quad \frac{\lambda x \text{most}(\text{PI}(\text{snack}))(\lambda y \text{like}(y)(x)) : l_\sigma}{d_\sigma \rightarrow l_\sigma} \frac{\rightarrow_{I.2}}{\lambda x \text{most}(\text{PI}(\text{snack}))(\lambda y \text{like}(y)(x)) :} \frac{\rightarrow_{\varepsilon, \forall \mathcal{E}[l_\sigma/X], \Rightarrow \alpha}}{\text{every}(\text{dog})(\lambda x \text{most}(\text{PI}(\text{snack}))(\lambda y \text{like}(y)(x))) : l_\sigma} \\
\text{every}(\text{dog})(\lambda x \text{most}(\text{PI}(\text{snack}))(\lambda y \text{like}(y)(x))) : l_\sigma \\
\text{like} : \\
\frac{s_\sigma \rightarrow d_\sigma \rightarrow l_\sigma}{\lambda y_1 \lambda x \text{like}(y_1)(x) : \Rightarrow \eta} \\
\frac{s_\sigma \rightarrow d_\sigma \rightarrow l_\sigma}{\lambda x \text{every}(\text{dog})(S) :} \quad \frac{\lambda x \text{every}(\text{dog})(S) :}{\forall X.(d_\sigma \rightarrow X) \rightarrow X} \quad \frac{\lambda x \text{every}(\text{dog})(S) :}{\forall X.(d_\sigma \rightarrow X) \rightarrow X} \quad \frac{\lambda x \text{every}(\text{dog})(S) :}{\forall X.(d_\sigma \rightarrow X) \rightarrow X} \\
\text{snack} : \quad \text{PI} : \\
\frac{(v_2 \rightarrow r_2) \rightarrow \forall X.(s_\sigma \rightarrow X) \rightarrow X}{\lambda R \lambda S \text{most}(R)(S) :} \quad \frac{(v_2 \rightarrow r_2) \rightarrow \forall X.(d_\sigma \rightarrow X) \rightarrow X}{\lambda Y \text{PI}(\text{snack}) :} \\
\frac{\lambda S \text{most}(\text{PI}(\text{snack}))(S) :}{\forall X.(s_\sigma \rightarrow X) \rightarrow X} \quad \frac{\lambda Y \text{PI}(\text{snack}))(Y) : l_\sigma}{\lambda Y \text{like}(y)(x) :} \frac{\rightarrow_{I.1}}{s_\sigma \rightarrow l_\sigma} \frac{\rightarrow_{\varepsilon, \forall \mathcal{E}[l_\sigma/X], \Rightarrow \beta}}{\text{most}(\text{PI}(\text{snack}))(\lambda y \text{every}(\text{dog})(\lambda x \text{like}(y)(x))) : l_\sigma} \\
\text{most}(\text{PI}(\text{snack}))(\lambda y \text{every}(\text{dog})(\lambda x \text{like}(y)(x))) : l_\sigma \\
\text{every}(\text{dog})(\lambda x \text{every}(\text{dog})(\lambda x \text{like}(y)(x))) : l_\sigma
\end{array}$$

Figure 3: Top: Surface scope (*every* > *most*). Bottom: Inverse scope (*most* > *every*)

2022). Thus, $L_RFG+Glue$ functions just like $LFG+Glue$ and prior results from the Glue literature remain intact.

It is important to note, though, that this also continues to allow L_RFG to have a sensible notion of *morphosemantics*, i.e. a direct morphology-semantics interface, because exponence in the Vocabulary is a function on a bundle of information that includes semantics. Thus, meaning can condition exponence. This is not really possible in the classic ‘Y model’ that DM inherits from Minimalism, because Phonological Form (PF) and Logical Form (LF) merely interpret the output of syntax and PF is never conditioned on LF directly:



I next turn to a reconsideration of morphosemantics in L_RFG in light of the moves made above.

4 A refinement of morphosemantics and semantic blocking in L_RFG

Let us define $\mathbf{MostInformative}_s$, the semantic version of $\mathbf{MostInformative}$, as in (15). This is a slight revision of the definition in Asudeh & Siddiqi (2022).

- (15) $\mathbf{MostInformative}_s$ (to be revised)
 Given two Vocabulary Items, α and β , such that $\alpha = \langle \langle \dots, F_1 \cup G_1 \cup I_1 \rangle, \dots \rangle$
 and $\beta = \langle \langle \dots, F_1 \cup G_2 \cup I_1 \rangle, \dots \rangle$,

$$\mathbf{MostInformative}_s(\alpha, \beta) = \begin{cases} \alpha & \text{if } \llbracket G_1 \rrbracket \subset \llbracket G_2 \rrbracket \\ \beta & \text{if } \llbracket G_2 \rrbracket \subset \llbracket G_1 \rrbracket \\ \perp & \text{otherwise} \end{cases}$$

According to this version of the principle, semantic informativeness is captured as entailment in the model. If VI α 's meaning (proof conclusion) entails β 's meaning, then choose α . If the reverse holds, choose β . Otherwise, $\mathbf{MostInformative}_s$ borks and the constraint is indecisive.

Even in this preliminary version, $\mathbf{MostInformative}_s$ solves a puzzle for DM. Namely, if two VIs are such that one is semantically contentless (its G set is

empty) and the other is not, but they are in competition as the exponenda of a single c-structural locus, **MostInformative**_s will always prefer the contentful option. A relevant case occurs in English, where third singular agreement loses out to tense marking. In other words, assuming that the past tense marker [-d] and the third singular marker [-s] are vying to be exponents of a single c-structure node, it is potentially a puzzle for standard DM as to why we get (16) as the expression of past tense in English rather than (17).

(16) Max yawned. = past

(17) Max yawns. ≠ past

This is a puzzle in standard DM because exponence is not conditioned by meaning, as discussed above. However the past tense morpheme [-d] in English states information about tense, which is contentful, but third singular agreement does not add any content of its own. Therefore, **MostInformative**_s will always pick [-d] to mark the past tense rather than [-s]. Another example is cases where an exponendum for number is in competition with an exponendum for gender. Grammatical gender is not contentful, but grammatical number is (at least for plurals). Therefore, the VI expressing number will be preferred over the VI expressing gender.

A deeper consideration of **MostInformative**_s concerns competition between two contentful VIs. That is, what if we compare two VIs that share a category and make no f-structural or i-structural distinctions from each other, but where the meaning computed for one VI entails the other? Such a case is provided by the inchoative/causative alternation in English (see also, Asudeh 2021), as in:

(18) On special occasions, the vase sits on the shelf. inchoative

(19) On special occasions, Max sets the vase on the shelf. causative

I have used the verbs *sit/set* here purposefully, because some DM analyses treat this as a case of *suppletion* with an external causativizing functional head. However, Asudeh, Melchin, et al. (2024) has argued against this sort of analysis.

Reasonable VIs for *sit* and *set* (recall that the present tense inflection is handled by Agr) are shown in (20) and (21). In this case, some event structure is useful in the meaning language, so I have adopted a version of event semantics:

$$(20) \langle [\sqrt{\quad}], (\uparrow \text{PRED}) = \text{'sit'} \rangle \xrightarrow{v} [\textit{sit}]$$

$$\lambda y \lambda e. \textit{sit}(e) \wedge \textit{patient}(e) = y :$$

$$(\uparrow \text{SUBJ})_{\sigma} \multimap (\uparrow_{\sigma} \text{EVENT}) \multimap \uparrow_{\sigma}$$

$$(21) \quad \langle [\sqrt{\quad}], \quad (\uparrow \text{PRED}) = \text{'sit'} \quad \rangle \xrightarrow{v} [\text{set}]$$

$$\lambda y \lambda x \lambda e. \text{sit}(e) \wedge \text{patient}(e) = y \wedge \text{agent}(e) = x :$$

$$(\uparrow \text{OBJ})_{\sigma} \multimap (\uparrow \text{SUBJ})_{\sigma} \multimap (\uparrow_{\sigma} \text{EVENT}) \multimap \uparrow_{\sigma}$$

Observe that the following entailment holds (by the standard logical rule of conjunction simplification):

$$(22) \quad \frac{\text{sit}(e) \wedge \text{patient}(e) = y \wedge \text{agent}(e) = x}{\text{sit}(e) \wedge \text{patient}(e) = y} \text{CS}$$

The linear logic terms are irrelevant to this entailment pattern. Nevertheless, there is another factor at play here. All else being equal, **MostInformative_s** should *always* prefer $[\text{set}]$ over $[\text{sit}]$. Moreover, the model-theoretic version of **MostInformative_s** in (15) potentially makes strange predictions about lexical entailments, unless further adjustments are made.⁷ For example, we would not want *apple* to block *fruit* simply because $[\text{apple}]_c \subset [\text{fruit}]$.

As far as the grammar is concerned, this is true, but grammar is not the only consideration in forming a linguistic expression. The speaker's intentions also matter. That is, presumably the speaker has some conceptual *message* in mind (Levelt 1989) that they wish to encode linguistically. The message is not part of the grammar per se, but the speaker must pick the right VIs to encode the intended message grammatically. Thus, if the speaker has a message about someone causing something to happen, they would presumably pick the causative VI. If they wish to leave the causer out, they will presumably pick the inchoative VI. In sum, the relationship between the message and the grammar is not a purely linguistic one: it involves the interface between thought and grammar. The hearer's job is much easier: they actually hear the speaker's expression and the job of the grammar in this direction is to pull up the right VIs to parse the expression. If the arguments present in the expression are such that only the inchoative makes sense (i.e. there is no causer mentioned in the expression and the subject is a patient), then the hearer must simply pick the right VI to accomplish the task. However this in turn makes it seem that **MostInformative_s** is simply switched off in parsing. This is undesirable, because we do not wish to postulate different grammars for production and parsing. In other words, we want grammar, as a knowledge system, to be *process neutral* (Pollard & Sag 1994).

Put simply, there has to be some aspect of the grammar that is sensitive enough to these conceptual differences to individuate them grammatically. The natural

⁷I thank one of the reviewers for their discussion of this point.

place to look in LFG, and thus $L_{\mathcal{R}}FG$, is the PRED feature, which represents syntactic predication in f-structure. Thus, I make the representational choice to distinguish the two *sit* values as *sit_{inch}* and *sit_{caus}*. The revised entries now look like this:

$$(23) \quad \langle [\sqrt{\quad}] \quad (\uparrow \text{PRED}) = 'sit_{inch}' \quad \rangle \xrightarrow{v} [sit]$$

$$\lambda y \lambda e. \mathbf{sit}(e) \wedge \mathbf{patient}(e) = y :$$

$$(\uparrow \text{SUBJ})_{\sigma} \multimap (\uparrow_{\sigma} \text{EVENT}) \multimap \uparrow_{\sigma}$$

$$(24) \quad \langle [\sqrt{\quad}] \quad (\uparrow \text{PRED}) = 'sit_{caus}' \quad \rangle \xrightarrow{v} [set]$$

$$\lambda y \lambda x \lambda e. \mathbf{sit}(e) \wedge \mathbf{patient}(e) = y \wedge \mathbf{agent}(e) = x :$$

$$(\uparrow \text{OBJ})_{\sigma} \multimap (\uparrow \text{SUBJ})_{\sigma} \multimap (\uparrow_{\sigma} \text{EVENT}) \multimap \uparrow_{\sigma}$$

We have distinguished these two VIs in terms of their PRED values, but the version of **MostInformative_s** in (15) is not sensitive to this distinction and the lexical entailment problem has still not been dealt with.

Given the assumption that PRED values will be the hook into the conceptual or message space, **MostInformative_s** must be reformulated to take this into account. Secondly, in order to avoid the lexical entailment problem, I recast the principle in terms of proof entailment in the meaning language:⁸

(25) **MostInformative_s** (revised)

Given two vocabulary items, α and β , such that

1. α 's PRED FN = β 's PRED FN
2. $\alpha = \langle \langle \dots, F_1 \cup G_1 \cup I_1 \rangle, \dots \rangle$
3. $\beta = \langle \langle \dots, F_1 \cup G_2 \cup I_1 \rangle, \dots \rangle$

$$\mathbf{MostInformative}_s(\alpha, \beta) = \begin{cases} \alpha & \text{if } \Gamma : G_1 \vdash_{\mathcal{M}} \Gamma' : G_2 \\ \beta & \text{if } \Gamma' : G_2 \vdash_{\mathcal{M}} \Gamma : G_1 \\ \perp & \text{otherwise} \end{cases}$$

The notation $\vdash_{\mathcal{M}}$ designates syntactic entailment in the meaning language. Thus, basic items like *apple* and *fruit* are no longer in competition, because neither's set of meaning terms entails the other's. Nevertheless, we preserve the observations about the past tense marker vs. the agreement marker, because the latter, if a contentless element, forms the bottom of the entailment lattice.

If we are considering entailments on the meaning side between VIs that satisfy the [PRED FN] requirement in (25), there are three basic cases to consider:

⁸It is necessary to use PRED FN (Crouch et al. 2011, Asudeh et al. 2013), because we care about the *type* of PRED concerned, not uniquely instantiated tokens designated by semantic forms.

1. $p \wedge q : g \vdash_{\mathcal{M}} p : g$
2. $p : g \vdash_{\mathcal{M}} p \vee q : g$
3. $p : g \vdash_{\mathcal{M}} p \vee (p \wedge q) : g$

Case 1 is the case where two VIs that satisfy the requirements are such that one VI's meaning term is a conjunction that includes the other's meaning term; in this case, the first VI's meaning term entails the second VI's and **MostInformative_s** chooses the first VI. Intuitively, the first VI contains more semantic information than the second VI. Case 2 is the case where two VIs that satisfy the requirements are such that one VI's meaning term is included in a disjunction that is the other VI's meaning term; in this case, the first VI's meaning term entails the second VI's and **MostInformative_s** chooses the first VI. Intuitively, a disjunction contains less information than either of its disjuncts. For example, the meaning constructor for *divinity* in Asudeh & Siddiqi (2022) expresses **holy** (or something like that), whereas *divineness* expresses **holy** or **good**; *divinity*'s meaning term thus entails *divineness*'s and *divinity* is correctly preferred to *divineness* by **MostInformative_s** when specifically expressing **holy**. Case 3 is the case where two VIs that satisfy the requirements are such that one VI's meaning term contains an obligatory meaning constructor and an optional meaning constructor. This again boils down to a disjunction, where the VI in question can express either just the obligatory meaning constructor or the conjunction of the obligatory meaning constructor and the optional one. Thus, as in the second case, the VI that just expresses only the particular information, i.e. just the obligatory meaning constructor, is chosen by **MostInformative_s**.

With respect to the *sit/set* issue, the revised principle's first condition—that the two competing VIs must have the same [PRED FN] values—means that inchoative *sit* and causative *set* are not in competition. However, forms like *divinity* and *divineness* (Asudeh & Siddiqi 2022) remain in competition, because they do have the same [PRED FN] (divine). Furthermore, it gives us an opportunity to form generalizations about what are now effectively subtypes of PREDs (e.g. '*sit_{inch}*' vs. '*sit_{caus}*') in terms of containment relationships captured by entailment. If we package the related information into macros, the generalizations start to naturally emerge:

$$(26) \text{ a. } \langle [\sqrt{\quad}] \quad (\uparrow \text{PRED}) = \text{'sit}_{inch}' \quad \rangle \xrightarrow{v} [\textit{sit}]$$

$$\lambda e.\textit{sit}(e) : (\uparrow_{\sigma} \text{EVENT}) \multimap \uparrow_{\sigma}$$

$$\text{@INCHOATIVE}$$

$$\begin{array}{l}
 \text{b. } \langle [\sqrt{\quad}] \quad (\uparrow \text{ PRED}) = \text{'sit}_{\text{caus}}' \quad \rangle \xrightarrow{v} [\text{set}] \\
 \lambda e. \text{sit}(e) : (\uparrow_{\sigma} \text{ EVENT}) \multimap \uparrow_{\sigma} \\
 @\text{INCHOATIVE} \\
 @\text{CAUSATIVE}
 \end{array}$$

$$\begin{array}{l}
 (27) \text{ a. } \text{INCHOATIVE} \quad \lambda P \lambda y \lambda e'. P(e') \wedge \text{patient}(e') = y : \\
 [(\uparrow_{\sigma} \text{ EVENT}) \multimap \uparrow_{\sigma}] \multimap \\
 (\uparrow \text{ SUBJ})_{\sigma} \multimap (\uparrow_{\sigma} \text{ EVENT}) \multimap \uparrow_{\sigma}
 \end{array}$$

$$\begin{array}{l}
 \text{b. } \text{CAUSATIVE} \quad \lambda P \lambda y \lambda x \lambda e'. P(y)(e') \wedge \text{patient}(e') = y \wedge \text{agent}(e') = x : \\
 [(\uparrow \text{ SUBJ})_{\sigma} \multimap (\uparrow_{\sigma} \text{ EVENT}) \multimap \uparrow_{\sigma}] \multimap \\
 (\uparrow \text{ OBJ})_{\sigma} \multimap (\uparrow \text{ SUBJ})_{\sigma} \multimap (\uparrow_{\sigma} \text{ EVENT}) \multimap \uparrow_{\sigma}
 \end{array}$$

The relations between different subtypes of PRED are thus captured in terms of containment relations between argument-structure macros (for further details of this approach to argument structure, see Asudeh & Giorgolo 2012). Nevertheless, as distinct concepts, sit_{inch} and sit_{caus} are not in an exponence competition with each other. It just depends on what you mean to say.

5 Conclusion

In §2, I provided a capsule introduction to $L_{\text{R}}\text{FG}$, a fusion of DM and LFG that provides a morphemic, realizational morphology (as in DM) for an LFG-like syntax and grammatical architecture. In §3, I further developed $L_{\text{R}}\text{FG}$ theory of c-structure targets for exponents (we call them *c-structure exponenda* or *c-exponenda*) matching with exponenda in the Vocabulary (we could call them *v-exponenda*). These are mapped to exponents in v-structure in the standard way from prior $L_{\text{R}}\text{FG}$ work. Lastly, in §4 I refined $L_{\text{R}}\text{FG}$'s theory of morphosemantics and morphosemantic competition. Much work remains to be done, but I hope this has served as an effective introduction to the possibilities of the theory.

Acknowledgments

I was lucky enough to work under Dick's supervision on the computational semantics project as part of XLE development at Xerox PARC. I often felt out of my depth, but Dick was patient and his explanations were always clear. I learned much from working with him at PARC and collaborating with him on a couple of papers. I consider Dick one of my core mentors, and I will always be grateful to him for that. I wouldn't be where I am without him, wherever that is.

Many thanks to Dan Siddiqi for comments and discussions of this paper. Thanks also to the L_RFG Lab (<https://lrfg.online>) for continuing fertile discussion and development of the theory. Lastly, thank you to the two reviewers; the reviews were very helpful and led me to reconsider certain aspects of the formalization. Any remaining errors are my own.

References

- Asudeh, Ash. 2006. Direct compositionality and the architecture of LFG. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*, 363–387. CSLI Publications.
- Asudeh, Ash. 2012. *The logic of pronominal resumption* (Oxford Studies in Theoretical Linguistics). Oxford University Press. DOI: 10.1093/acprof:oso/9780199206421.001.0001. <https://doi.org/10.1093/acprof:oso/9780199206421.001.0001>.
- Asudeh, Ash. 2021. The unrealized and the unheard. In I Wayan Arka, Ash Asudeh & Tracy Holloway King (eds.), *Modular design of grammar: Linguistics on the edge*, 22–42. Oxford University Press.
- Asudeh, Ash. 2022. Glue semantics. *Annual Review of Linguistics* 8. 321–341.
- Asudeh, Ash, Bronwyn Bjorkman, Frances Dowle, Neil Myler, Daniel Siddiqi & Lisa Sullivan. 2024. Fusional morphology, metasyncretism, and secondary exponence: A morphemic, realizational approach to Latin declension. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG24 conference*, 44–68. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/3>.
- Asudeh, Ash, Tina Bögel & Daniel Siddiqi. 2023. Modelling exponents. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG23 conference*, 23–44. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/2>.
- Asudeh, Ash, Mary Dalrymple & Ida Toivonen. 2013. Constructions with lexical integrity. *Journal of Language Modelling* 1(1). 1–54.
- Asudeh, Ash & Gianluca Giorgolo. 2012. Flexible composition for optional and derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG12 conference*, 64–84. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2012/lfg12.html>.

- Asudeh, Ash, Paul B. Melchin & Daniel Siddiqi. 2024. Constraints all the way down: DM in a representational model of grammar. In *Proceedings of the 39th West Coast Conference on Formal Linguistics (WCCFL)*, 468–476. Cascadilla Press. <http://www.lingref.com/cpp/wccfl/39/index.html>.
- Asudeh, Ash & Daniel Siddiqi. 2022. Realizational morphosemantics in L_RFG . In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG22 conference*, 20–40. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/1>.
- Asudeh, Ash & Daniel Siddiqi. 2023. Morphology in LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 855–901. Language Science Press. DOI: 10.5281/zenodo.10185976. <https://doi.org/10.5281/zenodo.10185976>.
- Asudeh, Ash & Daniel Siddiqi (eds.). 2026. *Constraint-based morphosyntax: The L_RFG approach*. Forthcoming. With contributions by Oleg Belyaev, Bronwyn Bjorkman, Tina Bögel, Frances Dowle, Michael Everdell, Paul B. Melchin, Neil Myler, Will Oxford, and Lisa Sullivan. Cambridge University Press.
- Bögel, Tina. 2015. *The syntax-prosody interface in Lexical Functional Grammar*. University of Konstanz. (Doctoral dissertation).
- Borer, Hagit. 2005a. *Structuring sense, vol. I: In name only*. Oxford University Press.
- Borer, Hagit. 2005b. *Structuring sense, vol. II: The normal course of events*. Oxford University Press.
- Borer, Hagit. 2005c. *Structuring sense, vol. III: Taking form*. Oxford University Press.
- Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-functional syntax*. 2nd edn. Wiley-Blackwell.
- Chomsky, Noam. 1970. Remarks on nominalization. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in English transformational grammar*, 184–221. Ginn & Company.
- Chomsky, Noam. 1995. *The Minimalist Program*. MIT Press.
- Crouch, Dick, Mary Dalrymple, Ronald Kaplan, Tracy Holloway King, John T. Maxwell III & Paula Newman. 2011. *XLE Documentation*. Palo Alto Research Center (PARC). <https://ling.sprachwiss.uni-konstanz.de/pages/xle/index.html>.
- Dalrymple, Mary (ed.). 1999. *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*. The MIT Press.
- Dalrymple, Mary, Ronald M. Kaplan & Tracy Holloway King. 2004. Linguistic generalizations over descriptions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG04 conference*, 199–208. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2004/lfg04.html>.

- Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell III & Annie Zaenen (eds.). 1995. *Formal issues in Lexical-Functional Grammar*. CSLI Publications.
- Dalrymple, Mary, John Lamping & Vijay Saraswat. 1993. LFG semantics via constraints. In *Proceedings of the sixth conference of the European chapter of the Association for Computational Linguistics (EACL) (EACL '93)*, 97–105. Association for Computational Linguistics. DOI: 10.3115/976744.976757. <https://doi.org/10.3115/976744.976757>.
- Dalrymple, Mary, John J. Lowe & Louise Mycock. 2019. *The Oxford reference guide to Lexical Functional Grammar*. Oxford University Press.
- Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge University Press.
- Everdell, Michael, Paul B. Melchin, Ash Asudeh & Daniel Siddiqi. 2021. Beyond c-structure and f-structure: On the argument-adjunct distinction in O'dam. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG21 conference*, 125–145. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2021/lfg21.html>.
- Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum & Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Harvard University Press.
- Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In Samuel Jay Keyser & Ken Hale (eds.), *The view from Building 20*, 111–176. MIT Press.
- Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40(3–4). 225–276.
- Haugen, Jason D. & Daniel Siddiqi. 2016. Towards a restricted realizational theory: Multimorphemic monolistemicity, portmanteaux, and post-linearization spanning. In Daniel Siddiqi & Heidi Harley (eds.), *Morphological metatheory*, 343–386. John Benjamins.
- Kaplan, Ronald M. 1987. Three seductions of computational psycholinguistics. In Peter Whitelock, Mary McGee Wood, Harold L. Somers, Rod Johnson & Paul Bennett (eds.), *Linguistic theory and computer applications*, 149–188. Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 337–367). Academic Press.
- Kaplan, Ronald M. 1989. The formal architecture of Lexical-Functional Grammar. *Journal of Information Science and Engineering* 5(4). Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 1–27), 305–322.
- Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 173–281. The MIT Press. Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 29–130).

- Kiparsky, Paul. 1973. "Elsewhere" in *phonology*. Indiana University Linguistics Club.
- Lapointe, Steven. 1980. *A theory of grammatical agreement*. University of Massachusetts, Amherst. (Doctoral dissertation).
- Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. MIT Press.
- Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4(2). 201–225.
- Melchin, Paul B., Ash Asudeh & Dan Siddiqi. 2020. Ojibwe agreement in Lexical-Realizational Functional Grammar. In Miriam Butt & Ida Toivonen (eds.), *Proceedings of the LFG20 conference*, 268–288. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2020/index.shtml>.
- Merchant, Jason. 2015. How much context is enough? Two cases of span-conditioned stem allomorphy. *Linguistic Inquiry* 46(2). 273–303.
- Pollard, Carl & Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. The University of Chicago Press & CSLI Publications.
- Ramchand, Gillian Catriona. 2008. *Verb meaning and the lexicon: A first phase syntax*. Cambridge University Press.
- Stump, Gregory T. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge University Press.
- Svenonius, Peter. 2016. Spans and words. In Daniel Siddiqi & Heidi Harley (eds.), *Morphological metatheory*, 201–222. John Benjamins.

Chapter 5

Glue Semantics for Hopf-Algebraic Minimalism

Avery Andrews^a

^aAustralian National University

Recent works by Marcolli, Chomsky and Berwick have developed the observation that the Hopf algebra of trees, originally developed to solve a problem in physics, has a surprisingly good fit to the basic ideas of the most recent version of Minimalist syntax. They have also proposed a number of ideas about how to do semantics, but these seem to depend on an incompletely worked out utilization of traces. Here I will suggest a way of using the glue semantics of Lexical-Functional Grammar to do semantic interpretation on Hopf-algebraic syntactic derivation without traces, accommodating the basics of quantifier scope without syntactic movement. A diagrammatic notation based on Construction Grammar is used for the glue proofs.

The purpose of this paper is to use the ‘glue semantics’ of Lexical-Functional Grammar (LFG) to provide a form of semantic interpretation for the new formulation of some of the basics of Minimalist syntax using Hopf algebras presented in Marcolli, Berwick, et al. (2023), Marcolli, Chomsky, et al. (2023a) and several lectures and lecture series, here collectively referred to as MCB.¹ Glue semantics is essentially Montague-style semantics attached to a subset of Girard’s Linear Logic in such a way that it can do compositional semantics on structures that are not trees, such as LFG’s f-structures, which fail to be trees due to containing reentrancies, and can furthermore induce scope relations on structures (such as, originally, LFG’s f-structures) that do not have enough hierarchical structure to directly support them.

The Hopf-algebraic version of Minimalism arises from the observation by Marcolli that the Hopf algebra of unordered trees, originally developed for dealing

¹These include the lecture Marcolli (2023a), the lecture series Marcolli (2023b) and the course materials Marcolli (2024), especially Marcolli et al. (2024).

with renormalization in Feynmann diagrams, provides the basic machinery for External and Internal Merge as proposed in the ‘New Minimalism’ as developed in syntax since 2013, recently presented in Chomsky et al. (2023), while not fitting so well with older versions, such as the one formalized by Stabler (2010). This apparent compatibility with the most recent version of Minimalism might of course be an accident, with no real significance for linguistics, but I think that it is nevertheless sufficiently intriguing to be worth looking at carefully.

The ingredients of this new mathematical formulation are:

- (1) a. The Hopf algebra of unordered trees, presented concisely by Foissy (2013).² This provides a concept of ‘workspace’ as essentially a set or multiset of trees,³ together with a ‘coproduct’ that dissects trees into an extracted part and a remainder. The coproduct provides the materials for Internal Merge, with removal of the extracted material from its original position. There is also an operation of disjoint union for workspaces, which plays a role in the formulation of the syntactic Merge operations.
- b. The B^+ operator, which combines a subset of trees in a workspace into a single tree. This does the tree assembly for syntactic Merge.
- c. Additional concepts to manage linearization and other parameters, that are not our concern here.

Appendix B provides a brief discussion to the basics of Hopf algebras including of trees, with further references.

Some features of New Minimalism that are covered by this formulation are:

- (2) a. Trees are (mostly) binary, are unordered, and only terminals have properties apart from the contents of their daughters.

²Foissy does not restrict the trees to binary, while for the most part, MCB want to do that, but allow non-binary trees in the process of ‘Form Set’ (Marcolli et al. 2024: 137-142). They argue that this is not actually Merge, but the argument is incomplete, because, for example we can presumably combine a trinary coordinate structure and a binary to produce a binary coordinate structure (*Tom, Mary and Alice, and Bill and Susan*, perhaps a guest list comprising a polycule and a couple). Whether the Hopf algebra itself should be restricted to binary trees only, or the evident restriction should be explained in terms of something else, such as Theta role assignment, deserves further thought.

³Which it depends on aspects of formalization; if trees are taken to be sets, then the workspace must be a multiset, since it must be able to contain copies, but if they are graphs, then it can be a set. The latter formulation seems to be the most popular, and will be assumed here, but we will also find a use for sets.

- b. Labels are found by searching for material on terminals, not on nonterminals.
- c. Merge is restricted to External and Internal, some other possibilities, such as sideways and countercyclic are excluded. (This is achieved by adding some parameters to the formulation, and taking them to 0 as a limit.)

This degree of convergence is rather intriguing, and makes the Hopf-algebraic formulation worth looking into.

Semantics has been considered in this framework in Marcolli, Chomsky, et al. (2023b) and other sources, including Marcolli et al. (2024), Marcolli (2024) and sources cited there. The latter two include a discussion of how to include conventional Montague-based formal semantics in the approach, adapting the Heim & Kratzer (1998) formulation of the syntax-semantics interface. However these accounts do not yet appear to provide an account of scope, either of elements that take wide scope *in situ*, such as quantifiers in English, nor those that get displaced to become adjoined to their scope domain, such as *wh*-words in English. However in Dalrymple et al. (1997), it was shown that ‘glue semantics’ can assign the correct scope interpretations in some complicated cases where classic ‘Quantifier Raising’ fails. Furthermore, Andrews (2007, 2008, 2019) argues that glue semantics can be adapted to give reasonably clean accounts of idiomatic/multiword expressions, as well as problems of semantically idiosyncratic government of cases and adpositions, another range of problems not considered in the currently envisioned treatments of semantics in Hopf-algebraic Minimalism.

For these reasons, I think it is worthwhile to see whether glue semantics can be applied to the Hopf-algebraic version of Minimalism.⁴ The version I will use is the ‘propositional glue’ of Andrews (2010), which dispenses with the use of linear universal quantification as standardly used in the treatment of scope (in the logic of semantic assembly in glue, not the NL semantics version of universal quantification), instead using the LFG version of ‘instantiation’, which we will come to see later. Gotham (2021) argues that propositional glue does not give as good a treatment of certain scope restrictions as the scope restrictions as the version with universal quantification, but propositional glue is a bit simpler than either first order glue (Kokkonidis 2008), or System F as used in the now standard ‘new glue’ of Dalrymple et al. (1999), and also constitutes a free symmetric monoidal closed category (Troelstra 1992: 81-92).

⁴See Gotham (2018) for its application to an older version of Minimalism.

Another feature of the approach presented here is that the semantics is applied incrementally, as the grammatical structure is produced by Merge, rather than to completed syntactic outputs as presented in Marcolli et al. (2024) and other sources cited above. This avoids the use of traces, whose functioning in semantics does not seem to be fully worked out, and it is in any event interesting to see whether interpretation can be done at all without them. A further difference is that in this version, we make no use of the concept of ‘head’ and therefore of ‘head functions’ defined over the syntax. These of course might start being used as more syntactic phenomena are considered.

I will finally introduce the presentational, but not substantive, innovation of presenting the glue proofs in a box notation inspired by that of construction grammar (CxG) as used in Kay & Fillmore (1999), although enhanced with a means to directly represent quantifier scope. The motivation for this is that many people find the deductive presentation of glue intimidating, and perhaps the CxG-inspired notation will make it easier. Another is that I think this partial resemblance to CxG is inherently interesting. For mainstream introductions to glue, see Dalrymple (2001), Asudeh (2022) and Asudeh (2023).

It is a feature of glue that semantic assembly can proceed autonomously, without constraints from syntax. I will therefore start by introducing glue and the box notation in that manner, and then show how to link glue assembly to the Hopf-algebraic syntax.

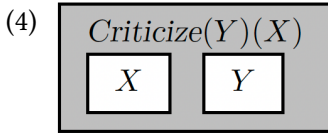
1 Glue by itself

1.1 Applying predicates to arguments

We begin with the apparently simplest semantic contributions from the point of view of compositional semantics, proper names, which, following CxG notation, we will represent as boxes containing a label for their meaning (whose connection to pragmatics is well-known to be extremely complicated). However we differ from standard CxG notation in having the boxes be shaded:

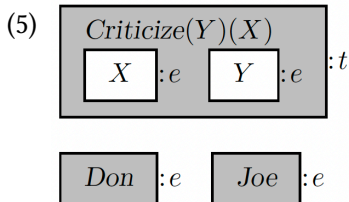
(3) 

For the next-most-complex kind of contribution, predicates with arguments, in addition to the shaded outer box, there are unshaded inner boxes, which contain upper case variables, and the meaning label repeats these variables as arguments to the predicate:

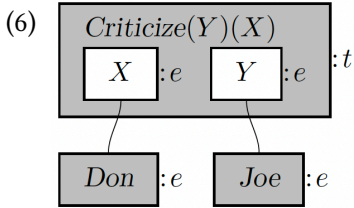


The shading represents ‘polarity’ in (intuitionistic implicational) proof-nets, as explained by de Groote (1999), with the difference from de Groote that we have taken the shaded boxes to be positive, representing the ‘presence of content’ to be sent elsewhere, the unshaded ones to be negative, representing an absence of content, to be remedied by the importation of content from a shaded, positive box. There is a further fundamental issue of whether the order of applications, as exists by necessity if curried functions as employed by Montague are used, has any linguistic meaning; I tend to take arguments from the 1970s and early 1980s, summarized and extended by Marantz (1984), as indicating that it does, with the less active arguments composed earlier (with a considerable range of variation, as discussed in the literature on Icelandic starting with Bernódusson (1982), and numerous discussions of double object constructions in many frameworks). Because of our use of upper case substitution variables, we can arrange the boxes so as to reduce crossing of the connecting lines, with their linear ordering having no significance, but we could also convert to an order-based notation (discussed in appendix A, and better if you want to write in semantic values with the substitutions applied).

The next elaboration over CxG is that each box has a type, which are the types as usual in basic formal semantics, using simple type theory. Conventionally, people start with types e for ‘entity’, and t for things at least somewhat like propositions (with truth values), and so shall we, at the beginning; after introducing them, people tend to omit them in practice. When we write them, we will write them to the right of the boxes, separated by colons (the types of all boxes are atomic, as discussed in the final section). So now the basic semantic ingredients for a sentence such as *Don criticized Joe* can be represented as:



It should now be clear that the technique of semantic assembly is going to be to connect shaded boxes to unshaded ones, matching the types. This gives us two ways of hooking up the boxes in (5), one of which is depicted below:



The links are called ‘axiom links’, because they represent the use of the (sole) axiom $a \vdash a$ in the Gentzen sequent calculus, the basis for proof-nets.

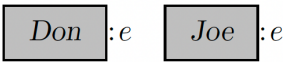
In fact, it appears that Kay and Fillmore came up with the basic idea of proof-nets for the easier half of intuitionistic implicational logic, but did not to my knowledge finish the formalization, which we do by presenting the rules governing legitimate hookups with axiom links (which allow all and only the valid sequent calculus deductions using the ‘implication left’ rule). To formulate the rules, we want another concept, the ‘dynamic graph’, introduced by Lamarche (1994), and used by de Groote (1999), with the directionality reversed. We follow de Groote, and, for us, the dynamic graph is composed of two kinds of oriented links:

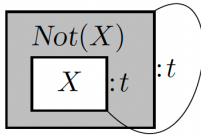
- (7)
- a. Axiom links, going from shaded to unshaded boxes.
 - b. Implicit links, going from unshaded boxes to their immediately containing shaded box (these links are explicit in the usual proof-net notation).

The rules are then as follows:

- (8)
- a. Each unshaded box must be axiom-linked to one and only one shaded box.
 - b. Each shaded box is axiom-linked to at most one unshaded box.
 - c. The dynamic graph forms a rooted (non-planar) tree.

The first two conditions seem to be implicitly obeyed in CxG, while the third rules out combinations such as:

(9a) 

(9b) 

There are other ways in which these configurations could be excluded, but (8c) is the choice that will be needed later.

1.2 Forming lambda-abstracts

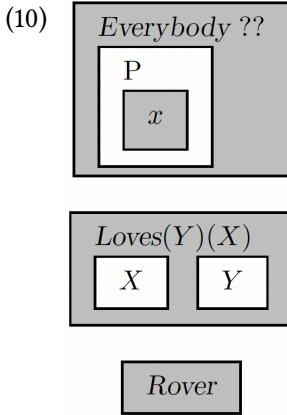
Montague analysed quantifiers such as *everyone* (including what linguists would be likely to call ‘quantified noun phrases’, such as *every sensible person*) as being of the type $\langle\langle e, t \rangle, t \rangle$, that is, predicates that apply to a complex argument of type $\langle e, t \rangle$ (e.g. *hops*) to produce a truth-value. A problem with this is that such quantifiers also appear in object position, sometimes in combinations such as *somebody loves everybody*, which are perceived as ambiguous. This problem is normally solved by some kind of rule of ‘Quantifier Raising’. Glue uses something that can be regarded as yet another version of this, but implemented in the semantics rather than in the syntax.⁵

In the present adaptation of the CxG box notation, the essential idea is to allow unshaded boxes to themselves contain shaded boxes. The semantic value of these ‘inner shaded boxes’ is a lower-case variable, that will wind up being bound by a lambda in the semantic formula.⁶

A preliminary version of the constructors for the meaning of *Rover loves everybody* would be (types omitted because in glue they usually are, when they can be understood from context):

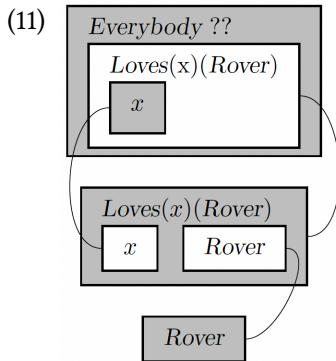
⁵Discussed in greater detail for earlier Minimalism by Gotham (2018).

⁶These are essentially ‘inner values’ in the algebraic correctness criterion of de Groote (1999), while the semantic labels without their arguments are ‘outer values’.



It is clear that to get the intended meaning, we will want to axiom-link the Rover box to the X box. Then to get a lambda abstract on x , we will want to link the inner shaded x box to the unshaded Y box. But what next?

Part of the answer is that first we connect the shaded *Loves* box to the unshaded *P* box, which we would expect to result in *P* being evaluated to $Loves(x)(Rover)$:



But what then do we put in for the ‘??’ in the *Everybody* box? The answer is that we put the value of the argument-box bound by the variable in its inner shaded box, so that the pre-substitution formula for ‘??’ is $\lambda x.P$, and the result of substitution is:

(12) $\lambda x.Loves(x)(Rover)$

The general rule is then that if an unshaded box contains a shaded box, then, in the argument position for the unshaded box contents, one writes the unshaded

box's upper case (naive substitution) variable, lambda-bound by the unshaded box's lower case variable (genuine lambda-calculus variable, not manageable by naive substitution). If the unshaded box contains more than one shaded box, then the content of the unshaded box must be specified as bound by all of the variables of the shaded boxes.

There is then a final clause to the rules governing axiom linking, which assures that this always makes sense:

- (8) d. The dynamic path for an inner shaded box must pass through its immediately containing unshaded box.

Note in particular that there is *no* direct dynamic path link from an inner shaded box to its immediately containing unshaded box; the path between them must consist of other links.

1.3 A bit more on types

Before moving on to the connection to syntax, it is useful to say a bit more about types, in particular, the relationship between semantic types and glue types. In linear logic by itself, atomic propositions belong to ('inhabit') basic types. In formal semantics, the basic types correspond to ontological categories such as 'entity' and 'proposition', and people usually want to have more than one of them, although the need for this is not entirely beyond question.⁷ From the basic types, we form derived types, corresponding to non-atomic propositions.

However in glue, linear logic is not by itself, but functioning as part of an interface to some kind of typed lambda calculus, possibly with extensions. This brings up the question of the relationship between the glue types and the semantic types. So far, I am aware of three possibilities:

- (13) a. There is only one basic glue type, but semantic composition is regulated by composition of the semantic types.
- b. The glue types are in one-to-one correspondence with the semantic types. This would include the possibility that in semantics also, there is only one basic type.
- c. The two systems are different, requiring a well-formed assembly to satisfy two different sets of type constraints at the same time.

⁷See for example Partee (2009) and Partee & Borschev (2004).

The first two options are identical in effect, as far as I can determine, so it does not matter which one we choose. The third is potentially different, for example we might have a basic glue type ‘predicate’ corresponding to the semantic type $\langle e, t \rangle$. However although this idea seems intriguing, I have not managed to derive any useful results from it. Therefore I will assume either one of the first two.

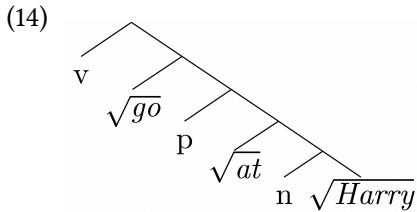
I conclude this subsection by noting that if we can retrieve the meanings of the words and idiomatic expressions in an utterance, we have a reasonable chance of retrieving the associated semantic contributions, and, especially with shorter utterances, might be able to work on the meaning without resorting to information from the syntax. However syntax is essential to comprehension in many situations, so in the next section, we show how it can be used to constrain semantic assembly, within the framework of Hopf-algebraic Minimalism.

2 Semantic assembly and External Merge

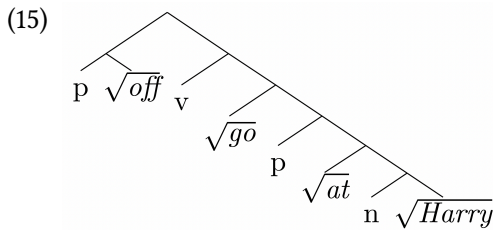
The basic idea here is that of Andrews (2007, 2008, 2019), to use configurations of syntactic items to license the introduction of semantic contributions. In these papers, this is accomplished by means of a ‘Semantic Lexicon’, consisting of Semantic Lexicon Entries (SLEs), which specify the interpretations of combinations of conventional lexemes and features, independently of a ‘Morphological Lexicon’ (which probably should have just been called ‘the morphology’), which specifies the morphological realization of combinations of the values of PRED (‘lexical items’) and inflectional features. Minimalism would use different mechanisms for packing together the features and lexical items to be realized in the morphology, which I will not discuss here. However for an example, the finite past form of *go* is *went*, across all of the idioms it appears in, such as *go off (at)* (with at least two meanings if there is no preposition) as well as in the single-word lexeme usage. I will illustrate the approach with the example *Meghan went off at Harry*, which contains some relevant syntactic problems in addition to an idiom/multiword expression (MWE).

Another respect in which this approach diverges from the Hopf-algebraic presentations so far (which do not provide fully specified trees with enough details to do syntax) is that we explicitly use categorizer formatives, such as *n* and *v*, as in various Minimalist works such as Harley (2014). This is helpful in formulating the SLEs, and also provides a clear distinction between ‘lexical items’ and ‘features’: the former co-occur with categorizers and are ‘roots’; the latter do not require categorizers, and what is involved in their distribution will not be considered here.

The first question is what to do about direct complements of verbs and prepositions; I will follow Harley (2014) in putting them as complements to the verb or preposition root under the appropriate categorizer, giving us (14) as the bottom-most portion of the structure:⁸



Our next problem is what do do about the particle *off*, and my proposal is to place it as a ‘specifier’ to the *v*, where indirect objects might also go:



Next, we have the problem of what to do about the ‘external argument’ *Meghan*. I am going to defer this, and first indicate how the internal argument *Harry* and various items (here, roots) that are responsible for the meaning of the predicate are handled.

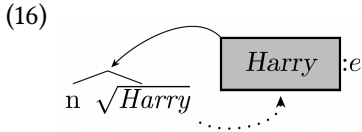
Andrews (2007, 2008, 2019) proposes that, in LFG, combinations of semantically interpretable attributes in *f*-structure trigger the introduction of a meaning contribution, and when they do that, they get ‘checked off’, so that they cannot do it again (either individually or in another combination). Furthermore, every so-interpretable attribute must get checked off (this sense of ‘interpretable’ is not necessarily the same as in Minimalism). Andrews (2019) proposes to formalize this by having interpretable feature instances in an *f*-structure contain a pointer, which must be set to a meaning-contribution, or a list of these.⁹ These pointers go from interpretable feature-values to meaning-contributions. The SLEs also

⁸Adopting the proposal of Jackendoff (1973) that particles are intransitive prepositions.

⁹Dalrymple (2001) discusses cases where it is useful to allow one item to introduce more than one meaning-contribution.

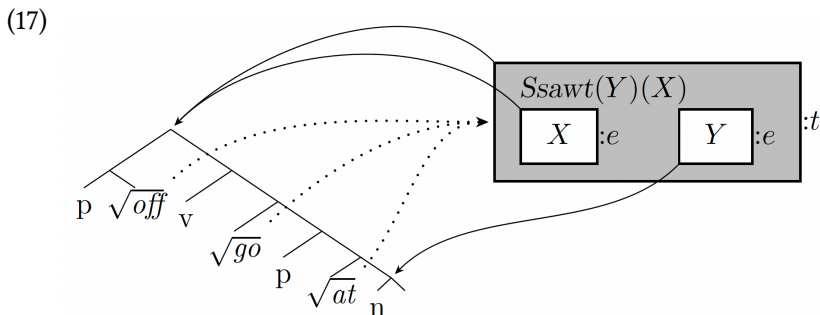
specify what I call ‘semantic assembly points’ in the syntactic structure, which allow the syntactic structure to constrain semantic assembly in a manner we will describe shortly.

The simplest form of SLE would be one for *Harry*:



Here we take the category features to be uninterpretable, which allows their presence to be referenced by more than one SLE, which provides a straightforward way to specify the syntactic categories of the complements to a root. The dotted line from the root $\sqrt{\text{Harry}}$ represents the ‘checking off’ link that will cause this instance of this root in the structure to be linked to this instance of a semantic contribution, while the solid links with arrowheads are to the semantic assembly points, which function in a way to be explained later. The arrowheads represent the fact that these links are many-to-one, from the boxes in the meaning-contributions to the positions in the syntactic structure, which is important later.

The entry for the MWE *go off at* is more complex. I take its meaning to be something along the lines of *X suddenly starts saying angry words to Y*, which I will abbreviate as *Ssawt*, and it will need pointers to three semantic anchoring points, for the *at*-object, the external argument, and the resultant meaning. For reasons to be explained shortly, it works out for the second two to point to the top of the tree, yielding the following SLE:

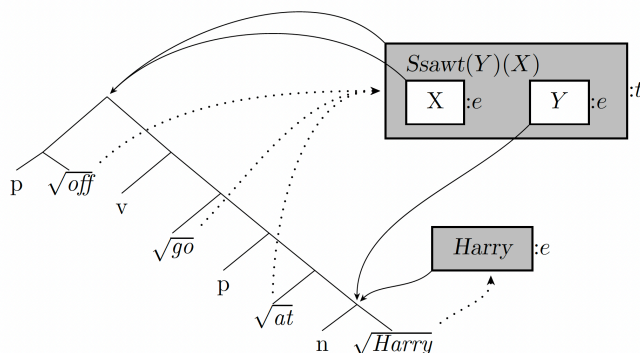


The tree on the left should be seen as a convenient and easy to read representation for a description of a tree in a tree-description language, similar in conception but different in details from the f-structure description language used in LFG. The

dotted line semantic instantiation links are redundant in the formulation of the SLE, because they will be supplied by convention to each interpretable feature in the portion of the tree that matches the description, pointing to the outermost box of the semantic contribution,¹⁰ but are important in the structure produced. Note how the ‘n’ at the bottom of (17) can be the same as that in (16), allowing the *Harry* SLE to provide a meaning for the complement of (17).

Next, consider what happens when we apply SLE’s to the full tree (15):

(18)

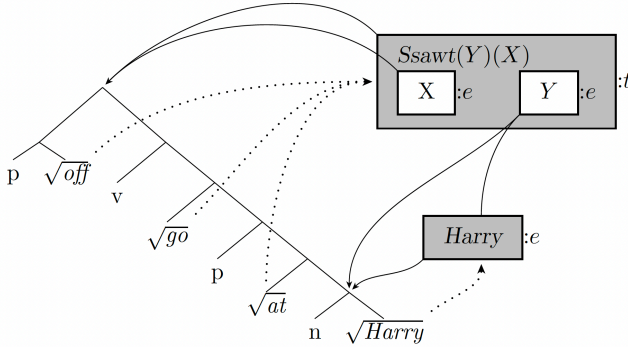


Given the SLEs introduced so far, this is the only way to attach SLEs so as to satisfy the requirement that all interpretable features get interpreted, so hooking them up with axiom links would be the next step.

The effect we want to produce is that we can axiom-link the *Harry* box to the *Y* inner box of the *Ssawt* meaning contribution, but not to its *X*-box. Visually, we can achieve this by imposing the requirement that in order to be axiom-linked, boxes must be connected by a solid arrow to the same semantic assembly point in the syntactic structure. Then this hookup is allowed, but the other conceivable one is not:

¹⁰In terms of function, an SLE such as (17) does essentially the same thing as a ‘meaning constructor’ in mainstream LFG glue, but the way it does it is different: it is matched against a generated structure (‘description by analysis’) as opposed to being included in a conventional lexical entry (‘codescription’), as discussed by Halvorsen & Kaplan (1988).

(19)



This gives us a visual story, which we can formulate as the following principle:

- (20) Axiom Link Principle: in order to add an axiom link between a shaded and unshaded box, both must be associated to the same semantic assembly point.

In terms of the deductive formulation of (propositional) glue, we want the assembly points to represent atomic propositions. This is straightforward in LFG, since these locations have always been thought of as having unique labels, but the environment here is a bit different, so we need to say more about how this can work.

The problem is that a derivation starts with a workspace consisting of a collection of atomic items (lexical items and grammatical features), which is the first in a sequence of workspaces, each created from the preceding one by an operation of External or Internal Merge, the last consisting of a single tree. Hopf algebras do not include any concept of persistent identity between nodes in successive stages of a derivation, and we do not want to add any such thing in any unnecessary way.¹¹ However we arguably need something of this nature for the terminals, since different instances of the same feature in a structure can have different meanings, e.g. one semantic plural and one *pluralia tantum* in *The investigators found some bathing trunks lying on the beach*. Therefore I will take from LFG the technique of assigning to each terminal, or perhaps only each interpretable terminal, a unique ‘instantiation index’ when it is taken from the lexicon and put in the first workspace of the derivation.

¹¹And whatever was attempted would have to be consistent with the fact that the multiplication for the Hopf algebra of trees is disjoint union (\sqcup in MCB), so that for a workspace W other than the empty one, it cannot be the case that $W \sqcup W = W$.

These are then packed into binary trees by a External Merge operation that combines the workspace items into trees, recursively, and Internal Merge, which uses the Hopf Algebra operation called the ‘coproduct’ to extract some members from the extant trees, and then further applications of the Merge operation to implement extraction.¹² These operations are explicitly formulated in such a way as to preclude anything like a ‘corresponding nodes’ relationship between nodes at one stage of the derivation and those of another, such as that proposed by Lakoff (1970) for Transformational Grammar.

However the instantiation indexes produced above make the terminals (or at least the interpreted ones) different, because these indexes have persistent identity through the derivation, and we can use them as ‘anchors’ to locate semantic assembly points via paths through the tree. If we suppose that the semantic assembly points are restricted to be non-terminals, we can specify them uniquely by means of upward paths from terminals and downward paths through nodes that have only one non-terminal daughter (recall that since the trees are unordered, there can be no specification of ‘right’ vs. ‘left’ daughter). A consequence is, for example, that if trees include n -ary branching nodes representing coordinate structure, an SLE will not be able to identify any of these daughters via a downward path, which would seem to be a good result.

Therefore, we can specify a semantic attachment point for a box with the index of a terminal that is connected to the meaning-contribution followed by a sequence of up or down arrows that uniquely specifies a relative position to that terminal. However once an axiom link between two boxes is created, it does not matter if the shared attachment point is removed. This corresponds to the fact that when one applies implication elimination ($A, A \multimap B \Rightarrow B$) in a proof, both the minor premise and the matching term in the other premise disappear. For formulations more general than propositional glue, currently either System F or first order, we can use the instantiation indexes combined with path specifications to provide the atomic symbols for the formulas.

There is however an issue: for an SLE comprising multiple interpreted items, which one(s) should we use as the starting point of the semantic assembly path(s)? There are several possible answers, and for the purposes of this paper, it does not matter which one is chosen. Some possibilities are:

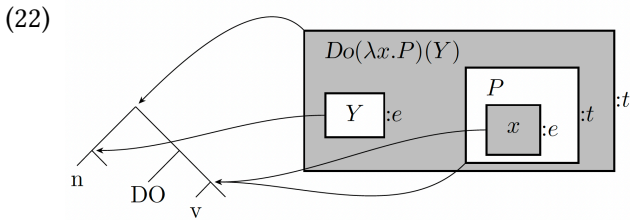
- (21) a. For each attachment point, the anchor that makes the path as short as possible.

¹²There is also ‘Form Set’, in Marcolli et al. (2024), which combines n workspace items into an n -ary tree, but this does not yet seem to be well-integrated into the system. For an account of the semantics of coordination in glue, see Asudeh & Crouch (2002a,b).

- b. Free choice
- c. On some basis, one of the interpreted terminals is determined as the ‘prime terminal’.

(21a) is in some sense the immediately simplest, (b) seems somewhat unsatisfactory, while in order to be satisfactory, (c) would require more work to justify criteria for identifying the prime terminal. However for present purposes, any of them would suffice.¹³

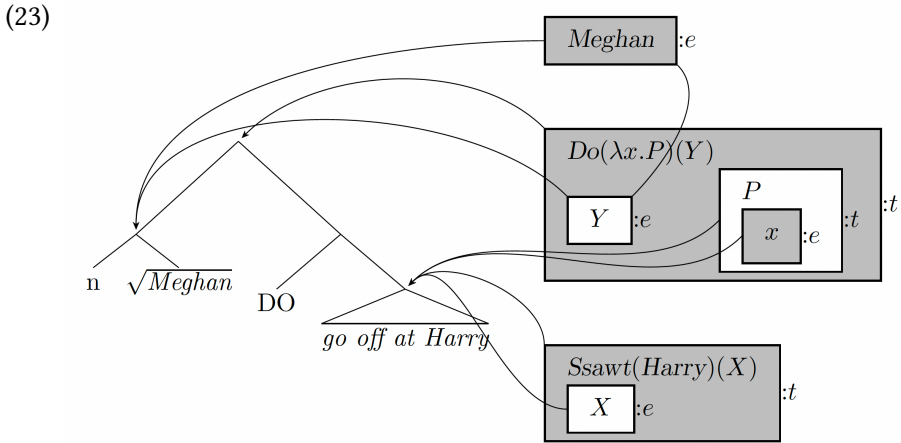
Returning to more concrete grammatical issues, we need a proposal for external arguments. The approach we will take here will be similar to the one in Wood & Marantz (2017) and much other Minimalist work, in which external arguments and a range of other items, such as applicatives, are introduced by an additional item adjoined to the verb phrase. In that paper, this item is a multifunctional element i^* , but since here we are considering only external arguments, we will call it *DO*. This will take the external argument as its specifier, and the verbal phrase as its complement. Its SLE looks like this:



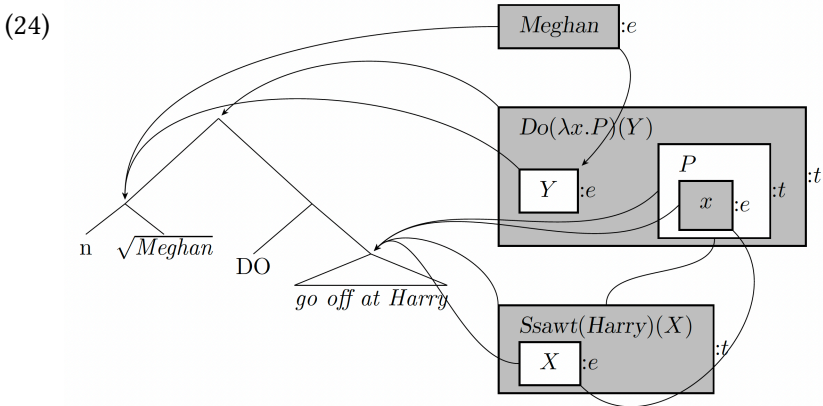
What substantive contribution *DO* makes to the meaning is a topic I will not look into here. (External arguments clearly have a strong association with agentivity, but there is likely more to it than that.)

The tree including *DO* and the external argument will then be (22) below, with the trivial axiom link for *Meghan* added, and the previously assembled structure presented as a block:

¹³There would also be the possibility of using the tree to construct a set in the obvious way (just like a decoration of a graph in the development of non-well founded set theory), with the instantiation indexes as atomic elements. However the timing of interpretation gets more critical, since if we extract something from a subtree, its associated set will change. Furthermore, if Σ is the set of sets for the trees in a workspace W , then it will also be the set of trees for the product of W by itself ($W \sqcup W$). This is not a problem for syntax, since the workspaces combined in the course of a derivation will always correspond to sets that are disjoint in the first place, since terminals are not copied.



So what about the P box and the predicate material? Here we have four arrows converging on one node. However two of them are of one type, and two of another, and, for each type, one comes from a shaded and the other from an unshaded box, so we can add 2 more links:



with the result that P gets evaluated as $Ssawt(Harry)(x)$, as an argument of DO it gets lambda-bound by x , and the DO meaning can then combine this with *Meghan* to a final result:

(25) $DO(\lambda x.Ssawt(Harry)(x))(Meghan).$

For the reason explained above, the semantic assembly will persist if changes are made to the internal structure of (25) by Internal Merge extracting things

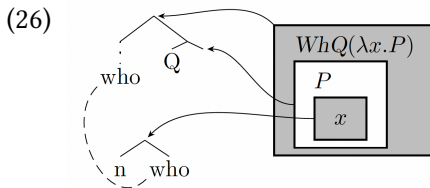
from it, but there are still some questions about the timing of the introduction of meaning contributions and semantic assembly.

The most obvious interpretation of Andrews' previous proposals is that an SLE is attached at some point after its syntactic description is satisfied. However it does not matter when; with the structure in (15), we could attach the *Harry* SLE either as soon as the object noun phrase assembled or at some later point. We will also need the attachment of SLEs to be optional, so that lexical items can get either a meaning determined by their local context or one that they acquire later. An acceptable structure will be one that has all of its interpretable elements interpreted, but the order of operations by which this is accomplished is taken to be irrelevant.

3 Internal Merge

We now consider Internal Merge (IM), using the example of constituent *Wh*-questions, where a questioned phrase appears to move to a position where it can take its scope as a complement. For this section, it is not necessary to use an MWE as the main predicate of the example sentence; a simple transitive verb is sufficient, so that is what we will use. The complexities of external arguments are also not relevant, so we will use *Who does Rover love*, with no representation of an external argument.

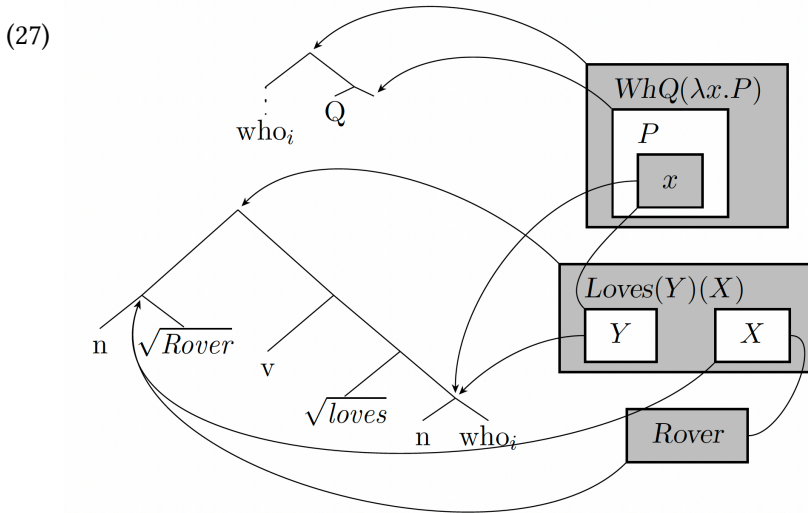
Our problem is how to get *who* installed both as the object (presumably lower) argument of *like*, and how to get it to move to the position of its scope. I will suppose that *wh*-questions have an operator Q in complementizer position, with the clause as complement and the *Wh*-phrase as specifier. The syntactic portion of the SLE will also consist of two distinct parts, with no relations between them, but sharing a single lexical item:



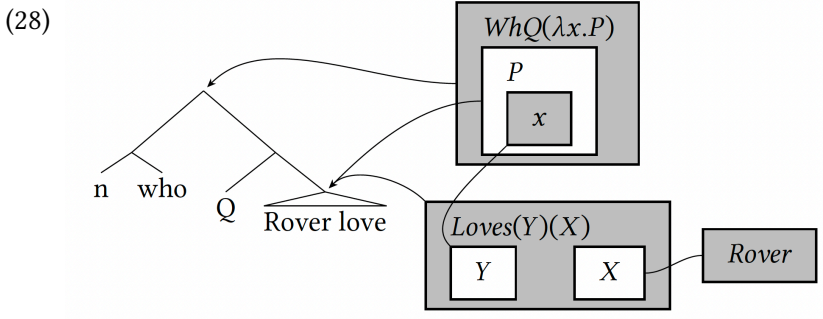
The dashed line connecting the two appearances of *who* is to indicate that they are the same syntactic object, including instantiation index, while the dotted line connecting the upper appearance to the tree-node above it indicates that it can

appear inside a containing phrase (pied piping). The nature of the constraints on this will not be considered here.

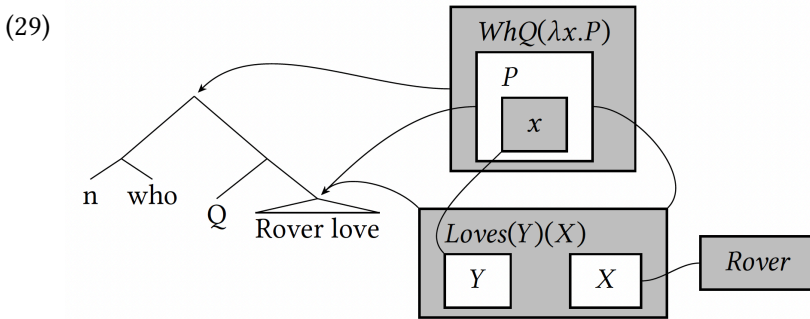
There is also an unshared item, Q . The proposal is that if two syntactic specifications are not connected by any tree-internal predicates, such as for example that something in one be the mother of something in the other, then they do not both have to be satisfied in the same workspace, but rather, possibly, in different ones, one derived from the other by External and Internal Merge operations. We can represent the pre-IM situation as follows, where the appearances of *who* are co-subscripted, to indicate that they are have the same instantiation index and are therefore the same item:



In order for the derivation to proceed, we have to Merge Q with the structure of (27) or with some larger one including (27) (to produce something like *Who do you think Rover loves?*). However once this is done, we can use IM to extract the questioned NP, at which point the rest of the SLE can apply, satisfying the positional requirements for the relative pronoun and providing an interpretation for the Q feature, and yielding:



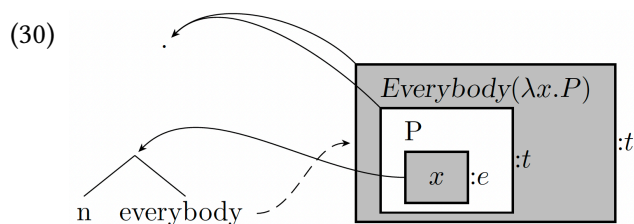
So now we can add the final axiom link to get a well-formed interpretation:



We now have an account for single *wh*-NPs that prepose to the position of their scope.

The other type we will look at here is *in situ*, exemplified by *wh*-NPs in Japanese and many other languages, as well as quantifiers in English. A type we will not consider here are multiple *wh*-questions, which can move one *wh*-phrase, all of them, or none of them; for an analysis of all these types within LFG+glue see Mycock (2006).

The account of *in situ* constructions, both *wh*-questions in Japanese and quantifiers in English, is a simpler version of the analysis above of *wh*-questions, on the basis of our earlier treatment of quantifiers in section 1. A possible SLE for *everybody* is (30), where the upper syntactic component of the SLE has no syntactic connections to the lower one:



Condition (8d) on acceptable hookups will cause the node of the upper part of the syntactic specification to be identified with some syntactic structure that includes the instance of *everybody* in the lower part of the syntactic structure, so that the x will be properly bound.

This has the behavior of the earliest version of quantification in glue, using universal linear quantification, as in Dalrymple et al. (1997), which does not include any syntactic restrictions on quantifier scope. If this analysis is accepted, the actual restrictions on quantifier scope that have been observed, as discussed for example in Gotham (2021), have to be treated as semantic in origin. This seems possible, but has not been worked out. Gotham provides a modification of the glue that allows them to be described in glue, but uses a deductive formulation, which I have not worked out how to implement in the box notation used here.

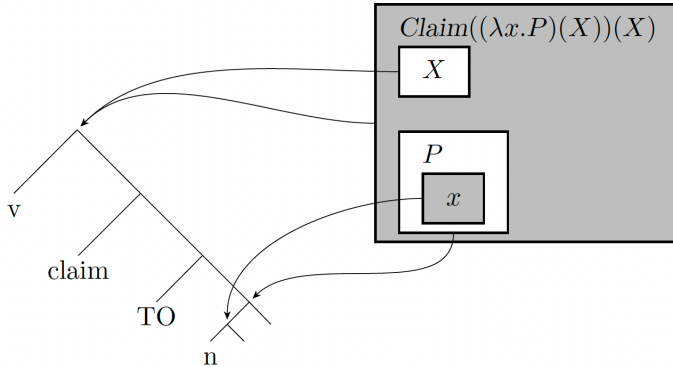
The case of *in situ wh*-phrases is similar to *in situ* quantifiers, with a difference that semantics is plausibly involved in determining their scope. For example narrow scope is strongly preferred or required for the quantifier in (31a), but wide scope is required for the *situ* echo question in (31b).

- (31) a. Somebody thinks that John gave a treat to every dog
 b. Somebody thinks that John gave a treat to WHICH dog??

I propose that this be accommodated by having the output type of the *wh*-constructions be different from that of the quantifiers: the latter would be of type $(e \multimap t) \multimap t$, while the type of the former would be $(e \multimap t) \multimap ct$, ' ct ' being the type of a 'conversational turn', only well-formed as a full independent conversational turn, expecting a response from the interlocutor.

A final observation I will make is that glue can implement a version of the obligatory control as movement of Hornstein (1999). Given the previous proposals, the SLE for *claim* might be:

(32)



The links from the unshaded box labelled P and the shaded box labelled x construe the TO -complement as the predicate $\lambda x.P$, which is the first argument of the meaning of *claim*. We assume that other factors, not discussed here, force the subject of the TO -complement to move to the regular subject position for external arguments, here treated as specifier of DO , whereupon the SLE for DO will deliver this meaning to the location linked to the unshaded box labelled X , so that it functions both as the agentive argument of *claim* and the (possibly derived) subject argument of the TO -complement. Glue semantics therefore allows us to implement Hornstein’s analysis of obligatory control without any additional mechanism of FORM COPY . The semantic analysis is a Hopf-algebraic Minimalist version of the LFG analysis of Asudeh (2005).

4 Conclusion

I have presented a technique for applying LFG’s glue semantics to Hopf-Algebraic Minimalist derivations. Unlike the proposal in Marcolli et al. (2024), it does not use traces, but Marcolli et al. (2024) do not give a full account of how traces would work for semantic interpretation. A significant difference between their formulation and this one is that this one uses the ‘deletion’ coproduct Δ^p (Marcolli et al. 2024: 29), which has the property of leaving a residue that is not binary branching, since the mother of an extracted node will be left with a single daughter if only one is extracted, as would be the case in a linguistically realistic example. This is not satisfactory if we insist on a Hopf algebra of binary trees, although the nonbranching node can be removed to produce the Δ^d coproduct. However the system is also supposed to contain an operation of ‘Form Set’, which combines k trees (assumed to be binary) into a k -ary branching tree, for arbitrary

k . Form Set is claimed not to be Merge (Marcolli et al. 2024: 141-142), on various grounds, but Merge must apply to trees containing k -ary subtrees, so there is no clear basis for avoiding subtrees where $k = 1$. The Δ^p version of the coproduct also has a relatively simple proof of coassociativity, provided by Foissy, and automatically has an antipode, and is therefore immediately a Hopf algebra, rather than requiring some further adjustments (Marcolli et al. 2024: 36).

Another issue, noted by Annie Zaenen, is how this approach is different from TAGs, which have already been combined with glue by Findlay (2019). The answer is that TAGs derive both the semantics and the overt forms from a single structure, subtrees spanning multiple lexical items, while in both the present proposal and LFG+glue, semantics and overt form come from different structures. However the present proposal and Andrews' version of LFG+glue share with Findlay's TAGs the property that semantic interpretation applies to constellations of items that can be regarded as 'lexical items' of some kind, but lacking certain features of overt lexical items, such as linear ordering.

Appendix A: From the box notation to conventional proof-nets

A conventional proof net presents a deduction as a pair of forests of (ordered/planar) syntax trees of formulas, the premises written to the left of a '⊢' (turnstile) symbol, the conclusions to the right, with the leaves connected by axiom links, obeying certain rules. For an intuitionistic deduction, there is only a single formula on the right, and in glue, we normally want an atomic formula in that position. In the version of de Groote (1999), the nodes of these trees are decorated with polarities, and also algebraic expressions which play a role in the algebraic correctness criterion. Perrier (1999) then replaces the abstract algebraic expressions with lambda-calculus formulas, producing a semantic reading procedure as well as another version of the correctness criterion (which guarantees that a 'proof structure' that satisfies it represents a class of valid deductions, and is therefore a proof-net). Beyond these publications, more on proof-nets can be found in Troelstra (1992), Crouch & van Genabith (2000), and Moot 2002, among many other sources.

The correspondence between the present box notation and conventional proof-nets is as follows, taking premises as having positive polarity (negative is also used in the literature, although I think positive is more intuitive for semantics):

1. Each box represents an atomic formula or a 'maximal comb' of implications, positive if shaded, negative if unshaded.

2. The outermost shaded boxes are the premises, the conclusion is implicit; one can think of the shaded box with no overt link as implicitly axiom-linked to a non-overt unshaded box representing the conclusion.
3. The correspondence to conventional proof-nets is simpler if we do not use the upper-case substitution variables, but instead list the sub-boxes in the same left to right order as the arguments they represent appear for a shaded positive box, or the lambda-bound variables for an unshaded negative box.

It is also worth observing that if the basic semantic values of the boxes (assigned as ‘outer values’ to the premises in the de Groote/Perrier formulation) are just atomic symbols, the ‘semantic reading’ procedure does nothing other than implement the correctness criterion: the original boxes and links are just as good as a notation as the expressions in (linear) lambda calculus that the semantic reading procedure produces.

Appendix B: A lightning sketch of Hopf algebras

Below is a rapid sketch of Hopf algebras (actually, only of bialgebras, the last step on the way to Hopf algebras), with their application to tree-based syntax, aimed at the level of somebody who might have done basic abstract algebra, but has not used it much. For more extensive discussions at various levels, see Ardila (2012), Foissy (2013), Marcolli et al. (2024) and further references cited there.

A Hopf algebra is a vector space with some additional structure:

- (33)
- a. a ‘product’ whereby the vectors can be multiplied
 - b. a ‘coproduct’ whereby the vectors can be in a sense disassembled
 - c. a ‘unit’ which maps scalars onto vectors, allowing them to function as if they were vectors
 - d. a ‘counit’ mapping the vectors onto the scalars
 - e. an ‘antipode’, which is a rather abstract sort of inverse, and is automatically present in the Hopf algebra we will be using

All this is subject to various rules.

If a vector space has a product and a unit obeying the rules, it is an ‘algebra’, in one of several uses of this term. An algebra in this sense that all readers

would have encountered is secondary school polynomials in variables such as x , y , etc. which can be multiplied by numbers (real, rational, or complex) with the results added together (e.g. $2xy + 14z^2$). A natural basis for this vector space is all products of the variables. Unlike the bases of the geometry-based vector spaces normally introduced in undergraduate abstract algebra courses, this basis is denumerably infinite if the set of variables is denumerable (but always infinite since the variables can be multiplied by each other without limit), but the vector space should contain only finite sums of the basis vectors multiplied by the scalars (otherwise certain things become much more difficult).

In the application to Minimalism:

- (34) a. The simplest usable scalars are the rational numbers (needed for the characterization of Minimal Search).
- b. The natural basis is workspaces, most easily thought of as sets of disjoint, graph-theoretical, unordered (planar) trees, generally referred to in the mathematical literature as forests.
- c. The product is disjoint union of workspaces/forests. This is commutative, although that is not required for algebras (but true by stipulation for the secondary school polynomials).
- d. The unit maps the scalar r onto $r\mathbf{1}$, where $\mathbf{1}$ is the empty workspace (the unit of the product). This product is often written as \sqcup . The secondary school polynomial algebra is different, since the variables have no substantive product, only a formal one.

As with the secondary school polynomial example, the vectors are constituted only by finite sums.

If we add to the algebra a coproduct and a counit, we get a bialgebra; for the bialgebra of trees, the coproduct takes each tree in the workspace, and produces every way of pulling (possibly total, or empty) subtrees out of that tree, paired with the remainder that is left. Unfortunately, the pairing takes the form of a tensor, which cannot be adequately explained here, but Gowers (2024) is a popular resource.

However for a very sketchy overview, suppose we have two vector spaces V and W , with bases E and F respectively. The tensor product $V \otimes W$ is the space of vectors we can construct by using the members $\langle e, f \rangle$ as a basis (these normally written as $e \otimes f$), and imposing equalities which have the following effects:

- (35) a. The function $\Phi : V \times W \rightarrow V \otimes W$ that takes the members of $V \times W$ onto their equivalence classes as determined by the equalities is bilinear (linear in each input considered separately while the others are held constant at arbitrary values).
- b. For any bilinear function $f : V \times W \rightarrow Z$, there is a unique linear function $\tilde{f} : V \otimes W \rightarrow Z$ such that $f = \tilde{f} \circ \Phi$.

(35b) is the ‘universal property’ of the tensor product, and can be seen as a way of saying that Φ is the most general bilinear function on $V \times W$.

A useful consequence is that if a binary operation on a vector space $\circ : V \rightarrow V \times V$ is bilinear, it corresponds to a unique linear mapping from $V \otimes V$, typically written as $\tilde{\circ}$, although people tend to not bother to write the tilde, since it can be supplied as needed from the context. So one might see $\sqcup : W \otimes W \rightarrow W$ for the product of the workspace algebra, since the product has to be bilinear if thought of as from $V \times W$.

Now we can look at a very simple case, what the coproduct we use here, written Δ^p , does to the tree $[a, b]$, representing trees as square bracketings (which need to be considered as unordered):

$$(36) \Delta^p([a, b]) = 1 \otimes [a, b] + [a, b] \otimes 1 + a \otimes [b] + b \otimes [a]$$

Applied to a forest rather than a single tree, the results of all these extractions are added together, and a single extraction can remove any number of items from any number of trees, as long as none of the removed items are contained in any others (a restriction intriguingly reminiscent of the concept of ‘analyzability’ in classic Transformational Grammar).

A small increase in the complexity of a tree or forest produces a big increase in the complexity of the coproduct, which needs to obey a condition of ‘coassociativity’, making the entire thing rather confusing; we will leave further exposition to the mathematicians. The counit also has to obey a compatibility condition with the coproduct, and if all the rules are satisfied, the result is a bialgebra. Finally, if there is a further item called an ‘antipode’, which exists automatically for the bialgebra of unordered trees with Δ^p as the coproduct (some of the other coproducts discussed in Marcolli et al. (2024) need some adjustments to get an antipode), the result is a Hopf algebra. However, so far, the antipode plays no role in this formalization of syntax.

Acknowledgments

The connection of this paper to Dick Crouch's work is that he has been a leader in introducing new ideas from logic and mathematics to linguistics, and here I try to apply one of the things he has worked on, glue semantics, to this new recent introduction.

I would like to acknowledge the very helpful comments of Ash Asudeh, Annie Zaenen, and an anonymous reviewer. I have also benefited from from being able to present this material to an online seminar conducted by Andras Kornai in 2023. Errors and obscurities are of course entirely mine.

References

- Andrews, Avery D. 2007. Generating the input in OT-LFG. In Jane Grimshaw, Jane Simpson, Tracy Holloway King, Joan Maling, Christopher D. Manning & Annie Zaenen (eds.), *Architectures, rules, and preferences: A festschrift for Joan Bresnan*, 319–340. CSLI Publications.
- Andrews, Avery D. 2008. The role of PRED in LFG+glue. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG08 conference*, 46–76. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2008/lfg08.html>.
- Andrews, Avery D. 2010. Propositional glue and the correspondence architecture of LFG. *Linguistics and Philosophy* 33. 141–170.
- Andrews, Avery D. 2019. A one-level analysis for Icelandic Quirky Case. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG19 conference*, 24–27. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2019/index.shtml>.
- Ardila, Frederico. 2012. *Hopf algebras*. <https://fardila.com/Clase/Hopf/lectures.html>.
- Asudeh, Ash. 2005. Control and semantic resource sensitivity. *Journal of Linguistics* 41(3). 465–511.
- Asudeh, Ash. 2022. Glue semantics. *Annual Review of Linguistics* 8. 321–341.
- Asudeh, Ash. 2023. Glue semantics. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 651–697. Language Science Press. DOI: 10.5281/zenodo.10185964. <https://doi.org/10.5281/zenodo.10185964>.

- Asudeh, Ash & Richard Crouch. 2002a. Coordination and parallelism in glue semantics: Integrating discourse cohesion and the element constraint. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG02 conference*, 19–39. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2002/lfg02.html>.
- Asudeh, Ash & Richard Crouch. 2002b. Derivational parallelism and ellipsis parallelism. In Line Mikkelsen & Christopher Potts (eds.), *WCCFL 21 proceedings*, 1–14. Cascadilla Press.
- Bernódusson, Helgi. 1982. *Ópersunólegar Setningar [impersonal sentences]*. University of Iceland. (MA thesis).
- Chomsky, Noam, T. Daniel Seely, Robert C. Berwick, Sandiway Fong, M.A.C. Huybrechts, Hisatsugu Kitihara, Andrew McInnerney & Yushi Sugimoto. 2023. *Merge and the strong Minimalist thesis*. Cambridge University Press.
- Crouch, Richard & Josef van Genabith. 2000. *Linear logic for linguists*. <http://www.coli.uni-saarland.de/courses/logical-grammar/contents/crouch-genabith.pdf>.
- Dalrymple, Mary (ed.). 1999. *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*. The MIT Press.
- Dalrymple, Mary. 2001. *Lexical Functional Grammar*, vol. 34 (Syntax and Semantics). Academic Press.
- Dalrymple, Mary, Vineet Gupta, John Lamping & Vijay Saraswat. 1999. Relating resource-based semantics to categorial semantics. In Mary Dalrymple (ed.), *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*, 261–280. The MIT Press.
- Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell III & Annie Zaenen (eds.). 1995. *Formal issues in Lexical-Functional Grammar*. CSLI Publications.
- Dalrymple, Mary, John Lamping, Fernando Pereira & Vijay Saraswat. 1997. Quantification, anaphora and intensionality. *Logic, Language and Information* 6. Appears with some modifications in Dalrymple (1999), pp. 39–90, 219–273.
- de Groote, Philippe. 1999. An algebraic correctness criterion for intuitionistic multiplicative proof-nets. *Theoretical Computer Science* 224(1-2). 115–134.
- Findlay, Jamie Y. 2019. *Multiword expressions and the lexicon*. Oxford University. (Doctoral dissertation).
- Foissy, Loïc. 2013. *An introduction to the Hopf algebras of trees*. <https://www2.mathematik.hu-berlin.de/~kreimer/wp-content/uploads/Foissy.pdf>.
- Gotham, Matthew. 2018. Making logical form type-logical: Glue semantics for Minimalist syntax. *Linguistics and Philosophy* 41(5). 511–556.

- Gotham, Matthew. 2021. Approaches to scope islands in LFG+glue. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG21 conference*, 146–166. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2021/lfg21.html>.
- Gowers, Tim. 2024. *How to lose your fear of tensor products*. Undated webpage accessed in 2024. <https://www.dpmms.cam.ac.uk/~wtg10/tensors3.html>.
- Halvorsen, Per-Kristian & Ronald M. Kaplan. 1988. Projections and semantic description in Lexical-Functional Grammar. In *Proceedings of the international conference on fifth generation computer systems*, 1116–1122. Also published in Dalrymple, Kaplan, Maxwell & Zaenen (1995), pp. 279–292. Institute for New Generation Computer Technology.
- Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40(3–4). 225–276.
- Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Blackwell.
- Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30. 69–96.
- Jackendoff, Ray. 1973. The base rules for prepositional phrases. In Stephen R. Anderson & Paul Kiparsky (eds.), *A festschrift for Morris Halle*, 345–357. Holt, Rinehard & Winston.
- Kay, Paul & Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The what’s X doing Y? construction. *Language* 75(1). 1–33.
- Kokkonidis, Miltiadis. 2008. First-order glue. *Journal of Logic, Language and Information* 17. 43–68.
- Lakoff, George. 1970. Global rules. *Language* 46. 627–639.
- Lamarche, Francois. 1994. *Proof nets for intuitionistic linear logic 1: Essential nets*. Technical Report, Imperial College London.
- Marantz, Alec. 1984. *On the nature of grammatical relations*. MIT Press.
- Marculli, Matilde. 2023a. *A mathematical model of syntactic merge*. Lecture at University of Utrecht, June 25, 2023. <https://www.youtube.com/watch?v=JPdjRaU4JNU&t=2027s>.
- Marculli, Matilde. 2023b. *A mathematical model of syntactic merge*. Lecture series at MIT. https://www.youtube.com/watch?v=7Gtpc_E1Tfw.
- Marculli, Matilde. 2024. *Mathematical models of generative linguistics (Ma191c spring 2024)*. CalTech course materials. <https://www.its.caltech.edu/~matilde/LinguisticsMa191Spring2024.html>.
- Marculli, Matilde, Robert Berwick & Noam Chomsky. 2023. *Old and new Minimalism: A Hopf algebra comparison*. <https://arxiv.org/abs/2306.10270>.
- Marculli, Matilde, Noam Chomsky & Robert Berwick. 2023a. *Mathematical structure of syntactic merge*. <https://arxiv.org/abs/2305.18278>.

- Marcolli, Matilde, Noam Chomsky & Robert Berwick. 2023b. *Syntax-semantics interface: An algebraic model*. <https://lingbuzz.net/lingbuzz/007696>.
- Marcolli, Matilde, Noam Chomsky & Robert Berwick. 2024. *Mathematical structure of syntactic merge*. Unpublished manuscript, CalTech. <https://www.its.caltech.edu/~matilde/MergeMCB-MITPress-LI.pdf>.
- Moot, Richard. 2002. *Proof-nets for linguistic analysis*. University of Utecht. (Doctoral dissertation).
- Mycock, Louise. 2006. *The typology of constituent questions: A Lexical-Functional Grammar analysis of 'WH'-questions*. University of Manchester. (Doctoral dissertation).
- Partee, Barbara H. 2009. Do we need two basic types? *Snippets: Special issue for Manfred Krifka* 20. 37–41.
- Partee, Barbara H. & Vladimir Borschev. 2004. Genitives, types, and sorts. In Jiyung Kim, Yury A. Lander & Barbara H. Partee (eds.), *Possessives and beyond: Semantics and syntax*, 29–43. UMass GSLA. <http://people.umass.edu/partee/Research.htm>.
- Perrier, Guy. 1999. Labelled proof-nets for the syntax and semantics of natural languages. *Logic Journal of the IGPL* 7. 629–654.
- Stabler, Edward P. 2010. Computational perspectives on Minimalism. In Cedric Boeckx (ed.), *Oxford handbook of linguistic minimalism*, 616–641. Oxford University Press.
- Troelstra, Anne S. 1992. *Lectures on linear logic*. CSLI Publications.
- Wood, Jim & Alec Marantz. 2017. The interpretation of external arguments. In Roberta D'Alessandro (ed.), *The verbal domain*, 255–278. Oxford University Press.

Part II

Reasoning and Discourse

Chapter 6

The Linguistic Discourse Model at 40

Livia Polanyi^a & Martin H. van den Berg^a

^aIndependent Scholar

In this paper, we report upon advances to the Linguistic Discourse Model (LDM). The LDM is a theory of the syntactic structure of discourse modeled as a parser that takes in linguistically and paralinguistically encoded semiotic signs exchanged in an Interactional Context by (real or modeled) participants building a discourse history with one another by communicating sequentially organized elementary pieces of conceptual information holding in some set of contexts. In the course of parsing, a representation of the discourse history is constructed in the form of a Discourse Parse Tree (DPT) constructed in accordance with the Right Frontier Constraint. The nodes of the DPT are Discourse Constituent Units (DCU) created through the process of computing the relationship of elementary Basic Discourse Units (BDUs) to DCUs already present on the right edge of the DPT using explicit rules. The LDM takes the point of view of the recipient and uses the information in the encoding form of BDUs as well as information about the nature and structure of higher level socially recognized and genre units to build the DPT. In doing so, it has no access to the intentions of the speaker or to the complicated, under-specified, and often contradictory “coherence” rules relied upon in S-DRT and Rhetorical Structure Theory. An account is given for why the Right Edge Constraint holds and a revision of some aspects of discourse parsing are suggested.

1 Introduction

The Linguistic Discourse Model (LDM) (Polanyi & Scha 1984, Polanyi 1985, 1988, 1996, Polanyi et al. 2003) is a theory of the syntactic structure of discourse, the level of language above the sentence. The LDM provides an account of discourse continuity by making explicit the underlying mechanisms of human language that determine if further development of a semantic context begun earlier can

Livia Polanyi & Martin H. van den Berg. 2025. The Linguistic Discourse Model at 40. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 111–136. Konstanz: PubliKon.

continue or not. Accounting for continuation creates a dynamic understanding of “where we are at” in talk or text and provides information vital to understanding the constraints determining interpretation of indexical expressions and other forms of reference to times, places and, critically, the form of anaphoric reference to previously introduced entities or coordinates.

Discourse is inherently interactive whether it emerges from several adults discussing foreign policy, a human-robot task session, a baby and mother at play, two parakeets mating, an author sitting in solitude writing an article, or a reader working their way through a mystery novel.¹ Each interaction establishes a common context in which participants encode, transmit, receive and decode meaningful signals directed at one another through gesture, speech or writing either face to face or physically or digitally removed. These exchanges of meaningful signals during interaction create *discourse histories* in which participants know what the current topic is and which topics that may have been suspended are available to develop further without re-establishing the temporal, spatial and entities involved.

Discourse is a formal term in our work and is defined as:

A phenomenon that emerges through real or modeled behavior among two or more entities at least one of whom believes them to be in a state of interaction with one another and that they are constructing a discourse history together through the exchange of meaningful linguistic and/or non-linguistically encoded signals.

The LDM, formalized as a discourse parser, enables the assignment of a structural representation to a discourse history as it dynamically unfolds. The parser takes as input sequentially ordered elementary semiotic signals encoded in language or gesture exchanged among interactants and, as the discourse history unfolds, analyzes their relationship to previously analyzed signals in order to assign a structural description to the emerging interaction in the form of a *discourse parse tree*. The tree is constructed in accordance with *Parsing Rules* based on *Discourse Relations* the LDM specifies.

¹In the case of written communication, the interaction is “modeled” in the sense that the author is addressing the message to a modeled reader, someone who may or may not exist but who nonetheless can be addressed as “you”. Similarly, in first person fictional accounts, the “I” used felicitously in the text need not refer in any way to the “I” who created it nor, in fact, could the temporal deictic “now” ever refer to a moment in the real world shared by the author and reader. Rather “now” in any written text always refers to a modeled moment in some modeled time/space location which convention decrees be interpreted as a time point shared by both author and recipient.

In this paper, we present an overview of the model. In section 2, the elemental units of discourse, functions and Basic Discourse Units are introduced. In section 3, an overview of LDM discourse parsing is presented followed by an example of an LDM parse in section 4. The paper continues in section 5 with a discussion of challenges to the present formulation of the theory presented by interruptions of one discourse history by another whether by participants attending to other matters in their immediate environment or by using direct speech to report what was said in another context. We conclude with a few observations in section 6.

2 Basic Discourse Units

In linguistics we are used to dealing with units both of content and function. Nouns and verbs typically encode content while connectives (“and”, “or”, “but”), determiners, vocatives, interjections, and logical operators and coordinating and subordinating conjunctions are functional units normally used to indicate relationships between the denotations of different constituents. In the LDM, we have adopted this distinction to characterize two types of units in an input string:

- *Basic discourse units (BDUs)* are the minimal units of content encoded in gesture or language indexed for context of occurrence. The content of a BDU is encoded in terms of dynamic semantics (Kamp 1981, Heim 1988, van den Berg 1996, Asher & Lascarides 2003). The context indices of BDUs include precise identification of the unique interactional, social, performative, syntactic, and semantic contexts in which they occur (see (1) below).
- *Function units* encode meta information about how previously occurring (or possibly subsequent) linguistic gestures relate structurally, semantically, interactionally or rhetorically to other units in the discourse or to information in or out of the context in which the discourse takes place.

An important class of English functional terms including “anyway”, “ok”, and “so” uttered with marked prosody are used to indicate shifts in discourse structural development, indicating that part of a discussion is completed or that a previously interrupted discussion will be resumed.²

²Some apparent content bearing phrases e.g. “as I was saying”, “getting back to what we were talking about” and “on the other hand” are also functional operators. At this point, it is fair to point out that implementations of planning discourse understanding systems based on Discourse Structures Theory (Grosz & Sidner 1986, Litman & Hirschberg 1990) and Rhetorical Structure Theory (Marcu 2000) relied heavily on “cue phrases” to segment discourse and build

Syntactically, linguistically encoded BDUs include clauses and other verb based structures that carry both content and indexical information explicitly tying content to a context of interpretation providing links for discourse reference and continuation. Linguistic structures such as gerunds, nominalizations and clefts while verbal in nature, do not independently carry enough indexical information to establish or access contexts available to be “continued” are thus not BDUs but are included in the BDU to which they are syntactically attached.

The core values of the indices are set by the interactional context defined relative to participants, time/place of interaction remaining constant throughout the construction of the discourse history. Social framing of interaction in Speech Events—conventional organized types of interaction often composed of sequentially organized constituent activities that assign roles to participants (Hymes 1974)—also serve to segment discourse into utterances belonging to an ongoing Speech Event activity or Speech Event constituent. Text or genre unit information also serves to segment a discourse history and can influence the encoding form used to refer to times, place and entities.

(1) BDU Feature Set

- Kaplan Contexts: Interactants, time, place
- Speech Event, Speech Event Constituent
- Genre constituent
- Surface string: surface ordering of words
- Phonological and phonetic features
- Lexicon and morphology
- Performance realization features: Prosody/Typography
- Gesture, facial expression, kinesics, proxemic features etc.
- Text internal indices: Genericity, modality, point of view, tense, aspect, factivity etc.
- Propositional content
- Syntactic structure

tree representations of them. Cue phrases, largely adopted from the book *Discourse Markers* (Schiffrin 1987), while useful are unfortunately of limited use since they occur relatively infrequently as do discourse level logical connectives (Miltsakaki et al. 2004). Discourse syntax, which in our view includes syntactically complex sentential information, makes use of the full linguistic and paralinguistic portfolio of lower level language structures as well as interactive performative features to construct a representation of how they relate to one another to enable anaphoric representation and interactive communication to take place.

- Semantic interpretation
- Information structure
- Topic/Focus/Background articulation
- Sentiment

During discourse parsing, a structural representation of the discourse history of all BDUs assigned to a given interaction by these indices is constructed. One component of this representation is the compositional constructed meaning of the discourse, but it is important to point out that that meaning is just one component of the representation and does not encode the structure of the discourse, as it does in DRT (Kamp 1981) based approaches to discourse analysis. In this paper, we will in general not discuss this semantic component, although some lexical semantic information is necessary in the discussion of the discourse construction rules in section 3.3.

3 Discourse parsing

The LDM constructs a hierarchical encoding of the discourse constituent structure. There is a fundamental difference between parsing sentences and parsing a discourse. A sentence is given to us in its entirety; it is a *bounded* object that we can inspect as a single complete unit and permits integration of non-contiguous constituents. A discourse, on the other hand, is a continuous flow of sequential informative structures. It is by its very nature *unbounded*. The parser cannot “see” the eventual end of the discourse history, because it has not occurred yet.³

This means the parsing has to be implemented by incrementally analyzing each BDU in order of occurrence. At any moment during parsing, the discourse history is a completely analyzed object. The structural analysis cannot depend on incoming information that may or may not arrive. The analysis of the discourse history must be able to stand on its own.

Unlike sentences in which all constituents are available to the parser from the beginning of the parsing process and the parse is therefore only complete when each constituent has been assigned its proper structural relationship to every

³Of course, as we know well from conversation, an utterance can always be added at what might be thought to be the end of any discussion. Similarly, interactants may add information about a topic that seems to no longer be relevant because the flow of talk has moved on. What is key to point out is that the placement of that information will be marked explicitly as “off topic” by the encoding form used including re-nominalization, change in prosody and other paralinguistic behavior, topic/focus and the use of functional operators.

other constituent, the parse of a discourse is structurally “complete” at every moment in the parsing process of the interaction.

Each new BDU is added to the end of the discourse history which consists of all analyzed constituents that end where the discourse ends. A new BDU can only be added to those final constituents of the discourse so far. Let’s illustrate what it means to be at the end of the discourse history with a simplified example.⁴ This example was chosen because each sentence provides additional information about some aspect of the content of the immediately preceding sentence. Each sentence consists of one clause and can be interpreted as a single BDU.

- (2) ^{a1}Japanese people eat noodles all the time. ^{a2}They eat them for lunch or for a light snack. ^{a3}They are served hot in soup or cold like a salad. ^{a4}When they are served in soup, ^{a5}vegetables, tofu and meat may be added.

For the moment, assume the parser encodes the elaboration of one unit on another by creating a constituent containing both. This leads to the following constituent structure.

- (3) ^{a1}[Japanese people eat noodles all the time. ^{a2}[They eat them for lunch or for a light snack. ^{a3}[They are served hot in soup or cold like a salad. ^{a4}[When they are served in soup, ^{a5}[vegetables, tofu and meat may be added.]^{a5}]^{a4}]^{a3}]^{a2}]^{a1}

There are several constituents (marked by named brackets *a1...a5*) at the end of the discourse history. In fact, we could continue each constituent with different additional elaborations or continuations, such as these:

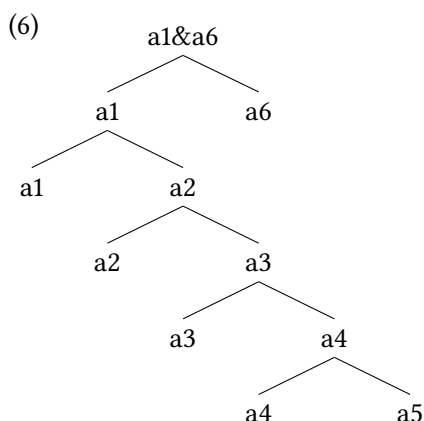
- (4) a. An egg is also often added (adding to *a5*)
b. When served as a salad, they often come with a dipping sauce. (added to *a4*)
c. There are different kinds of noodles (added to *a3*)
d. They are also sometimes eaten for dinner (added to *a2*)
e. Chinese people also eat noodles (added to *a1*)

⁴Example text adapted from a slightly longer text that was published on the Web as a student webpage produced in the Fall 2000 class, “Introduction to Japanese Culture” given by Professor Shigeru Miyagawa at MIT.

In the original text, the next sentence was in fact (4.e), and the resulting constituent structure ended with only two constituents, (4.e), the last unit that was just added and the whole discourse (a_0 below).

- (5) a_0 [a_1 [Japanese people eat noodles all the time. a_2 [They eat them for lunch or for a light snack. a_3 [They are served hot in soup or cold like a salad. a_4 [When they are served in soup, a_5 [vegetables, tofu and meat may be added. a_5] a_4] a_3] a_2] a_1 a_6 [Chinese people also eat noodles] a_6] a_0

In other words, after attaching (4.e), if we were to attach a further sentence, it would either have to be about both Japanese and Chinese noodle eating practices, or Chinese noodles only. Of course, continuation about other noodle eating cultures or about something altogether different is also possible. This bracketing can be encoded as a tree structure, and the constituents that are open for attachment are those that are on the right edge of the tree.



Looking at the tree, it becomes clear that only elements that are on the right edge of the tree, corresponding to spans that are still open, are available for attachment. This is the explanation for the Right Edge Constraint originally proposed by Polanyi in 1985 which is one of the basic assumptions of the LDM and most other theories of discourse structure (Polanyi 1985, 1988, Grosz & Sidner 1986, Marcu 2000).⁵ It is a direct consequence of the fact that discourse is comprehensible at any time during production, ruling out insertion after the fact of elements somewhere in the middle.

⁵The right-edge constraint has been widely discussed in discourse parsing literature (Sassen & Kühnlein 2005, Cristea 2005, Webber 2006, Holler & Irmen 2007, Asher 2008).

3.1 Discourse relations

Unlike Rhetorical Structure Theory (Mann & Thompson 1988, Marcu 2000), Structured Discourse Representation Theory (Asher & Lascarides 2003) or Discourse Structures Theory (Grosz & Sidner 1986), the tree is not an encoding of rhetorical relations or the “intentions” of a speaker but is simply a representation of the syntactic structure of discourse. The LDM discourse parse trees consist of labeled nodes, each of which is annotated with the syntax and semantics of the utterance recursively constructed from the information in the child nodes according to the node-type assigned during parsing.

There are only three basic node types: *Subordination* (S), *Coordination* (C) and *N-ary* (N).⁶ Each node has a full set of features associated with it constructed from the features of the children. While the leaf nodes of the tree are BDUs, these constructed nodes are referred to as *Discourse Constituent Units* (DCUs).

Subordinations encode interruptions to actual or possible coordinations, include elaborations in which the daughter nodes provides more information about the parent subordinating node, explanations, examples, parentheticals and other comments and asides. Subordinated BDUs are attached as daughter nodes under the interrupted element.

The DCU to which the subordination is attached is annotated with the information present at the node to which it is subordinated. The meaning of the subordinated node is not added to the dominating S-node.

Coordinations are list structures. In coordinations all children bear the same relationship to the parent node. Narratives are a special subtype.

The meaning of a C node is calculated from the meanings of all the children and the features of the C node include all information the child nodes have in common including all contextual indices.

N-aries, are language and culture dependent unlike subordinations and coordinations. They are ad-hoc constructions that differ from language to language and from social group to social group. N-aries include the re-analysis of complex sentences as mini-discourses calculated from the meanings of all the children, if-then structure, sentences with preposed modifiers, conventional expressions

⁶Besides these basic node-types, we will sometimes distinguish operator (O) nodes. Operators are non-propositional elements such as connectives (“and”, “or”, “but”) and elements like “if”, “when”, “during the vacation” that inform the construction and choice of the constructed node. Although we sometimes explicitly show them in the tree, they are more often incorporated in one of the leaf nodes. So “and” may just give rise to a coordination, and be left as part of the utterance that follows it, but “during the vacation” will be annotated. This is done to make the trees manageable, not to make any theoretic point.

that are bounded in their scope like question-answer pairs and greetings as well as Speech Events and Genre Units.

N-aries are bounded units with defined constituents that are described by their own grammatical rules.

3.2 The LDM Discourse Parse Tree

LDM Discourse Parse Trees (DPT) are recursively constructed by analyzing utterances and attaching them to the tree at some node at the right edge of the existing tree. In practice, when the discourse consists of sentences that may be more complex, the sentences are first parsed and then attached to the discourse tree by the main clause. The steps are as follows (cf. Polanyi et al. 2004b).

1. Parse an incoming sentence using a sufficiently rich syntactic theory (we have used LFG based parsing before Kaplan & Bresnan 1982, Butt et al. 1999).
2. Transform the resulting sentential parse tree into a small discourse tree.
3. Attach that discourse tree to the right edge of the preceding parse tree. This involves (a) deciding where to attach, and (b) what kind of attachment this is (see 3.1 below).⁷
4. Assign appropriate semantic representations to the new nodes.

The result of the first two steps is a chunking of the sentence into Basic Discourse Units (BDUs) which contain atomic propositional information, and Discourse Operators (DOs), which are non-propositional connectives, modals, and others, and combining them into a small sentential discourse tree. The third and fourth steps incorporate the sentential discourse tree onto the right edge of the existing tree. In Polanyi et al. (2004b), an algorithm was proposed for choosing the location of the attachment point on the parse tree, for determining the node-type and for interpreting the newly attached node as part of the resulting tree.

⁷We should point out that initially in designing the discourse parsing algorithm we were flummoxed about how to treat complex sentences and developed a set of procedures that took the sentence as input and using rules of discourse parsing such as those described below, built a small discourse tree and attached the BDUs making up the sentence to the DPT incrementally. Dick Crouch, observing the convoluted processing pointed out that if the sentence were parsed with a sentential parser and the main clause attached to the DPT at an appropriate node, any subordinated clauses would already be in the correct structural position and those which were not blocked by island constraints would be automatically available along the right edge of the tree where they would be available to serve as points of attachment to incoming BDUs.

Note that there is some superficial similarity between discourse parsing in the LDM and semantic discourse parsing in DRT (Discourse Representation Theory: Kamp (1981), Kamp & Reyle (1993), van Eijck & Kamp (2011)), where first a (preliminary) DRS (Discourse Representation Structure) for an incoming sentence is calculated, and then merged with the DRS of the preceding text.⁸ The difference is that in DRT only one overall structure is constructed, the DRS of the whole text. This DRS encodes both the syntax and the semantics of the discourse, as well as often pragmatic information.⁹

In contrast, the LDM parse tree is similar to a derivation tree. Each node preserves the still active (accessible) parts of the semantic representation of the underlying span and corresponding contextual indices, decoupling syntactic structure (which remains a tree) from more complex semantic structure like long-distance binding of anaphors and contextual restrictions of quantifiers.

It is important to point out the DCUs are all first class computable nodes that all have at least some of the features as BDUs. In fact, there is no reason to distinguish BDUs which are the leaf nodes of the tree from constructed nodes on the tree and in the rest of this discussion we will use the term DCU to include BDUs (DCU_{new}) and will distinguish a new node DCU_c from an existing node on the right edge DCU_{re} .

3.3 Parsing rules

Parsing a discourse consists of sequentially attaching each DCU_{new} to the right edge of the tree. As pointed out in 3.2, this involves two steps. Deciding to which node on the right edge to attach and determining the content of the node selected. To select an attachment point, we must scan each DCU_{re} and compare the indexical and content information of each node beginning with the Kaplan Contexts to the Kaplan Context of the incoming unit,¹⁰ Speech Event, Speech Event constituent, and then Genre Units and Genre Constituent indices are scanned. If they do not match, the grammars associated with each will determine if these DCUs have been completed, and thus attachment will be to a node dominating them or, if not completed, attachment will be to a unit subordinated to it. Once this process is complete and candidate nodes have been chosen, the indexical and

⁸Cf. van der Sandt (1992) for clear examples of this process.

⁹Cf. Asher (2008) for a discussion on why in S-DRT, which encodes structural, semantic, rhetorical and pragmatic information in one representation of the discourse, the right frontier is problematic.

¹⁰Interactional indices that do not match are not part of the construction of the same discourse history and thus will not be attached to one another.

lexical semantics, plus all aspects of the encoding forms and information structure of the candidates are compared to the DCU_{new} and a node dominating the incoming and previously attached DCU will be chosen or created. For details of the latter, see Polanyi et al. (2003, 2004b).

The following rules describe the result of creating a new node DCU_c from an existing node on the right edge DCU_{re} by attaching a new DCU_{new} to it.

3.3.1 Subordination

Subordinations are nodes where a subordinated (embedded) constituent is interpreted in the context of a subordinating (governing) node but does not contribute to the parent node. These includes elaborations, parentheticals, interruptions, and examples.

The meaning of an S node is that of the subordinating node. The information in the subordinated node is not available in the S node. In other words, the newly constructed node DCU_c will inherit the features of DCU_{re} . Because the information in a subordinated node is not merged with the subordinating node, the subordinated node need not inherit the complete context of the subordinator. Some key rules for constructing subordinate nodes are as follows (the subordinated parts are underlined for ease of understanding).

- The DCU_{new} has a different modality than the DCU_{re} , e.g. the DCU_{new} is realis and DCU_{re} is irrealis, or the DCU_{new} is irrealis and DCU_{re} is realis.

John could have been a writer. He is the winner of two amateur contests.

- The DCU_{new} is generic and DCU_{re} is specific or the DCU_{new} is specific and DCU_{re} is generic.

Japanese people eat noodles. My friend Toru had Udon yesterday and liked it a lot.

- The grammatical subject of the DCU_{new} is the grammatical object of the DCU_{re} (“sentence promotes, discourse demotes”).

Japanese people eat noodles. Noodles are often served in soup.

- Grammatical roles of the DCU_{new} are subcases of the parallel grammatical roles of the DCU_{re} .

John and Mary are great mathematicians. John works on differential algebra.

- The subject of the DCU_{new} is part of the subject of the DCU_{re} .
The car seems to have a problem. The tire looks a little flat.
- The DCU_{new} is a parenthetical or footnote, and the DCU_{re} does not have any daughters. Parentheticals are attached at the bottom of the right edge.
John came over yesterday. What an idiot. He usually comes on Wednesdays.

3.3.2 Coordination

Coordinations are subdivided into lists of states of affairs or attributes and Narratives.

Recognizing and Parsing Lists

- The units have similar predication of their subjects.

John likes apples. He hates pears.

The resulting DCU_c will identify the subject discourse referent and introduce an implementation dependent encoding of what the units have in common. We will use the notation [John likes apples]⊕[He hates pears] to denote the resulting representation at the C-node.¹¹ With the right ontology, the result is [*John ±Like FRUIT*].

- The units have verbs that are synonyms or antonyms, or the DCU_{new} has a VP anaphor compatible with the verb of the DCU.

John likes apples. Mary likes pears.

The constructed DCU_c will introduce discourse referents for the ordered sets {*John, Mary*} and {*apples, pears*}, and will establish a liking relationship between the parts. The resulting coordinating constructed DCU_c will have a semantic representation that conjoins the representations of the meanings of the DCU_{re} and DCU_{new} under a C-node that includes all indices and content information they have in common. With a similar ontology as before, this results in [John likes apples]⊕[Mary likes pears] = [PERSON likes FRUIT] as the semantic content of the C node.

¹¹This operator is called the *most specific common denominator* and constructs a result that preserves as much as possible of the combined contents of its arguments (van den Berg & Prüst 1990).

Recognizing and parsing narratives A narrative is a list of eventive clauses in which the order of presentation in the flow of discourse recapitulates the ordering of events in the semantic representation of the text (Labov & Waletzky 1965, Labov 1972, Polanyi 1989). The units of the list each express specific events with identical tenses. The resulting constructed DCU_c is annotated with a semantic representation that conjoins the representations of the meanings of the DCU_{re} and DCU_{new} , and additionally imposes a condition such that the event of the DCU_{new} is the next event after the event corresponding to the DCU_{re} .

3.3.3 N-ary

N-aries are subdivided into inter-sentential and extra-sentential units with two or more constituents required for the N-ary to be complete. The nature and order of the constituents required to create a complete N-ary are specified by discourse level grammatical rules and may be flagged by external clues, whether actual grammatical constructions in the intra-sentential cases, or specific forms of expression or actions for inter-sentential Interactional, Speech Event and Genre units. How constructed nodes are formed is defined in all cases by the actual meaning of the N-ary combination.

Recognizing and Parsing Intra-sentential N-aries

- Preposed constructions that position the discourse in time, space or other dimension.

During the Middle Ages, Reims was an important city.

While in Paris, John had a lot of fun.

According to Bob, the capital of Italy is Milan.

In these cases, the constructed DCU_c has an accessible semantic content that is identical to that of the main clause with contextual indices set by the modifying preposed unit.

- Logical connectives within a sentence

If John arrives before six, he can join us for dinner.

Recognizing and Parsing Inter-sentential N-aries

- Question/Answer

Who came to the party yesterday? Bill did.

The resulting constructed node DCU_c has as its accessible semantic content the combination of question and answer, i.e. *Bill came to the party yesterday.*

- Repair

John is coming tomorrow. Sorry, today.

Repair N-aries are a meta-instruction requiring replacement of an update to a semantic representation of a DCU_{re} by an update deriving from a DCU_{new} . “Sorry” is a functional operator signaling the Repair. The semantic content of the resulting DCU_c will be equivalent to *John is coming today.*

- Logical connectives

Let’s assume John can come. He will bring the cake.

- Greetings and other Adjacency Pairs (Schegloff & Sacks 1973)

Hi John, how are you doing. I’m great, how are you?

Greetings set the participants of an interaction, and set up the relevant indices so the DPT can be constructed. The constructed DCU_c is an essentially empty unit with all indices set to the right values.

- Genre Meta Structure (Story, Technical Paper, Lecture, etc.): Different genres can have different specialized rules associated with them, and the weights of rules might be a function of the genre too. However, the start of a discourse in a particular genre itself implies the setting of relevant indices and the expectation of a particular structure. For example, a technical paper N-ary will consist of a number of daughters, including title material, abstract, several sections, conclusions, and references (some of which may be optional).
- Speech Events have specialized rules of formation as well.

A Doctor-Patient Interaction which assigns the hierarchical roles of Doctor and Patient to participants with incumbent responsibilities and deference requirements has the form:

- *Greetings*
- *Problem*
- *Introduction*
- *Examination*

- *Discussion of Diagnosis*
- *Further Action*
- *Parting*

While a Storytelling constituent of a conversation has three parts (Polanyi 1989):

- *Entrance talk* in which a content relative to a topic currently on the floor or relevant to the concerns of the interactants is introduced relative to a past time potential story world (*I had a fabulous time at work today.*)
- *Story* describing events in the past time (fictional or non-fictional content) and
- *Exit Talk* in which the events in the story world are made relevant to the *current interactional context*.¹²

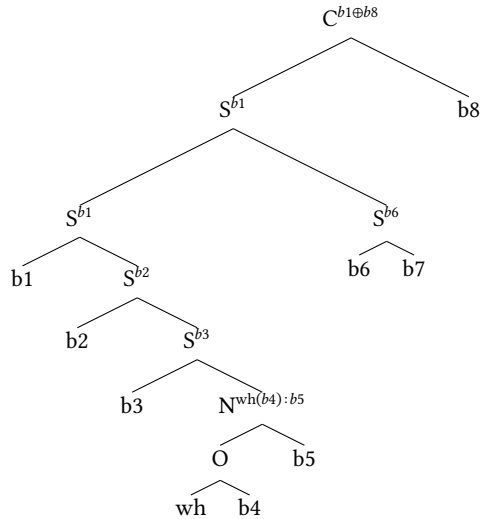
4 Putting it all together

We can now return to a version of our original example, and show how it is actually analyzed.¹³

- (7) ^{b1}Japanese people eat noodles all the time. ^{b2}They eat them for lunch or for a light snack. ^{b3}They are served hot in soup or cold like a salad. When ^{b4}they are served in soup, ^{b5}vegetables, tofu and meat may be added. ^{b6}They eat several types of noodles in Japan. ^{b7}Udon, for example, are thick white noodles made from wheat flour. ^{b8}Chinese people eat a lot of noodles, too.

¹²The Penn Discourse Treebank (Miltsakaki et al. 2004) provides a rich corpus of data to investigate using ML techniques to determine surface features that could be used to identify the constituents of n-ary structures that would otherwise have to be inferred.

¹³An in-depth analysis of a slightly different version of this text is found in Polanyi et al. (2003).



The tree is interpreted as follows.

1. After the first sentence ($b1$) introduces noodle eating in general, $b2$ gives an example of when those noodles are eaten. It constitutes a subordination specifying information about instances of noodle eating. S^{b1} and the superscripts on the nodes in the tree denote the content that is still available at that level.
2. The BDU $b3$ promotes the object of the previous sentence to subject, which according to a rule of discourse parsing triggers a subordination, and gives two instances of how noodles are served.
3. The operator “when”, marked explicitly on the tree to highlight it,¹⁴ selects, through $b4$, one of the two cases from $b3$ and gives additional information ($b5$, again a subordination).
4. The next sentence $b6$, like $b2$, elaborates on $b1$, constituting an independent subordination to it.
5. $b7$ gives information about Udon, one type of noodles eaten in Japan, resulting in $b7$ to be subordinated¹⁵ to $b6$.
6. Finally, $b8$ pops all the way up to the main level and is coordinated to $b1$. The resulting C-node has as its semantic contents $b1\oplus b8$, which, depending

¹⁴Cf. footnote 6.

¹⁵a fact reinforced by the “for example”, an operator we left out of the tree for clarity

on the implementation, will correspond to something like *Asian people eat Noodles*

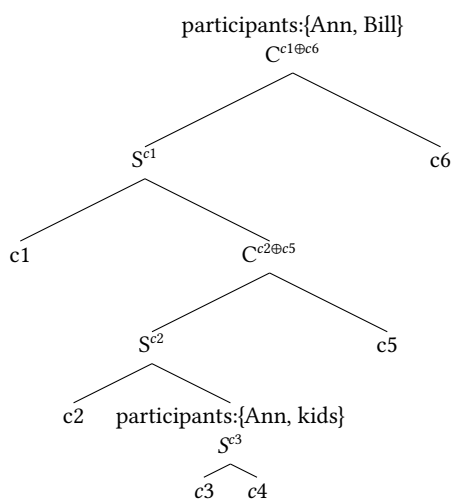
5 Continuing previous embedded contexts: A challenge to the LDM

In this section, we discuss two cases that illustrate problems for a naive application of the rules we have. In section 5.1 we consider the case of interruptions. What should an analysis look like if there is a sudden change of participants during a discourse? In section 5.2, we point out some difficulties that arise when quoting one discourse from within another.

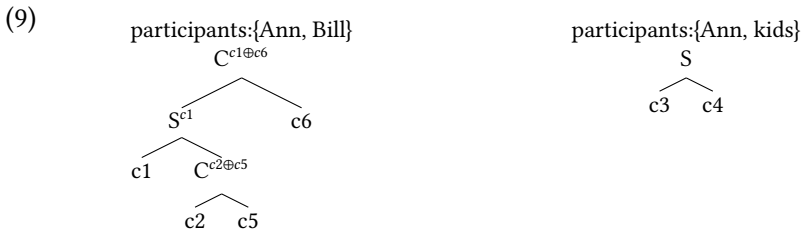
5.1 Changing the context

DPTs are interpreted relative to a particular context, characterized by indices like (story-) world, time and place, participants, roles of the participants, etc. When nodes are added to the tree, they are interpreted in the context of the node they are attached to. As a rather extreme example, take the following conversation (adapted from Polanyi & Scha (1983), Polanyi (1996)), of a conversation between Ann and Bill that is interrupted by Ann admonishing some children.

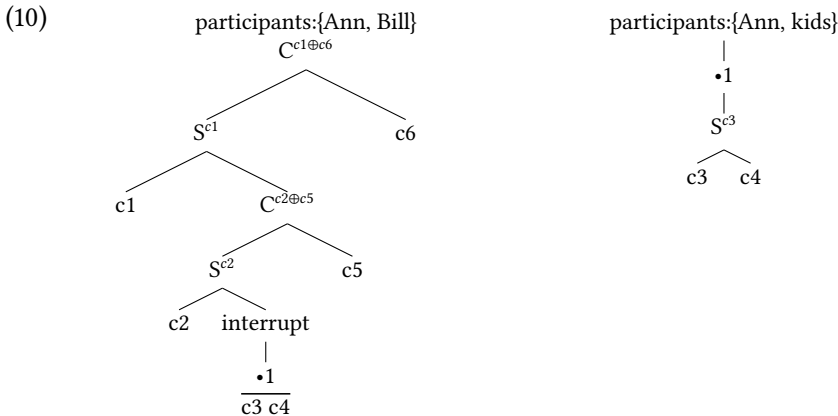
- (8) ^{c1}John came to the door. ^{c2}He always leaves the groceries at the backdoor.
^{c3}Stop that, you kids! ^{c4}leave the dog alone! ^{c5}I put them away immediately.
^{c6}Anyway, then he went home.



In this analysis, the conversation between Ann and Bill encoded as a straightforward discourse tree, gets interrupted by Ann turning to the children and addressing them. This embedded conversation is encoded as a subordination, because it does not contribute to the conversation between Ann and Bill. This how we talked about it in earlier papers, but we now believe it is not the right analysis. One alternative would be to analyze these two conversations as two trees, completely independent of each other.



However that seems to ignore relevant aspects of the interaction between Ann and Bill, namely that she is not interacting with him. Because the conversation between Ann and Bill is truly interrupted by the conversation with the kids, it cannot continue until the interruption is broken off.¹⁶ To accommodate the complexity of the double interaction, we have adopted the practice of marking the location of the interruption and linking it to separate tree.

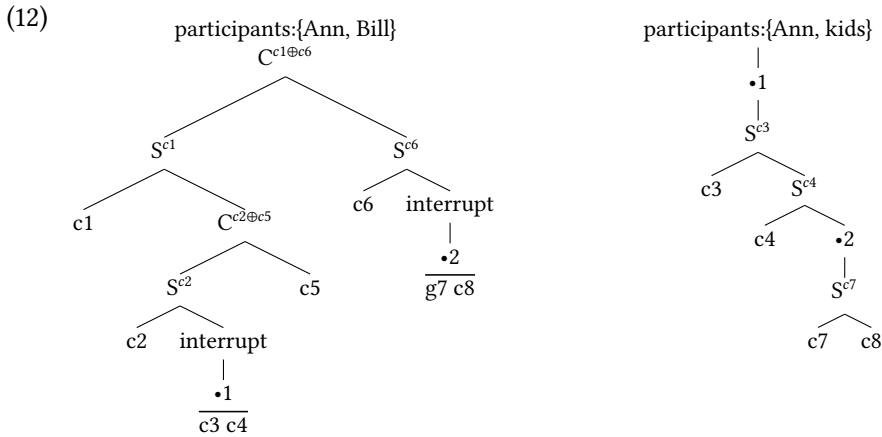


Having pointers from one node to another tree makes it possible to encode more complicated cases such as (11) in a more transparent fashion.

¹⁶By, for example, a change in body orientation towards Bill. Note that information about body-position is explicitly included in the feature set for each BDU.

- (11) c^1 John came to the door. c^2 He always leaves the groceries at the backdoor.
 c^3 Stop that, you kids! c^4 leave the dog alone! c^5 I put them away immediately.
 c^6 Anyway, then he went home. c^7 Kids, what did I just say? c^8 Stop bothering him!

In this new notation, the main tree treats interruptions as subordinations, while the interruptions themselves stand on their own as fully independent separate trees.¹⁷



5.2 Grappling with quotations

The correct characterization of direct speech is a well-known, highly debated problem,¹⁸ one we will certainly not take on except to point out how one discourse is referred to from within another. This already poses a significant challenge to the LDM.

For our discussion here, we assume quotes are discourse operators that change the speaker index in the context of interpretation of the quoted material.

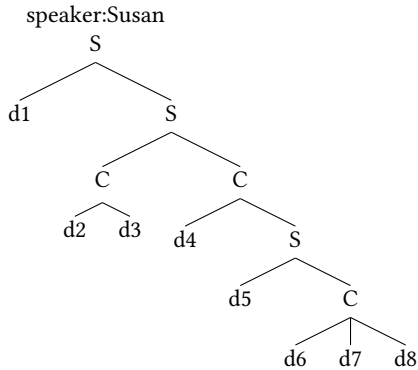
- (13) **Susan talking to Ann**

d^1 I've been getting back into astronomy a bit more recently. d^2 I talked to John about the Hubble Constant, and d^3 he told me about the strange discrepancy between the two different ways of measuring it. d^4 He is not

¹⁷We believe that this is essentially the analysis of this example that Mark Steedman in a personal communication argued for some years ago using TAGs instead of LDM trees.

¹⁸Cf. The Stanford Encyclopedia of Philosophy entry on Quotation by Cappelen et al. (2023) for an introduction to some of the issues.

normally someone I go to discuss things, ^{d5}but in this case he was the right person to go to. ^{d6}I'm just a regular physicist, but ^{d7}he specialized in astrophysics, so ^{d8}I knew he'd have some good ideas about the subject.

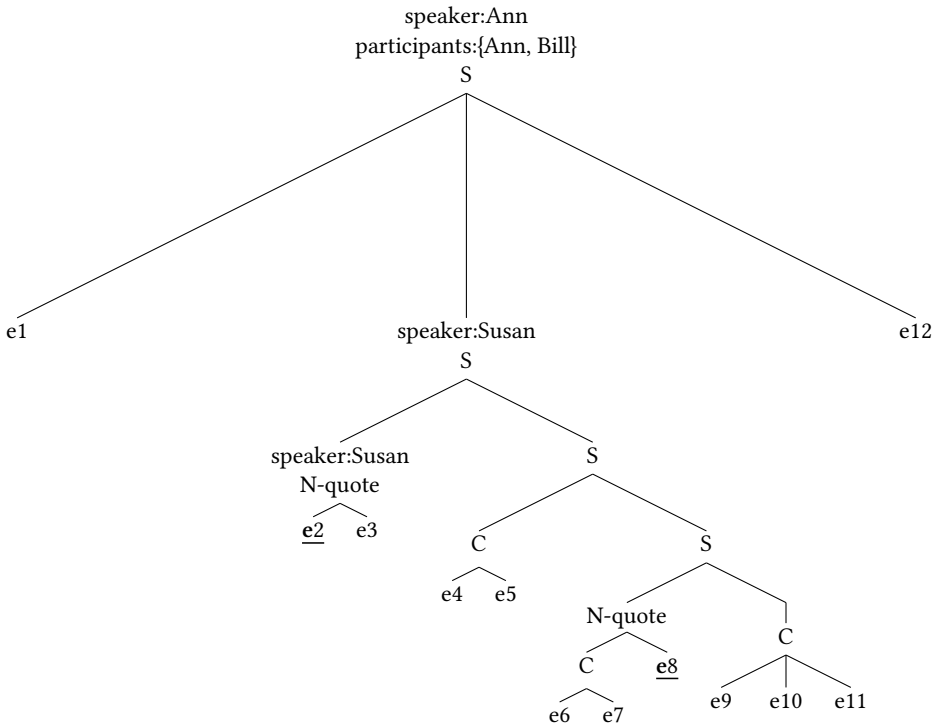


(13) is a rather straightforward text. Susan describes talking to John about an interesting astronomy topic, even though, as she reminds us, she normally does not ask for his opinion about issues like that. The structure, therefore, is straightforward as well.

When this story is retold by Ann in (14), the resulting structure is similar, with some additional N-ary nodes assigning Susan as the speaker for the quoted text.

(14) Susan's story as retold by Ann to Bill

^{e1}I had an interesting conversation. ^{e2}Susan told me yesterday, "^{e3}I've been getting back into astronomy a bit more recently," "^{e4}I talked to John about the Hubble Constant, and ^{e5}he told me about the strange discrepancy between the two different ways of measuring it." "^{e6}He is not normally someone I go to discuss things, ^{e7}but in this case he was the right person to go to," ^{e8}she explained, "^{e9}I'm just a regular physicist, but ^{e10}he specialized in astrophysics, so ^{e11}I knew he'd have some good ideas about the subject." ^{e12}Anyway, nice talk.



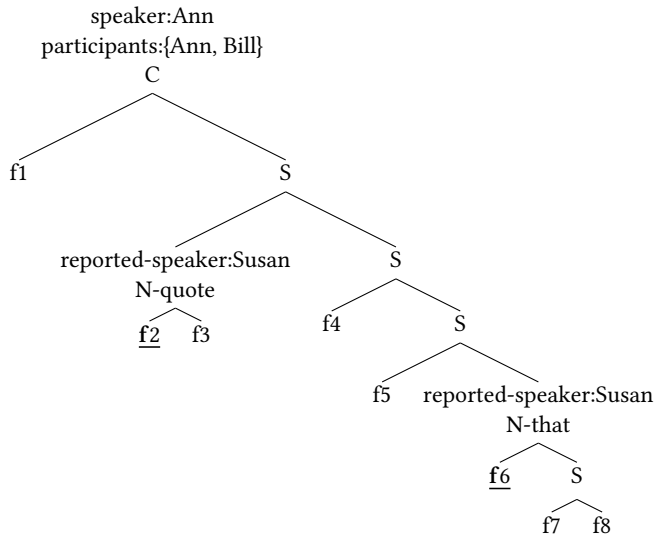
e2 and *e8*, which specify Susan as the speaker, are treated as modal modifiers changing the speaker index from Ann, who is telling the story, to Susan. Susan is inherited as a speaker by the subordinated nodes following it. The second N-node subordinates *e6*, *e7*, *e9*, *e10*, and *e11* as an explanation.

Note that there is no change in participants of this story. Ann is still talking to Bill. However for the quoted direct speech, some indices are changed. An index like “I” in reported speech refers to the reported speaker, not to the actual speaker reporting the speech.

Now, let’s look at the situation when only part of the story is quoted:

(15) **Ann telling Bill about Susan’s story**

^{f1}I asked Susan about the Hubble constant issues. ^{f2}She said, “^{f3}I talked to John about the Hubble Constant.” ^{f4}I was surprised because ^{f5}I expected her to have some ideas of her own about it, ^{f6}but she said, “^{f7}he specialized in astrophysics, and ^{f8}I knew he’d have some good ideas about the subject.”



The quoted sentences originate from the original discourse (13), but in (15) we do not have enough information about the accessibility of referents to resolve the pronouns *he* in *f7* and *f8* to refer back to *John* in *f3*. Where in the case of interruptions (section 5.1) the interrupting tree was available to us to do anaphora resolution, this is not the case here. To refer back to *John* in *f3*, we will need a different mechanism. The tree as presently constructed does not guarantee the correct resolution of the anaphora. We will have to trust that Ann is cooperative.

6 Conclusions

In this paper, we have provided an overview of the Linguistic Discourse Model, and provided an explanation of the Right Edge Constraint, a well known and widely accepted property of formal discourse analytic theories. We have described how, in contrast to the boundedness of sentences, which allows for a more complex local structure by allowing access to the analyzed object as a whole, the unboundedness of discourse gives rise to a more restricted space of syntactic relationships. Segmentation of the large scale discourse structure into discourse constituents and restricting continuation only to constituents at the end of the discourse history as constructed so far create a tree structure that only allows right edge attachments.

However, interruptions do not fit well in the one tree view of the discourse. Even though from the perspective of the embedding discourse, an interruption

looks like a subordination, as we have seen in (11), the interruption may need to be analyzed as a separate tree. Finally, direct speech raises even more problems, ones that cannot be addressed through the construction of a separate tree.

The LDM has proved useful both in demystifying some linguistic and literary phenomena and, as implemented computational programs, has powered applications that address relevant practical problems. For example, the Right Edge Constraint provides insight into the resolution of discourse anaphora, while by providing structural descriptions of texts in Mocho (Polanyi 1996), a Mayan language, as well as texts in Japanese and Korean, the LDM has explained the distribution of some otherwise “mystery particles” in those languages. Of more applied interest, LDM analysis has provided a mechanism to describe the discourse patterns of voluble Alzheimer’s patients in a way which connects their verbal output to cognitive disability—put simply, they cannot form lists and once they PUSH to an embedded constituent they can not POP. In addition, a patented system based on the LDM was designed and provisionally implemented that provided support for students to improve their writing skills at the discourse level. This educational application uses implemented computational summarization methods based on the LDM (Polanyi et al. 2004a) to guide students to organize the flow of sentences and topics in their papers. Hopefully, when the needle moves back from purely statistical AI applications in NLP and towards curiosity about language above the sentence in linguistics, more applications of theories such as the LDM will be of interest again.

There is certainly still a lot of work left to do and a lot left to hope for.

References

- Asher, Nicholas. 2008. Troubles on the right frontier. In Anton Benz & Peter Kühnlein (eds.), *Constraints in discourse*, vol. 2 (Pragmatics & Beyond New Series), 29–52. John Benjamins Publishing Company. DOI: 10.1075/pbns.172.02ash. <https://doi.org/10.1075/pbns.172.02ash>.
- Asher, Nicholas & Alex Lascarides. 2003. *Logics of conversation*. Cambridge University Press.
- Butt, Miriam, Tracy Holloway King, María-Eugenia Niño & Frédérique Segond. 1999. *A grammar writer’s cookbook*. CSLI Publications.
- Cappelen, Herman, Ernest Lepore & Matthew McKeever. 2023. Quotation. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford encyclopedia of philosophy*, Summer 2023. Metaphysics Research Lab, Stanford University.
- Cristea, Dan. 2005. The right frontier constraint holds unconditionally. In *Proceedings of the multidisciplinary approaches to discourse 2005 (MAD’05)*.

- Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell III & Annie Zaenen (eds.). 1995. *Formal issues in Lexical-Functional Grammar*. CSLI Publications.
- Grosz, Barbara J. & Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3). 175–204. <https://aclanthology.org/J86-3001>.
- Heim, Irene. 1988. *The semantics of definite and indefinite noun phrases* (Outstanding Dissertations in Linguistics). Garland Publishing. <http://semanticsarchive.net/Archive/Tk0ZmYyY/dissertation.pdf>.
- Holler, Anke & Lisa Irmen. 2007. Empirically assessing effects of the right frontier constraint. In António Branco (ed.), *Anaphora: Analysis, algorithms and applications*, 15–27. Springer.
- Hymes, Dell H. 1974. *Foundations in sociolinguistics: An ethnographic approach*. University of Pennsylvania Press. <https://archive.org/details/foundationsinsoc0000hyme>.
- Kamp, Hans. 1981. A theory of truth and semantic representation. In Jeroen A. G. Groenendijk, Theo M. V. Janssen & Martin B. J. Stokhof (eds.), *Formal methods in the study of language*, vol. 1, 277–322. Mathematisch Centrum.
- Kamp, Hans & Uwe Reyle. 1993. *From discourse to logic: Introduction to model theoretic semantics of natural language, formal logic and discourse representation theory*. Kluwer Academic Publishers.
- Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 173–281. The MIT Press. Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 29–130).
- Labov, William. 1972. The transformation of experience in narrative syntax. In William Labov (ed.), *Language in the inner city*, 352–389. University of Pennsylvania.
- Labov, William & Joshua Waletzky. 1965. Narrative analysis: Oral versions of personal experience. In *Essays on the verbal and the visual arts*, 3–38. University of Washington Press.
- Litman, Diane & Julia Hirschberg. 1990. Disambiguating cue phrases in text and speech. In *COLING 1990 volume 2: Papers presented to the 13th international conference on computational linguistics*. <https://aclanthology.org/C90-2044>.
- Mann, William C. & Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text — Interdisciplinary Journal for the Study of Discourse* 8(3). 243–281. DOI: [doi:10.1515/text.1.1988.8.3.243](https://doi.org/10.1515/text.1.1988.8.3.243). <https://doi.org/10.1515/text.1.1988.8.3.243>.
- Marcu, Daniel. 2000. *The theory and practice of discourse parsing and summarization*. MIT Press.

- Miltsakaki, Eleni, Aravind Joshi, Rashmi Prasad & Bonnie Webber. 2004. Annotating discourse connectives and their arguments. In *Proceedings of the workshop frontiers in corpus annotation at HLT-NAACL 2004*, 9–16. Association for Computational Linguistics. <https://aclanthology.org/W04-2703>.
- Polanyi, Livia. 1985. A theory of discourse structure and discourse coherence. In Paul D. Kroeber, William H. Eilfort & Karen L. Peterson (eds.), *Papers from the general session at the 21st meeting of the Chicago Linguistics Society*. Chicago Linguistic Society.
- Polanyi, Livia. 1988. A formal model of discourse structure. *Journal of Pragmatics* 12. 601–638.
- Polanyi, Livia. 1989. *Telling the American story: A structural and cultural analysis of conversational storytelling*. MIT Press.
- Polanyi, Livia. 1996. *The linguistic structure of discourse*. Tech. rep. CSLI Publications.
- Polanyi, Livia & Remko Scha. 1983. On the recursive structure of discourse. In Konrad Ehlich & Henk C. van Riemsdijk (eds.), *Connectedness in sentence, discourse and text*, 141–178. Tilburg University.
- Polanyi, Livia & Remko Scha. 1984. A syntactic approach to discourse semantics. In *10th international conference on computational linguistics and 22nd annual meeting of the Association for Computational Linguistics (ACL)*, 413–419. Association for Computational Linguistics. DOI: 10.3115/980491.980576. <https://aclanthology.org/P84-1085>.
- Polanyi, Livia, Martin van den Berg & David Ahn. 2003. Discourse structure and sentential information structure: An initial proposal. *The Journal of Logic, Language and Information* 12(3). 337–350.
- Polanyi, Livia, Martin Henk van den Berg, Gian Lorenzo Thione, Christopher Culy & David Ahn. 2004a. Hybrid text summarization: Combining external relevance measures with structural analysis. In *ACL2004 workshop text summarization branches out*, 51–55.
- Polanyi, Livia, Martin Henk van den Berg, Gian Lorenzo Thione, Christopher Culy & David Ahn. 2004b. Sentential structure and discourse parsing. In *ACL2004 workshop on discourse annotation*, 51–55. <https://aclanthology.org/W04-1009/>.
- Sassen, Claudia & Peter Kühnlein. 2005. The right frontier constraint as conditional. In Alexander F. Gelbukh (ed.), *Computational linguistics and intelligent text processing international conference (CICLing) proceedings*, vol. 3406 (Lecture Notes in Computer Science), 222–225. Springer. DOI: 10.1007/978-3-540-30586-6_23. https://doi.org/10.1007/978-3-540-30586-6_23.

- Schegloff, Emanuel A. & Harvey Sacks. 1973. Opening up closings. *Semiotica* 8(4). 289–327. DOI: doi:10.1515/semi.1973.8.4.289. <https://doi.org/10.1515/semi.1973.8.4.289>.
- Schiffrin, Deborah. 1987. *Discourse markers* (Studies in Interactional Sociolinguistics). Cambridge University Press.
- van Eijck, Jan & Hans Kamp. 2011. Discourse representation in context. In Johan van Benthem & Alice ter Meulen (eds.), *Handbook of logic and language*, 181–252. Elsevier. DOI: 10.1016/B978-0-444-53726-3.00003-7. <https://doi.org/10.1016/B978-0-444-53726-3.00003-7>.
- van den Berg, Martin H. 1996. *Some aspects of the internal structure of discourse – the dynamics of nominal anaphora*. Institute for Logic Language & Computation, University of Amsterdam. (Doctoral dissertation).
- van den Berg, Martin H. & Hub Prüst. 1990. Common denominators and default unification. In Ton van der Woude & Wietske Sijtsma (eds.), *Computational linguistics in the Netherlands: Papers from the first CLIN meeting*. University of Utrecht.
- van der Sandt, Rob A. 1992. Presupposition projection as anaphora resolution. *Journal of Semantics* 9(4). 333–377. DOI: <https://doi.org/10.1093/jos/9.4.333>. <https://academic.oup.com/jos/article-pdf/9/4/333/9836990/333.pdf>.
- Webber, Bonnie. 2006. Accounting for discourse relations: Constituency and dependency. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*. CSLI Publications.

Chapter 7

From Network Theory to Linguistic Insights

Alexandros Tantos^a & Kosmas Kosmidis^a

^aAristotle University of Thessaloniki

In this paper, we argue that using network analysis to map discourse representations offers a robust method for revealing intricate properties of discourse structures. By examining the Discourse GraphBank corpus, we applied key network indices to identify distinct profiles within discourse networks. We discovered prevalent network motifs, notably the feed-forward loop, which resembles mechanisms in biological signaling, enhancing signal control and amplification. Conversely, we found linear chains and common cause patterns to be antimotifs, characterized by their simplicity and lack of functional complexity. Additionally, the dual cause pattern showed variability, acting as both a motif and an antimotif, suggesting its role in complex interaction coordination. These insights underscore the functional implications of network motifs in discourse and suggest potential avenues for refining our understanding of discourse coherence and structural dynamics.

1 Introduction

Understanding the coherence patterns in discourse is a pivotal research goal in the computational analysis of discourse structure. Such patterns are crucial for addressing the broader question of how discourse coherence and cohesion are constructed and realized within the language (Zhang et al. 2021, Ma et al. 2022, Jiang et al. 2022, Gupta et al. 2022, Wu et al. 2022, Xiang & Wang 2023) and for the development of advanced natural language processing (NLP) applications, as highlighted by numerous studies (Webber & Joshi 2012, Lai & Tetreault 2018, Narasimhan & Barzilay 2015, Jiang et al. 2021, Liu et al. 2019, van Noord et al. 2018).

Alexandros Tantos & Kosmas Kosmidis. 2025. From Network Theory to Linguistic Insights. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 137–163. Konstanz: PubliKon.

Recent trends in NLP have shifted towards training Transformer models (Vaswani et al. 2017, Lin et al. 2021, Reimers & Gurevych 2019) and creating probing tasks to indirectly explore how discourse coherence and cohesion can be traced through the layers of a deep neural network. Despite these advancements, the interpretability of semantic properties directly from these models remains elusive. However, in this paper we suggest that exploring traditional discourse semantic theories through rigorous network analytic tools may significantly enhance our understanding of underlying discourse structures (Hobbs 1985, Grosz & Sidner 1986, Polanyi & Scha 1984, Polanyi 1985, 1988, Asher 1993, Kuppevelt 1995, Asher & Lascarides 2003, Knott 2007, Rohde et al. 2018).

Specifically, we employ network analysis techniques to examine how discourse structures can be represented as networks to reflect the flow of information in the discourse and reveal hidden discourse properties. Central to our network analysis is the role of network edges, which encapsulate the links or relationships between different discourse units, highlighting their importance in mapping the dynamics of discourse interactions.

This study builds upon previous work by Tantos & Kosmidis (2023), where a network theory approach was applied to the discourse-annotated C58 and STAC corpora. That work introduced the use of network indices to compare different types of discourse structures, such as single-author written texts and multiparty dialogue. In contrast, the current study focuses on the Discourse GraphBank corpus (Wolf & Gibson 2005), exploring the presence and significance of network motifs and antimotifs in a broader linguistic context. Notably, while the methodology overlaps, this work emphasizes motif discovery across discourse networks and reveals new insights, particularly regarding the variability of the dual cause pattern. Thus, while the earlier study analyzed and compared two types of corpora in different languages, our paper explores more complex and varied discourse structures through a single, larger and more diverse dataset.

Identifying distinct network patterns within discourse representations in annotated corpora indicates that speakers likely utilize implicit strategies in their discourse engagement. This implies that speech acts are interconnected in foreseeable ways. Additionally, identifiable variances and similarities in the network patterns can aid in developing theoretical insights pertinent to various types of discourse.

Section 2 describes the principles behind the types of discourse units and discourse relations adopted in the GraphBank corpus. Section 3 offers a brief overview of the compilation and annotation process of GraphBank. Section 4 provides a detailed analysis of (a) the mapping between discourse representations and networks, (b) the profile of discourse representations through key network

indices and (c) the presence of motifs and antimotifs for three-node subgraphs for the discourse networks of our datasets. The last section, Section 5, sums up the findings of this study and offers a series of theoretically driven reflections related to the types of restrictions on discourse inference and interpretation imposed by the discourse structure.

2 Discourse relations and discourse units

The literature features numerous formal theories of discourse representation (Segmented Discourse Representation Theory (SDRT; Asher 1993, Asher & Lascarides 2003), Rhetorical Structure Theory (RST; Mann & Thompson 1988), the Linguistic Discourse Model (LDM; Polanyi et al. 2004), the Discourse GraphBank model (Wolf & Gibson 2005) and Discourse Lexicalized Tree Adjoining Grammar (DLTAG; Creswell et al. 2003)), which guide the annotation of corpora according to various models such as Segmented Discourse Representation Theory (SDRT), Rhetorical Structure Theory (RST), the Linguistic Discourse Model (LDM), the Discourse GraphBank model, and Discourse Lexicalized Tree Adjoining Grammar (DLTAG). While each theory offers unique perspectives on defining and identifying discourse relations, their foundational goal is similar: they organize discourse hierarchically around elementary discourse units (EDUs).

Despite theoretical divergences in defining discourse relations and their characteristics, for the purpose of this paper, such debates are secondary. The focus here is on examining discourse structures as networks, disregarding the specific types of discourse relations forming these networks. All theories recognize the need for a discourse segmentation algorithm to divide the discourse into EDUs, generally agreeing on the segmentation criteria.¹

2.1 Elementary discourse units

The EDUs function as linguistic utterance nodes within the discourse graph and serve as the fundamental components of discourse representation. The complex semantic and pragmatic relationships between the EDUs are delineated through various discourse relations, e.g. *Elaboration*, *Explanation*, *Result* and *Narration*.

¹SDRT employs a second type of node on a discourse graph, the Complex Discourse Unit (CDU) that is also admitted in the Discourse GraphBank corpus. Following Asher et al. 2011, a CDU represents a semantically coherent group of EDUs that collectively serve as an argument to a discourse relation between this group and another EDU. For more details as to the definition and role of CDUs refer to Asher et al. 2011 and Tantos & Kosmidis 2023.

Setting aside the theoretical differences previously discussed, another critical aspect of classifying discourse relations is determining whether an EDU is considered more or less important, or even of equal importance, compared to another EDU to which it is related. Based on the relative importance of the related EDUs and the theoretical framework, various discourse theories recognize discourse relations as either symmetrical or asymmetrical, or as coordinating or subordinating (Asher & Vieu 2005). Although the terms symmetrical/asymmetrical and coordinating/subordinating appear to be grounded in a shared theoretical assumption regarding the importance of related EDUs, the concept of importance is approached differently across theoretical perspectives. This can lead to opposite classifications of discourse relations with the same essence. For example, *Narration* and *Continuation* in SDRT are considered coordinating relations, whereas a relation with similar intent and content, such as *Temporal Sequence (temp)* in RST, is viewed as asymmetrical, with the event of the first EDU acting as the Satellite and the event of the second EDU as the Nucleus.

However, beyond the linguistic contexts that need to be studied further so that the linguistic classification tests such as the ones suggested by Asher & Vieu (2005) for categorizing discourse relations along the above-mentioned importance axis can be confirmed and/or refined, this paper abstracts away from the distinction between subordinating and coordinating relations and inspects deeper properties of discourse development that would probably play a role in revealing important aspects of the essence of discourse's information flow. We propose that analyzing network motifs and other descriptive network indices provides a structured method to uncover these deeper properties, particularly leveraging the directionality of discourse edges as discussed in Section 4 for the Discourse GraphBank corpus. Subsequently, the distinction between subordinating/asymmetrical and coordinating/symmetrical relations can be reassessed in light of the insights gained from this exercise.

3 The Discourse GraphBank corpus

The Discourse GraphBank is one of the few highly qualitative discourse-annotated corpora developed to facilitate systematic analyses of text coherence. Its texts are annotated with 63 coherence relations, and it adopts a modular design, allowing information to be stored in separate, linked files for ease of future expansion, such as adding parts of speech or syntactic structures without needing to edit existing files. Each text was annotated independently by two annotators, achieving an inter-annotator agreement of 88.45%, with a kappa of 84.63%,

indicating high reliability of the annotations without systematic discrepancies related to text length or the complexity of the coherence graph.

This database is positioned as a significant advancement over the RST Discourse Treebank, which utilizes a tree graph-based annotation scheme that fails to account for the crossed dependencies observed in coherence structures, a limitation addressed in our database design. Discourse structures are represented in graphs with directed labeled arcs (representing the coherence relations) between nodes in an ordered array (representing the discourse segments in a text).

4 Networks and connectivity

Networks, also known formally as graphs, comprise sets of nodes and edges (Barabási & Albert 1999, Newman et al. 2001, Newman et al. 2006, Watts & Strogatz 1998). Nodes are representations of objects, while edges signify the relationships between these objects. For instance, in a social network, nodes are individuals linked by edges if they are acquainted. In a biological protein interaction network, nodes represent proteins that are connected through biochemical interactions if they physically interact (Milo & Alon 2002). Graphs can be either undirected or directed. A directed graph, also called a digraph, is a graph in which the edges have a direction. Pictorially, this is indicated with an arrow on the edge. If v and w are vertices, an edge is an unordered pair $[v, w]$, while a directed edge is an ordered pair, (v, w) or (w, v) . The directed edge (v, w) is drawn as an arrow from v to w . Of course, it is possible to treat a digraph as an undirected network and still obtain valuable information.

In the context of our study, a discourse representation is translated to a network, as it can be mapped onto a graph where nodes represent utterance labels, and edges signify discourse relations between these labels. This approach differs significantly from other types of linguistic networks previously explored (Choudhury & Mukherjee 2009, Mehler et al. 2016, Mehler 2008), which typically focus on the relationships between words within sentences. Moreover, our graphs are unweighted, i.e. all edges are treated as equal, and directed, emphasizing the directional nature of discourse relations and contain no multilinks or self-loops.

Networks have proven to be an essential tool in applied mathematics, offering a versatile framework to describe a wide array of physical systems (Song et al. 2005, 2006, Barzel & Barabási 2013, Lacasa et al. 2008, D'Souza et al. 2019, Clauset et al. 2008, Bashan et al. 2012, Battiston et al. 2016). A recent search in the Dimensions bibliographic database indicated that over 4300 articles published in 2023 featured the term “Complex Network” in their title or abstract.

4.1 From discourse representations to directed discourse networks

In our study, we leverage the annotated data from the Discourse GraphBank corpus to construct directed discourse networks. Each node in the network represents a specific discourse unit as identified by the annotators within the text. Our methodology adheres to the guidelines set forth in the corpus documentation, which involves parsing each line of annotated data to determine the structural relationships between these units. For instance, an entry in the dataset might display discourse units and their discourse relations in a format like “6 6 7 7 temp” taken from the annotated version of the file #52 in the Discourse GraphBank corpus, where the first pair of numbers indicates the source nodes and the second pair the target nodes of the discourse connections. Under our established rules, this configuration directs the creation of edges from node 6 to node 7 for relating them with the *Temporal Sequence* relation. In other words, in (1), the less important or satellite EDU, labeled 6, describes the event of *having a pampered childhood* that precedes the event of *becoming orphaned* described in the more important or nucleus EDU, labeled 7. Therefore, directionality in the newly-created networks is mapped directly to the directionality from a satellite to a nucleus in the discourse representations of the Discourse GraphBank corpus, following the principles of RST.

- (1) 6. During the sixties and seventies we had a pampered childhood.
 ...
 7. And then very abruptly at the beginning of the eighties, we were orphaned.

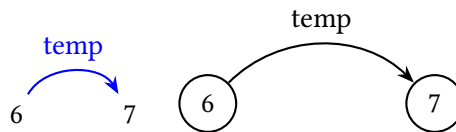


Figure 1: RST-like Discourse GraphBank representation on the left. Network representation of the RST-like Discourse GraphBank representation on the right.

In more complex scenarios, such as ‘64 67 55 57 examp,’ which corresponds to (2) from file #1 in Discourse GraphBank, the braces involve multiple nodes representing CDUs. The mapping of the RST-like representation to the network representation is straightforward. The first EDU in the sequence 64–67, 64, is independently connected to the preceding EDUs 55, 56, and 57 with the discourse relation

of ‘example.’ The same applies to the rest of the nodes in the sequence 64–67.² This meticulous parsing extends across the dataset, resulting in a comprehensive network that graphically depicts the directional flow of discourse throughout the document.

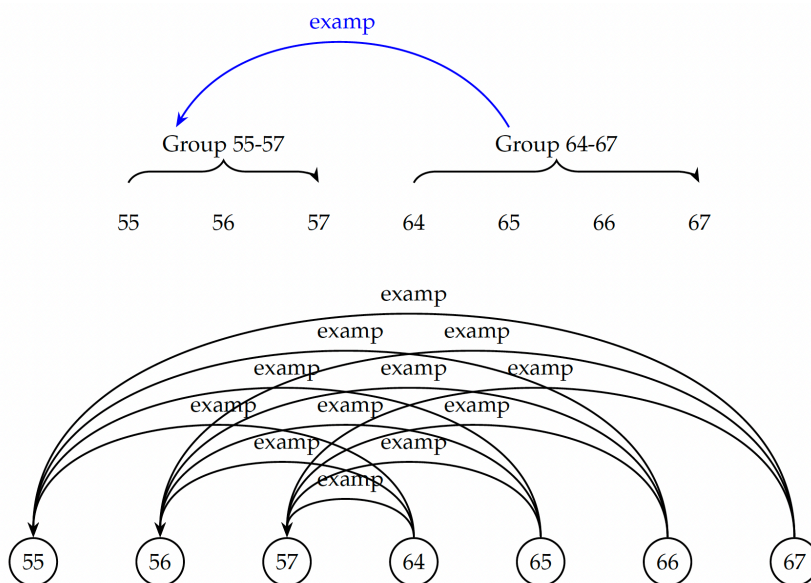


Figure 2: Two Representations: RST-like representation (Top) and Network-based representation (Bottom) of the discourse relations in (2).

- (2) 55. “At least our students will be aware of the enhancements and developments in this area, and
 56. when they make decisions as engineers and managers,
 57. they will take this into consideration,”
 58. she said.
 ...
 64. A doctoral candidate is developing computer software that will allow the system to diagnose itself.
 65. The system could be programmed to recognize when a component is not working, and

²There are more discourse relations between these nodes that we did not include for the sake of clarity, i.e. so that our network representation remains legible.

66. then decide to switch to manufacturing a product that does not involve the disabled component,
67. much as human managers would react.”

Moreover, our analysis is versatile in accommodating symmetrical discourse relations, such as parallelism and contrast, by adjusting the directional nature of the network links accordingly. Particularly, for these discourse relations, we treat the connections between nodes as bidirectional, thereby recognizing the reciprocal nature of these relationships.

4.2 Key network indices and their relevance to discourse

In network studies, various quantities are of interest to our research. Some are applicable to both directed and undirected networks, while others are specific to directed networks. Typically, the examination of a directed network begins by considering it as an undirected graph. Initially, the fundamental properties of the undirected graph are determined, followed by an analysis of the quantities related to the directionality of the graph’s links. This method is the approach adopted in this study. When analyzing discourse networks as undirected graphs, we focus on the following five indices:

- (a) The network’s size is quantified by two measures: the number of nodes, $nsize$, and the number of edges, $nedge$. These metrics provide a fundamental understanding of the discourse network’s extent.
- (b) The edge fraction f_e is defined as the ratio of the actual number of edges $nedge$ to the maximum possible number of edges ne_{max} , expressed by the equation:

$$f_e = \frac{nedge}{ne_{max}} = \frac{nedge}{\frac{(nsize)(nsize-1)}{2}} \quad (1)$$

This ratio, where $ne_{max} = \frac{(nsize)(nsize-1)}{2}$ is the maximum number of undirected edges in a network of $nsize$ nodes, ranges from 0 to 1. A low f_e value suggests sparse connectivity in the discourse graph, while a value near one indicates a highly interconnected structure.

- (c) Degree centrality, a key metric for nodes in the network, represents the count of edges connected to a node. High degree centrality in discourse nodes, or utterance labels, suggests a pivotal role in the development of

the discourse. Below we analyze the network's degree distribution by examining the maximum, minimum, and average node degrees, denoted as k_{max} , k_{min} , and k_{av} , along with the standard deviation of these degrees, k_{std} .

- (d) The average clustering coefficient C is calculated by averaging the local clustering coefficients. The local clustering coefficient for a node i is:

$$C(k_i) = \frac{t_i}{[k_i(k_i - 1)/2]} \quad (2)$$

Here, t_i represents the number of triangles connected to node i , normalized by the maximum number of such triangles. Values of C close to 1 indicate a network where two nodes connected to a common third node are likely to be connected to each other, typical of social networks. The clustering coefficient measures the likelihood that two nodes, both connected to a common third node, are also directly connected to each other in a given network. It is possible to construct networks with the same number of nodes and edges but with significantly different clustering coefficients. This makes the clustering coefficient an important metric for understanding network topology.

- (e) The degree assortativity coefficient r , which measures the tendency of nodes to connect with other nodes of similar or dissimilar degrees, is defined as:

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_q^2} \quad (3)$$

In this equation, q_k is the distribution of the remaining degree of the nodes, calculated from the degree distribution p_k . The term e_{jk} represents the joint probability distribution of the remaining degrees of two connected vertices. Assortative mixing, indicated by r values close to 1, occurs when high-degree nodes predominantly connect to other high-degree nodes, while disassortative mixing, indicated by r values close to -1, involves high-degree nodes connecting to low-degree nodes.

In the Appendix, Table 1 provides basic statistics for several real-world networks, detailing numerical estimates of crucial properties such as clustering and assortativity. Real-world networks exhibit considerable variation in their clustering coefficients. As observed in Table 1, social networks generally exhibit higher clustering coefficients. For instance, the film actor collaboration network shows a clustering coefficient (C) of 0.78, while a network of physicist collaborations

records a C of 0.56, which are typical values for social networks. In contrast, technological and biological networks usually have lower values; for example, the power grid network has a clustering coefficient of just about 0.08. In the context of studying discourse networks, a high clustering coefficient might be a significant indicator for assessing discourse coherence and the degree of connectivity within discourse structures. However, confirming such a hypothesis would necessitate verifying whether discourse networks with high clustering coefficients are also perceived as highly coherent by participants.

Regarding the assortativity property, Table 1 reveals that social networks typically exhibit assortative mixing ($r > 0$), a phenomenon well recognized by sociologists who attribute it to homophily – the tendency of individuals to associate with others who are similar to themselves. This concept of assortative mixing is also occasionally seen in the opposite form as disassortative mixing, where individuals tend to connect with others who are dissimilar. This pattern of assortative or disassortative mixing can be observed in certain nonsocial networks as well. For example, papers in a citation network generally cite other papers within the same field more frequently than those in different fields, and web pages often link to other pages in the same language. The assortativity coefficient r , studied here, quantifies this type of homophily, measuring the *degree* of assortativity within the network. It allows for the analysis of nodes in two categories: those that are highly connected and those that are poorly connected. The coefficient r then measures the propensity of nodes to connect within their own group.

Focusing on discourse networks, a highly assortative discourse graph could indicate a strong connection among highly important nodes. This pattern might suggest that a few, or even a single, common topic or argument is being predominantly supported by various influential utterances within the discourse. Conversely, a disassortative discourse network could point to a lack of connectivity among key nodes, which might be interpreted as a form of mutual avoidance among nodes with high degree centrality.

4.3 Network indices of Discourse GraphBank

Figure 3 shows, in the form of histograms, the distribution of the basic properties of 135 networks derived from dialogues that were included in GraphBank. The largest of these networks contains 138 nodes while the smallest contains only 6 nodes. Similarly, concerning the edges, the largest graph contains 264 edges while the smallest only 6. GraphBank networks are rather sparse as the mean value of the fraction of edges is 0.06. There are, however, some exceptions as the maximum value of the fraction of edges among the 135 networks equals 0.40.

Most of the networks have a low clustering coefficient (mean value of C is 0.27). The mean assortativity coefficient of the 135 networks is very close to zero. There are, however, networks that are highly disassortative (the minimum assortativity coefficient r observed is equal to -0.527) and networks that are highly assortative (maximum $r = 0.483$).

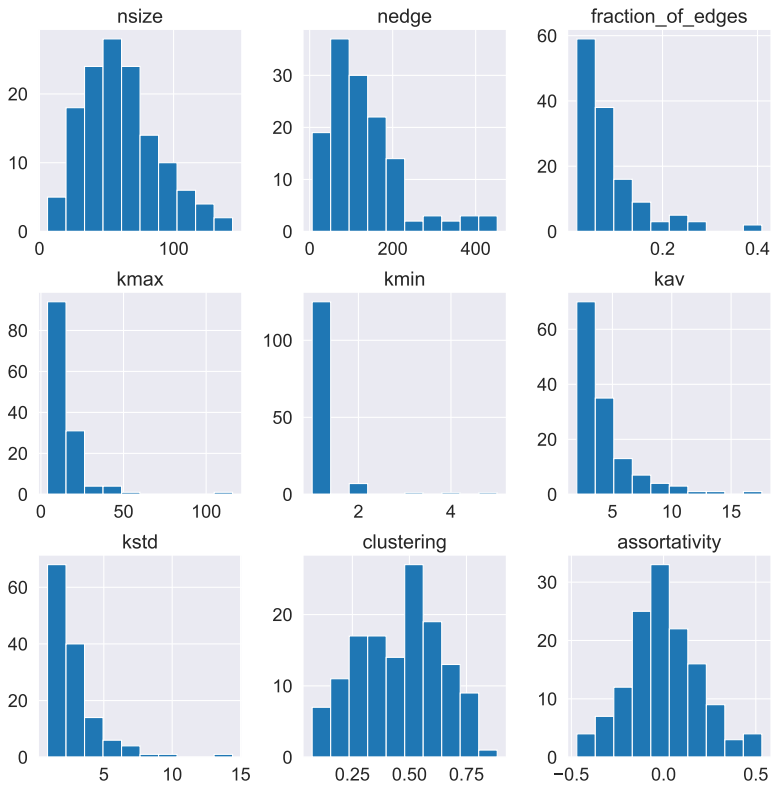


Figure 3: Histograms of basic undirected network properties: Number of nodes $nsize$, Number of edges $nedge$, Fraction of edges i.e. the ratio of $nedge$ to the maximum possible number of edges, Mean degree k_{av} , degree standard deviation k_{std} , Clustering coefficient C , Assortativity coefficient r .

4.4 Network motifs and discourse structure

A key non-trivial measure used to characterize directed networks is the presence or absence of network motifs, as identified in studies such as those by Milo & Alon (2002) and Shen-Orr et al. (2002). Network motifs are essentially recurring simple patterns within a network. To identify these patterns, the network is analyzed for specific n -node subgraphs, where $n = 3$ in our case. Studying higher-order patterns, specifically those with $n \geq 4$, is theoretically possible but practically challenging due to their complexity. For instance, there are 199 possible configurations for 4-node subgraphs, 9,364 configurations for 5-node subgraphs, and so on (Alon 2019). Furthermore, to count the occurrences of a specific 4-node subgraph within a graph of 1600 nodes, one would need to examine all $\binom{1600}{4}$ node combinations, a task that involves checking on the order of 10^{11} combinations, making the computation impractically demanding. Therefore, research predominantly focuses on three-node motifs, which have proven to be particularly valuable in studies of biological networks (Bullmore & Sporns 2009, Boyer et al. 2005, Barabási et al. 2011, Marr et al. 2010, Klosik et al. 2014, Beber et al. 2012, Shen-Orr et al. 2002).

These 3-node subgraphs are referred to as patterns and are labeled with the letter S followed by an index number. We will analyze five distinct three-node subgraph patterns, designated as follows:

- **S1 – Linear Chain:** This pattern shows three nodes arranged sequentially. Node 1 connects to Node 2, which then connects to Node 3, forming a direct, linear path from the first to the last node.
- **S2 – Feed Forward Loop:** This motif features three nodes where Node 1 connects to both Node 2 and Node 3. Additionally, Node 3 also connects back to Node 2, forming a feedback loop within the structure.
- **S3 – Cycle:** This configuration consists of three nodes arranged in a cyclical pattern. Node 1 connects to Node 2, Node 2 connects to Node 3, and Node 3 connects back to Node 1, creating a closed loop with no starting or ending point.
- **S4 – Dual Cause:** In this pattern, Node 3 acts as a central point of influence, connecting to both Node 1 and Node 2. This creates two distinct directional paths leading out from Node 3 to the other two nodes.
- **S5 – Common Cause:** This pattern comprises three nodes, with Node 3 connecting to both Node 1 and Node 2. Unlike in the Feed Forward Loop

(S2), Node 1 and Node 2 do not directly connect to each other, positioning Node 3 as the common source influencing both other nodes.

Between the five possible three-node subgraphs in Figure 4, S3 is excluded, since discourse structure graphs are not allowed to include loops that could relate a node back to itself. Naturally, discourse advancement forbids this type of subgraph and, thus, S3 is excluded from our analysis (although it is not analyzed, we included it in the graphs in Section 5 and in Table A2 with the mean pattern values; see the discussion below).

The frequency of these patterns in the network is compared to their occurrence in suitably randomized networks. From this comparison, z-scores are calculated to determine the likelihood of a particular subgraph appearing by chance. Z-scores higher than 2 indicate that a particular pattern is frequently present in the actual network, and it is unlikely that this is due to pure chance, since it is infrequent in the randomized graphs. Such a pattern is termed a network motif. It is equally important to locate patterns that are missing from the original network while they appear in the randomized graphs. Such an absence would indicate that there is a mechanism prohibiting their appearance. We identify them when their associated z-score is less than -2 . Such patterns are termed antimotifs, i.e. patterns that are rarer than at random, and may be equally important for revealing aspects of the discourse structure. Thus, we must emphasize the difference between a pattern and a motif. For a pattern to be termed a network motif, it is required that it appears much more often or rarely than it would, on average, appear in random networks.

Figure 4 illustrates the five potential three-node subgraphs (patterns) where two of three edges are directed. An important finding in biological networks was that the feed-forward loop (S2 in Figure 4) is a prevalent motif in the Transcriptional Regulatory Network (TRN) of *E.coli* (Alon 2019). These subgraphs, however, cannot be directly applied to the discourse graphs defined by the annotation principles of the Discourse GraphBank corpus, which recognizes two different types of nodes: EDUs and CDUs.

Below, we provide a brief explanation of the significance of 3-node subgraphs in discourse networks. A more comprehensive discussion can be found in Tantos & Kosmidis (2023).

- **S1: Linear Chains**

S1 represents linear chains of utterances connected by discourse relations, which can be either symmetrical or asymmetrical. An example of a linear chain pattern is *53 53 54 54 temp / 54 54 55 55 temp*, as shown in (3) from

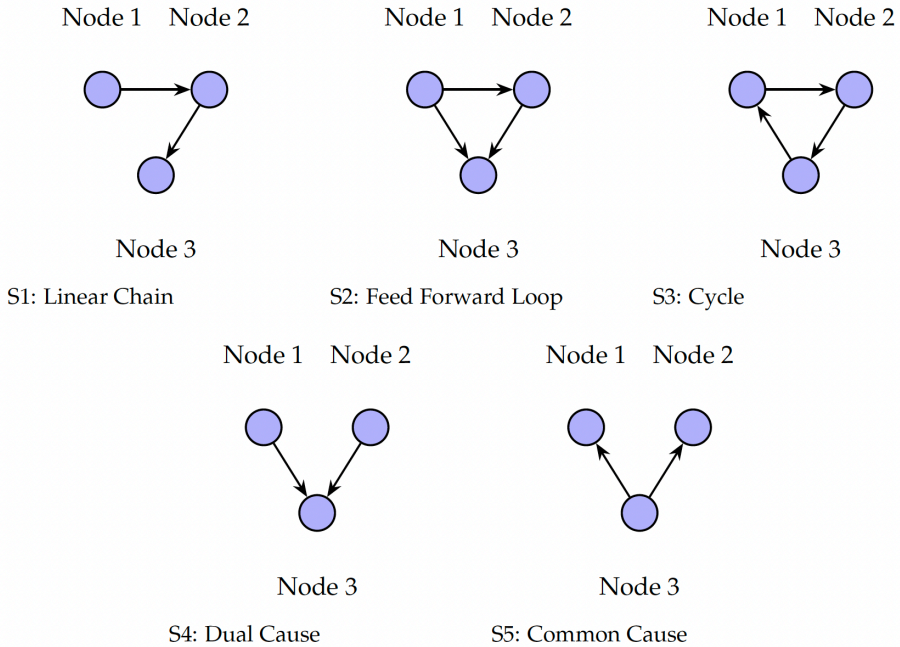


Figure 4: The S1–S5 network patterns

file #110. In this case, all three nodes are temporally related in a sequential manner through the temporal sequence relation. However, triplets of nodes are not always connected by temporal sequence relations, as one might assume. For instance, in (4), based on the annotation *6 6 7 7 attr / 7 7 1 1 elab-det* from file #99, the linear chain pattern still holds.³

- (3) 53. He studied at a glass-cutting school in Poland
 54. then spent 20 years cutting glass there.
 55. He and his wife and son, Matthew, now 10, immigrated to the United States in April 1983, two weeks before huge antigovernment protests at the Lenin shipyards.

- (4) 1. The commission trying to find a way to eliminate the huge federal budget deficit has the go-ahead from President-elect Bush to start

³*attr* corresponds to the attribute relation between a source EDU and a target EDU, while *elab-det* refers to detail elaborations, where the satellite EDU provides details about an entity involved in the event described by the nucleus EDU.

putting its proposals in writing.

...

6. Commission co-chairman Drew Lewis said

7. the panel would be able to meet the deadline, ...

- **S2: Full Connectivity**

S2 illustrates fully-connected discourse subgraphs where all three nodes are interconnected. Such full connectivity is often interpreted as indicating maximum discourse coherence, in line with the idea that higher network connectivity is directly associated with enhanced coherence. In (5), a passage taken from file #1 of the corpus, annotated as *24 24 23 23 elab-det / 24 24 26 26 par / 26 26 23 23 elab-det*, maps to the S2 pattern, where EDU 24 is a satellite to both 23 and 26 through *elab-det* and *par* respectively. Additionally, 26 and 23 are connected through a parallelism relation. While it may seem plausible to view pattern S1 as a mere subgraph of S2, this interpretation overlooks the critical distinction between the two. The additional link in the feed-forward loop of S2 signifies a deliberate choice by the speaker to enhance the connectivity between discourse elements, indicating a more tightly woven structure. In contrast, the absence of this link in S1 suggests a different, less interconnected discourse configuration, reflecting a more linear progression. The presence or absence of links directly correlates with the level of discourse coherence, where S2's full connectivity leads to stronger semantic relationships, while S1 reflects a simpler, more loosely connected discourse structure. This difference highlights that S1 cannot be viewed simply as a subset of S2, as each represents a distinct discourse strategy.

(5) 23. Students learn to program a computer and automated machines linked to it in a complete manufacturing operation -

24. retrieving raw materials from the storage shelf unit,

...

26. lifting and placing the parts in position with the robot's arm;

- **S4: Central Nodes and Indirect Connections**

S4, known as the dual cause pattern, represents discourse subgraphs where the central node crucially links the first and last nodes. This central node may contribute to enhancing discourse coherence, though its function differs from the nucleus-satellite relationships described in RST. In S4, the

nodes may be more distantly connected due to the possibility of either direct or indirect links between the central node and the other two nodes. An example of this can be found in (6) which corresponds to the annotation *12 12 11 11 elab-det / 13 13 11 11 elab-pers* in file #1 of GraphBank.⁴

- (6) 11. Ms. Alptekin has created “a model for similar laboratories ...
12. Students can see how the factory of the future operates.”
13. The Society of Manufacturing Engineers has recognized Ms. Alptekin as an Outstanding Young Manufacturing Engineer.

- **S5: Prominent First Nodes**

S5, or the common cause pattern, corresponds to discourse graphs where the first node connects directly to the second and third nodes via two outgoing edges. Therefore, the first node may be thought of being more prominent in the triplet, fostering indirect connections between the second and third nodes through itself. In (7) a short passage from the file #101 of the Discourse GraphBank is included where the S5 pattern may be traced due to the annotation *3 3 2 2 temp / 3 3 6 6 temp*.

- (7) 2. Siegel spent Wednesday night in the Lieber Correctional Institution in Ridgeville.
3. But he was released ...
6. Siegel was then sent back to Charleston County,

In summary, we delineated the four network patterns of interest and discussed the associated properties of the corresponding discourse graphs, such as connectivity, coherence, and the roles of central or prominent nodes. While these patterns may provide a framework for analyzing discourse coherence, further empirical validation is necessary.

4.5 Network motifs in Discourse GraphBank

The distribution of z-scores for the five fundamental network patterns, described in Figure 4, is presented in Figure 5. Specifically, we counted the number of

⁴*Elab-det* and *elab-pers* are two subclasses of Elaboration. *Elab-det* provides information about an entity that cannot be categorized under other Elaboration subclasses, whereas *elab-pers* offers details about a person involved in the event described by the Nucleus.

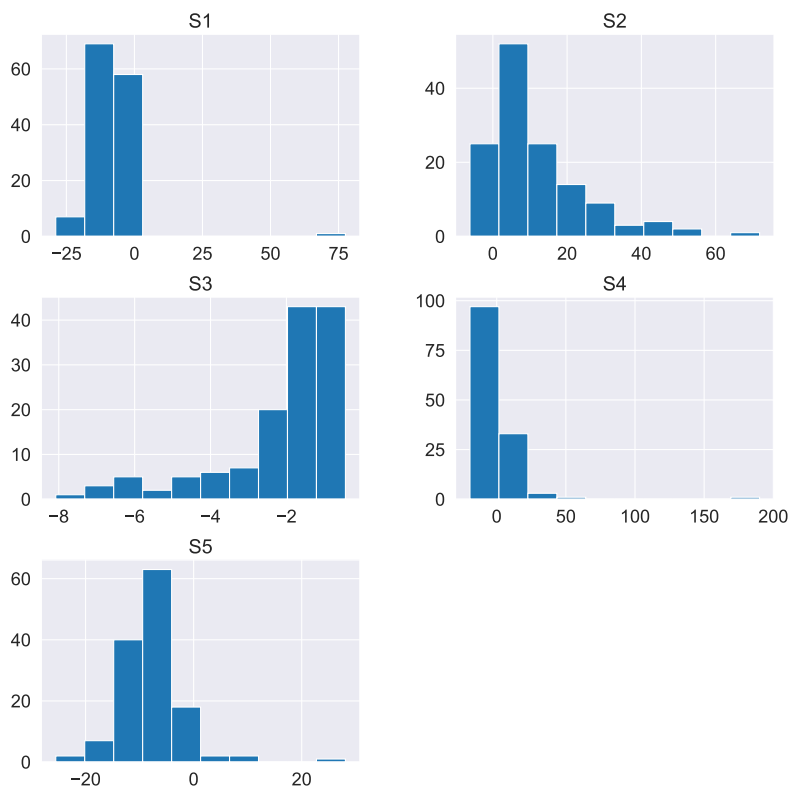


Figure 5: GraphBank Networks – Histograms of network motif z-scores. Three node subgraph patterns are denoted as follows: S1 – linear chain, S2 – the feed forward loop, S3 – a cycle, S4 – dual cause, S5 – common cause

$S_i, i = 1..5$ patterns in each of the 135 GraphBank discourse networks. Next, using the same number of nodes and edges as in the actual networks, we generated 100 random networks and measured the occurrence of the same S_i patterns in each of them. A z-score was calculated for each pattern in each actual network as follows: the number of occurrences of pattern S_i in the actual network is denoted by n_i . Let μ and σ represent the mean and standard deviation of S_i occurrences in the randomized networks, respectively. The z-score is computed using the formula $z = \frac{n_i - \mu}{\sigma}$. Figure 5 shows that the feed-forward loop, S2, is clearly a network motif

in a significant number of the 135 GraphBank networks, with several cases exhibiting very high z-scores (in some instances, exceeding 20!). The discovery that the feed-forward loop, a well-known motif in biological networks, also appears in linguistic networks is certainly noteworthy.

It is also noteworthy that *S1* and *S5*, which are the linear chain and common cause patterns depicted in Figure 4, are consistently identified as antimotifs, as evidenced by their large negative z-scores in the distribution. Additionally, the dual cause pattern *S4* presents an interesting mixed behavior. In some cases, it acts as an antimotif, with several GraphBank networks showing z-scores as low as -10 . Conversely, in fewer instances, *S4* appears as a motif, with z-scores exceeding 20. This variability in the structural occurrence of *S4* highlights the complex interplay of network dynamics within the GraphBank discourse networks. This phenomenon calls for further detailed and qualitative analysis to fully understand the factors driving such diverse manifestations of this pattern.

To elucidate the previous discussion, we present in Figure 6 a heatmap of motif z-scores for all 135 GraphBank networks. This visualization employs a color-coded system to distinguish between antimotifs, motifs, and neutral patterns. Green shades indicate z-scores less than -2 , marking the presence of antimotifs—patterns that occur less frequently than expected by chance, implying they may have a minimal role in the network's structure or functionality. Yellow shades are used for z-scores ranging between -2 and 2 , denoting patterns of neutral significance, which are neither overrepresented nor underrepresented significantly. Blue shades signify z-scores greater than 2 , identifying motif—patterns that appear more often than what random chance would predict, suggesting their significant involvement in the network's structure or function. Although the full names of most networks are not included on the y-axis, all 135 networks are represented in colors along the heatmap. This focuses the viewer's attention on the broad patterns and distributions of motifs and antimotifs rather than individual network identities. The plot visually confirms that *S1* and *S5*, the linear chain and common cause patterns respectively shown in Figure 4, are clearly antimotifs, as indicated by the green color predominating in the first and last column of Figure 6. Additionally, it illustrates that the feed-forward loop *S2* is distinctly a network motif, evidenced by the blue color saturating the *S2* column.

All numerical simulations and data analyses in this study were conducted on a workstation equipped with two Intel Xeon Gold 6140 processors, totaling 72 CPU cores, provided by the MSc program in “Computational Physics” at the Physics Department of Aristotle University of Thessaloniki. Network motifs were identified using custom Python code. Basic network properties were analyzed using the Python modules NetworkX (Hagberg et al. 2008) and rustworkx (Treinish

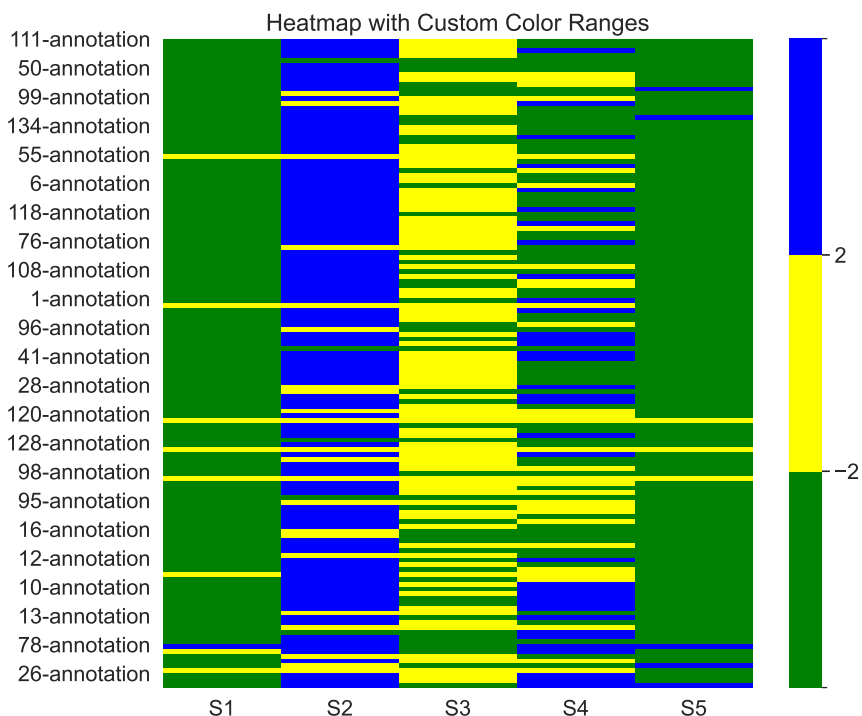


Figure 6: Heatmap of motif z -scores across 135 networks, using a color-coded scheme to distinguish between antimotifs, motifs, and neutral patterns according to their statistical significance. Green signifies z -scores below -2 , marking antimotifs.

et al. 2021). The code was parallelized with the Python module Dask (Dask Development Team 2016), which is crucial for this project as searching for network motifs is computationally demanding, even in moderately sized networks. For instance, analyzing a network with 10^3 nodes can take about 64 hours on a Xeon Gold 6140 processor. Therefore, without parallelization, analyzing the 135 networks from the Discourse GraphBank would be impractical. Additionally, data preprocessing was carried out using Python and R on a PC equipped with an Intel i9 processor with 8 cores.

5 Discussion

In this study, we employed network analysis techniques to explore the deep properties of discourse structures using the Discourse GraphBank corpus. Our anal-

ysis aimed to discern the prevalence and significance of network motifs across a diverse spectrum of Discourse GraphBank networks.

Our results showcased the feed-forward loop *S2* as a motif in many of the networks analyzed, with certain instances exhibiting extraordinarily high z-scores, some exceeding 20. This specific three-node configuration, where one node triggers an activation sequence through two subsequent nodes, mirrors similar patterns observed in biological signaling pathways. Such structures are celebrated for their ability to provide nuanced control over temporal signaling and amplifying responses, underscoring their evolutionary and functional significance.

In contrast, our analysis pinpointed the linear chain and common cause patterns as definitive antimotifs, marked by significant negative z-scores. The linear chain is characterized by a sequential connection of nodes, offering minimal structural complexity. Similarly, the common cause pattern, which connects two nodes to a shared intermediary, lacks the intricate connectivity that typifies more dynamically functional motifs. The identification of these patterns as antimotifs aligns with expectations based on their structural simplicity, reinforcing the notion that more complex network features are typically associated with enhanced functional capabilities.

Moreover, the dual cause pattern *S4* displayed intriguing variability in its role across the networks. In some instances, *S4* appeared as an antimotif with z-scores reaching as low as -10, indicating its infrequent occurrence compared to random expectation. However, in other cases, it emerged as a prominent motif, with z-scores above 20. This pattern, which features two independent nodes influencing a common third node, suggests a potential mechanism for integrating multiple signal pathways or feedback loops, enhancing the network's ability to coordinate complex interactions. Such a pattern might be instrumental in systems requiring robust control mechanisms to synchronize input signals effectively.

This nuanced exploration not only reaffirms the feed-forward loop's pivotal role within biological and potentially linguistic frameworks but also highlights the diverse functional implications of simpler and more complex network motifs in discourse analysis. The observed variability in the dual cause motif underscores the need for further research into how such patterns may impact the dynamics and functionality of complex network systems, particularly within the realm of linguistic discourse. We strongly believe that this investigation is crucial for understanding the relationship between these network patterns and the concepts of coherence and cohesion in discourse structures.

This study is closely related to previous work by Tantos & Kosmidis (2023), where network theory was applied to the C58 and STAC corpora. While the methodologies in both studies share a common foundation in network analysis,

our current investigation builds upon and extends this earlier work by applying these techniques to a new corpus—the Discourse GraphBank. Unlike the earlier study, which focused on comparing discourse structures from different types of corpora, our analysis dives deeper into the identification and significance of motifs and antimotifs across a larger and more complex set of discourse networks.

Notably, both studies highlight the presence of certain network motifs, such as the feed-forward loop. However, this paper also sheds light on the variability of the dual cause pattern, which was not a key focus in Tantos and Kosmidis (2023). These differences suggest that network-based discourse analyses exhibit a broader range of structural dynamics, warranting further exploration of how such patterns influence discourse coherence and cohesion. Our findings, therefore, offer a new perspective on the role of network motifs in discourse networks, contributing to the growing understanding of how discourse structures resemble networks and enriching discourse analysis with network-analytic tools.

6 Appendix

Table 1: Basic statistics for a number of real networks: The properties presented here are Number of nodes *nsize*, Number of edges *nedge*, clustering coefficient *C* and assortativity coefficient *r*. Data taken from Newman (2010).

Network	Nsize	Nedge	C	r
Film actors	449,913	25,516,482	0.78	0.208
Company Directors	7673	55,392	0.88	0.276
Math coauthorship	253,339	496,489	0.34	0.120
Physics coauthorship	52,909	245,300	0.56	0.363
Biology coauthorship	1,520,251	11,803,064	0.60	0.127
Email address books	16,881	57,029	0.13	0.092
Student dating	573	477	0.001	-0.029
WWW nd.edu	269,504	1,497,135	0.29	-0.067
Roget's thesaurus	1022	5103	0.15	0.157
Internet	10,697	31,992	0.039	-0.189
Power grid	4941	6594	0.080	-0.003
Train routes	587	19,603	0.69	-0.033
Software packages	1439	1723	0.082	-0.016
Software classes	1376	213	0.012	-0.119
Electronic circuits	24,097	53,248	0.030	-0.154
Peer-to-Peer network	880	1269	0.011	-0.366
Metabolic network	765	3686	0.67	-0.240
Protein interactions	2115	2240	0.071	-0.156
Marine food web	134	598	0.23	-0.263
Freshwater food web	92	997	0.087	-0.326
Neural network	307	2359	0.28	-0.226

References

- Alon, Uri. 2019. *An introduction to systems biology: Design principles of biological circuits*. CRC Press.
- Asher, Nicholas. 1993. *Reference to abstract objects in discourse*. Springer.
- Asher, Nicholas & Alex Lascarides. 2003. *Logics of conversation*. Cambridge University Press.

- Asher, Nicholas, Antoine Venant, Philippe Muller & Stergos Afantenos. 2011. Complex discourse units and their semantics. In Nicholas Asher (ed.), *CID 2011—constraints in discourse*. Agay.
- Asher, Nicholas & Laure Vieu. 2005. Subordinating and coordinating discourse relations. *Lingua* 115. 591–610.
- Barabási, Albert-László & Réka Albert. 1999. Emergence of scaling in random networks. *Science* 286. 509–512.
- Barabási, Albert-László, Natali Gulbahce & Joseph Loscalzo. 2011. Network medicine: A network-based approach to human disease. *Nature Reviews Genetics* 12. 56–68.
- Barzel, Baruch & Albert-László Barabási. 2013. Universality in network dynamics. *Nature Physics* 9. 673–681. <https://doi.org/10.1038/nphys2741>.
- Bashan, Amir, Ronny P. Bartsch, Jan W. Kantelhardt, Shlomo Havlin & Plamen C. Ivanov. 2012. Network physiology reveals relations between network topology and physiological function. *Nature Communications* 3. <https://doi.org/10.1038/ncomms1705>.
- Battiston, Stefano, J. Doyne Farmer, Andreas Flache, Diego Garlaschelli, Andrew G. Haldane, Hans Heesterbeek, Cars Hommes, Carlo Jaeger, Robert May & Marten Scheffer. 2016. Complexity theory and financial regulation. *Science* 351. 818–819. <https://doi.org/10.1126/science.aad0299>.
- Beber, Moritz E., Christoph Fretter, Shubham Jain, Nikolaus Sonnenschein, Matthias Müller-Hannemann & Marc-Thorsten Hütt. 2012. Artefacts in statistical analyses of network motifs: General framework and application to metabolic networks. *Journal of the Royal Society Interface* 9. 3426–3435.
- Boyer, Lorie A., Tong Ihn Lee, Megan F. Cole, Sarah E. Johnstone, Stuart S. Levine, Jacob P. Zucker, Matthew G. Guenther, Roshan M. Kumar, Heather L. Murray, Richard G. Jenner, David K. Gifford, Douglas A. Melton, Rudolf Jaenisch & Richard A. Young. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122. 947–956.
- Bullmore, Ed & Olaf Sporns. 2009. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* 10. 186–198.
- Choudhury, Monojit & Animesh Mukherjee. 2009. The structure and dynamics of linguistic networks. In *Dynamics on and of complex networks*, 145–166. Springer.
- Clauset, Aaron, Christopher Moore & M.E.J. Newman. 2008. Hierarchical structure and the prediction of missing links in networks. *Nature* 453. 98–101. <https://doi.org/10.1038/nature06830>.

- Creswell, Cassandre, Katherine Forbes, Eleni Miltsakaki, Rashmi Prasad, Aravind K. Joshi & Bonnie Webber. 2003. Penn discourse treebank: Building a large scale annotated corpus encoding DLTAG-based discourse structure and discourse relations. In *Proceedings of the Penn Discourse Treebank*.
- D'Souza, Raissa M., Jesus Gómez-Gardeñes, Jan Nagler & Alex Arenas. 2019. Explosive phenomena in complex networks. *Advances in Physics* 68. 123–223. <https://doi.org/10.1080/00018732.2019.1650450>.
- Dask Development Team. 2016. *Dask: Library for Dynamic Task Scheduling*. Software package. <https://docs.dask.org/en/stable/>.
- Grosz, Barbara J. & Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3). 175–204. <https://aclanthology.org/J86-3001>.
- Gupta, Kuhika, Joseph Ripberger, Andrew Fox, Hank Jenkins-Smith & Carol Silva. 2022. Discourse network analysis of nuclear narratives. In *Narratives and the policy process: Applications of the narrative policy framework*. Pressbooks.
- Hagberg, Aric, Pieter Swart & Daniel S. Chult. 2008. *Exploring network structure, dynamics, and function using NetworkX*. Tech. rep.
- Hobbs, Jerry. 1985. *On the coherence and structure of discourse*. Tech. rep. CSLI Research report number 85–37.
- Jiang, Congcong, Tiejun Qian & Bing Liu. 2022. Knowledge distillation for discourse relation analysis. In *Proceedings of the companion proceedings of the web conference 2022*, 210–214.
- Jiang, Lanlan, Shengjun Yuan & Jun Li. 2021. A discourse coherence analysis method combining sentence embedding and dimension grid. *Complexity*. 6654925. <https://doi.org/10.1155/2021/6654925>.
- Klosik, David F., Stefan Bornholdt & Marc-Thorsten Hütt. 2014. Motif-based success scores in coauthorship networks are highly sensitive to author name disambiguation. *Physical Review E* 90. 032811.
- Knott, Alistair. 2007. Coherence in natural language: Data structures and applications. *Computational Linguistics* 33. 591–595.
- Kuppevelt, Jan Van. 1995. Main structure and side structure in discourse. *Linguistics* 33. 809–833.
- Lacasa, Lucas, Bartolo Luque, Fernando Ballesteros, Jordi Luque & Juan Carlos Nuño. 2008. From time series to complex networks: The visibility graph. In *Proceedings of the national academy of sciences of the United States of America*, vol. 105, 4972–4975. <https://doi.org/10.1073/pnas.0709247105>.
- Lai, Alice & Joel Tetreault. 2018. Discourse coherence in the wild: A dataset, evaluation and methods. In *Proceedings of the 19th annual SIGdial meeting on discourse and dialogue*, 214–223. <https://doi.org/10.18653/v1/W18-5023>.

- Lin, Tianyang, Yuxin Wang, Xiangyang Liu & Xipeng Qiu. 2021. *A survey of transformers*. <https://arxiv.org/abs/2106.04554>.
- Liu, Jiangming, Shay B. Cohen & Mirella Lapata. 2019. Discourse representation structure parsing with recurrent neural networks and the transformer model. In *Proceedings of the IWCS shared task on semantic parsing*, 429–439. <https://doi.org/10.18653/v1/W19-1203>.
- Ma, Yuhao, Jian Zhu & Jie Liu. 2022. Enhanced semantic representation learning for implicit discourse relation classification. *Applied Intelligence* 52. 7700–7712.
- Mann, William C. & Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text — Interdisciplinary Journal for the Study of Discourse* 8(3). 243–281. DOI: [doi:10.1515/text.1.1988.8.3.243](https://doi.org/10.1515/text.1.1988.8.3.243). <https://doi.org/10.1515/text.1.1988.8.3.243>.
- Marr, Carsten, Fabian J. Theis, Larry S. Liebovitch & Marc-Thorsten Hütt. 2010. Patterns of subnet usage reveal distinct scales of regulation in the transcriptional regulatory network of *Escherichia coli*. *PLoS Computational Biology* 6. e1000836.
- Mehler, Alexander. 2008. Large text networks as an object of corpus linguistic studies. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook of the science of language and society*, 328–382. De Gruyter.
- Mehler, Alexander, Andy Lücking, Sven Banisch, Philippe Blanchard & Barbara Frank-Job. 2016. *Towards a theoretical framework for analyzing complex linguistic networks*. Springer.
- Milo, Ron & Uri Alon. 2002. Network motifs: Simple building blocks of complex networks. *Science* 298. 824–827. <https://doi.org/10.1126/science.298.5594.824>.
- Narasimhan, Karthik & Regina Barzilay. 2015. Machine comprehension with discourse relations. In *Proceedings of the 53rd annual meeting of the Association for Computational Linguistics (ACL) and the 7th international joint conference on natural language processing (volume 1: long papers)*, 1253–1262. <https://doi.org/10.3115/v1/P15-1121>.
- Newman, M. E. J., Steven H. Strogatz & Duncan J. Watts. 2001. Random graphs with arbitrary degree distributions and their applications. *Physical Review E* 64. 026118. <https://doi.org/10.1103/PhysRevE.64.026118>.
- Newman, Mark. 2010. *Networks: An introduction*. Oxford University Press.
- Newman, Mark, Albert-László Barabási & Duncan J. Watts. 2006. *The structure and dynamics of networks*. Princeton University Press.
- Polanyi, Livia. 1985. A theory of discourse structure and discourse coherence. In Paul D. Kroeber, William H. Eilfort & Karen L. Peterson (eds.), *Papers from the general session at the 21st meeting of the Chicago Linguistics Society*. Chicago Linguistic Society.

- Polanyi, Livia. 1988. A formal model of discourse structure. *Journal of Pragmatics* 12. 601–638.
- Polanyi, Livia, Chirs Culy, Martin van den Berg, Lorenzo Thione & David Ahn. 2004. A rule based approach to discourse parsing. In *Proceedings of the 5th SIGdial workshop on discourse and dialogue at HLT-NAACL*, 108–117.
- Polanyi, Livia & Remko Scha. 1984. A syntactic approach to discourse semantics. In *10th international conference on computational linguistics and 22nd annual meeting of the Association for Computational Linguistics (ACL)*, 413–419. Association for Computational Linguistics. DOI: 10.3115/980491.980576. <https://aclanthology.org/P84-1085>.
- Reimers, Nils & Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, 3982–3992. <https://doi.org/10.18653/v1/D19-1410>.
- Rohde, Hannah, Alexander Johnson, Nathan Schneider & Bonnie Webber. 2018. Discourse coherence: Concurrent explicit and implicit relations. In *Proceedings of the 56th annual meeting of the Association for Computational Linguistics (ACL) (volume 1: long papers)*, 2257–2267. <https://doi.org/10.18653/v1/P18-1210>.
- Shen-Orr, Shai S., Ron Milo, Shmoolik Mangan & Uri Alon. 2002. Network motifs in the transcriptional regulation network of escherichia coli. *Nature Genetics* 31. 64–68.
- Song, Chaoming, Shlomo Havlin & Hernán A. Makse. 2005. Self-similarity of complex networks. *Nature* 433. 392–395. <https://doi.org/10.1038/nature03248>.
- Song, Chaoming, Shlomo Havlin & Hernán A. Makse. 2006. Origins of fractality in the growth of complex networks. *Nature Physics* 2. 275–281. <https://doi.org/10.1038/nphys266>.
- Tantos, Alexandros & Kosmas Kosmidis. 2023. From discourse relations to network edges: A network theory approach to discourse analysis. *Applied Sciences* 13(12). 6902.
- Treinish, Matthew, Ivan Carvalho, Georgios Tsilimigkounakis & Nahum Sá. 2021. *Rustworkx: A high-performance graph library for Python*. <https://arxiv.org/abs/2110.15221>.
- van Noord, Rik, Lasha Abzianidze, Antonio Toral & Johan Bos. 2018. Exploring neural methods for parsing discourse representation structures. *Transactions of the Association of Computational Linguistics* 6. 619–633. https://doi.org/10.1162/tacl_a_00241.

- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser & Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st international conference on neural information processing systems*, 6000–6010. Curran Associates Inc.
- Watts, Duncan J. & Steven H. Strogatz. 1998. Collective dynamics of ‘small-world’ networks. *Nature* 393. 440–442.
- Webber, Bonnie & Aravind Joshi. 2012. Discourse structure and computation: Past, present and future. In *Proceedings of the ACL-2012 special workshop on rediscovering 50 years of discoveries*, 42–54.
- Wolf, Florian & Edward Gibson. 2005. Representing discourse coherence: A corpus-based study. *Computational Linguistics* 31. 249–287. <https://doi.org/10.1162/0891201054223977>.
- Wu, Changxing, Liuwen Cao, Yubin Ge, Yang Liu, Min Zhang & Jinsong Su. 2022. A label dependence-aware sequence generation model for multi-level implicit discourse relation recognition. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, 11486–11494.
- Xiang, Wei & Bang Wang. 2023. A survey of implicit discourse relation recognition. *ACM Computing Surveys* 55. 1–34.
- Zhang, Yingxue, Fandong Meng, Peng Li, Ping Jian & Jie Zhou. 2021. Context tracking network: Graph-based context modeling for implicit discourse relation recognition. In *Proceedings of the 2021 conference of the north american chapter of the Association for Computational Linguistics: Human language technologies (ACL-HLT)*, 1592–1599.

Chapter 8

Modal and Higher Order Logical Reasoning with SUMO

Adam Pease^a

^aNaval Postgraduate School, Monterey

Prior work in modal logic has not examined how to integrate multi-modal reasoning with a large common sense knowledge base. We have been exploring how to translate the axioms in the Suggested Upper Merged Ontology (SUMO) into several different logics that can be processed with different automated theorem provers. This paper discusses how a multi-modal logic can be used with a comprehensive ontology. We describe the techniques needed to automatically translate formulas in the theory for use in automated inference.

1 Introduction

Inference with modal and higher order logics has a long history of foundational work (Fitting & Mendelsohn 1998, Blackburn et al. 2001, Goldblatt 2003), but the practical illustrations of their formulas have most often been limited to proofs of correctness or complexity (Kripke 1959), exploring the relationships among various axioms, or relatively small problems where all or most of the axioms are relevant to the particular problem being solved (Raths & Otten 2012). More recently work has been done in embedding several different modal operators in a single higher order logic framework, expanding the scope of possible inferences (Benzmüller & Paulson 2013).

If we wish for work in these expressive logics to have applicability to general purpose question answering and problem solving, in the context of explainable AI, then much larger and more comprehensive axiom sets are needed, where necessarily few of those axioms are employed for answering any given question. Such an exercise should also elucidate how different modalities can be used

in common sense question answering, and how the interactions among these modalities and first order logical expressions characterising our world govern which axioms to include in a single large theory.

For these reasons we have been exploring how to translate the axioms in the Suggested Upper Merged Ontology (SUMO) (Niles & Pease 2001, Pease 2011) into several different logics that can be processed with different automated theorem provers. This paper discusses how a multi-modal logic can be used with automated theorem proving together with a comprehensive ontology, describing the techniques needed to automatically translate formulas in the theory for use in automated inference.

2 Background

The overall goal of the SUMO effort is to capture as much knowledge as possible, in as formal and computable a language as possible. Over the course of the long-term effort since its origin in the year 2000, it was expected that there would be advances in logic and in automated theorem proving (ATP), and that the collection of knowledge should not be limited by the current state of the art. At the start of the effort, the semantics of first order logic were well-known and studied. While it was possible at that time to perform automated inference in first order logic, the implementation of efficient ATP with equality, enabled in part by the superposition calculus (Bachmair et al. 1992) was still new. While logicians had developed classical higher order logic (Church 1940, Henkin 1950, Andrews 1972) well before the SUMO effort, and automatic (Andrews et al. 1996) and interactive higher order logic provers (Paulson 1986) existed as well, it would still be some years before the ATP community reached a standard for a classical higher order language (Benzmüller et al. 2008). There also has only more recently been work to create a quantified multi-modal logic embedding in higher order logic (Benzmüller & Paulson 2013). While it might appear risky to define axioms in SUMO without a fully specified semantics, it has proven possible to refine and correct the body of knowledge as better support for more expressive logics has come about (Schulz et al. 2017, Pease & Schulz 2022). Had we limited the knowledge collected to what could be subjected to efficient inference at the time, much useful knowledge would have been missed.

The more expressive the logic used, the greater opportunity to create general axioms which have more inference power or what has been called a larger inferential closure. While the number of axioms in SUMO beyond first order is small compared to the largest category of ground binary relationships, they are con-

siderably more productive when part of a chain of inference. At the time of this writing there are the following statistics for SUMO (Table 1).

Table 1: SUMO Statistics

Total Terms	Total Axioms	Total Rules
16129	227412	6955

Relations: 1713

Ground tuples:	220371	Rules/Implications:	6955
of which are binary:	150609	of which are horn:	2311
of which are <i>arity</i> > 2:	69848	first-order:	5103
		temporal:	789
		modal:	257
		epistemic:	87
		other higher-order:	827

There are a few dozen relations in SUMO that take formulas as arguments:¹

- KappaFn
- ProbabilityFn
- attitudeForFormula
- believes
- causesProposition
- conditionalProbability
- confersNorm
- confersObligation
- confersRight
- considers
- containsFormula
- decreasesLikelihood
- deprivesNorm
- describes
- desires
- disapproves
- doubts
- entails
- expects
- hasPurpose
- hasPurposeForAgent
- holdsDuring
- holdsObligation
- holdsRight
- increasesLikelihood
- independentProbability
- knows
- modalAttribute
- permits
- prefers
- prohibits
- rateDetail
- treatedPageDefinition
- visitorParameter

¹The full axiomatization of each can be seen by entering the case-sensitive term in the “KB term” field at <https://sigma.ontologyportal.org:8443/sigma/Browse.jsp?kb=SUMO&lang=EnglishLanguage>

Over the years, we have created translators (Trac et al. 2008, Benzmüller & Pease 2010, Pease 2023)² from SUMO’s authoring language of SUO-KIF (Pease 2009) to languages in the TPTP (thousands of problems about theorem proving) family of languages (Sutcliffe 2010), as languages of increasing expressiveness have become available and supported in Automated Theorem Proving (ATP). The first translation effort was to the TPTP First Order Form (FOF) language. The first step was simply to convert the syntax of the two languages, as SUO-KIF uses a prefix notation for logic, built on LISP S-expression syntax and TPTP uses a combination of prefix and infix notation adhering to Prolog syntax. Since initial uppercase identifiers in Prolog syntax are interpreted as variables, we add a lowercase prefix to all constant terms and an uppercase prefix to all variables.

For a very simple axiom in SUMO like:

```
(=>
  (and
    (connected ?X ?Y)
    (part ?Y ?Z))
  (connected ?X ?Z))
```

it becomes in TPTP FOF:

```
fof(kb1,axiom,(![V__X,V__Y,V__Z]: (
  (s__connected(V__X,V__Y) & s__part(V__Y,V__Z)) =>
  (s__connected(V__X,V__Z))))).
```

SUO-KIF has only the basic logical operators as primitives (and or not => <=> forall exists equal). All other constant terms must be axiomatized in SUO-KIF to state their intended semantics explicitly. This causes an issue in that TPTP FOF requires that constant term and relation symbols are disjoint but SUMO refers to relation symbols as arguments in defining their types. Each relation symbol is given a suffix in translation when used as an argument to distinguish it from the same relation symbol, when used as a relation. For example we state that a wheel is a part of a particular car `s__part(s__MyWheel, s__MyCar)` but if we want to say that the `part` relation takes an instances of type `Object` as its first argument, we append a `_m` suffix as in `s__domain(s__part__m, 1, s__Object)`.

SUO-KIF has existential and universal quantifiers but for convenience, any variable that is not quantified is assumed to be universally quantified at the outermost scope.

²Implemented as part of the Sigma Knowledge Engineering Environment and found at <https://github.com/ontologyportal/sigmakee/tree/master/src/java/com/articulate/sigma/trans>

2.1 Type guards

All SUMO relations have a type signature. In order to enforce that signature, the translation involves adding type guards to formulas. During translation, all the type restrictions on variables are collected, with respect to which relation they appear in. If multiple restrictions apply, then the most restrictive type wins out. For example, say that B is a subclass of A, and C is a subclass of B. Further say that a relation *r1* is specified as having a first argument that is an instance C and another relation *r2* has a first argument of an instance of B. Finally, say that we have a formula with a variable ?X that is given as the first argument to *r1* and the first argument to *r2*. Then ?X would be given the type restriction of being required to be an instance of C, since C is a more restrictive type than B.

The part and connected relations both take two instance of the class Object as their arguments, so our example becomes:

```
fof(kb1,axiom,(![V__X,V__Y,V__Z]: (
  (s__instance(V__X,s__Object) &
   s__instance(V__Y,s__Object) &
   s__instance(V__Z,s__Object)) =>
  ((s__connected(V__X,V__Y) & s__part(V__Y,V__Z)) =>
   (s__connected(V__X,V__Z)))))).
```

An additional issue is that SUMO makes use of relation variables – variables in the relation position. While this might indicate use of a logic beyond first order, it can be given a restricted first order interpretation, which is what we do. The first order interpretation involves treating formulas with a relation variable as an axiom schema, where we create a new axiom for each relation in SUMO that can be substituted for a relation variable. Note that in a first order interpretation, we do not quantify over all possible relations, which would include ones that have not already been defined in SUMO.

```
(=>
  (instance ?REL TransitiveRelation)
  (=>
    (and
      (?REL ?INST1 ?INST2)
      (?REL ?INST2 ?INST3))
      (?REL ?INST1 ?INST3)))
```

becomes:

```
fof(kb2,axiom(![V__INST1,V__INST2,V__INST3]:(  
  (s__instance(V__INST1,s__Object) &  
    s__instance(V__INST2,s__Object) &  
      s__instance(V__INST3,s__Object)) =>  
    (s__instance(s__part,s__TransitiveRelation) =>  
      ((s__part(V__INST1,V__INST2) &  
        s__part(V__INST2,V__INST3)) =>  
          s__part(V__INST1,v__INST3))))))
```

This axiom is also generated automatically, with the appropriate type guards, for all other `TransitiveRelations`, such as `subclass`, `larger`, `sister` etc.

An additional language feature in SUO-KIF is that of the `VariableArityRelation`. We have a similar treatment to that of relation variables in that axioms with a variable arity relation can be treated as a schema, and expanded automatically. This is a restricted interpretation, but one which has been an acceptable limitation in practice, at least in our work so far. Each formula containing a variable arity relation becomes a set of formulas, expanding the arities up to an arbitrary limit of seven arguments and again employing type guards.

2.2 Row variables

Variables called *row variables*, and designated with a leading ‘@’ character, denote a list of several variables. This is a necessary language feature given that we have variable arity relations, and that we can have predicate variables, which means that in some axioms we do not know how many arguments a relation has.

If there is more than one row variable in a formula, then they expanded independently, so there will be 7^N copies of the axiom, where N is the number of row variables. An example of how variable arity relations and row variables are translated from SUO-KIF to TPTP FOF is as follows. We start with the source axiom in SUMO:

```
(=>  
  (partition @ROW)  
  (and  
    (exhaustiveDecomposition @ROW)  
    (disjointDecomposition @ROW)))
```

Partitions of classes are exhaustive and disjoint. A partition can have a variable number of arguments. We compute type guards for the variables and one axiom expands into several axioms for all the possible numbers of arguments for the variable arity relations.

```

fof(kb3,axiom(![V__ROW1,V__ROW2,V__ROW3]:(
  (s__instance(V__ROW1,s__Class) &
   s__instance(V__ROW2,s__Class) &
   s__instance(V__ROW3,s__Class)) =>
  (s__partition(V__ROW1,V__ROW2,V__ROW3) =>
   (s__exhaustiveDecomposition(V__ROW1,V__ROW2,V__ROW3)
    (s__disjointDecomposition(V__ROW1,V__ROW2,V__ROW3)))))))

```

```

fof(kb4,axiom(![V__ROW1,V__ROW2,V__ROW3]:(
  (s__instance(V__ROW1,s__Class) &
   s__instance(V__ROW2,s__Class) &
   s__instance(V__ROW3,s__Class) &
   s__instance(V__ROW4,s__Class)) =>
  (s__partition(V__ROW1,V__ROW2,V__ROW3,V__ROW4) =>
   (s__exhaustiveDecomposition(V__ROW1,V__ROW2,V__ROW3,V__ROW4)
    (s__disjointDecomposition(V__ROW1,V__ROW2,V__ROW3,V__ROW4)))))))

```

etc.

2.3 Additional translations

The next translation effort for SUMO was an initial attempt at classical typed higher order logic, with the TPTP THF language (Benzmüller & Pease 2010, 2012). We sketched an approach for translating modalities, including implementing some examples of hand translation. The implementation of this work in Sigma was restricted and did not handle all modalities. However, it provided a foundation for our present work.

We also implemented a translation (Pease 2023) to Typed First order form with Arithmetic (TFA) (Sutcliffe et al. 2012). TFA requires that we state the types of variables in the syntax of the language, rather than just as additional literals that constrain the meaning of a formula, as in SUO-KIF. There are only a few types: numbers, booleans (notated in TFA as type “\$o”) and everything else (type \$i). Users can also define custom types. Numbers are further divided into rationals (\$rat), reals (\$real) and integers (\$int). This causes some complexity in the SUMO/SUO-KIF translator since SUMO has subclasses of integers such as PositiveInteger. Typed higher order form (THF) language and its implementations in E and Vampire do not currently allow us to combine numbers and arithmetic functions with THF.

3 Modal logic axioms

Modal logic was originally developed to handle the notions of necessity and possibility (called *alethic* operators). While these operators can be used in a propositional modal logic, in this paper we are concerned with their use in first and higher order logic. Expanding first order logic with these two new operators supports additional statements that conform to natural language intuitions but using formal logic. The same approach to adding new modal operators was later expanded to addresses other modalities including knowledge, belief, and temporal relationships.

There have come to be a set of modal axioms that researchers have argued apply to some or all modalities. In this section we examine the modal axioms developed for modalities in SUMO prior to undertaking a systematic analysis. We also discuss development of a set of problems to test whether particular axioms should hold for different modalities.

Our goal in this work is not to rediscover modal axioms that have already been specified, but rather to provide a more comprehensive set of axioms (both modal and non-modal) in which we can test how different modalities can co-exist and not conflict. While many different modal operators and sets of associated axioms have been given in the literature, these have not to our knowledge been part of a large body of common sense knowledge. Having a much wider set of possible test cases and contexts in which to test modalities should provide additional guidance for how to create a comprehensive system for practical reasoning.

Necessity is normally given the symbol \Box and possibility the symbol \Diamond . They are unary operators that hold over a formula. So “John necessarily likes sushi.” would be $\Box \text{likes}(\text{john}, \text{sushi})$. The necessity operator states that something must be true in all worlds accessible from the current world. The possibility operator holds true over only some accessible worlds. Necessity and possibility operate similarly to quantifiers (\forall and \exists , respectively) over possible worlds. They have the same relationship as our usual first order quantifiers:

$$\Diamond\phi =_{def} \neg\Box\neg\phi \tag{1}$$

just as $\exists\phi =_{def} \neg\forall\neg\phi$. A set of modal axioms (termed ‘D’, ‘K’, ‘T’, ‘B’, ‘4’, ‘5’ etc.) have been developed in prior literature and different sets may be assumed for particular work in modal logic. We do not however limit ourselves to these axioms, they are only a starting point. In particular, different axiom sets have been proposed in deontic logic (Gabbay et al. 2013).

A paradox such as the Gentle Murderer, or Forrester’s paradox (Sinnott-Armstrong 1985) depends upon certain standard modal axioms from a logic of

possibility and necessity also being true in a logic of obligation. In order to provide flexibility for different axiomatizations for different modalities, we use a general modal relation in SUMO termed `modalAttribute`. For each modal attribute related to a formula, we can provide a different set of modal axioms.

Due to the fact that there are several modalities, and we claim that the axiomatizations of those modalities are different, we introduce a parameterized modal operator M for our notation in this paper. M has subscripts for each operator, and later we will see that it has additional subscripts for its parameters, such as the agent for which the particular modality holds. Many of the standard modal axioms are already in SUMO, for example, axiom ‘D’:

```
(=>
(modalAttribute ?FORMULA Necessity)
(modalAttribute ?FORMULA Possibility))
```

which we may notate more formally as $M_n P \rightarrow M_p P$, meaning that the modal operator of necessity M_n applied over a formula P implies that what the formula states is not just necessary but also possible, i.e. M_p applies to P .

We have ‘K’, the distribution axiom:

$$(M_n(P \rightarrow Q)) \rightarrow (M_n P \rightarrow M_n Q) \quad (2)$$

‘T’ – reflexivity:

$$(M_n P) \rightarrow P \quad (3)$$

axiom ‘4’:

$$(M_n P) \rightarrow (M_n(M_n P)) \quad (4)$$

axiom ‘5’:

$$(M_p A) \rightarrow (M_n(M_p A)) \quad (5)$$

axiom ‘B’:

$$P \rightarrow (M_n(M_p P)) \quad (6)$$

axiom ‘D’:

$$(M_n P) \rightarrow (M_p P) \quad (7)$$

3.1 Epistemic operators

To have a comprehensive theory of world knowledge, we need to include operators for the knowledge and beliefs of agents, so this is part of SUMO. Some previous work in modal logic has treated knows as the analogue to necessity and believes as the analogue to possibility.

We augment our notation here with a subscript for the agent, so $M_{ba}P_1$ signifies that an agent a believes formula P_1 . $M_{ka}P_1$ signifies that an agent a knows formula P_1 .

$$(M_{ba}P_1 \wedge M_{ba}P_2) \rightarrow (M_{ba}(P_1 \wedge P_2)) \quad (8)$$

$$(M_{ba}P_1 \vee M_{ba}P_2) \rightarrow (M_{ba}(P_1 \vee P_2)) \quad (9)$$

If you believe P1 and believe that believing P1 allows you to believe P2, then it makes sense that you would believe P2.

$$(M_{ba}P_1 \wedge (P_1 \rightarrow P_2)) \rightarrow (M_{ba}P_2) \quad (10)$$

However is less clear that its logical equivalent should hold.

$$(M_{ba}P_1 \wedge (\neg P_1 \vee P_2)) \rightarrow (M_{ba}P_2) \quad (11)$$

If I believe P1, it would not make sense to believe that either (or both) not P1 or P2 is true.

‘K’ the distribution axiom is generally considered to hold for epistemic logic:

$$(M_{ka}(P_1 \rightarrow P_2)) \rightarrow (M_{ka}P_1 \rightarrow M_{ka}P_2) \quad (12)$$

as does axiom ‘4’:

$$(M_{ka}P_1 \rightarrow (M_{ka}(M_{ka}P_1))) \quad (13)$$

Take axiom ‘T’: $\Box P \rightarrow P$ and replace modal box for necessity with ‘knows’:

$$M_{ka}P \rightarrow P. \quad (14)$$

Axioms ‘T’, ‘K’, and ‘4’ appear valid in an epistemic logic. If we add ‘believes’ as the analogue to possibility, axiom ‘B’ does not make sense. If P is true, that does not mean one knows that one believes P . Axiom ‘D’ is intuitively appropriate with epistemic operators — if one knows P then one also believes P . Axiom ‘5’

appears valid – if one believes P then one knows that one believes P . Axioms 4 and 5 could be called ‘introspection’ axioms for the epistemic (‘knows’) and doxastic (‘believes’) operators.

Axiom ‘D’ is in SUMO for epistemic and doxastic operators.

$$(M_{ka}P) \rightarrow (M_{ba}P). \quad (15)$$

as is axiom ‘5’.

$$(M_{ba}P) \rightarrow (M_{ka}(M_{ba}P)). \quad (16)$$

4 Deontic logic

It is typical to map obligation to necessity and permission to possibility. There are however critiques to this notion of what is called standard deontic logic (McNamara & Van De Putte 2022). ‘K’ could be true – it is obligatory that if you steal money you steal a small amount of money, therefore if you are obliged to steal money then you are obliged to steal a small amount of money (at least several times over). Or should we say if one is obliged to steal money (and one does) that one has in fact also stolen a small amount of money (several times over). We give M subscripts $_o$ for obligation, $_p$ for permission etc.

Let p be the proposition that one steals money. We should further clarify the standard example that p should be the proposition that one steals a large amount of money, since if p stands only for the general proposition of stealing money, maybe it is a tiny or insignificant amount of money. So p stands for the proposition that one steals a large amount of money and let p' be the proposition that one steals a small amount of money. $M_o p \rightarrow M_o p'$ should at the very least only hold true if p' is a “smaller” version of p . Put another way, if we treat stealing money as a set of many small stealing events, such as every occurrence of a bank robber stuffing a packet of bills in his duffel bag, p' should be a subset of that set.

Axiom ‘T’ does not appear to be true of obligation – if you ought to do P that does not mean P is true since people do not always do what they are supposed to. Axiom ‘4’ is questionable – if you ought to do P does that mean you ought to ought to do P ? Axiom ‘B’ is problematic – if P is true then ought it to be permissible to P ? If someone committed murder, that does not mean that one ought to be permitted to commit murder. Axiom ‘D’ however appears to make sense – if you ought to do P then you should be permitted to do P , as if the obligation and permission are conferred by the same authority, which we return to later.

We have this in two forms in SUMO. The relation of `confersNorm` expresses that one agent confers a formula (?F) with an associated normative force:

```
(=>
  (confersNorm ?E ?F Obligation)
  (confersNorm ?E ?F Permission))
```

A simpler form of relation simply states that a formula has a given normative force, with the authority of that force unspecified.

```
(=>
  (modalAttribute ?FORMULA Obligation)
  (modalAttribute ?FORMULA Permission))
```

and a related axiom is

```
(=>
  (deprivesNorm ?E ?F Permission)
  (deprivesNorm ?E ?F Obligation))
```

Axiom ‘5’ seems problematic – if you have permission for *P* does that mean you are obliged to have permission? It may be dangerous to have obligation and permission without an authority – what does it really mean that *P* is permitted? If there is no authority, there can be no notion of permissibility, just possibility. Permission and obligation are always with respect to an authority. Stated in natural language: “He was permitted to do X (by authority Y).” Once relativized to an authority, ‘5’ makes a bit more sense, but only if we have a hierarchy of authority – if entity *X* gives you permission to do *Y* then maybe some parent authority *A* should have obliged *X* to give you permission for *Y*, but that starts to sound rather contrived. To support the notion of an authority conferring or depriving norms, SUMO has the modal operators `confersNorm` and `deprivesNorm`.

Another element needed for a deontic theory is a restriction on the formulas that are the arguments to a deontic operator. An agent cannot be obligated by another agent to, for example, make $2 + 2 = 5$. An agent can only be obligated by another agent to perform some action. However, an immoral agent might compel another to do or not do something it has no control over, in order to be certain to apply sanction, such as with a totalitarian regime making it impossible for an opposition figure to escape punishment under inconsistent laws.

We need many more examples in a common, computable, multi-modal framework to make sure our intuitions accord with a particular axiomatization, and

that the axiomatization is consistent across all modalities, and in conjunction with the other formulas that we need to state about the world. Having a much larger inventory of practical axiomatizations of aspects of the world should help to highlight different scenarios where our modal axioms might be true or false.

5 Temporal logic

SUMO has Allen's 13 temporal relations between intervals (Allen 1984) and functions that allow specification of calendar dates and times. The function `WhenFn` allows us to relate temporal specifications to `Processes`. All the preceding can be handled in first order. However, we also have the relation `holdsDuring`, which relates a `TimePosition` (points and intervals) to a `Formula` that is true at that time position.

One system of temporal model logic is that of Prior (1957). There are four modal operators in Prior's system

- P: "It was the case that..." (P stands for "past")
- F: "It will be the case that..." (F stands for "future")
- G: "It always will be the case that..."
- H: "It always was the case that..."

We take 'P' and 'F' to be "It was/will be the case (at one point) that...". These can be handled in SUMO with the functions `FutureFn` and `PastFn` along with a deictic³ (Lyons 1977) `Now`. For example, "It was the case that John had a blue car."

```
(exists (?C ?T)
  (holdsDuring ?T
    (and
      (during ?T (PastFn Now))
      (instance ?C Automobile)
      (possesses John ?C)
      (attribute ?C Blue))))
```

³A deictic is a linguistic token that has meaning only in the context of a particular utterance, such as "here" or "now".

Axiom ‘F’ is handled by instead using (FutureFn Now) and ‘G’ and ‘H’ by having the holdsDuring over the entire (PastFn...) or (FutureFn...) rather than just being during the period. FutureFn and PastFn are functions that have a single TimePoint as an argument and which denote a TimeInterval that spans the time from the infinite past until the argument for PastFn and from the argument to the infinite future for FutureFn.

6 Translation of modalities

We have started our exploration of modalities by employing Kripke’s *possible worlds* semantics (Kripke 1980).⁴ However, we have also created some test cases in Neighborhood semantics (Pacuit 2017)

We relax the requirement (per Scott 1970) of having the “normal” axiom ‘K’. We use Kripke’s notion of an *accessibility relation* between possible worlds. The intuition is that modalities select sets of worlds that are related to other worlds. A benefit of this approach is that we can state some formulas involving modalities in first order logic, as well as using them in higher order logic.

We employ the standard relational semantics for modal logic. A relational model is a tuple $\mathfrak{M} = \langle W, R, V \rangle$ where: W is a set of possible worlds, R is a binary relation on W (which we call the *accessibility relation* between worlds), and V is a valuation function which assigns a truth value to each pair of an atomic formula and a world, (i.e. $V : W \times F \rightarrow \{0, 1\}$ where F is the set of atomic formulae).

Many modal operators in SUMO use the relation modalAttribute, which takes a Formula and a modal operator as arguments, as we have seen in some of the discussion above. There are an additional number of modalities that take a parameter of some sort, and a formula. These are considers, sees, believes, knows, holdsDuring, and desires. holdsDuring is different than the others in that its parameter is a TimePosition rather than an Agent.

Each of the modalities constrain the set of worlds in which a formula is true. We add a world parameter to every logical literal. Start with the SUO-KIF formula:

```
(authors ThomasEdison HarryPotterSorcersStone)
```

that makes the questionable assertion that Thomas Edison authored a Harry Potter book. However:

⁴The notion of possible worlds was first conceived by Leibniz and also developed by Kripke’s contemporary, Lewis (Lewis 1986)

```
(believes JohnSmith
  (authors ThomasEdison HarryPotterSorcersStone))
```

is a bit more reasonable, i.e. that merely some deluded person believes he wrote that book. So in the set of worlds corresponding to what John Smith believes, Edison wrote the book. We translate this into an implication (involving an accessibility relation between worlds and adding a world argument for every logical literal in the formula) that constrains the worlds in which this is true.

```
(accreln believes JohnSmith ?W ?W2) =>
  (authors ThomasEdison HarryPotterSorcersStone ?W2))
```

We can also compose modalities or nest them. Let's say that John only believes this statement on Tuesdays (simplifying away for the purpose of this example that it should just be Tuesdays during his adult lifetime or similar restrictions).

```
(=>
  (instance ?T Tuesday)
  (holdsDuring ?T
    (believes JohnSmith
      (authors ThomasEdison HarryPotterSorcerersStone))))
```

This becomes:

```
(=>
  (instance ?T Tuesday)
  (=>
    (accreln holdsDuring ?T ?W ?W2)
    (=>
      (accreln believes JohnSmith ?W2 ?W3)
      (authors ThomasEdison HarryPotterSorcerersStone ?W3))))
```

and in THF (leaving aside for now details like term name prefixing, for clarity):

```
thf(kb5,axiom,
  (![W:w,W2:w,W3:w,X:$o,Y:$o,A:$i]:(
    ((instance @ T @ tuesday) =>
      ((accreln @ holdsDuring @ T @ W @ W2) =>
        ((accreln @ believes @ johnSmith @ W2 @ W3) =>
          (authors @ thomasEdison @ harryPotterSorcerersStone @ W3))))))).
```

Now let's return to the relation `modalAttribute` which takes two arguments, a `Formula` and a `NormativeAttribute` in SUMO. We provide a new accessibility relation to handle one fewer argument, called `accrelN`. Axiom 'D' becomes:

```
thf(ax204,axiom,
  ( ! [F:(w > $o),W1:w,W2:w] : (
    ((accrelN @ Necessity @ W1 @ W2) => F) =>
    ((accrelN @ Possibility @ W1 @ W2) => F) ) ) ).
```

Note that `Formula F` has the type of worlds-to-booleans since, in a given world, a formula is true or not. This is a change from the original SUMO types, before the Kripke encodings were introduced.

6.1 Rigidity

Do we assume that all entities designate the same object in all possible worlds (LaPorte 2022)? This is referred to the *rigidity* of entities. In our first experiment (Brown et al. 2023) we took the most careful approach that everything, including the designation of entities, is relative to a world. Encoded in this way, every constant c becomes a function of a world $c(w)$. This restriction has not led us so far to discover any contradictions in SUMO. So for reasons of clarity and computational efficiency, we have relaxed this approach and assume that entities are rigid. Paraphrasing Kripke, “Gandhi still refers rigidly to something that would not exist if Gandhi had never been born in some world.” (Kripke 1980)

A related question is which relationships are rigid. Although one might imagine a world in which arithmetic does not work in the normal way, it seems unlikely that there would be much practical utility in allowing such a world. So we do not add a world argument to arithmetic expressions. However, it is reasonable to imagine a world in which Thomas Jefferson was only a farmer and never a US president, or where basketball star Michael Jordan was taller or heavier. We also assume that classes and class membership are rigid across worlds. Venus is always a planet and not a donut. Actor Clint Eastwood is always a Human and not an orangutan. Note however that SUMO is strict about modeling qualities that can change during the lifetime of an entity as relationships to instances of attributes. `ClintEastwood` is always a Human, but sometime a mayor and not an actor. So his attribute of being an actor is stated with respect to a possible world, but his being an instance of the class Human is not.

More work remains to see which of the ~1500 relations in SUMO are rigid and which require parameterization with a world argument. For now, we add a world argument to every statement that is not arithmetic or a subclass relationship.

7 Test problems

We have just a handful of test problems at the moment,⁵ including one that is designed not to get an answer/refutation. We are using “Vampire 4.8 HO - Sledgehammer schedules (2023-10-19)” and “E 3.0pre008 Shangri-La (57b1a04ffd1869871-b2c863334343811aea78b68)” on an 8 core Lenovo laptop. However the problems are all quite small so the execution times are minimal (Table 2).

Table 2: Test problem execution times (in seconds)

Problem	Vampire	Eprover
TQM2.tff	0.026	0.008
TQM3.tff	correct, saturation	correct, saturation
TQM3.thf	0.015	0.033
TQM6g.tff	0.008	0.291
TQM6g.thf	0.094	0.35
TQM7.tff	0.011	0.009
TQM7.thf	0.01	0.01

- Problem TQM2 encodes that “Mary knows that Bill likes Sue and Bob likes Joan. Does Mary know that Bill likes Sue? Does Bill like Sue?”
- Problem TQM3.tff encodes that “Bill believes that the Earth is flat. Does Bill know that the Earth is flat?” TQM3.thf adds “If you circumnavigate the Earth and it is flat you will die. Joe circumnavigates the Earth. Does Bill believe Joe will die?”
- Problem TQM6g encodes that “Bill believes things until they are contradicted and then believes the latest thing. Bill sees Mary wearing a red hat (after he sees her wearing a green hat). What color does Bill believe Mary is wearing (now)?”. The THF version of this problem is able to state the first sentence generally, but in typed first order form (TFF) it has to be specific to the wearing of colored hats.
- Problem TQM7 is from the QMLTP problems⁶ and encodes “if a second class flight to Paris necessarily costs \$70.00 and I necessarily fly to Paris second class, what is the necessary cost of my flight?”. The THF and TFA versions are similar to each other except that we can have numbers in the TFA version but the THF version uses constant terms.

⁵TQM* problems at <https://github.com/ontologyportal/sumo/tree/master/tests/TPTP>

⁶<http://www.iltp.de/qmltp/problems.html>

8 Higher order interpretations

One advantage of moving to a THF translation of SUMO is that we would have less preprocessing to do for axioms with type guards, predicate variables and row variables. We can refer to relations directly as objects in the logical language and directly use SUMO's domain relation that specifies argument types. We could simplify this even in TFF, as long as we rename relations (which we do with a suffix '*_m*') when they appear as arguments. To implement this approach in THF however we would still need numbers, which are not yet part of THF, or at least encode the first few integers as symbols such as *n_1*, *n_2* etc. as well as inequalities.

In TFF/TFA (as well as FOF) we have to preprocess all implications to determine the implied types of variables based on their use in relations, as described in section 2.1 to include type guards.

```
tff(kb1,axiom,(![V_X:$i,V_Y:$i,V_Z:$i,T_X1:$i,T_Y1:$i,T_Y2:$i]: (
  (s_domain(s_connected_m,1,T_X1) & s_instance(V_X,T_X1) &
  s_domain(s_connected_m,1,T_Y1) & s_instance(V_Y,T_Y1) &
  s_domain(s_connected_m,1,T_Z1) & s_instance(V_Z,T_Z1) &
  s_domain(s_part_m,1,T_Y2) & s_instance(V_Y,T_Y2) &
  s_domain(s_part_m,1,T_Z2) & s_instance(V_Z,T_Z2)) =>
  (s_connected(V_X,V_Y) & s_part(V_Y,V_Z)) =>
  (s_connected(V_X,V_Z))))).
```

In THF we do not need to instantiate predicate variables, which significantly reduces the size of the resulting translation from SUMO, since we do not need potentially hundreds of copies of the same formula, such as the axiom of transitivity, for every transitive relation (again assuming that numbers will eventually be supported in THF). In the formula below note that we can refer to the argument types of a predicate variable (*s_domain @ R @ 1 @ V_T1*) where that variable is a relation, appearing later as (*R @ V_INST1 @ V_INST2*)...

```
thf(kb1,axiom,(![V_INST1:$i,V_INST2:$i,V_INST3:$i,
  V_R:$i,V_T1:$i,V_T2:$i,V_T3:$i]: (
  ((s_domain @ R @ 1 @ V_T1) & (s_instance @ V_INST1 @ V_T1) &
  (s_domain @ R @ 2 @ V_T2) & (s_instance @ V_INST2 @ V_T2) &
  (s_domain @ R @ 1 @ V_T1) & (s_instance @ V_INST2 @ V_T1) &
  (s_domain @ R @ 2 @ V_T3) & (s_instance @ V_INST3 @ V_T3)) =>
  ((s_instance @ R @ s_TransitiveRelation) =>
```

$$(((R @ V_INST1 @ V_INST2) \& \\ (R @ V_INST2 @ V_INST3)) \Rightarrow \\ (R @ V_INST1 @ V_INST3))))$$

9 Conclusions and future work

We have sketched an approach to take SUMO's modal axioms and embed them in a multi-modal logic framework in TFF and THF. Our next step is to complete the implementation of an automated translator for SUMO that generates TFF and THF, according to what we have specified here. The present work is a means to an end of being able to do practical reasoning with knowledge about beliefs, obligations, temporal qualification etc. Once we have an automated translation in place that embodies the manual translations presented, we will attempt to create more tests to determine if we can develop a practically useful system of modal operators and axioms that is also logically consistent. Since up until this point we have only been able to do small scale proofs with a portion SUMO's modal constructs and other knowledge, a fully automated framework should allow us to make more rapid progress and provide a higher level of assurance that the resulting system is logically consistent.

10 Acknowledgments

Several discussions with Richard Crouch over the years, especially with regards to the possibility of embedding modalities in first order, and motivated by representations of the semantics of natural language, have formed the germ of this effort, and we are honored to celebrate Dr. Crouch's achievements with our addition to this volume.

The approach to Kripke semantics in this translation of SUMO, as well as the new higher order translation of SUMO into THF, are thanks to Chad Brown. Thanks also to Martin Suda, Chad Brown and an anonymous reviewer for their very helpful comments on a draft of this paper. Any errors in the text are however due solely to the author.

References

Allen, James F. 1984. Towards a general theory of action and time. *Artificial Intelligence* 23. 123–154. <https://api.semanticscholar.org/CorpusID:22263624>.

- Andrews, Peter B. 1972. General models and extensionality. *Journal of Symbolic Logic* 37. 395–397.
- Andrews, Peter B., Matthew Bishop, Sunil Issar, Dan Nesmith, Frank Pfenning & Hongwei Xi. 1996. TPS: A theorem-proving system for classical type theory. *Journal of Automated Reasoning* 16(3). 321–353.
- Bachmair, Leo, Harald Ganzinger, Christopher Lynch & Wayne Snyder. 1992. Basic paramodulation and superposition. In *Proceedings of the 11th international conference on automated deduction*, vol. 607 (Lecture Notes in Computer Science), 462–476. Springer-Verlag. DOI: 10.1007/3-540-55602-8_185.
- Benzmüller, Christoph & Lawrence C. Paulson. 2013. Quantified multimodal logics in simple type theory. *Logica Universalis* 7(1). 7–20. DOI: 10.1007/s11787-012-0052-y.
- Benzmüller, Christoph & Adam Pease. 2010. Progress in automating higher-order ontology reasoning. In Boris Konev, Renate Schmidt & Stephan Schulz (eds.), *Workshop on practical aspects of automated reasoning (PAAR-2010)*. CEUR Workshop Proceedings.
- Benzmüller, Christoph & Adam Pease. 2012. Higher-order aspects and context in SUMO. In Jos Lehmann, Ivan José Varzinczak & Alan Bundy (eds.), *Special issue on reasoning with context in the semantic web*, vol. 12-13. Science, Services & Agents on the World Wide Web.
- Benzmüller, Christoph, Florian Rabe & Geoff Sutcliffe. 2008. THF0 – the core of the TPTP language for higher-order logic. In Alessandro Armando, Peter Baumgartner & Gilles Dowek (eds.), *Automated reasoning*, 491–506. Springer.
- Blackburn, Patrick, Maarten de Rijke & Yde Venema. 2001. *Modal logic* (Cambridge Tracts in Theoretical Computer Science 53). Cambridge University Press.
- Brown, Chad E., Adam Pease & Josef Urban. 2023. Translating SUMO-K to higher-order set theory. In *International symposium on frontiers of combining systems*. <https://api.semanticscholar.org/CorpusID:258685309>.
- Church, Alonzo. 1940. A formulation of the simple theory of types. *Journal of Symbolic Logic* 5. 56–68.
- Fitting, Melvin & Richard L. Mendelsohn. 1998. *First-order modal logic*. Kluwer. DOI: doi:10.1007/978-94-011-5292-1.
- Gabbay, Dov M., John F. Horty, Xavier Parent, Ron van der Meyden & Leendert van der Torre. 2013. *Handbook of deontic logic and normative systems*. College Publications. <https://api.semanticscholar.org/CorpusID:60312009>.
- Goldblatt, Robert. 2003. Mathematical modal logic: A view of its evolution. *Journal of Applied Logic* 1(5). 309–392. DOI: <https://doi.org/10.1016/S1570->

- 8683(03)00008 - 9. <https://www.sciencedirect.com/science/article/pii/S1570868303000089>.
- Henkin, Leon. 1950. Completeness in the theory of types. *Journal of Symbolic Logic* 15. 81–91.
- Kripke, Saul A. 1959. A completeness theorem in modal logic. *Journal of Symbolic Logic* 24(1). 1–14. DOI: 10.2307/2964568.
- Kripke, Saul A. 1980. *Naming and necessity*. Harvard University Press. <https://books.google.com/books?id=9vvAlOBfq0kC>.
- LaPorte, Joseph. 2022. Rigid designators. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford encyclopedia of philosophy*, Winter 2022. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2022/entries/rigid-designators/>.
- Lewis, David K. 1986. *On the plurality of worlds*. Wiley-Blackwell.
- Lyons, John. 1977. Deixis, space and time. In *Semantics*, 636–724. Cambridge University Press.
- McNamara, Paul & Frederik Van De Putte. 2022. Deontic logic. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford encyclopedia of philosophy*, Fall 2022. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2022/entries/logic-deontic/>.
- Niles, Ian & Adam Pease. 2001. Toward a standard upper ontology. In Chris Welty & Barry Smith (eds.), *Proceedings of the 2nd international conference on formal ontology in information systems (FOIS-2001)*, 2–9.
- Pacuit, Eric. 2017. *Neighborhood semantics for modal logic*. Springer.
- Paulson, Lawrence C. 1986. Natural deduction as higher-order resolution. *The Journal of Logic Programming* 3(3). 237–258. DOI: [https://doi.org/10.1016/0743-1066\(86\)90015-4](https://doi.org/10.1016/0743-1066(86)90015-4). <https://www.sciencedirect.com/science/article/pii/S0743106686900154>.
- Pease, Adam. 2009. *SUO-KIF reference manual*. Retrieved 20 June 2020. <https://github.com/ontologyportal/sigmakee/blob%20/master/suo-kif.pdf>.
- Pease, Adam. 2011. *Ontology: A practical guide*. Articulate Software Press.
- Pease, Adam. 2023. *Converting the suggested upper merged ontology to typed first-order form*. <https://arxiv.org/abs/2303.04148v1>.
- Pease, Adam & Stephan Schulz. 2022. Contradiction detection and repair in a large theory. In *The international FLAIRS conference proceedings*.
- Prior, Arthur Norman. 1957. *Time and modality: Being the John Locke lectures for 1955–6 delivered in the University of Oxford*. Oxford University Press.
- Raths, Thomas & Jens Otten. 2012. The QMLTP problem library for first-order modal logics. In Bernhard Gramlich, Dale Miller & Uli Sattler (eds.), *International joint conference on automated reasoning (IJCAR)*, 454–461. Springer.

- Schulz, Stephan, Geoff Sutcliffe, Josef Urban & Adam Pease. 2017. Detecting inconsistencies in large first-order knowledge bases. In *Proceedings of CADE 26*, 310–325. Springer.
- Scott, Dana. 1970. Advice on modal logic. In Karel Lambert (ed.), *Philosophical problems in logic: Some recent developments*, 143–173. Springer.
- Sinnott-Armstrong, Walter. 1985. A solution to Forrester’s paradox of gentle murder. *Journal of Philosophy* 82(3). 162–168.
- Sutcliffe, Geoff. 2010. The TPTP world – infrastructure for automated reasoning. In Edmund M. Clarke & Andrei Voronkov (eds.), *LPAR*, vol. 6355 (Lecture Notes in Computer Science), 1–12. Springer. DOI: http://dx.doi.org/10.1007/978-3-642-17511-4_1. <http://dx.doi.org/10.1007/978-3-642-17511-4>.
- Sutcliffe, Geoff, Stephan Schulz, Koen Claessen & Peter Baumgartner. 2012. The TPTP typed first-order form with arithmetic. In Nikolaj Bjørner & Andrei Voronkov (eds.), *LPAR*, vol. 7180 (LNCS), 406–419. Springer. http://dx.doi.org/10.1007/978-3-642-28717-6_32.
- Trac, Steven, Geoff Sutcliffe & Adam Pease. 2008. Integration of the TPTPWorld into SigmaKEE. In *Proceedings of IJCAR ’08 workshop on practical aspects of automated reasoning (PAAR-2008)*. CEUR Workshop Proceedings.

Chapter 9

Variation in Natural Reasoning

Robin Cooper^a

^aUniversity of Gothenburg

Natural reasoning is the kind of reasoning and argumentation that we engage in in everyday situations involving natural language. Following the logical tradition, it is often assumed, in both theoretical and computational linguistics, that this kind of inference is determined by a fixed set of rules or principles and that there is a fixed decidable answer to the question of whether a conclusion follows from certain premises. Here I argue that it could be advantageous to regard natural inference as subject to variation in the sociolinguistic sense. People reason according to certain patterns depending on the situation they find themselves in, their state of mind or the persona that they wish to present. Similarly, people judge the personae of others based in part on the kind of inferences they draw. I consider the implications of this for how we collect data about reasoning and I present two small steps in conceptual work which might help us deal with variation in reasoning: *topoi* in the sense developed by Ellen Breitholtz and types with witness conditions.

1 Natural reasoning

By “natural reasoning and argumentation” I understand the kind of reasoning and argumentation that we engage in in everyday situations involving natural language. In such situations we sometimes employ ancillary languages or representations such as mathematics, logic and diagrams, usually in addition to natural language. Intuitively, everyday natural reasoning without such an explicit ancillary representation does not involve any other language than the natural language, even though many linguistic theories propose an abstract formal representation (such as a “logical form”). Naively, natural reasoning in everyday conversation without ancillary languages involves reasoning with the help of

the structure of the natural language itself and no other representation, an intuition which we find expressed in the literature on natural logic. Perhaps there is also natural reasoning without language: a transition from one brain state to another possibly even without any kind of external stimulus, linguistic or not. However, we will not be discussing this kind of reasoning in this paper.

How does this intuitive notion of natural reasoning compare with the kind of inference that has been discussed in computational linguistics? Back in the early heady days of computational semantics, we “knew” what a natural language inference was and produced the FraCaS test suite (Cooper et al. 1996). This was a product of the FraCaS (A Framework for Computational Semantics) project.¹ It was written by linguists, based on results of theoretical semantics and belongs to a more innocent time when inference in natural language seemed more straightforward than it does today. Later in the history of NLP came the task of Recognizing Textual Entailment (RTE) which to some of us came as a bit of a shock since it involved dealing with conclusions that people actually draw rather than what actually follows. The problem is that people seem to draw inferences which do not actually follow in any logical sense (and also fail to draw inferences that do follow). Dick Crouch, Lauri Karttunen and Annie Zaenen (Zaenen et al. 2005) pointed out the problems of trying to extract a coherent general notion of inference on the basis of RTE data. Among other things, they rightly pointed out that many of the inferences people draw depend on “real world knowledge” in addition to what is presented in the text and it is very difficult to circumscribe the task and draw a line between what is presented in the text and general knowledge. Bernardy & Chatzikyriakidis (2020) explore this by re-annotating the RTE, asking annotators to fill in missing premises when an inference depends on information not explicitly stated in the text. The issue is also addressed by Amanaki et al. (2022). A problem that I would like to focus on in this paper is that different people have different knowledge and assumptions and that this can affect which inferences they draw.

These days in computational linguistics we tend to talk more generally in terms of Natural Language Inference (NLI) rather than textual entailment and a number of NLI datasets are available. The problem of different people with different knowledge and assumptions is perhaps reflected in the brittleness of systems trained on one NLI dataset and tested on another (Talman & Chatzikyriakidis 2019). What kind of world knowledge can be involved in making an inference? It turns out that it can be very specific and local to the utterance situation or the situation described. (1) is an example from the SNLI dataset and is discussed by Talman & Chatzikyriakidis (2019).

¹on which both Dick Crouch and I worked

- (1) a. A man with a tattoo on his arm staring to the side with vehicles and buildings behind him.
- b. A man with no tattoos is getting a massage.
- c. Label: *contradiction*

According to the dataset the pair (1a) and (1b) is labelled as a contradiction. However, those of us with a logic orientation and even only slight knowledge of predicate calculus would recognize this pair to be consistent with each other. We know that, whereas (2a) and (2b) are contradictory, (2c) and (2d) are not contradictory.

- (2) a. $P(a)$
- b. $\neg P(a)$
- c. $\exists x P(x)$
- d. $\exists x \neg P(x)$

So, how could it be that the pair (1a) and (1b) got the label “contradiction” on the basis of human annotation? As competent speakers of (and reasoners using) English we all know that if you describe a picture as (a picture of) *a man with a tattoo*, then there is one salient person in the picture. If there had been more than one they would have been mentioned. Once we know that the original task generating these examples involved creating captions for pictures we can make the inference that just one man is being described and that both expressions describe the same man. One and the same man cannot both have a tattoo and have no tattoos. Thus, there is a contradiction with respect to the picture described. Of course, if the picture had been captioned in the first place as *a man with a tattoo and a man with no tattoos* that would have been perfectly consistent — but it would be a picture of two men. Note that it is possible to reason like this without ever having seen the original picture that was captioned. What is crucial is the specific local knowledge that the examples were generated by a captioning task and some general common sense knowledge of what humans take to be sufficient descriptions of pictures in captioning tasks. If the label annotators had been told that the original task was simply to note something about the picture, not necessarily a description of the complete picture, then I guess that the annotators would not have been so ready to mark these two descriptions as a contradiction. In general, it seems that a system trained on one NLI dataset cannot be guaranteed to be reliable on another dataset because the nature of the

annotation task has major implications for what inferences are drawn. Perhaps much of the variation in inference across datasets can be attributed to this.

The lesson from this is *not* that the labellers were wrong or incompetent, but rather that the labelling takes crucial background information into account and that it is then unsafe to take the results from a data collection with certain background assumptions and map these onto a relation between the *forms* of natural language expressions ignoring the background information. That way, a model trained on one dataset may have unexpectedly low success when tested on another dataset. Even though it is commonly accepted that NLI and what is traditionally thought of as logical inference are not the same, we still tend to make the assumption that there is a single answer to the question “What inferential relation do these expressions of natural language stand in?” as there would be in a classical proof theoretic approach to logical inference. In reality, however, the inferential relations relating to “conclusions that people would draw” seem to involve not just relations between expressions but also contextual parameters and background assumptions. In principle, one might think, this could be reduced to a logical approach to inference. All we need to do is spell out the hidden premises corresponding to contextual parameters and background assumptions. In practice, however, discovering the hidden premises seems to be intractable. It is not, for example, the case that human reasoners can spell out exactly the set of relevant hidden premises that license a given inference on a given occasion. For this reason, I want to suggest that we take inspiration from variationist sociolinguistics.

2 Variation

Variation in sociolinguistics is often related to pronunciation and morphology. For example, Labov (2012) studies Obama’s pronunciation of the verbal *-ing* morpheme (e.g. *grilling* vs. *grillin*) in different kinds of contexts. He found, for example, that Obama’s use of the *-in*’ was much more frequent in an informal situation at a barbecue than in his acceptance speech for the Democratic nomination where he used *-ing* almost exclusively. Burnett (2019) gives a game-theoretic way of modelling inferences that might be drawn from these different uses. For example, *-in*’ can be associated with informality and friendliness but might also give the impression of being less educated or authoritative than *-ing*. Such inferences underlie the *social meaning* associated with utterances. There is no suggestion in this discussion that Obama is deliberately manipulating this aspect of his speech in order to obtain the effect or even that he is conscious of the difference in his

speech on the different occasions. Similarly, hearers of Obama's speech at these events may well draw the conclusion that he is friendly or authoritative without being aware of which aspects of his speech are leading them to this conclusion. They do not, in general, have the kind of preoccupations that linguists might have. In natural inference it seems often to be the case that the extra premises that are leading us to make an inference or not in different circumstances are not always consciously available to us and even if you ask somebody why they drew a certain conclusion, they may try to reconstruct reasons that do not correspond to the facts. Let us consider some (constructed) examples.

In (3) we see an example of natural inference where it is easy to see that different people could draw different conclusions from the same utterance or from similar observations of a situation.

(3) **A says:** (or **B and C observe:**) The lights are on and there is music coming from the house.

B concludes: Somebody is in the house.

C concludes: The people in the house left the lights and the music on when they left.

The variation here may have to do with the fact that *B* and *C* have different knowledge about the situation. For example, *C* might have seen everybody in the house leave. However it could also be that *C* is the kind of person who leaves lights on when they go out so that other people will draw the conclusion that there is somebody in the house. (4) is an example where different opinions or prejudices can lead people to draw different conclusions on the basis of the same circumstances.

(4) **Policeman A says:** (or **Policemen B and C observe:**) The guy driving the Mercedes is Black.

Policeman B concludes: The guy driving the Mercedes is likely to have stolen it.

Policeman C concludes: There is not enough evidence to stop and search the driver of the Mercedes

The variation here has to do not so much with *B* and *C* having different knowledge about the situation at hand but rather having different general beliefs about what is likely to be true. In the case where the premise is spoken, the fact that a particular person is saying it can also affect the inference which is drawn. (5) is an example where the variation might depend on current temporary aspects of the reasoner's mental state.

(5) **Child A says:** I just heard a strange noise at the back door.

Parent B (watching a horror movie) concludes: Someone is trying to get into the house.

Parent C (completing a meditation session) concludes: The sound was made by the wind at the back door.

In this case one can imagine that it is the emotional state of the reasoner which is involved in determining the variation. This example shows that the same person can draw different inferences from similar facts at different times without in general changing their opinion, just as Obama can sometimes say *grilling* and sometimes *grillin'* without this changing anything about the grammatical resources he has available for English.

One might think that examples (3)–(5) are beyond the scope of NLI since they depend on the knowledge, opinions or moods of the reasoners. I would, however, want to include these as examples of natural inference and there are examples which are included in the standard canon of natural language inference which also exhibit variation. Cases et al. (2019) discuss the likelihood of drawing inferences associated with implicative verbs as in (6).

- (6) a. I was able to speak $\overset{?}{\Rightarrow}$ I spoke
b. I was able to speak but I didn't know what to say $\overset{?}{\Rightarrow}$ I spoke
c. I was able to speak to her about my paper $\overset{?}{\Rightarrow}$ I spoke to her about my paper
d. I had a chance to speak to her about my paper $\overset{?}{\Rightarrow}$ I spoke to her about my paper

In contrast, if we negate the antecedent of the examples in (6) we get a clear inference without variation to the negation of the consequent as shown in (7).

- (7) a. I wasn't able to speak \Rightarrow I didn't speak
b. I didn't have a chance to speak to her about my paper \Rightarrow I didn't speak to her about my paper

Cases et al. (2019) suggest assigning signatures as in (8) to implicative verbs indicating the likelihood of concluding the truth of the complement or its negation from the positive or negative sentences.

- (8) a. *be able*: .9 | –
 b. *have a chance*: .4 | –

(8) indicates that there is a high likelihood of drawing the positive conclusion from a positive sentence containing *be able* but only a moderate possibility of drawing this conclusion in the case of *have a chance*. The negation of both verbs leads to a 100% probability that the negation of the complement will be concluded, indicated by the ‘–’ in the signatures.

We have to be careful to be clear about what kind of probability is involved here. Three possibilities are given in (9).

- (9) a. Probability that the inference will be intended by the speaker or understood by the hearer, e.g. from *I was able to speak*, there is a .9 probability that the (categorical) inference that I spoke will be drawn
 b. Inference that the conclusion is true with the probability, e.g. from *I was able to speak*, we can draw the conclusion that there is a .9 probability that I spoke
 c. 9 out of 10 speakers would judge that there is an entailment relation here

Cases et al. (2019) are clear that it is (9a) that is intended. This corresponds to what I have been calling variation in this paper.

3 Developing a strategy

What is a good strategy for dealing with this kind of data? One option is to ignore variation and restrict ourselves to the kind of “logical” inference that was treated in the FraCaS test suite. This would mean giving up on RTE, NLI or any kind of natural inference. Possibly this was the view represented by Zaenen, Karttunen and Crouch back in 2005 (Zaenen et al. 2005). I am not sure that this is a realistic option in 2024, though. It seems like we have already discovered too much of interest that needs to be explored.

A second option is to try to figure out a way of doing probabilistic variational reasoning on the basis of the kind of NLI datasets that we have now. This would involve answering the question of what probability there is that somebody will draw a particular conclusion given a certain premise. It is not obvious, though, that we would get meaningful probabilities if we do not take account of the kind

of people who are making the inference and in which contexts they are doing so. The current standard of aggregating data rather than keeping track of variation revealed, for example, in annotator disagreement, does not bode well for obtaining the kind of data we need. This complaint is closely related to the discussion to be found recently in the perspectival approach to machine learning in NLP (Cabitza et al. 2023). A related approach is to allow multiple labels in datasets, with or without associated probabilities (Chen et al. 2020, Kogkalidis et al. 2024).

A third option is to bite the bullet and try to treat natural inference in terms of certain people in certain contexts and in certain mental states drawing different conclusions from the same premises. It is not obvious how to do this based on standard methods for learning from datasets involving aggregation which smooths out all the variation in the data which has been collected. Perhaps the work on perspectival NLP will enable us to create new NLI datasets that will reflect this variation. The numbers that Cases et al. (2019) use to represent variation in inferences involving implicative verbs were informally computed on the basis of Google searches. As far as I know there is still an open question on what exactly would be a reliable general method of extracting data concerning this kind of variation in inference.

One might think, given the names, that natural logic (van Benthem 1986, Moss 2010, 2015) should be helpful in gaining an understanding of natural reasoning. It has also become important in work on inference in computational linguistics in a thread of research started by MacCartney & Manning (2007) and MacCartney (2009). The central idea in natural logic is that reasoning is defined on the syntactic structure of natural language expressions rather than, say, a representation of the meaning of these expressions in some formal language such as “logical form”. This seems to me to embody an important insight about natural reasoning: that we exploit the structure of natural language in natural reasoning rather than translate it first into some other representation and then reason with that. This is not to say that we do not in addition use other representations such as diagrams and mathematical formulae to aid our reasoning, or even that there might be reasoning at the neural level which does not rely on any explicit external representation. However when we reason naturally using natural language, then the first natural hypothesis is that it is the structure of the natural language itself that supports this reasoning. This basic intuition was central in Montague’s work (Montague 1974). His slogan “English as a formal language” meant that a natural language could be given a semantics in the same way as a formal language and that you did not first have to translate it into a formal language in order to define its semantics. Even when Montague used his intensional logic as a convenient representation for the model theoretic semantics, he was careful

to embed this in a general theory (expressed in his paper ‘Universal Grammar’) which ensured that the logical representation was an eliminable notational convenience rather than a necessary requirement for being able to relate a semantics to the syntax of the natural language. Thus, in this way Montague was an early proponent of natural logic, even though Montague was concerned with the model theory of natural language rather than the more proof theoretic approach which is normally associated with the term “natural logic”.

Ambiguity presents something of a challenge for the natural logic approach and natural logicians sometimes abandon complete naturalness saying that reasoning is defined on disambiguated languages “close” to the structure of natural language (see, for example, Muskens 2010). Among the reasons for this can be quantifier scope ambiguity and lexical ambiguity. Montague elegantly treated this problem with respect to quantifier scope by using his unambiguous derivation trees as the “natural structure” rather than the strings corresponding to the actual natural language expression. If there is variation in natural language inference as we are suggesting, then this is presumably going to require similar techniques in natural logic as ambiguity. Indeed, one might see variation in inference as a species of ambiguity.

Another issue raised by variation in inference is that of decidability. One interest in natural logic is to provide decidable fragments of natural language (Moss 2015). An argument for decidability in natural language is perhaps the reliability and certainty with which people can draw inferences. However this argument is undermined if we introduce variation into natural reasoning. There are also other reasons to be suspicious of decidability for natural language at least in the area of lexical semantics. Consider the three words: *child*, *youth*, *adult*. Is a 12-year-old a youth? Is an 18-year-old a child or an adult? A great deal of law involves making precise concepts which are vague in everyday language. This involves making precise inferences that seem to be left unclear (undecidable?) in everyday language. We can even talk about the different inferences that can be drawn in different contexts, as in (10).

(10) Legally, an 18-year-old is an adult, but in many ways they are still children.

This suggests that, while there may be decidable fragments of natural language, these would be stripped of the variation which we are concerned with here. While we may want to say that natural languages give us the tools to build decidable languages for particular purposes (following the kind of suggestion in Cooper & Ranta 2008), experience with the law suggests that we often do not

find it easy to eliminate all vagueness and ambiguity in real life cases of even moderate complexity.

To my knowledge, variability is not discussed in natural logic approaches and this is perhaps not surprising given the above discussion. However, it seems that a natural logic approach could be helpful if we could find the right way to combine a notion of variation with a notion of decidability.

If we decide that our strategy is the third option of trying to analyze natural inference in terms of variation, we are left with the problem we mentioned above of how to collect data. It might involve more than just keeping track of annotator variation and building a model of it as in the perspectival approach to NLP. Perhaps the original data gathering involved needs to be oriented towards the suspected use of variation. For example, Obama's use of *ing/-in'* in different types of situations as discussed above would need a specially targeted dataset. It seems likely that if we want to explore variation in inference we would need to do something similar and not treat this variation as noise in the data. NLI datasets are in any case standardly specially designed based on some specific task, such as captioning, in order to reveal inference, since inference is not readily discoverable in general text or dialogue datasets. The problem is that inferences available in one task may not transfer to another task. It is not clear what the ideal dataset for looking at variation in inference would be: perhaps a longitudinal study of individuals and inferences they make relative to their prior experience and the current situation in which they find themselves. This seems pretty implausible.

Perhaps sociolinguistics can give us a hint on how to deal with this in a tractable way. While the reality is that each individual may behave in a unique way depending on their previous experience and the situation in which they find themselves, perhaps there is a middle ground between trying to account for this and the standard practice of aggregating all the data which suppresses the interesting variation. This would involve looking at data from people with similar backgrounds and in similar situations. In third wave sociolinguistics (Eckert 2012) the notion of *persona* plays an important role. For example, using *-ing* can be associated with a formal persona, educated and in command of the situation. Using *-in'* can be associated with a more informal, relaxed and friendly persona. Different people can be associated with different personae and people may project different personae in different situations. Associating data with different personae could get us away from the extreme aggregation that is often involved in creating datasets but not commit us to separating out each individual/situation pair that would be intractable. We instead try to answer the question: what phenomena are associated with which personae? In sociolinguistics personae are involved in *social meaning*. Speakers attempt to present personae and one task of a hearer

is to extract social meaning by recognizing the persona that the speaker is trying to present, or in more practical terms, perhaps, by estimating a distribution over the personae that the speaker might be presenting. Burnett (2019) offers a treatment of this kind of process in game theoretical terms.

4 Towards an account of natural reasoning

From this discussion, it seems clear that an account of natural reasoning should contain the following three ingredients:

- an account of “local” reasoning
- incorporation of probability
- a way of blending machine learning and logical techniques

Locality is needed because we need to account for reasoning not in terms of a monolithic logical system but in terms of people drawing different conclusions in different contexts. This is central to an account of variation in reasoning. Probability seems necessary because we will not be able to give a categorical account of exactly what inferences will be drawn in what context but we might be able to say something about the probability that a given inference will be drawn given a certain type of context. In a computational setting, if we are to learn from datasets in anything like the way we have been discussing and yet give an account of which inferences are being drawn, it seems like it would be useful to have a good way of combining machine learning with logical techniques.

I do not pretend to have a complete working account of natural inference, but I would like to present a couple of things that I have been involved in that may make a contribution. The first is a notion of *topos* to account for local reasoning. The second is a notion of types with witness conditions for blending machine learning and logic in a probabilistic approach. In the remainder of this section I will discuss these two notions in turn.

4.1 Enthymemes and topoi

Breitholtz (2020) develops notions of *enthymeme* and *topos* derived originally from Aristotle, as discussed for example in the *Rhetoric*. For Aristotle, an enthymeme is an argument that can be used in rhetoric to convince someone of a conclusion. It need not be logical or necessary. It can be defeasible. These

features put it in contrast with Aristotle's notion of syllogism. Different enthymemes need not be consistent with each other. You can argue one way or another and come to different, even contradictory, conclusions. Examples of enthymemes which Aristotle discusses (from Kennedy's translation of Aristotle's *Rhetoric*, Aristotle 2007) are given in (11).

- (11) a. [he] is sick, for he has a fever (*Rhetoric*, I.2.18)
b. it is a sign of fever that somebody breathes rapidly (*Rhetoric*, I.2.18)

A slightly more logical rendering of these enthymemes is given in (12).

- (12) a. he has a fever \Rightarrow he is sick
b. he breathes rapidly \Rightarrow he has a fever

(12a) is, according to Aristotle, an instance of an "irrefutable sign" and (12b) is an instance of a "refutable" sign. That is, (12b) is a defeasible inference whereas (12a) is not. Aristotle's discussion of (12b) suggests that the interpretation of enthymemes as syllogisms with missing premises that occurs in the modern literature is not quite correct. It is not clear what premise you could add to (12b) to make it into a syllogism. Aristotle seems to have been more interested here in what we would now call "defeasible reasoning".

A *topos* is a "place" to keep an enthymeme. Breitholtz (2020) interprets this as a more general principle of reasoning which can be used to underpin enthymemes. She treats *topoi* as the same kind of mathematical object as an enthymeme. *Topoi* are more general and can be specified in various ways to become an enthymeme used as an argument. Thus, *topoi* have the same properties as enthymemes as discussed above. Whereas the enthymemes in (12) are meant to represent particular arguments about a particular person in context referred to by *he*, corresponding *topoi* would be more general rules that could be applied to any particular person as represented in (13).

- (13) a. for any x , x has a fever $\Rightarrow x$ is sick
b. for any x , x breathes rapidly $\Rightarrow x$ has a fever

Many of Aristotle's *topoi* are very general and it is not always obvious how to construe them by something as simple as universal quantification over variables. An example is (14).

- (14) *the more and the less*: if something is not the fact where it would be more [expected], it is clear that it is not a fact where it would be less (*Rhetoric*, 2.23.4)

An enthymeme underpinned by this topos would be (15).

- (15) If you can't run a half marathon, you can't run a marathon

Breitholtz (2020) relates topoi to coherence. She cites (16), an example from Walker (1996).

- (16) Let's take Walnut St. It's shorter.

In order to interpret (16) as a coherent discourse with the second sentence being a reason to take Walnut St., we need a topos along the lines of "if some route is shorter than another, it is preferable to take the shorter one". A topos like this is readily available. There are, however, other topoi relating to which routes are preferable, for example, faster routes, non-motorway routes, scenic routes and so on. Breitholtz (2020) points out that (17) requires us to accommodate a normally less available topos.

- (17) Let's take Walnut St. It's longer.

A topos involving longer routes being preferable perhaps requires a background where you need exercise or do not want to arrive too early. Such a topos is, of course, not consistent with one that says that shorter routes are preferable. Topoi do not necessarily belong to a consistent logical system.

Notionally, Breitholtz's characterization of topoi is related to topoi as discussed by Ducrot (1988) and elsewhere in French linguistics.² Each agent has their own (not necessarily consistent) collection of topoi to draw on (or perhaps several different collections associated with various types of context). Part of the characterization of a persona is a distribution over topoi the agent is likely to use. Intuitively, this says that part of the characterization of a persona is the way you might expect a person to argue. Different kinds of people argue in different ways and will argue differently (perhaps trying to present a different persona) in different contexts depending in part on their assumptions about the people they are talking to. Noble et al. (2020) relate these ideas to game theoretic aspects of dialogue. One can, for example, estimate distributions over personae given the

²Breitholtz (2020) also discusses relationships to Relevance Theory (Sperber & Wilson 1986).

topoi an agent uses or estimate distributions over topoi to use given possible personae an agent might have.

Breitholtz (2020) models topoi in terms of types within the framework of TTR (a type theory with records).³ A topos is a function from an object (situation) of a given type to a *type*.⁴ Thus, a topos is a kind of dependent type. The general form of a topos is (18), where $T((r))$ indicates that T is partly determined by r (textually, this means the r occurs within T).⁵ The ‘.’ here is the standard dot used in some notations of the λ -calculus.

$$(18) \lambda r:T_1 . T_2((r))$$

That is, a function whose domain is objects, r , of type T_1 , and which, when applied to r returns a type which depends on r . An example is given in (19).

$$(19) \lambda r: \left[\begin{array}{l} x:Ind \\ c_1:breathe_rapidly(x) \end{array} \right] . [c_2:has_fever(r.x)]$$

Intuitively, (19) represents a function which given a situation, r , in which there is an individual, x , who is breathing rapidly, returns the *type* of situations in which the individual x from r has a fever. (19) uses record types. A witness for the record type (20a) would be a record (20b) where $a : Ind$ and $s : breathe_rapidly(a)$. Similar records containing additional fields with different labels would also be of the same type. See Cooper (2023) for a detailed precise account of the kind of record types we are using.

$$(20) \text{ a. } \left[\begin{array}{ll} x & : \quad Ind \\ c_1 & : \quad breathe_rapidly(x) \end{array} \right]$$

$$\text{ b. } \left[\begin{array}{ll} x & = \quad a \\ c_1 & = \quad s \end{array} \right]$$

If r is such a record, we can refer to the values associated with its labels by projection using a dot notation, for example, $r.x = a$. This notation is used in connection with the label ‘ c_2 ’ in (19).

What you do with the type returned, T_2 , depends on action rules associated with the topos, appropriate to the context, such as those actions given in (21).

³See Cooper (2023) for the most complete account of TTR. Cooper & Ginzburg (2015) provides a more brief overview.

⁴The use of a function here relates to the logical view of inference rules as functions, cf. the Curry-Howard correspondence.

⁵This is an adaptation of the standard logical notation as in $\varphi(x)$ to mean an expression φ containing x . We use double parentheses to distinguish this from other uses of single parentheses such as function application.

- (21) a. conclude that s is of type T_2 ($s : T_2$), where s is what you applied the topos to
- b. conclude that there is some s , such that $s : T_2$ (that is T_2 is instantiated (witnessed, inhabited) or true)
- c. create an object (a situation) of type T_2 , i.e. perform an action

Sometimes the same topos can be used with different action rules. They represent an abstract association between types which can be exploited by different people in different situations. Thus, they seem the right kind of tool to model variation in reasoning.

Dogwhistles are an example of how agents can exploit variation in natural reasoning. Dogwhistle events are those where a dialogue participant exploits that fact that different agents interpreting an utterance will draw different conclusions (cf. Khoo 2017). Breitholtz & Cooper (2021), working within the GRIPES project,⁶ discuss the example in (22) constructed on the basis of the dogwhistle phrase *inner city* much discussed in the dogwhistle literature. For example, Henderson & McCready (2024: p. 1) introduce it as a primary example of a dogwhistle with the following text in the introduction to their book:

On a 2014 radio program, Representative Paul Ryan said the following.

(1) We have got this tailspin of culture, in our inner cities in particular, of men not working and just generations of men not even thinking about working or learning the value and the culture of work.

He was criticized shortly after by fellow Representative Barbara Lee for making a “thinly veiled racial attack.” This is because the phrase inner-city is code or euphemism for Black neighborhoods (especially stereotypically racialized views of such neighborhoods). Many people heard Paul Ryan say:

(2) We have got this tailspin of culture, in our Black neighborhoods in particular, of men not working and just generations of men not even thinking about working or learning the value and the culture of work.

This example illustrates the notion of a dogwhistle—that is, language that sends one message to an out-group while at the same time sending a second (often taboo, controversial, or inflammatory) message to an in-group.

⁶www.gu.se/en/research/gripes-gothenburg-research-initiative-for-politically-emergent-systems

The example we present here is based on Paul Ryan’s remarks though with a much simpler structure for expository reasons. (22) presents three messages that an agent may wish to send.

- (22) a. $\varphi_{\text{inner_cities}}$ — We need to solve the problem with inner cities
b. $\varphi_{\text{social_support}}$ — We need to solve the problem with social support for poverty in inner cities
c. φ_{lazy} — We need to solve the problem with lazy out of work Black populations in inner cities

Now suppose that an agent A (a politician), in conversation with B and C , wants to communicate $\varphi_{\text{social_support}}$ to B and φ_{lazy} to C . A assumes that B has a persona which disposes them to use the topoi represented informally as (23a), whereas C is disposed to apply (23b).

- (23) a. $\varphi_{\text{inner_cities}} \rightarrow \varphi_{\text{social_support}}$
b. $\varphi_{\text{inner_cities}} \rightarrow \varphi_{\text{lazy}}$

By sending the message $\varphi_{\text{inner_cities}}$, A can simultaneously cause B and C to draw different conclusions without revealing to B and C that this has happened. That is, we can exploit variation in natural reasoning in order to send coded messages to target groups within an audience.

Topoi provide us with a way of talking about variation in reasoning. Different agents may have different topoi available and the same agent can have different topoi in focus at different times.

4.2 Types with witness conditions

Another way of introducing variation is to allow types to be related to classifiers which may be retrained with new data over the course of time. This is something which can also be combined with a probabilistic version of TTR (Cooper et al. 2015). In probabilistic TTR we replace the standard $a : T$ for “ a is of type T ” with $p(a : T)$, “the probability that a is of type T ” and $p(T)$, “the probability that there is something of type T ”. (Note that $p(a : T) \leq p(T)$.) Complex types such as $T_1 \wedge T_2$, $T_1 \vee T_2$ are treated according to adaptations of the standard Kolmogorov rules for connectives in probability theory. Probabilistic TTR does not of itself specify what kind of probability it is used for. Thus, for example, it could be used for any of the three probabilities in (9).

In the theoretical statement of TTR basic types are related to sets of witnesses creating a structure which is similar to a sorted first order model. In the Python implementation `pyttr`⁷ (Cooper 2019) types are associated with witness conditions which allow for more flexibility and dynamic development. For basic types the witness conditions can be arbitrary procedures which may be external to TTR. These procedures can be learnt dynamically. Several witness conditions can be associated with a single type: a witness is recognized as being of the type if one of the witness conditions is successful. In the probabilistic version of `pyttr`, the probability of an object being a witness for the type is the maximum of those probabilities returned by a collection of witness conditions. Larsson (2020) presents the beginnings of a formal theoretical treatment of witness conditions. The fact that we allow arbitrary external procedures to be used in the characterization of witness conditions means that we can, for example, use classifiers trained by machine learning techniques to characterize which objects are of a type. This is exploited, for example, by Matsson (2018). Thus, while TTR itself does not give us a way to create classifiers, the notion of witness condition gives us a way to use classifiers to characterize the witnesses of a type.

To make the difference between the theoretical presentation and the implementation more concrete we present the original theoretical version of the characterization of a system of basic types in (24).

(24) A *system of basic types* is a pair:

$$\mathbf{TYPE}_B = \langle \mathbf{Type}, A \rangle$$

where:

1. **Type** is a non-empty set
2. A is a function whose domain is **Type**
3. for any $T \in \mathbf{Type}$, $A(T)$ is a set disjoint from **Type**
4. for any $T \in \mathbf{Type}$, $a : \mathbf{TYPE}_B \ T$ iff $a \in A(T)$

(25) shows how witness conditions work in the non-probabilistic version of `pyttr` by giving an example of a type of real numbers implemented as floating point numbers.

⁷<https://github.com/GU-CLASP/pyttr>

```
Real = BType('Real')
Real.learn_witness_condition(lambda n: isinstance(n, float))
print(Real.query(0.5))
print(Real.witness_cache)
```

(25) True
[0.5]

(26) shows how to do the same example in the probabilistic version of pyttr.

```
def RealClassifier(n):
    if isinstance(n, float):
        return PConstraint(1)
    else:
        return PConstraint(0)
Real = Type('Real')
Real.learn_witness_condition(RealClassifier)
show(Real.query(.6))
```

(26) '1.0'

A slogan we could associate with this approach is “Types as classifiers” modelled on the slogan “Words as classifiers” introduced by Schlangen et al. (2015). It would be more accurate, though to say “witnesses for types can be characterized by classifiers, among other things”. That is, a witness condition can call a classifier since it is any arbitrary procedure which returns an appropriate value. Also, a type can have more than one witness condition as mentioned above. The set of witness conditions may be a mixture of classifiers and other procedures. This is one way in which this system allows us to compare the results of classifiers with logically oriented procedures. In addition, the method `learn_witness_condition()` can be used dynamically on a type that has already been defined, thus allowing for variation in what counts as a witness over time.

5 Conclusion

In this paper we have looked at natural reasoning and compared it with natural logic and NLI. We have proposed that there is a kind of variation in natural reasoning which is similar to the variation discussed in sociolinguistics. We have considered the implications of this for how we collect data about reasoning and we have presented two small steps in conceptual work which might help us deal with this phenomenon: topoi and types with witness conditions. However, there is still a lot of work to be done connecting this to learning algorithms that could handle variation in natural reasoning.

Acknowledgments

Dick Crouch and I first crossed paths when we worked on the FraCaS (A Framework for Computational Semantics) project in the mid-nineties. Among other things, we worked on the FraCaS test suite, a collection of constructed inference examples which were meant to provide a benchmark for natural language systems. Over the years we have both maintained an interest in natural language inference, though in different contexts and in different ways. I offer some recent thoughts of mine on the matter in honour of Dick and to thank him for his contribution to the field over many years.

A version of this paper was presented at the workshop Natural Logic Meets Machine Learning (NaLoMa) III, at ESSLLI 2022. I am grateful for comments from the participants of this workshop and also from Ellen Breitholtz, Stergios Chatzikyriakidis, Peter Sutton and two anonymous referees. This work was supported in part by a grant from the Swedish Research Council (VR project 2014-39) for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg and in part by a grant from the Marianne och Marcus Wallenbergstiftelse for the project Gothenburg Research Initiative for Politically Emergent Systems (GRIPES).

References

- Amanaki, Eirini, Jean-Philippe Bernardy, Stergios Chatzikyriakidis, Robin Cooper, Simon Dobnik, Aram Karimi, Adam Ek, Eirini Chrysovalantou Giannikouri, Vasiliki Katsouli, Ilias Kolokousis, Eirini Chrysovalantou Mamatzaki, Dimitrios Papadakis, Olga Petrova, Erofilii Psaltaki, Charikleia Soupiona, Efrosyni Skoulataki & Christina Stefanidou. 2022. Fine-grained entailment: Resources for Greek NLI and precise entailment. In Jonne Sälevä & Constantine Lignos (eds.), *Proceedings of the workshop on dataset creation for lower-resourced languages within the 13th language resources and evaluation conference (LREC)*, 44–52. European Language Resources Association. <https://aclanthology.org/2022.dclrl-1.6>.
- Aristotle. 2007. *On rhetoric: A theory of civic discourse*. 2nd edn. Translated with introduction, notes, and appendices by George A. Kennedy. Oxford University Press.
- Bernardy, Jean-Philippe & Stergios Chatzikyriakidis. 2020. Improving the precision of natural textual entailment problem datasets. In *Proceedings of the twelfth language resources and evaluation conference (LREC)*, 6835–6840.

- Breitholtz, Ellen. 2020. *Enthymemes and topoi in dialogue: The use of common sense reasoning in conversation*, vol. 41 (Current Research in the Semantics/Pragmatics Interface). Brill.
- Breitholtz, Ellen & Robin Cooper. 2021. Dogwhistles as inferences in interaction. In *Proceedings of the reasoning and interaction conference (ReInAct 2021)*, 40–46. ACL. <https://aclanthology.org/2021.reinact-1.6.pdf>.
- Burnett, Heather. 2019. Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy* 42(5). 419–450.
- Cabitzza, Federico, Andrea Campagner & Valerio Basile. 2023. Toward a perspectivist turn in ground truthing for predictive computing. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, 6860–6868. DOI: 10.1609/aaai.v37i6.25840. <https://ojs.aaai.org/index.php/AAAI/article/view/25840>.
- Cases, Ignacio, Clemens Rosenbaum, Matthew Riemer, Atticus Geiger, Tim Klinger, Alex Tamkin, Olivia Li, Sandhini Agarwal, Joshua D. Greene, Daniel Jurafsky, Christopher Potts & Lauri Karttunen. 2019. Recursive routing networks: Learning to compose modules for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies (ACL-HLT), volume 1 (long and short papers)*, 3631–3648. Association for Computational Linguistics. DOI: 10.18653/v1/N19-1365. <https://aclanthology.org/N19-1365>.
- Chen, Tongfei, Zhengping Jiang, Adam Poliak, Keisuke Sakaguchi & Benjamin Van Durme. 2020. Uncertain natural language inference. In Daniel Jurafsky, Joyce Chai, Natalie Schluter & Joel Tetreault (eds.), *Proceedings of the 58th annual meeting of the Association for Computational Linguistics (ACL)*, 8772–8779. Association for Computational Linguistics. DOI: 10.18653/v1/2020.acl-main.774. <https://aclanthology.org/2020.acl-main.774>.
- Cooper, Robin. 2019. Types as learnable cognitive resources in PyTTR. In Cleo Condoravdi & Tracy Holloway King (eds.), *Tokens of meaning: Papers in honor of Lauri Karttunen*. CSLI Publications.
- Cooper, Robin. 2023. *From perception to communication: A theory of types for action and meaning*. Open access. Oxford University Press. <https://global.oup.com/academic/product/from-perception-to-communication-9780192871312>.
- Cooper, Robin, Dick Crouch, Jan van Eijck, Chris Fox, Josef van Genabith, Jan Jaspars, Hans Kamp, David Milward, Manfred Pinkal, Massimo Poesio & Steve Pulman. 1996. *Using the framework: FraCaS deliverable D16*. Tech. rep. <https://gu-clasp.github.io/multifracas/D16.pdf>.
- Cooper, Robin, Simon Dobnik, Shalom Lappin & Staffan Larsson. 2015. Probabilistic type theory and natural language semantics. *Linguistic Issues in Language*

- Technology* 10(4). 1–45. <https://journals.colorado.edu/index.php/lilt/article/view/1357/1199>.
- Cooper, Robin & Jonathan Ginzburg. 2015. Type theory with records for natural language semantics. In Shalom Lappin & Chris Fox (eds.), *The handbook of contemporary semantic theory*, 2nd edn., 375–407. Wiley-Blackwell.
- Cooper, Robin & Arne Ranta. 2008. Natural languages as collections of resources. In Robin Cooper & Ruth Kempson (eds.), *Language in flux: Dialogue coordination, language variation, change and evolution*, vol. 1 (Communication, Mind and Language), 109–120. College Publications.
- Ducrot, Oswald. 1988. Topoi et formes topiques. *Bulletin d'études de la linguistique française* 22. 1–14.
- Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in the study of variation. *Annual Review of Anthropology* 41. 87–100.
- Henderson, Robert & Elin McCready. 2024. *Signaling without Saying: The Semantics and Pragmatics of Dogwhistles* (Oxford Studies in Semantics and Pragmatics 17). Oxford: Oxford University Press.
- Khoo, Justin. 2017. Code words in political discourse. *Philosophical Topics* 45(2). 33–64.
- Kogkalidis, Konstantinos, Stergios Chatzikyriakidis, Eirini Giannikouri, Vasiliki Katsouli, Christina Klironomou, Christina Koula, Dimitris Papadakis, Thelka Pasparaki, Erofilia Psaltaki, Efthymia Sakellariou & Charikleia Soupiona. 2024. OYXOY: A modern NLP test suite for Modern Greek. In Yvette Graham & Matthew Purver (eds.), *Findings of the Association for Computational Linguistics: EACL 2024*, 311–322. Association for Computational Linguistics. <https://aclanthology.org/2024.findings-eacl.21>.
- Labov, William. 2012. *Dialect diversity in America: The politics of language change*. University of Virginia Press.
- Larsson, Staffan. 2020. Discrete and probabilistic classifier-based semantics. In *Proceedings of the probability and meaning conference (PaM 2020)*, 62–68.
- MacCartney, Bill. 2009. *Natural language inference*. Stanford University. (Doctoral dissertation).
- MacCartney, Bill & Christopher D. Manning. 2007. Natural logic for textual inference. In *Proceedings of the workshop on textual entailment and paraphrasing*, 193–200. Association for Computational Linguistics.
- Matsson, Arild. 2018. *Implementing perceptual semantics in type theory with records (TTR)*. Masters in Language Technology. University of Gothenburg. (MA thesis).

- Montague, Richard. 1974. *Formal philosophy: Selected papers of Richard Montague*. Edited and with an introduction by Richmond H. Thomason. Yale University Press.
- Moss, Lawrence S. 2010. Natural logic and semantics. In Maria Aloni, Harald Bastiaanse, Tiki de Jager & Katrin Schulz (eds.), *Logic, language, and meaning: Selected papers from the seventeenth Amsterdam colloquium*, 84–93. Springer.
- Moss, Lawrence S. 2015. Natural logic. In Shalom Lappin & Chris Fox (eds.), *The handbook of contemporary semantic theory*, 2nd edn. Wiley-Blackwell.
- Muskens, Reinhard. 2010. An analytic tableau system for natural logic. In Maria Aloni, Harald Bastiaanse, Tiki de Jager & Katrin Schulz (eds.), *Logic, language, and meaning: Selected papers from the seventeenth Amsterdam colloquium*, 104–113. Springer.
- Noble, Bill, Ellen Breitholtz & Robin Cooper. 2020. Personae under uncertainty: The case of topoi. In Christine Howes, Stergios Chatzikiyiakidis, Adam Ek & Vidya Somashekarappa (eds.), *Proceedings of the probability and meaning conference (PaM 2020)*, 8–16. Association for Computational Linguistics. <https://aclanthology.org/2020.pam-1.2>.
- Schlangen, David, Sina Zarriess & Casey Kennington. 2015. *Resolving references to objects in photographs using the words-as-classifiers model*. <https://arxiv.org/abs/1510.02125>.
- Sperber, Daniel & Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Blackwell.
- Talman, Aarne & Stergios Chatzikiyiakidis. 2019. Testing the generalization power of neural network models across NLI benchmarks. In Tal Linzen, Grzegorz Chrupala, Yonatan Belinkov & Dieuwke Hupkes (eds.), *Proceedings of the 2019 ACL workshop BlackboxNLP: Analyzing and interpreting neural networks for NLP*, 85–94. Association for Computational Linguistics. DOI: 10.18653/v1/W19-4810. <https://aclanthology.org/W19-4810>.
- van Benthem, Johan. 1986. *Essays in logical semantics*. Springer. 240 pp. https://www.ebook.de/de/product/7614100/johan_van_benthem_johan_van_benthem_essays_in_logical_semantics.html.
- Walker, Marilyn A. 1996. Limited attention and discourse structure. *Computational Linguistics* 22(2). 255–264.
- Zaenen, Annie, Lauri Karttunen & Richard Crouch. 2005. Local textual inference: Can it be defined or circumscribed? In *Proceedings of the ACL workshop on empirical modeling of semantic equivalence and entailment*, 31–36. Association for Computational Linguistics.

Part III

Lexicon

Chapter 10

Out-Applicatives are Concealed Comparatives

Cleo Condoravdi^a & Paul Kiparsky^a

^aStanford University

The prefix *out-* attaches to nouns, adjectives, and verbs to form transitive verbs. We argue that the resulting verbs describe an event with the subject argument as a participant and have a uniformly comparative meaning, namely that the degree of the event along some contextually fixed dimension of comparison exceeds a standard identified by the object argument. The object can identify the standard in one of three ways: (a) directly by describing it, (b) by establishing an event of the same type as the verbal event but with the object as a participant, or (c) by establishing a modal domain of worlds in which there is an event of the same type as the verbal event with the subject argument as a participant. We show that these varieties of comparison have equivalent counterparts in overt *more/-er than* comparatives and in *too/not enough* comparatives.

Our analysis depends on deriving *out*-verbs by prefixing *out-* to a base with its complete argument structure and Theta-grid. We present evidence that supports this morphological derivation over Ahn's (2022) syntactic derivation, which gets the argument structure of *out*-verbs entirely from the prefix *out-* independently of the base's argument structure. Our analysis explains in particular why the base of *out*-cannot be an obligatorily transitive verb. We defend our unified analysis against previous proposals, in particular that *out*-verbs introduce a causative or "competition" macro-event and lie on an "interpretational cline" between comparative and resultative meanings (Kotowski 2021, 2023) and that some of them constitute "a separate (though perhaps related) phenomenon ... with a different (perhaps resultative) meaning" (Ahn 2022). We argue that the causative, resultative, and competition interpretations are inferential.

1 The prefix *out-*

The prefix *out-* turns intransitive verbs, adjectives, and nouns into transitive verbs.¹

- (1) a. A baby caribou can outrun its mother by the age of three days.
- b. Andrew Cuomo was outsmarted by bartender Alexandria Ocasio-Cortez.
- c. CBS certainly outgeneraled and outadmiraled its rivals.

*Out-*prefixation is productive. Consider inventive nonce words like (2):

- (2) a. some Dutch ladies, out-gutturaling even the Swiss themselves (OED)
- b. A Parsee team to outcricket an English eleven! (OED)
- c. We must out-Game-of-Thrones him by making powerful alliances.
- d. It was all this big elaborate joke to her, out-Southerning the folks who made a big deal about such stuff.
- e. People tend to surround themselves with people who don't outskill them when they outteam them. (OED)
- f. I had the glory of out-long-wording both parties. (OED)
- g. On the inside, Okinawa sweet potatoes can outpurple an eggplant. (OED)
- h. It is no longer Navratilova's world when she can not only be outplayed but also outsyllabled. (OED)
- i. He was sure their clients out-healed, out-helping-handed, overall out-charitied their competing charities.

¹Unless otherwise specified, all examples are from the Internet. They are sometimes abbreviated for readability.

These verbs express a comparison on some quantitative or qualitative scale between two events² – one in which the underlying subject argument is a participant and a similar one in which the underlying object argument is a participant (passivized in (2h)). The scale can be inferred from the base, possibly with contextual information (e.g. *outcompete* = *outdo in competition*). When the base is a noun or adjective, as in (2), the event itself must be inferred. Language users converge intuitively on the interpretation of creative uses of *out-*. Without additional context, *out-long-wording* is understood to mean “uttering more long words” – not learning more of them, or avoiding more of them, for example. *Out-syllabled* in the context of (2h) means “containing more syllables”, not “uttering more syllables”. *Out-Game-of-Thronesing* means acting more deviously in the manner of *Game of Thrones*, not watching more episodes of it on TV.

Such verbs are numerous and freely coined, yet morphologically and syntactically highly constrained. *Out-* cannot be prefixed to adverbs, conjunctions, or prepositions, or to obligatorily transitive verbs, or to verbs that require an obligatory predicate complement:

- (3) **outsoon*, **outquickly*, **outbecause*, **outunder*, **outnotwithstanding*, **outresemble*, **outinvolve*, **outhave*, **outconsider*, **outentail*, **outimply*, **outdespise*, **outdetach*, **outrepair*, **outenjoy*, **outreplace*, **outrecommend*, **outdeply*, **outbe*, **outseem*, **outshould*, **outcan*

For example, **John outresembled Fred* and **Biden outwas Trump confused* are ungrammatical, or even uninterpretable.

We will argue that *out-*verbs are morphologically derived by attaching the prefix *out-* to the base to form a transitive verb. The resulting verbs preserve the arguments and Theta-roles of the base, if it has any, and have a comparative meaning. The prefix and base are combined in the following way. If the base has neither a subject nor an object argument, as in the case of nouns and adjectives, two arguments are added to form a transitive verb, e.g. *outmuscle*, *outfox*, *outdistance*, *outsmart*, *outmodern*. If the base is an intransitive verb, its subject argument is retained, and an object is added, e.g. *outjump*, *outsing*, *outwrestle*. The base then identifies the compared events, and the arguments identify their respective subject roles. If the base is a transitive verb, its object argument must be omitted. So *outwrite*, *outdrink*, *outeat*, *outresearch*, *outearn*, *outcompete*, *outproduce*, *outspend*, *outlobby*, *outgoogle*, *outcook*, *outhit*, *outpunch*, *outgive*, *outperform*, *outinfluence*, *outdonate* are derived from the objectless versions of the base

²We use *event* in the broader sense of ‘eventuality’, which includes states and processes. The events could of course be part of a larger event, as in *Fischer outplayed Spassky in their 1972 match*.

verbs. If a transitive base cannot be interpreted without its object, and the object cannot be ellipsed without a parallel context (what we will call a *rigid* object argument), it cannot be prefixed with *out-* at all (Bresnan 1980). For example, *John outwrites Bill* is grammatical for the same reason that *John writes (well)* and *John is a writer* are grammatical, and **John outresembles Bill* is ungrammatical for the same reason that **John resembles* and **John is a ressembler* are ungrammatical. For analogous reasons *out-* cannot be prefixed to verbs with rigid predicate complements, such as *be, seem, appear*: *Biden was more confused than Trump* cannot be expressed as **Biden outwas Trump confused*.

The prohibition that bars prefixation of *out-* to transitive verbs is not arbitrary. It is the consequence of a grammatical/semantic bottleneck. *Out-* requires two arguments, one for the comparandum (the event that is compared), the other for the comparand (the event that it is compared to). However a rigidly transitive base has an additional obligatory argument. So *out-comparatives* from rigid transitives would have three obligatory direct arguments. If one is left out, the sentence is uninterpretable: **John outadmires Mary* makes no sense because *Mary* cannot be *both* the admirer who is outdone by John *and* the admiree. However the syntax has no room for all three. **John outadmires Mary than Fred* is syntactically ill-formed. So is **John outadmires Mary Fred*, for the first object of a ditransitive must be a recipient-type argument. Alternatives like *John outadmires Mary with/for Fred* are excluded because the prepositions cannot convey the intended semantic relation.

The pattern in (4) reflects the same bottleneck:

- (4) a. (i) We drink wine. (ii) *We outdrink wine.
 b. (i) We drink. (ii) We outdrink you. (iii) *We outdrink you wine / wine you.

Contrasts like (5), between *appreciate* (i) ‘to value’, and (ii) ‘to increase in value’ are also predicted.

- (5) a. (i) We appreciate you/* \emptyset . (ii) *We outappreciate you/ \emptyset .
 b. (i) Real estate has appreciated. (ii) Real estate has outappreciated commodities.

There is however one class of rigidly transitive verbs to which *out-* can be prefixed. These are verbs that require measure phrases as complements, such as *weigh, yield, gross, spend*. To the extent that measure phrases are DP-like, in that they can have articles (at least indefinite ones) and adjectival modifiers, we can consider them objects and classify such verbs as transitive.

- (6) a. (i) Fred weighs a good 200 pounds. (ii) He outweighs Max by an estimated 80 pounds.
- b. (i) An acre of wheat yields an average of 40 bushels. (ii) Fallow fields outyield nonfallow fields.

However measure phrase complements are peculiar objects in many respects. They are nonreferential, cannot have definite articles (**Fred weighs the good 200 pounds*), resist passivization (**Two years were lasted by the war*, **Three days were stayed by the guests*), and cannot serve as the obligatory object argument of *out-*verbs.

- (7) a. **The famine outlasted two whole years.*
(OK: The famine outlasted the war.)
- b. **The guests outstayed two months.*
(OK: They outstayed their welcome.)
- c. **Harris outraised \$1 million.*
(OK: Harris outraised Trump.)

With comparatives, measure phrases provide a differential expressed by prepositional phrases. The same is seen with *out-*, as illustrated in (6a). The reason why *out-* can be prefixed to verbs like *weigh* is that their measure phrase objects can be rephrased as prepositional measure phrases, leaving the object position free for the comparand's subject, as in the (ii) example in (6a). However, as illustrated in (7), measure phrases cannot be objects of *out-*applicatives, presumably because of their syntactically defective nature.³

The highly regimented morphology and syntactic uniformity of *out-*verbs contrasts with their semantic flexibility. Many *out-*verbs are polysemous. Their interpretation depends on how the dimension of comparison is specified and on how the direct object determines the threshold for the comparison. Consider the sentence *Mary outgrew her stoner boyfriend*. It can be an event comparison (8a), a degree comparison (8b), and – in this case the most accessible reading – a modal degree comparison (8c).

- (8) Mary outgrew her stoner boyfriend.

³As discussed in section 3, degree descriptions, by contrast, are perfectly good objects for *out-*applicatives.

a. *Event comparison*

- i. ‘Mary grew faster than her stoner boyfriend did.’
- ii. ‘Mary grew by a larger amount than her stoner boyfriend did.’

b. *Degree comparison*

‘Mary grew bigger than her stoner boyfriend.’
[Can be true even if her boyfriend did not grow at all, or even shrank.]

c. *Modal degree comparison*

‘Mary grew too old (or too mature) for her stoner boyfriend.’ = ‘Mary reached an age (or level of maturity) greater than one in which she could have (or wanted to have) the stoner as her boyfriend.’

We will refer to the first and second terms of a comparison as the *comparandum* and the *comparand*, and for want of a better term, to the added direct object argument introduced by *out-* as the *applicative* argument. The applicative argument determines the type of the comparand and specifies a degree which sets a threshold that is exceeded by the comparandum.

The addition of the applicative argument, the prohibition of transitive bases, and the special behavior of measure phrases seen in (6) and (7), all confirm that *out-*applicatives are covert comparatives. *Out-*applicatives are of several types. In the following sections we lay out how these comparative meanings come about.

2 Event comparatives

In the *event comparison* reading exemplified by (8a), the subject and object arguments are participants in two events of the same type – in this case, growing events. The *out-*verb then expresses a comparison between the maximum degree of two events along some dimension (such as time, distance, quality, or number) that is compatible with the base predicate. The event whose participant is the object argument sets a threshold for the event whose participant is the subject argument – here the amount of growth or the rate of growth. See the additional examples in (9).

- (9) a. She is estimated to outearn her husband by \$4 million. [= earn more money]
- b. The stores in Paris outdraw both the Louvre and the Eiffel Tower. [= draw more customers]

10 *Out-Applicatives are Concealed Comparatives*

- c. She could outdraw any member of the gun club. [= draw her gun faster]
- d. The sun outshines the moon by a ratio of 1 to 400,000.
- e. Stocks outperformed bonds.
- f. The Trump administration outdoes itself on climate change denial.
- g. Hispanic women outvoted Hispanic men 52% to 48%.

A characteristic of event comparatives is that they imply the existence of two events whose degree along some dimension is compared – a comparandum and a comparand with distinct participants. They have equivalent overt comparative paraphrases with *more/-er*.

- (10) a. Smith far outdistanced Jones. [= covered more distance]
- b. Invasive Mahonia plants outnumber their native relatives. [= are more numerous]
- c. Hyperdominant arboreal species outgrow others in the rainforest whenever space opens up. [= grow faster / more vigorously]
- d. Mary outbid John in the auction. [= bid more]
- e. The planet outlived its sun by a million years. [= lived longer]
- f. Kim easily outclassed everyone in dancing. [= showed more class]
- g. The sun outshines the moon by a ratio of 1 to 400,000. [= shone more strongly]
- h. Do-nothing House of Representatives outdoes itself again. [= does even less = is even more inactive (than usual)]
- i. Hispanic women outvoted Hispanic men. [= voted more]

It appears that each reading of an event-type *out*-comparative replicates a reading of an explicit verbal comparative. For example, (10i) has the same readings as *Hispanic women voted more than Hispanic men voted*:

- (11) a. Proportional reading: The proportion of Hispanic women that voted was greater than the proportion of Hispanic men that voted.

- b. Cardinality reading: The number of Hispanic women that voted was greater than the number of Hispanic men that voted.
- c. Frequency reading: Hispanic women voted more often than Hispanic men voted.

Moreover, the meanings of verbal comparatives can correspond to nominal comparatives based on their arguments. For example, *More Hispanic women voted than Hispanic men did* can mean both (11a) and (11b).

In addition, *out*-applicatives obey the *commensurability condition*, which has been shown to hold for overt comparatives (Kennedy 1997, Doetjes 2010). The commensurability condition, which requires that the compared events be of the same type and that the participants engage in the same type of activity in them is mirrored exactly by *out*-applicatives.

The parallelism between *out*-applicatives and overt comparatives goes still deeper. When compared events are not identical but sufficiently similar, an indirect interpretation can be constructed. For example, although volcanoes do not run, *John ran faster than the volcano* is a felicitous way of comparing the speed of John's running with the speed of the lava flow. It is licensed by a coercion mechanism that replaces the predicate of the comparandum with a superordinate predicate applicable to both events. Exactly the same coercion process also applies in *out*-applicatives, such as *John was able to outrun the volcano*. When the subject and object cannot be participants of an event of the same type, commensurability can be achieved, as in overt comparatives, by replacing the predicate of the comparandum with a superordinate predicate of which both are hyponyms. The parallelism between covert and overt comparatives with respect to this commensurability strategy is illustrated in (12).

- (12) a. (i) John outran the eagle =
 John ran faster than the eagle (moved/flew/*ran).
 (ii) *John outflew the eagle =
 *John flew faster than the eagle (moved/flew/ran).
 b. (i) The eagle outflew John =
 The eagle flew faster than John (moved/ran/*flew).
 (ii) *The eagle outran John =
 *The eagle ran faster than John (moved/flew/ran).
 c. (i) Can you outrun a volcano? =
 Can you run faster than a volcano (*runs/flows)?
 (ii) *Can a volcano outflow you? =
 *Can a volcano flow faster than you?

- d. (i) The shark outswam the boat =
 The shark swam faster/better/longer than the boat (moved/swam).
 (ii) *The boat outswam the shark =
 *The boat swam faster/better/longer than the shark (moved/swam).

The (i) sentences are acceptable, while their (ii) counterparts are degraded to an extent that may vary between speakers, but is invariant across covert and overt comparatives. The hallmark of coercion is that it arises when the straightforward reading is unavailable for pragmatic reasons or due to conflict with background knowledge. For example, *John outplayed Mary* means ‘John played better/louder than Mary played’ unless background knowledge tells us that Mary did not play, in which case we try to accommodate the interpretation in the best possible way, perhaps by inferring that she may have sung. When no such hypernymic relation between the compared events is found, coercion fails. For example, **John outplayed the beer* is anomalous to the extent that we cannot find a V for ‘John played better/more than the beer V-ed’ (*performed/*tasted...).

A particularly complex case of coercion is (13):

- (13) I will outlive this danger, and outdie it, if I can. (C. Dickens, *Bleak House*, 1853)

Here *outlive* means ‘live longer than’ and *outdie* means ‘die later than’. In effect, *live* and *die* are coerced into the same type of gradable predicate.

Assuming the degree-based analysis of comparatives, we can characterize the event-comparative reading of *out-* as in (14), where Q_S is a contextually recoverable gradable property of events relative to a scale S , and P is an event predicate corresponding to the verbal predicate or an appropriate hypernym.

- (14) $\lambda P \lambda y \lambda x \lambda e. \exists d [P(e, x) \wedge Q_S(e, d) \wedge d > \text{Max}(\lambda d'. \exists e' [P(e', y) \wedge Q_S(e', d')])]$

3 Degree comparatives

Whereas event comparatives compare two events of the same type, degree comparatives, such as (8b), assert that the event denoted by the verbal predicate has a contextually salient gradable property to a degree of an event that exceeds a threshold which is identified by the object argument, either directly or via coercion. For example, in the degree comparative reading of (8b) *Mary outgrew*

her stoner boyfriend, Mary has to grow bigger than her boyfriend even if the boyfriend did not grow at all, and even if he shrank.

In the simplest kind of degree comparative, the object argument is just a degree-denoting expression that sets the threshold that is exceeded by the event described by the verbal predicate. The examples in (15) have a natural reading as degree comparatives, as the overt comparative paraphrases confirm.

- (15) a. Active managers seek to outperform their benchmarks. [= perform better than their benchmark = reach a level that is above the benchmark level]
- b. A 5-point means the team has to outscore the handicap. [= to score more than the handicap = get a score that is higher than the benchmark score]
- c. How can house prices outgrow the average income? [= become greater than the average income]
- d. Graduates from elite schools outperformed the norm in small companies.
- e. We have then outperformed and outachieved another person's level.

The interpretation is contextually dependent. For example, the meaning of (16) depends on whether the targeted goal of BAT level water consumption is understood as a minimum or maximum level.

- (16) NLMK Group outachieved the BAT level of water consumption as early as 2010.

In the more complex types of cases represented in (17), the object does not denote a degree directly, but contextually by coercing the object into a degree via a measure function that is based on the same scale as the scale of the gradable property that the described event is inferred to have.

- (17) a. I have outwalked the furthest city light. [= walked farther than the furthest city light] – Robert Frost
- b. My brave little doggy is trying to outbark the thunder but he's not winning.
- c. the Mac screen isn't strong enough to outshine the daylight.

- d. the joyful sense of safety to the hundreds on board found vent in a shout that out-shouted the storm. (OED)

We can characterize the degree comparative reading of *out* when the direct object is a degree denoting description as in (18a), and when the direct object is individual-denoting, as in (18b), where μ_S is a measure function based on the same scale as the scale of the gradable property Q_S .

- (18) a. $\lambda P \lambda y \lambda x \lambda e. \exists d [P(e, x) \wedge Q_S(e, d) \wedge d > y]$
 b. $\lambda P \lambda y \lambda x \lambda e. \exists d [P(e, x) \wedge Q_S(e, d) \wedge d > \mu_S(y)]$

4 Modal comparatives

In reading (8c), the object argument likewise serves to specify a degree that Mary's growth exceeded, but this time it is specified *modally* in terms of Mary's preferences. Although it is the most complex reading of all, it happens to be the one that *outgrowing the boyfriend* most immediately evokes, by far the most frequent use of *outgrow*, and the most creative and possibly most productive type of *out-comparative*.

This type of interpretation comes about when the applicative argument refers to something that cannot be a participant of an event of the same general type as the event of the main predicate, even by predicate coercion to a hypernym. It determines the type of the comparand event indirectly by establishing a modal domain of worlds in which there is an event of the same type as the comparandum event. Modal comparatives compare the degree along some dimension of an event in the *actual* world with the same kind of event in a domain of *alternative* worlds. Hence the subject argument is a participant of the event in both types of worlds. The applicative argument is not itself a participant; it just determines the threshold indirectly by providing a modal domain. We refer to this as the *modal comparative*.

The modal domain consists of the set of worlds compatible with some kind of content associated with the direct object. That content must contain information that has an implication about the degrees to which the comparand event has the contextually recoverable gradable property.

We can distinguish two cases of modal comparatives. In one case, the object argument denotes entities associated with informational content, either directly (for example, *predictions* or *expectations*), or indirectly (for example, *actuarial tables* or *polls*).

- (19) a. We have outrun our goal. [= run farther than we intended to run]
b. At 100, Carter has outlasted all expectations. [= has lived longer than anyone could have expected that he would live]
c. Trump inflates his record to see himself outperforming history [= performing better than what is expected on the basis of history]
d. Europe has outperformed the requirements of the existing UN sanctions regime. [= imposed sanctions to a greater degree than the UN requires.]

In (20a), since polls are not performers, the regular event comparative reading ‘performed better than the polls performed’ is pragmatically odd, and *the polls* is understood as specifying an expected or predicted degree of performance. Similarly, in (20b), where the regular event comparative reading is odd because actuarial tables do not live.

- (20) a. The candidates outperformed the polls [= performed better than what is expected on the basis of the polls ≠ *performed better than the polls performed]
b. John outlived the actuarial tables. [= lived longer than the actuarial tables predicted he would live, ≠ *lived longer than the actuarial tables lived]

The ambiguity of (21) further illuminates the difference between event readings and modal readings:

- (21) He far out-performed his critics.
a. Event comparison: He performed better/more than his critics performed. [comparison of actual performances]
b. Modal comparison: He performed better than his critics thought/ predicted he would/could perform [comparison of actual and potential performances].

The modal domain can also come about through a goal-oriented construal of the modality. In that case the modal domain consists of worlds where a salient goal is achieved, and which are exactly like the actual world with respect to a set

of facts. The modal domain in effect specifies the prerequisites on the degrees to which the comparandum event has the contextually determined gradable property for the goal to be realized. The applicative argument helps to determine the relevant goal, or a relevant fact that figures in what is required for the goal to be achieved.⁴

- (22) a. The plant outgrew its pot. [= grew too big for its pot (to fit its pot). = grew bigger than the maximum size it could be in order to fit into its pot ≠ *grew bigger than its pot grew.]
- b. Smith outlived his money by decades. [= lived longer than the maximum time he could afford to live = lived longer than the maximum time he could support himself with the money he had ≠ *lived longer than his money lived.]
- c. The global economy is outgrowing the capacity of the earth to support it. [= growing too big for the capacity of the earth to support it = bigger than what the maximum size the earth can sustain.]
- d. You can't outrun a crappy diet. [= You cannot run more intensively than your crappy diet requires of you to run to stay fit. ≠ *run more intensively than your crappy diet runs.]
- e. You can't outcompute the market's vagaries. [= You can't compute better than the vagaries of the market require you to compute (in order to make a profit) ≠ *compute better than the vagaries of the market compute.]
- f. ...out-witting and out-sneaking the fierce-looking rat guards. [= sneaking with more skill than the minimum amount of skill needed to evade the rat guards.]

Pots do not grow, requirements do not perform, but they identify a degree which sets the threshold for the growth and performance. Modal comparatives based on informational direct objects have regular *more/-er* paraphrases, but those that are based on a goal-oriented construal of the modality can be paraphrased by *too* and *not enough* constructions, whose modal comparative readings have been established by Nadathur (2019).

⁴Examples (22d,e,f) are from Kotowski (2023).

5 The uniqueness and origin of *out*-comparatives

In aggregate, the *out*-applicative is the only morphological construction of its kind in English, and it is a rarity in the languages of the world (we are not aware of any other instances of it). However, it shares almost every one of its individual properties with at least one other construction of English.

Homonymous with *out*-applicatives are locative verbs like *outsource*, *outflow*, *outhaul*. In earlier English this class included numerous compounds of verbs with the normally separable particle *out*- such as *outjeer* ‘jeer out’, most of which are now obsolete. Kotowski (2023) makes a case for unifying *out*- in a larger comparative/competitive/locative scalar category. In this paper we treat applicative and locative *out*- as functionally distinct morphemes. Unlike applicative *out*-, locative *out*- does not verbalize nouns, but nominalizes verbs (*outlook*, *output*, *outcome*, *outgrowth*, *outreach*) or adjectivalizes them (*outdoor*, *outboard*) or else preserves the base’s category (*outpatient*, *outbound*). In any case, locative *out*- verbs have no comparative meaning.⁵

The decline of locative *out*-verbs coincides with the rise of *out*-applicatives beginning at the turn of the 17th century.

Applicative *out*- is not the only category-changing prefix that makes transitive verbs by adding syntactic subject and object arguments if the base does not have them. Others include privative *de*-, repetitive *re*-, *dis*- (*disillusion*, *disrobe*), *en*- (*encourage*, *encircle*, *ensure*, *endanger*, *envision*), *be*- (*befriend*, *bedew*, *behead*). Old English and Middle English had a small number of covert comparative verbs, derived not with *out*- but with *over*-, combined with verbs (never nouns or adjectives) into inseparable prefixed verbs. The most frequent of these was *oferlyfan*, a counterpart of Latin *supervivere* ‘outlive’, ‘survive’, and probably a calque from it, like French *survivre* and German *überleben*.

- (23) a. Ðet lond ... wes becueden Osbearte..., gif he Cyneðryðe oferlifde.
the land was bequeathed to Osbeart..., if he Cyneðryð outlived
‘The land was bequeathed to Osbeart, if he were to outlive Cyneðryð.’
(early OE, Kentish dialect)

⁵A tricky case is *outflank*, as in *The cavalry outflanked the enemy*. It implies no event comparison (the enemy could be stationary, or even asleep), nor a degree comparison (*the enemy* cannot be construed as a threshold-setting measure phrase), and there seems to be no modal interpretation that could yield an indirect comparison between two events. To *flank* something is to be located at its flank, and to *outflank* it is to move past its flank. Its extended use in the sense “to defeat” (as in *a determined majority of the council may outflank the direct legislation supporters*, OED) would then be a metaphorical use based on the locative meaning.

10 Out-Applicatives are Concealed Comparatives

- b. Seint Bryde ... overlevede him by sixty zere. (1387)
'St. Brigid survived him by 60 years.'
- c. Wilfridus ... overpasseþ Beda his tyme (1387)
'Wilfrid lives past Bede's time'
- d. Howe lange Marie ouerlyved hire sons ascensioune. (c1429)
'How long Marie lived past her son's ascension.'
- e. zif deþe wolde resayve mede, Many wolde oþer ouerbede. (c1400)
'if death would receive a reward, many would overbid each other.'
(Cf. German *überbieten*)

The same prefix *over-* was used to make verbs of excessive degree, such as *overflow*, *overeat*, *overcharge*, *overestimate*, *overflow*, *overgrown*, which are still in use.

- (24) a. þæt land middeweard oferfleow mid fotes þicce flode. (Orosius)
'the middle of the land overflowed with a foot-deep flood.'
- b. Whereas there be some men which ouerreach and goe beyond this marke. (1569)
- c. God would haue a Sabbath kept; they ouer-keep it. (1608)

Along with its antonym *under-*, comparative *over-* is still in use. Unlike *out-*applicatives, they are not category-changing and they are exclusively modal and paraphrasable by *too (much)* and *not enough*. They apply a measure function to an eventuality and assert that its degree value exceeds or falls short of a threshold, which may be conventional or contextually determined or specified by a measure phrase. Compare the meanings in (25):

- (25) a. Brussels overregulates us. [= Brussels regulates us too much – one event]
- b. Brussels out-regulates London. [= Brussels and London regulate – two events]
- c. John oversold Bill. [John touted Bill too much – one event]
- d. John outsold Bill. [John sold more than Bill sold – two events]

Over- and *under-* when used in this sense are freely added to transitive bases and do not transitivize intransitive ones: *oversleep*, *overreact*, *overprotect*, *overkill*.

Like all event-constructing derivatives from nominal bases, *out-*applicatives restrict their nominal base to its canonical function.

- (26) If an action is named after a thing, it involves a canonical function of the thing (cf. Kiparsky 1997)

Consider “putting” verbs like *saddle* and *corral*.

- (27) a. They saddled the horse.
b. They corralled the horse.

(27a) does not mean just ‘They put a saddle on the horse.’ It means that it is done in the way that saddles are meant to be used. For example, it cannot describe an act of loading a basket onto the horse’s back for purposes of transporting it, or putting a saddle on the wrong part of the horse’s anatomy. And (27b) does not mean just ‘They put the horse in the corral.’ It must be done in the canonical way that corrals are intended for: for one thing, the horse must be alive.

In sum, all individual features of applicative *out-*, including its affixation to nouns, adjectives, and verbs, its verbalizing and transitivizing function, and its comparative semantics, have existed since Old English times. *Only their combination is new.*

There are very few instances of *out-*applicatives before the end of the 16th century. The earliest are *outlive*⁶ (replacing older *overlive*) and *outproffer* ‘outbid’ and *outproffer*.⁷ Palsgrave’s 1530 textbook and lexicon of French cites five of them, glossed with French counterparts that seem to be otherwise unattested, at least in the relevant meaning. Examples from Palsgrave, *Lesclarcissement de la langue francoyse*, 1530.

- (28) a. I OUTCRY. *Je forcrie*, prim. conj. Lette hym crye as loude as he wyll, yet I wyll **outcrye** hym: *quil crie aussi hault quil voudra, je le veulx forcrier* or **oultrecrier**.
b. I OUTEATE. *Je formangeus*, verbum defectivum, conjugate in *je mangeus*, I eate. *My horse wyll outete such four jades as thyne is: mon cheval formangera quatre telles charoignes quest la tienne.*

⁶In cas hereafter it happen You...to outleve our...Sovereigne Lord (1472).

⁷Eyther **out proferyd** other (1513) ‘each outbid the other’.

- c. I out go one, I go faster than he can go. Je vas plus viste, conjugate in «I go». Though thou be goynge an hour afore me, yet I wyll **out go** the: *combien que tu te mets a cheminer une heure deuant moy si te **oultrepasseray** je, or si yray je plus viste que toy.*
- d. I OUTRYDE. *Je oultre chevauche*, prim. conj. and *je surpasse en chevaulchant*. Take as swyfte a geldinge as thou canste fynde and I holde the twenty nobles I **outryde** the: *prens aussi viste hongre que tu peulx trouver, et je gaige a toy vingt angelotz que je te **surpasseray en chevaulchant**.*
- e. I OUTRYME. *Je outre rysme*, prim. conj.

We suspect that the prefixed French verbs such as *forcrier* and *oultrecrier* that Palsgrave uses to render *out*-applicatives are calques on the English verbs. Calques and code-switching in the French spoken in England occur from the fourteenth century on (Rothwell 2000).⁸

The extension of *out*-comparatives to their modern range and their systematic differentiation from *over*-comparatives seem to have been introduced by Shakespeare. In particular, he is the first author to coerce *out*-comparatives to commensurability⁹ and the first to feature denominal and deadjectival *out*- (*out-Herod*, *outvenom*, *outvillain*, *outpeer*, *outnight*, *outlustre*, *out-tongue*, *outbrave*, *outworth*, *outcrafty*). In fact, it may have been this category-changing force that drove the spread of *out*-applicatives, as the pairing of deverbal *overstare* and deadjectival *outbrave* in (29a) suggests. Shakespeare also introduced *out*- in cognate object constructions of the type (29b) and (29c), crafted the gorgeous line (29d), and used an *out*-applicative to boast of his prowess in (29e).

- (29) a. I would *ore-stare* the sternest eyes that looke:
Out-braue the hart most daring on the earth. (*Merchant of Venice*, ca. 1596)
- b. Our prayers do outpray his. (*Richard II*, ca. 1595)
- c. it out Herod's Herod (*Hamlet*, ca. 1600)
- d. My selfe could else outfrowne false Fortunes frowne. (*King Lear*, ca. 1605)
- e. Not marble, nor the guilded monument,
Of Princes shall out-liue this powrefull rime. (*Sonnet 55*)

⁸We are grateful to Eve Clark and Carol Harvey for this information.

⁹Marke ... *How he outruns the wind* (*Venus & Adonis*).

6 Preservation of the base's argument structure

Previous treatments of *out*-verbs recognize the semantic heterogeneity of *out*-verbs, but theorize it in different ways. Most of them concentrate on event comparatives like (8a), and to the extent that they treat plain degree and modal comparatives like (8b,c) at all, they classify them as resultatives. In a tightly reasoned syntax-centered analysis of event comparative *out*-applicatives in the framework of Distributed Morphology, Ahn (2022) argues that they have a comparative meaning and that their argument structure is independent of the base's argument structure, and comes entirely from the prefix *out*-. He touches only briefly on the degree and modal event *out*-applicatives, suggesting that, contrary to what we have argued here, they are not comparatives, but have a different, perhaps resultative meaning.

Ahn's derivation of *out*-applicatives merges *out*- with a minimal VP that has no superordinate syntactic structure, erasing the argument and introducing two new arguments which do not saturate any Theta-roles of its base. Rather, the base of *out*- constrains the dimension of comparison semantically, and the types of events in which the two arguments of *out*- can be interpreted as participating in is constrained by conceptual/pragmatic reasoning in which syntactic or semantic structure plays no role (Ahn 2022 p. 431).

As one piece of evidence for this analysis Ahn cites *out*-verbs like *outrun*, which seem to have a subject of a type that the basic verb does not allow, in which case it would have to be introduced by *out*-.

- (30) a. Aircraft carriers can outrun almost any other boat.
b. *Aircraft carriers can run.

These cases require a closer look. The verb *run* denotes not only a specific type of gait (moving quickly on foot with both feet leaving the ground at each step), but many other kinds of continuous and typically rapid movement, including the movement of ships and other conveyances, machines, liquids, fingers over a surface, etc. The problem with (30b) is that it is nonsensical on the 'gait' reading and vacuous on the general 'movement' reading. Adding a manner or direction to modify the movement makes (30b) contentful and acceptable, as examples like (31) (from the Internet) testify:

- (31) a. The best of the Soviet submarines can run as fast under water as the Navy's fastest ships can move on the surface.

- b. U.S. aircraft carrier runs away when encountering CCP military exercise? (heading)
- c. ...when the ship runs from deep into shallow water the kinetic energy stored in the entire hull converts itself into suction-energy.
- d. If you think for a minute that a big ocean liner can run up alongside the wharf at any port she happens to come to...
- e. The ship runs from the Texas-Mexico border all the way around to the Florida Keys.

Since *outrun* selects the same types of subjects as *run*, we conclude that it retains the subject of its base just like other *out*-verbs, exactly as our analysis predicts.

Ahn's second example of *out*-contributing a subject argument that supersedes the base verb's selectional requirements is *outrain*. The ungrammaticality of (32a) seems to show that weather verbs like *rain* have neither a subject argument nor an object argument. If so, then the subject and object arguments of *outrain* (32b) must be introduced by *out*-.

- (32) a. *Atlanta rained. *It rained Atlanta.
- b. Atlanta also out-rained Seattle in 1922 and 1923.

This is actually consistent with our analysis. As we have already shown, *out*-supplies a subject to a subjectless base, as well as an object to an objectless base. That is how it makes transitive verbs from nouns and adjectives, which have neither, e.g. *outpace*, *outmuscle*, *outnumber*, *outsmart*, *out-Einstein*. So on the supposition that *rain* has neither a subject nor an object, *out*-endows it with both.

That said, a deeper dive into *rain* (and other weather verbs) yields a more interesting picture. First, sentences like (32b) *Atlanta out-rained Seattle* are not accepted by all speakers. Secondly, they are matched with (33b)-type examples, which are also acceptable to some but not all speakers.

- (33) a. Scotland rains more and Ireland rains more than Scotland.
<https://bulbagarden.net/threads/random-messages.102920/page-2408>
- b. Most of Scotland rains all the time, so an electric tumble dryer is a necessity.
<https://www.vacuumland.org/cgi-bin/TD/TD-VIEWTHREAD.cgi?24213>

- c. Seattle rains more days than it doesn't.
<https://www.jessikachristinephoto.com/blog-1/urban-couple-photos-in-seattle>
- d. Even in the summer, Ireland rains a lot.
<https://travellingjezebel.com/one-day-in-dublin/>
- e. British Columbia's West Coast rains more than 50
<https://www.instagram.com/reel/C7fFSTxyME6/>

We have found that speakers who accept sentences like (32b) also tend to accept sentences like (33), and allow *rain* with a locative subject as long as a measure phrase is present or implied, both in the simple verb (33) and in *out*-verbs like (32b). *Rain* would then not only be consistent with our thesis, but support it positively against the alternative that *out*- has no access to the base's argument structure. However, to repeat, even if *rain* were argumentless, *out*- would turn it into a transitive verb, as it turns nouns and adjectives into transitive verbs. On that supposition, no co-variation between (33) and (32b) would be expected.

In any case, neither *outrun* nor *outrain* show that *out*-verbs do not inherit the base verb's subject argument but get their subject argument from *out*-. That leaves Ahn's claim without supporting evidence. The weaker claim that *out*-applicatives inherit only the base verb's subject argument but not its object argument would not be consistent with Ahn's way of implementing the argument-structural impoverishment of the input by restricting it to a small syntactic constituent, for there is no constituent that includes the base verb and its subject but not its object.

Counterevidence to Ahn's claim that *out*- discards the base's argument structure comes from the abovementioned generalization that it cannot be added to rigidly transitive verbs like *resemble*, *detach*, and *have*. This generalization requires access to the full argument structure of the base from which *out*-applicatives are derived.

The contrast between *buy* and *sell* is another telling piece of evidence:

- (34) a. (i) This book outsells the Bible. (ii) *This book outbuys the Bible.
- b. (i) This book sells (well). (ii) *This book buys (well).

Sentence (34a) follows directly from (34b) as long as *out*- can see that transitive *sell* has a middle counterpart but *buy* does not, as our analysis predicts it can. (On the other hand, *buy* is not a rigid transitive, so we can have both *Do-it-yourselfers*

buy from us, and therefore also *Do-it-yourselfers outbuy professionals 3-to-1*, as predicted.) In the derivation of (34ai), transitive *sell* is detransitivized to the middle *sell* and then retransitivized by *out-*. It could then be re-detransitivized again by passivization, e.g. *The Bible has never been outsold by any other book*.

If argument structure is not visible to *out-*affixation, we lose the explanation for this data, and specifically for the difference between *buy* and *sell* in (34). The reason is that every buying event is also a selling event and vice versa, and so the difference between *buy* and *sell* lies not in the semantics but in their argument structure – that is, in the way the semantic roles are associated with the syntactic arguments. Ahn (p. 458) notices this issue and proposes a somewhat problematic workaround to deal with it.

7 Inferential meaning: resultativity and competition

*Out-*prefixation has the same category-changing and transitivity function as the resultative construction, e.g. *to floor an opponent*, *to down a drink*, *to up the ante*, *to cellar the wine*. Capitalizing on this similarity, McIntyre (2003) and Nagano (2011) claim that all *out-*verbs are resultatives. In their analysis, the prefix does not introduce a comparative, but a sub-event that the object argument participates in, which is caused by an event that the subject argument participates in, creating a complex resultative predicate. The object argument is “outdone” or “beaten” as an effect of the subject argument’s involvement in the event denoted by the verbal predicate. The resultative interpretation can moreover be understood as an actual or implicit *competition* in which the subject argument is the winner. We think that the resultative and competition readings are context-dependent inferences (Barsalou 1982).

Resultatives are usually regarded as semantically complex predicates with two sub-events, in which the sub-event that the object argument participates in is caused by the sub-event that the subject argument participates in. For example, a resultative analysis of *Fred outdrank Stan* would be that there was an event of drinking by Fred which resulted in Stan being defeated. As Kotowski (2021) notes, this does not really capture the full meaning of the sentence, in particular that Stan was also drinking, and that Fred drank more than Stan did. Nor does it fit (8b,c), of course. Kotowski (2021, 2023) reviews and criticizes the comparative and the resultative analyses of *out-*verbs, concluding that the verbs lie on a cline between multiple related meanings, where the range of possible interpretations is context- and participant-bound rather than primarily driven by the base to which *out-* is prefixed.

According to Kotowski's (2021) analysis of *out-V*, the interpretation of the construction involves a *competition*: the subject argument surpasses the object argument on some dimension, thereby defeating it. The competition can be explicit:

- (35) a. He was sure their clients out-healed, out-helping-handed, overall out-charitied their competing charities.
- b. It is routine for presidential candidates to try to "out-pro-Israel" each other.
- c. trying to out-right-wing each other
- d. to bolster political legitimacy by "out-Indianing" their opponents through support of local cultural institutions.

Such resultative and competition meanings as *out*-applicatives may have, we argue, are inferential. For example, (36a) (from a sports page) tells you not only that the Vikings rushed, passed, and threatened more than the Bears did, but that they competed in a game, and that the Vikings won.

- (36) a. The Vikings outrushed, outpassed and outthreatened the Bears.
- b. Sebastian Coe outran Roger Bannister.
- c. The pronghorn outdistanced the cheetah.

(36b) implies that Coe and Bannister were vying to reach a common goal at the same time, rather than just moving independently at different times; the background knowledge that they actually never competed or raced at the same time makes it anomalous. (36c) suggests that the pronghorn was chased by the cheetah and escaped it handily, rather than just by the skin of its teeth. Even though a competition scenario is far-fetched for *outlast* and *outweigh* in (6), they still seem to convey that the quantities involved, and the differences between them, are somehow significant, in a way that their explicit comparative counterparts *last longer* and *weigh more* do not.

However, these are defeasible implicatures, for *out*-verbs can be used when there is no competition, as with non-sentient participants, and where the difference in degree between the events can be explicitly specified as negligible:

- (37) a. this bat species is one of the smallest, growing only big enough to barely outweigh one US nickel.

- b. The utilization of external calcium ions at the neuromuscular junction is restricted to a brief period which barely outlasts the depolarization of the nerve ending.
- c. Of particular interest are the systems close to the tipping point, when dispersion interactions barely outweigh or approach the strength of the other interactions.

If the degree of the comparandum event is understood in context to also be the degree required for some salient goal to be realized, and the degree of the comparandum exceeds the degree of the comparand, which is the modal threshold, then the basic comparative reading can inferentially yield both a resultative implication and an additional “competition” implication, namely that the subject argument prevails over the applicative argument. The contrast between (38a) and (38b) illustrates how the extra implicature depends on the volitionality of the *out*-verb’s object.

(38) a. John outran Bill.

- (i) Plain comparative: John ran faster than Bill.
- (ii) Comparative with resultative implication: John had the goal of beating Bill by running faster/farther/better than him, and he succeeded.
- (iii) Comparative with resultative and competition implication: John and Bill each had the goal of beating the other by running faster/farther/better than them, and John succeeded.

b. Indiana Jones outran the boulder.

- (i) Comparative: Indiana Jones ran faster than the boulder.
- (ii) Comparative with resultative implication: Indiana Jones had the goal of escaping the boulder by running faster than it, and he succeeded.
- (iii) Comparative with resultative and competition implication (unavailable): #Indiana Jones and the boulder each had the goal of running faster than the other, and Indiana Jones succeeded.

8 Conclusions

We have ferreted out at least some of the regularities behind the productive use of the prefix *out-* and charted this little morpheme's vast semantic range. Non-locative *out-*verbs express a comparison between two degrees along the same dimension, in which the event with the subject argument as a participant exceeds a threshold identified by the object argument. Like overt comparatives, they allow measure phrases that specify a differential. They are derived by prefixing *out-* to nouns, adjectives, or verbs, and the prefixation outputs a transitive verb. The base provides, directly or indirectly, a dimension of comparison, and its argument structure determines the availability of *out-*prefixation. If the base is a noun or adjective, a subject and object argument are supplied, and if it is an intransitive verb, an object is supplied. If the verb is rigidly transitive, in the sense that it is uninterpretable without its object, *out-* cannot be prefixed, unless the object is a measure phrase which can be turned into a differential specification rendered as a *by*-phrase.

By taking into account inferential meaning, we found that all *out-*applicatives share the common thread of comparative semantics, corresponding both to that of ordinary *more/-er* comparatives and modal *too/enough* comparatives. This is a challenge to the idea that morphology is syntax, or a spellout of syntax. The reason is that plain and modal comparatives have not (at least so far) been shown to be syntactically derived from a single base structure. An added advantage of our analysis is that it allows *out-* to be added to a base with its argument structure intact, whereas Ahn's syntactic derivation requires discarding the base's argument structure, with the untoward empirical consequences discussed in section 6.

That said, our own analysis of *out-*applicatives still falls short of a fully unified account of their systematic ambiguities. Perhaps an underspecification approach, or alternatively the technique of ambiguity packing (Crouch 2005, 2006) could pave the way to further progress.

Acknowledgments

Thanks to Chris Piñón and two anonymous reviewers for their careful reading of a draft, and to Ranjani Iyer for help in the early stages.

References

Ahn, Byron. 2022. Mapping *out-* argument structure. *Syntax* 25. 417–465.

- Barsalou, Lawrence W. 1982. Context-independent and context-dependent information in concepts. *Memory & Cognition* 10(1). 82–93.
- Bresnan, Joan. 1980. Polyadicity. In Teun Hoekstra, Harry van der Hulst & Michael Moortgat (eds.), *Lexical grammar*, 97–121. Revised as Joan Bresnan, ed., 1982. *The Mental Representation of Grammatical Relations*, Chapter 3, 149–172. MIT Press. Foris.
- Crouch, Richard. 2005. Packed rewriting for mapping semantics to KR. In *Proceedings of the sixth international workshop on computational semantics (IWCS-6)*, 103–114.
- Crouch, Richard. 2006. Packed rewriting for mapping text to semantics and KR. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: variations on a theme by Ronald M. Kaplan*, 389–416.
- Doetjes, Jenny. 2010. Incommensurability. In *Logic, language and meaning: 17th Amsterdam colloquium revised selected papers*, 254–263. Springer.
- Kennedy, Chris. 1997. Comparison and polar opposition. In *Semantics and linguistic theory*, vol. 7, 240–257.
- Kiparsky, Paul. 1997. Remarks on denominal verbs. In Alex Alsina, Joan Bresnan & Peter Sells (eds.), *Argument structure*. CSLI.
- Kotowski, Sven. 2021. The semantics of English *out*-prefixation: A corpus-based investigation. *English Language and Linguistics* 25(1). 61–89.
- Kotowski, Sven. 2023. Modeling locative prefix semantics. A formal account of the English verbal prefix *out*. *Morphology* 33(2). 115–152.
- McIntyre, Andrew. 2003. Preverbs, argument linking and verb semantics: Germanic prefixes and particles. In Geert Booij & Jaap Van Marle (eds.), *Yearbook of morphology 2003*, 119–144. Springer.
- Nadathur, Prerna. 2019. *Causality, aspect, and modality in actuality inferences*. Stanford University. (Doctoral dissertation).
- Nagano, Akiko. 2011. The right-headedness of morphology and the status and development of category-determining prefixes in English. *English Language and Linguistics* 15(01). 61–83.
- Rothwell, William. 2000. Aspects of lexical and morphosyntactical mixing in the languages of medieval England. In D. A. Trotter (ed.), *Multilingualism in later medieval Britain*. D. S. Brewer.

Chapter 11

Revisiting the Light Verb Jungle: How Good Is GPT at Hacking Away?

Annette Hautli-Janisz^a, Erisa Bytyqi^a, Pranshu Gupta^a & Miriam Butt^b

^aUniversity of Passau ^bUniversity of Konstanz

This paper examines how well current state-of-the-art Large Language Models fare with respect to providing information on verb-verb (V-V) complex predicates as found in Urdu and Hindi. These represent a complex linguistic phenomenon that is comparatively well understood from a theoretical perspective and for which we can rely on previous computational work to provide benchmarks we can test against. On the whole our results (unsurprisingly) show that GPT-4 fares better than GPT-3.5. However, we also found differences due to the language the prompt was given in that go counter to current findings in the literature.

1 Introduction

Complex predicates pose a challenge for both theoretical and computational models of language, because two or more individual verbs together form a construction that is syntactically complex on the one hand, but functionally behaves like a simple verb on the other. This is difficult to model theoretically and it also poses a challenge for computational approaches, because the detection and analysis of multiword expressions continues to pose a myriad of computational challenges (Constant et al. 2017, Baldwin & Kim 2010).

In this paper, we explore the capabilities of the new generation of Large Language Models (LLMs) to identify and analyse complex predicates. We focus particularly on V-V complex predicates in Urdu/Hindi, i.e. combinations of a main

verb and a light verb to form a single verbal predication, because these are comparatively well understood from a theoretical perspective (Butt 1995, 2010, Hautli-Janisz 2015) and have been studied in previous computational work (Ahmed & Butt 2011, Butt et al. 2012, Ahmed et al. 2012, Sulger & Vaidya 2014, Hautli-Janisz et al. 2015, Vaidya et al. 2019). The paper focuses on aspectual and permissive complex predicates (CPs) discussed in Butt (1995) and, more recently, in ‘The Light Verb Jungle: Still Hacking Away’ (Butt 2010), as well as CPs representing complex motion events (Hautli-Janisz 2013, 2015).

In order to assess the linguistic capabilities of LLMs, we systematically investigate the performance of two versions of GPT, GPT-3.5 (Open AI 2024a) and GPT-4 (Achiam et al. 2023), with respect to identifying and analysing three different types of V-V CPs in Urdu/Hindi. We do this by varying the prompts of the models along two axes: First, we prompt for different linguistic aspects of CPs, from syntactic properties (grammaticality judgments, part-of-speech tagging and word order permutations) to more lexical semantic aspects (translations and minimal pair generation). Second, motivated by previous work which has shown that prompting in the language under investigation yields better performance than prompting in English (Cao et al. 2023), we vary the prompt language between English with Romanized Urdu, Urdu (with Arabic script), Hindi (with Devanagari script) and Hindi (with Roman script). The results show that both GPT-3.5 and GPT-4 exhibit a varying zero-shot performance across tasks, with the only generalizable conclusion being that GPT-4 outperforms GPT-3.5. In contrast to previous work, we do not find consistent evidence that the language of the prompt has an impact on the quality of the responses, even though there is a tendency that prompts in Hindi (in Devanagari or Roman script) yield higher performance; in some cases, for instance for grammaticality judgments, these do not. We find initial evidence that chain-of-thought prompting brings GPT-3.5 performance on par with GPT-4 performance.

2 Related work

Linguistic capabilities assessment using Large Language Models A significant amount of work in computational linguistics and Natural Language Processing (NLP) is currently dedicated to assessing the linguistic capabilities of Large Language Models (LLMs), with studies aiming to identify to what extent linguistic structures are salient in the models and what prompts work best at getting meaningful responses. Kocoń et al. (2023) use 25 NLP tasks, from entailment and natural language inference to word sense disambiguation and grammaticality

checks, to compare the performance of ChatGPT and GPT-4 against state-of-the-art NLP models, showing that LLMs do not perform as well as current state-of-the-art models with respect to these tasks. In a similar vein, Dentella et al. (2023) show that for low-frequent grammatical constructions such as the avoidance violation (‘Dogs dogs dog dog dogs.’) and agreement attraction (‘The key to the drawers are on the table.’), GPT-3 has an error rate of almost 80% for English constructions. Cai et al. (2024) compare the performance of ChatGPT and Vicuna and show that LLMs reuse sentence structure, disambiguate words and can represent verbal event structure. Riemenschneider & Frank (2023) show that LLMs are able to capture linguistic properties of Ancient Greek texts. Hautli-Janisz & Bytyqi (2024) show that neither GPT-3.5 nor GPT-4 exhibit high or consistent performance in fallacy identification, i.e. faulty argumentation that violates normative principles of argumentation. On the other hand, in areas like persuasive essay writing, Herbold et al. (2023) show that GPT-3.5 and GPT-4 both outperform human competence by a significant margin (with GPT-4 significantly better than GPT-3.5 performance).

With respect to multilingualism (both GPT models are English-centric but will accept and output content in several languages), previous research shows that while the performance of the models (e.g. GPT-3, PaLM) does not drop significantly when the input is in the target language(s) of choice, it fails to deliver an improvement over using the prompt or input in the source language, English. Shi et al. (2022) present the MGSM benchmark, created by manually translating 250 grade school math problems sourced from the GSM8K dataset and their intermediate reasoning steps into ten typologically different languages. While the paper shows that multilingual chain-of-thought (CoT) prompting is extremely effective, including for low-resource languages, it still does not (in most cases) surpass English CoT.

Overall, drawing general conclusions on the linguistic capabilities of LLMs is difficult, because the field is using a number of very different tasks, datasets and benchmarks. However, this is the first step in being able to assess the capabilities of these models at all, before we can arrive at more systematic ways to probe them. The present paper makes a contribution in this area by investigating a complex linguistic phenomenon in two closely related low-resource languages.

Complex predicates in computational linguistics A significant body of work on complex predicates in computational linguistics has revolved around the identification and classification of CPs. Mukerjee et al. (2006) detect CPs in Hindi through POS alignment between Hindi and English, and Sinha (2009) mines

Hindi CPs in parallel data. Ahmed & Butt (2011) distinguish noun + verb complex predicates from individual verbs and their arguments. Butt et al. (2012) use bigram collocations to filter complex predicates from regular, tensed verb phrases. CPs are also present in other languages, for instance Brazilian Portuguese, where Sanches Duran et al. (2011) use POS patterns to identify and analyse CPs. For English, Jagfeld & van der Plas (2015) use distributional semantics and parallel data to find synonyms for CPs.

One area where CPs are crucial for performance is machine translation, because a complex predicate in the source is often translated in part by non-verbal lexemes in the target. Pal et al. (2011) show that an integration of CPs improves a Bengali-English machine translation (MT) system, parallel to the finding of Barančiková & Kettnerová (2017), who propose a dictionary of paraphrases of Czech CPs. Another line of work is concerned with lexical resource development for CPs, paving the way for more robust syntactic and semantic parsing. Hwang et al. (2010) provide the first multilingual annotation scheme for labeling CPs in corpora, Hendrickx et al. (2010) develop a corpus for Portuguese CPs, and Ahmed et al. (2012) present a reference bank for complex predicates in Urdu, followed by a VerbNet integration of Urdu verbs including complex predicates (Hautli-Janisz et al. 2015) and a PropBank resource for Urdu CPs (Anwar et al. 2016).

Despite the above research, complex predicates are still under-resourced given their ubiquity crosslinguistically and despite a recent large-scale COST action looking at the analysis and detection of multiword expressions, including complex predicates, from a computational perspective.¹ In this paper we provide a first capability assessment of LLMs on the task of identifying and analyzing complex predicates.

3 Complex predicates

This section briefly introduces the types of V-V Urdu/Hindi complex predicates we investigate with respect to the capabilities of LLMs.

3.1 Aspectual complex predicates

A ubiquitous part of Urdu/Hindi is the use of V-V combinations that Butt (1995) dubs *Aspectual Complex Predicates*. These consist of two verbs, where the main predicational element appears in the stem or bare form and the other verb carries tense/aspect and number and gender inflection, as shown in (1).

¹<https://typo.uni-konstanz.de/parseme/>

- (1) a. nadya=ne xat lk^h li-ya
 Nadya.F=Erg letter.M.Nom write take-Perf.M.Sg
 ‘Nadya wrote a letter (completely).’
- b. nadya=ne makan bana di-ya
 Nadya.F=Erg house.M.Nom make give-Perf.M.Sg
 ‘Nadya built a house (completely, for somebody else).’
- c. ram ga ut^h-a
 Ram.M.Sg.Nom sing rise-Perf.M.Sg
 ‘Ram sang out spontaneously (burst into song).’
- d. nadya has paɾ-i
 Nadya.F.Sg.Nom laugh fall-Perf.F.Sg
 ‘Nadya burst out laughing.’

The second inflected verb is referred to as a *light verb* since it does not predicate fully. Rather, as illustrated in the examples above, it acts as a predicate modifier (Butt & Geuder 2001). One consistent effect is that the event described (writing, making, singing or laughing in the examples above) must always be interpreted as having been completed (Singh 1991, Butt & Ramchand 2005). Other effects vary according to which light verb is used. For example, the light verb versions of ‘rise’ and ‘fall’ contribute a sense of spontaneity and inception of the event, while, the light verb version of ‘give’ can convey that the action is to benefit a third party.

All light verbs in these Aspectual Complex Predicates also always have a main verb use in the language. There is a set of about 24 verbs that can be used in this construction, with the most common being ‘take’, ‘give’, ‘go’, ‘fall’ and ‘hit’ (see Hook (1974) for a comprehensive overview).

3.2 Permissive complex predicates

Permissive complex predicates work slightly differently from aspectual CPs. As the name suggests, they are used to express the permission to perform an event. The light verb involved in these CPs is ‘give’ and it carries the tense/aspect inflections, as shown in (2).

- (2) a. anjum=ne saddaf=ko xat lk^h-ne di-ya
 Anjum.F=Erg Saddaf.F=Dat letter.M.Nom write-Inf.Obl give-Perf.M.Sg
 ‘Anjum let Saddaf write a letter.’

b. anjʊm=ne saddaf=ko tʃitt^hi lk^h-ne di
 Anjum.F=Erg Saddaf.F=Dat note.F.Nom write-Inf.Obl give-Perf.F.Sg
 ‘Anjum let Saddaf write a note.’

c. kɪs=ne kʊtte=ko g^har ke andar
 who.Obl=Erg dog.M.Obl=Dat house Gen.Obl inside
 a-ne di-ya?
 come-Inf.Obl give-Perf.M.Sg
 ‘Who let the dog come into the house?’

The permissive light verb in this case is combined with a verb that is in the infinitive. Generally, when verbs embed an infinitive verb, that verb inflects for number and gender. In the permissive, however, the infinitive is always in an invariant oblique form. Syntactic tests show that the permissive is not an instance of a control construction (Butt 1995, 2010). Instead the two verbs combine together to form a single predicate. This can be seen, among other tests, by the fact that the inflected light verb agrees with ‘note’ in (2b), showing that this argument is functioning as the syntactic object of ‘give’ and not just of ‘write’.

3.3 Motion CPs

A lesser known construction, but one that occurs fairly frequently, is that of complex predicates of motion, where two motion verbs combine in a verbal phrase to express a complex motion event. Superficially, these CPs are parallel to aspectual CPs in that at most two (motion) verbs are put in sequence, the first (main) motion verb in the base form, complemented by a finite (light) motion verb as the second. The constructions in (3) exemplify the phenomenon.

(3) a. sanḍ gayõ=ki revaṛ bar^h doṛ-a
 ox.M.Sg.Nom cow.F.Pl.Obl=Gen.F herd.M.Sg advance run-Perf.M.Sg
 ‘The ox charged into a herd of cows.’

b. sanḍ hamare makan=mẽ g^hʊs cal-a
 ox.M.Sg.Nom Pron.1.Pl.Obl.Gen house.M.Sg=Loc enter move-Perf.M.Sg
 ‘An ox got into our house.’

c. ʊs=ki ank^hõ=mẽ ansu a b^har-e
 Pron.3.Sg=Gen.F tear.F.Pl=Loc tear.M come advance-Perf.M.Pl
 ‘Tears welled up in her eyes.’

- d. $g^h\text{o}\text{r}\text{a}$ dor $b^h\text{a}\text{g}\text{-a}$
 horse.M.Sg.Nom run run-Perf.M.Sg
 ‘The horse ran away.’

In all of the above examples, the combination of two motion verbs yields a complex motion event, for instance combining $b\alpha\text{r}^h\text{-na}$ ‘to advance’ and $\text{d}\text{o}\text{r}\text{-na}$ ‘to run’ as in (3a) yields the bounded-path interpretation ‘running to’. The combinatorial possibilities are not restricted to directional motion verbs combining with manner of motion verbs as in (3a) and (3b), but constructions with both verbs denoting direction ($a\text{-na}$ ‘to come’ and $b^h\text{a}\text{r}\text{-na}$ ‘to advance’ in (3c)) or manner ($\text{d}\text{o}\text{r}\text{-na}$ ‘to run’ and $b^h\text{a}\text{g}\text{-na}$ ‘to run’ in (3d)) are also possible. Moreover, different valencies do not prevent motion verbs from combining, as shown in (3b) with the transitive verb $g^h\text{u}\text{s}\text{-na}$ ‘to enter’ and the intransitive $\text{c}\alpha\text{l}\text{-na}$ ‘to walk’, which combine to mean ‘walking/getting into’.

3.4 Dataset

The dataset underlying our investigation was created through a manual collection of CP constructions from Butt (1995) and Hautli-Janisz (2015), comprising 50 sentences that cover the three types of complex predicates described above. Table 1 gives an overview of the distributions of CP types and (un)grammaticality of the examples. Incorporating the ungrammatical sentences in the study allows for a more precise investigation into whether the LLM simply treats two adjacent verbs as a complex predicate, or whether there is a more abstract notion of complex predication visible in the responses. One instance of ungrammaticality is the incompatibility of main and light verb, for instance the main verb ($b\text{e}\text{t}^h\text{-na}$ ‘to sit’ cannot combine with the light verb $\text{p}\alpha\text{r}\text{-na}$ ‘to fall’).

Table 1: Distribution of complex predicate type and grammaticality in the dataset

	grammatical	ungrammatical	total
aspectual CPs	14	8	22
permissive CPs	6	3	9
motion CPs	17	2	19

The sentences are valid in both Urdu and Hindi, i.e. the same grammaticality and analysis hold in both Urdu and Hindi and can be prompted for as such. While no vocabulary or construction issues had to be accounted for, there might be minor differences in the Roman transcription, for instance the Roman Urdu *us*

nE ‘he Erg’ is transcribed as *usne* in Hindi Roman script. The dataset is available at <https://git.fim.uni-passau.de/NetteHJ/cps-urdu-hindi-llm-assessment>.

4 Probing the GPT models

For probing the LLMs, we formulate five different tasks that tap into syntactic and lexical semantic properties of complex predicates. Regarding syntax, we formulate three tasks, namely grammaticality judgments for grammatical and ungrammatical examples, part-of-speech (POS) tagging of the verbal complex and word order permutations (Urdu/Hindi is a free word-order language in that major constituents can scramble, but CPs exhibit some constraints on their distribution in the sentence). For prompting for the lexical semantic properties, we approximate GPT’s capabilities by querying for translations into and from English and for minimal pair generation in the source language (Urdu or Hindi). The latter task allows us to assess whether the representation of the meaning of complex predicates in GPT adheres to what we know of their linguistic properties (compositional where the light verbs contribute event structural modifications to the event denoted by the main verb). In the following we describe the set-up of the prompts, namely zero-shot and one-shot prompting (§4.1), and briefly discuss the general prompting parameters used in the study (§4.2).

4.1 Prompt engineering

The mode of communication with LLMs is natural language, i.e. tasks are formulated as queries in natural language, are processed by the model and are then responded to with human-like generated text. Prompt engineering is the process of conditioning a language model (Kojima et al. 2022) by identifying the set-up of the prompt that yields the highest performance. So far, there are only a small number of systematic studies on which kinds of prompts yield reliably good performance for individual tasks (e.g. Schulhoff et al. (2024)), meaning that the field is facing a trial-and-error scenario for tasks that lie outside of the spectrum of previous tasks. Another downside of this “Ask and you shall receive” mode of communication is that it gives rise to non-comparable results, because the formulation of the prompt has a direct effect on the quality of the generated responses. In addition, models change over time, leading to results that are difficult to replicate. None of these issues are alleviated in the present paper. However we try to be as comparable as possible to previous work in linguistic capability assessment by using the best practices described in OpenAI (2024b) in interacting with GPT-3.5 and GPT-4. We also probe the models regarding their performance in the

face of multilingual input, i.e. we keep the set-up of the prompt constant and only vary the language of the prompt.

4.1.1 Zero-shot prompting

In general, a prompt contains an instruction and an example that the instruction is supposed to be carried out on. With a *zero-shot prompt*, an LLM is given a task without any prior examples or context that could help in carrying out the task. An example of this is given in Figure 1, which shows the instructions and the example that constitutes the prompt *P* for minimal pair generation, i.e. the model is asked to generate a minimal pair for the example sentence in the prompt which contains a complex predicate. The instruction has three elements: (a) the persona that we ask the model to adopt (e.g. ‘You are a linguist specialising in the South Asian language Urdu/Hindi [...]’) – a strategy that is recommended in OpenAI’s engineering recommendations; (b) background information on CPs (‘A frequent linguistic phenomenon are complex predicates, which according to Miriam Butt (1995) are formed when two or more predicational elements enter into a relationship...’); (c) the details regarding the task (‘Provide a sentence which is similar in meaning and also contains a complex predicate [...]’). We then add the example sentence containing the CP in triple single quotes.

P: You are a linguist specialising in the South Asian language Urdu. A frequent linguistic phenomenon are complex predicates, which according to Miriam Butt (1995) are formed when two or more predicational elements enter into a relationship of co-predication. Each predicational element adds arguments to a monoclausal predication. Provide a sentence that is similar to the sentence in triple single quotes ‘‘‘anjum nE xat likh liyA’’’ that also uses a complex predicate.

R: ‘‘‘Anjum ne kitaab parh kar samajh liya’’’ In this sentence, the complex predicate ‘‘parh kar samajh liya’’ is formed by the two predicational elements ‘‘parh kar’’ (reading) and ‘‘samajh liya’’ (understood), where the subject Anjum is the agent of both actions.

Figure 1: Zero-shot prompt *P* for minimal pair generation and response *R* by GPT-3.5.

The response *R* in Figure 1 is partly incorrect: Only *samajh liya* (lit. ‘understand take’) is a complex predicate of the aspectual type, *parH kar* (lit. ‘read having’) is

formed with the conjunctive participle *kar* and represents a subordinated clause. The information that Anjum is the agent of both events is correct, but is not relevant to the task of the prompt. For the evaluation of the performance of the model, this response is considered incorrect.

Given our multilingual setting and our aim to compare the performance of the models with respect to different language input, prompt *P* in Figure 1 contains English instructions and Romanized Urdu for the example, based on the transliteration guidelines of the Urdu ParGram project.² This prompt is translated into Urdu (in the Perseo-Arabic script used to write Urdu), Hindi (in Devanagari script) and Hindi (Roman script). The translations and transliterations have been done manually by two native speakers, who have consulted grammar books to translate terminology like ‘monoclausal’ and ‘complex predicate’.³ The different language variants of the prompt in Figure 1 are given in Appendix A.

For the four other tasks, namely translation to English, POS tagging of the verbal phrase, grammaticality judgments and word order permutations, we reduce the instruction of prompt *P* to only contain (a) the persona and (c) the details of the task, not including (b) background information on complex predicates. For the word order permutation task, Figure 2 illustrates prompt *P* and response *R*: Here, response *R* is correct – the reordered words obey the constraints put forth by both Urdu and Hindi. See Appendix B for the zero-shot prompts for grammaticality judgment and POS tagging. Again, this prompt, as all other prompts, is translated and transliterated into the four language variants (Urdu Arabic, Urdu Roman, Hindi Devanagari, Hindi Roman).

4.1.2 One-shot prompting

One-shot prompting as introduced by Brown et al. (2020) is a technique that is generally considered to perform better than zero-shot prompting. The one-shot prompt contains an exemplary analysis for the task that is expected to be conducted by the model. Prompt *P* in Figure 3 shows the structure of the one-shot prompt for minimal pair generation, an extension of the zero-shot prompt in Figure 1. In addition to the instructions of the zero-shot prompt, the one-shot prompt includes an example sentence where we explicitly identify the complex predicate. The assumption is that this additional information will make it easier for the model to generate a minimal pair example.

²https://ling.sprachwiss.uni-konstanz.de/pages/home/pargram_urdu/

³See, for example <https://archive.org/details/page1-converted-compressed/page/392/mode/2up> for Hindi.

P: You are a linguist specialising in the South Asian language Urdu. Your task is to change the word order in the sentence in triple single quotes ```anjum nE likHnE dI saddaf k0 ciTHI''' so that the sentence is still grammatical.

R: The sentence ```anjum nE likHnE dI saddaf k0 ciTHI''' can be reordered while maintaining grammatical correctness as follows: ```saddaf k0 anjum nE ciTHI likHnE dI'''.

Figure 2: Zero-shot prompt *P* for word order permutation and response *R* by GPT-4.

In this case, response *R* is correct: The verb *likHvAyA* ‘make write’ is an indirect causative which also conveys the meaning of making somebody do something. Even though the verb has a causative morpheme and no separate light verb, the morphological causative in Urdu/Hindi has been shown to function as a complex predicate with respect to argument structure composition (Butt 1998, Butt & King 2006, Butt et al. 2009). However Figure 3 also illustrates one of the problems of LLMs: the structure of the responses can be flawed, because it stops mid-sentence. It is unclear what triggers this, because reprompting yields the full response. For evaluation, we treat response *R* as correct, because the information it contains so far is valid.

4.2 General prompting parameters

Prompt engineering not only includes setting up the queries/prompts for the model, but also requires attention to several other, more technical, parameters. These parameters control the length of the response, its randomness and the frequency penalty that prevents repetitive strings of words. We restrict the length of the generated responses to 128 tokens. The temperature, the parameter that controls the randomness or creativity of the generated text, was set to zero, increasing the chance of obtaining more deterministic answers, i.e. the same input yields the same response from the model upon multiple re-promptings. To further help with the determinism of the generated answers, we also set the seed parameter to the same random number.

P: You are a linguist specialising in the South Asian language Urdu. A frequent linguistic phenomenon are complex predicates, which according to Miriam Butt (1995) are formed when two or more predicational elements enter into a relationship of co-predication. Each predicational element adds arguments to a monoclausal predication. In the following sentence in triple single quotes, ```anjum nE saddaf k0 ciTHI likHnE dI''`, the complex predicate in triple single quotes is ```likHnE dI''`. Provide a sentence which is similar in meaning and also contains a complex predicate.

R: `anjum nE saddaf k0 ciTHI likHnE dI` ```In the sentence you provided "anjum nE saddaf k0 ciTHI likHnE dI" which translates to "Anjum allowed Saddam to write the letter" the complex predicate is "likHnE dI" (allowed to write). A similar sentence with a complex predicate that conveys a similar meaning would be: "anjum nE saddaf k0 kHaT likHwAyA."` This sentence translates to "Anjum made Saddam write the letter" where `likHwAyA` serves as the complex predicate indicating the causative action of making someone

Figure 3: One-shot prompt *P* for minimal pair generation and response *R* by GPT-4.

5 Results

In this section we report on the results obtained from prompting GPT-3.5 and GPT-4 along the two axes of setting up the prompt: first, by varying the language of the prompt and second, by querying for different aspects of complex predication. Both the GPT-3.5 and GPT-4 models have a tendency to generate lengthy answers, at times irrelevant to the task, or get ‘lazy’ and not finish the answer. The ‘laziness’ issue was addressed by using the **gpt-4-0125-preview** model, developed by OpenAI to reduce the instances in which the model does not complete a task. Even though we do not task the model to provide us with an explanation of the judgment in the present study, in many cases it is nevertheless provided. We therefore accept the response as is, but the evaluation is strict in that both the explanation and the provided example have to be correct in order for the response to be judged as correct. That way we provide a realistic assessment of the GPT capabilities, because ultimately we want to establish how reliable the models are without additional human interference (such as needing to determine the

correct and incorrect portions of the GPT response). The main metric we use to report performance is accuracy, i.e. the fraction of correct predictions made by GPT-3.5 and GPT-4 in comparison to the gold standard we assembled from the literature. We report F1 scores only for the grammaticality judgments where we request a specific label (True vs. False).

The remainder of this section is structured as follows: We first present the results for the tasks associated with the syntactic properties of complex predicates, namely grammaticality judgments, POS tagging and word order permutation (§5.1) and then look at the results of the lexical semantic analysis, namely translation and minimal pair generation (§5.2). We conclude the section with an error analysis (§5.3).

5.1 Assessing syntactic properties of CPs

Grammaticality judgments For getting the grammaticality judgments we use zero-shot prompts that request a binary decision from the model, namely ‘True’ (sentence is grammatical) and ‘False’ (sentence is ungrammatical) (see the prompt in Figure 8 in Appendix B). We then match the labels returned by the models against the gold standard.

The results in Table 2 show that prompting with English instructions and assigning GPT the role of an expert for Urdu yields consistently and significantly worse results than if we prompt using Urdu Arabic instructions and examples (‘Urdu Arab’). The difference is significant, e.g. illustrated by the F1-score of 0.0 for the English and 0.82 for the Urdu prompt in GPT-3.5 in Table 2. Moreover, the results show that the performance difference between GPT-3.5 and GPT-4 is largest when prompting in English (F1-score of 0.0 for GPT-3.5 vs. 0.65 for GPT-4), compared to prompting in Urdu where the F1-score is much higher and less different (F1-score of 0.82 for GPT-3.5 vs. 0.83 for GPT-4).

Table 2: Performance regarding grammaticality judgments across models, assigning GPT the role of an Urdu language expert.

	GPT-3.5				GPT-4			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
English	0.26	0.0	0.0	0.0	0.56	0.8	0.54	0.65
Urdu Arab	0.7	0.72	0.95	0.82	0.76	0.86	0.81	0.83

If we state in the zero-shot prompt that the model is assumed to take on the role of a language expert in Hindi and translate the prompt into Hindi written in Devanagari (‘Hindi Dev’) and its Romanized transliteration (‘Hindi Rom’),

the drop in performance is striking, especially that of GPT-3.5, where the translated/transliterated prompts seem to have a crippling effect on the performance, see Table 3. We can therefore only partly confirm previous findings that show that using the language under investigation as the prompt language increases performance (Cao et al. 2023).

Table 3: Performance regarding grammaticality judgments across models, assigning GPT the role of a Hindi language expert.

	GPT-3.5				GPT-4			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
English	0.72	0.83	0.78	0.81	0.7	0.76	0.84	0.81
Hindi Dev	0.48	0.66	0.48	0.56	0.58	0.8	0.58	0.6
Hind Rom	0.12	0.56	0.12	0.2	0.7	0.74	0.7	0.71

Interestingly, the significantly better performance of the Urdu prompts in comparison to the Hindi prompts goes against previous work on multilingual LLM assessment, where Hindi outperforms Urdu in almost all tasks (Lai et al. 2023). Moreover, given the significant difference in size between the two models (175 billion parameters for GPT-3.5 vs. 1 trillion parameters for GPT-4), it is surprising that the quality in the responses of GPT-4 is not consistently better than the responses of GPT-3.5 across languages.

Part-of-speech tagging For part-of-speech tagging, we use the zero-shot prompt in Figure 11 in Appendix B that requests the identification of all verbs in the sentence. Again we vary the role of the model as either an Urdu or Hindi language expert. The response is only judged as being correct if both verbal elements in the complex predicate are tagged as such.

The results in Table 4 shows that overall performance for this task is higher than for the grammaticality judgment. Again GPT-4 outperforms GPT-3.5 across all prompt languages. However, prompting in English with GPT-4 is in this case the most successful strategy, contrary to what we find in the grammaticality task above (Urdu scored higher) and previous work in the literature (prompting in the language under investigation improves performance).

Word order permutation For word order permutation, we use the zero-shot prompt in Figure 2 and request a change in word order in the sentence containing the complex predicates while retaining grammaticality. As can be seen in Table 5, prompting in Urdu instead of English (assigning the role as an Urdu expert) does not yield substantial improvement, especially not when using GPT-4. However,

Table 4: Accuracy scores for POS tagging assigning the models the role of an Urdu language expert ('English', 'Urdu Arab') and Hindi ('English', 'Hindi Dev' and 'Hindi Rom').

	GPT-3.5	GPT-4
	Acc	Acc
English	0.66	0.96
Urdu Arab	0.72	0.76
English	0.9	0.94
Hindi Dev	0.7	0.86
Hindi Rom	0.48	0.9

GPT-4 again performs significantly better than GPT-3.5. For Hindi, the situation is slightly different. Interestingly, prompting in English (assigning the role as a Hindi expert) with the Roman Hindi example results in the best performance. Translating the prompt into Hindi, either in Roman script or in Devanagari, does not yield better results. However, when comparing the results for Urdu and Hindi, the Hindi prompts yield a much higher performance than prompting in Urdu. The gap in performance between GPT-3.5 and GPT-4 persists.

Table 5: Accuracy scores for grammaticality judgments across models and prompt languages, assigning GPT the role of an Urdu language expert (upper two rows) and Hindi (lower three rows).

	GPT-3.5	GPT-4
	Acc	Acc
English	0.16	0.56
Urdu Arab	0.38	0.54
English	0.64	0.94
Hindi Dev	0.42	0.84
Hindi Rom	0.56	0.88

Overall, the only consistent behavior in the syntactic investigation is that GPT-4 outperforms GPT-3.5 in accuracy. The other aspects of the evaluation yield varied model behavior, in particular concerning whether the prompting language affects the quality of the results, whether the information in the prompt has consistent impact and whether the script of the input in Hindi has consequences for the quality. This is confirmed in the next section where we investigate whether GPT can capture lexical semantic properties of CPs.

5.2 Assessing lexical semantic properties of CPs

Translation This task evaluates the ability of the models to consider the CPs as a unit in translation, with Urdu/Hindi as the target and English as the source and vice versa. For aspectual and motion verb CPs this means that the main and light verbs do not get translated individually and literally into English, but for instance into a main verb and an adverb. From English into Urdu/Hindi, we assume the opposite: tensed verbs and adverbial clauses are translated into complex predicates. For permissive CPs with the light verb *dena* ‘to give’, the translation has to contain the permissive construct ‘let somebody do something’. The zero-shot prompt assigns the model the role of an Urdu or Hindi native speaker (Tables 6 and 7, respectively) and requests a translation of a source sentence into the target language, adding the instruction that the response should be in Roman letters throughout.

While there are a number of metrics in MT to check the appropriateness of a translation, e.g. the BLEU score, our focus lies on making sure that the complex predicate is appropriately translated. This means that the event structure contribution of the light verb must be captured on the linguistic surface of the target. Tables 6 and 7 show the results for the Urdu-English pairs and the Hindi-English pairs, respectively. Again, the only commonality between all settings is that GPT-4 outperforms GPT-3.5. Translations between English and Hindi are judged more appropriate than translations between Urdu and English. Interestingly, if English is used as the source and Hindi as the target, the translations are more appropriate, irrespective of whether the prompt was written in English, Hindi Devanagari or Hindi Roman script.

Table 6: Accuracy scores for the translation task across models and prompt languages (English vs. Urdu), assigning the model the role of an Urdu language expert.

	GPT-3.5	GPT-4	GPT-3.5	GPT-4
	English		Urdu	
	Acc	Acc	Acc	Acc
English-Urdu	0.48	0.70	0.52	0.82
Urdu-English	0.46	0.54	0.44	0.74

Minimal pair generation The generation of the minimal pairs of sentences, i.e. sentences that are similar in meaning to the prompted sentence and also contain a complex predicate, is the only task where we compare not only across prompt

Table 7: Accuracy scores for the translation task across models and prompt languages (‘English’ vs. ‘Hindi Dev’ vs. ‘Hindi Rom’), assigning the model the role of a Hindi language expert.

	GPT-3.5	GPT-4	GPT-3.5	GPT-4	GPT-3.5	GPT-4
	English		Hindi Dev		Hindi Rom	
	Acc	Acc	Acc	Acc	Acc	Acc
English-Hindi	0.70	0.98	0.74	0.94	0.70	0.98
Hindi-English	0.74	0.84	0.72	0.92	0.80	0.88

language and models, but also across prompt settings, namely zero-shot vs. one-shot prompting. The hypothesis is that one-shot prompting improves response quality in comparison to zero-shot prompts.

Here we only focus on the results for Hindi, provided in Table 8. The performance of GPT-3.5 and GPT-4 is surprisingly high, across both prompt settings and input languages. Only for Hindi in Roman script does the performance drop, otherwise the performance is excellent. This is partly in line with previous work which has shown that one-shot prompting is more reliable than zero-shot prompting (Brown et al. 2020).

Table 8: Accuracy scores for minimal pairs of complex predicates across models and prompt languages (‘English’ vs. ‘Hindi Dev’ vs. ‘Hindi Rom’), assigning the model the role of a Hindi language expert.

	GPT-3.5	GPT-4	GPT-3.5	GPT-4	GPT-3.5	GPT-4
	English		Hindi Dev		Hindi Rom	
Zero-shot	0.67	0.34	1	1	0.67	0.67
One-shot	0.34	1	0.67	0.67	1	1

In sum, probing for the lexical semantic properties of complex predicates by asking for translations and minimal pairs from GPT yields more reliable judgments than when probing the models for syntactic patterns. In the next section we provide a number of instances where judgments were incorrect, either in terms of label assigned or the explanation given.

5.3 Error analysis

The design of the prompts provides minimal opportunity for filler content in the response in relation to the tasks, i.e. the model responses can usually be unambiguously categorized as correct or incorrect. However, issues arise in some cases,

for instance in the prompt response in Figure 4 – the label provided is correct, but the model adds to the response text that is supposed to explain the decision, but the reason as to why the sentence is ungrammatical is not captured, i.e. just transliterating the input sentence in Urdu script does not make a difference in its grammaticality.

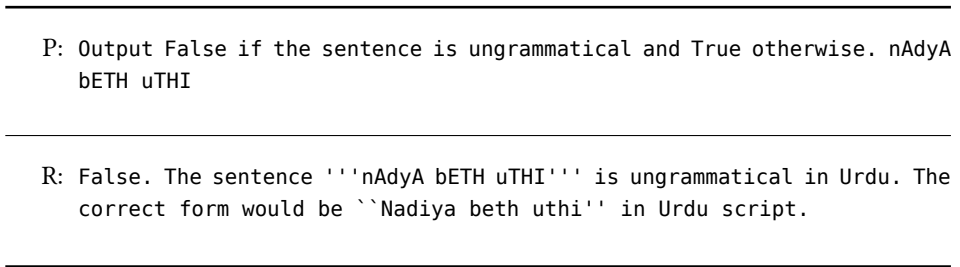


Figure 4: Zero-shot prompt P for grammaticality and response R by GPT-4.

For other responses the pattern is reversed, i.e. the model provides the incorrect label for the sentence under investigation but generates an explanation that is correct. This inconsistency is exemplified in Figure 5, where the generated grammatical sentence is the same that it was prompted with, which was judged as being ungrammatical.

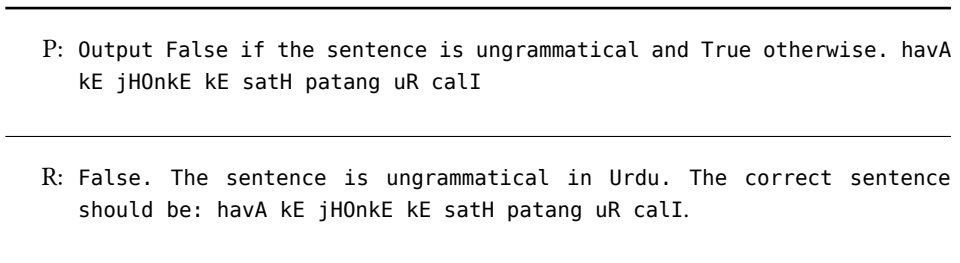


Figure 5: Zero-shot prompt P for grammaticality and response R by GPT-4.

For grammaticality judgments in Hindi, the performance of GPT-3.5 significantly declines when the prompt is transliterated to Hindi in Roman script. In fact, 84% of the output is incoherent, meaning that either the label or the explanation is incorrect, with the consequence that these responses are all treated as incorrect. An example is given in Figure 6 where the response is translated as ‘True: Anjum let Saddaf leave – this sentence is grammatical. False: Anjum

let Saddam leave — this sentence is ungrammatical’. This response is treated as incorrect in our evaluation.

P: Output False if the sentence is ungrammatical and True otherwise. Anjum ne Saddam ko jaane diya

R: saty: Anjum ne Saddam ko jaane diya - yah vaaky vyaakaranik hai. asaty: Anjum ne Saddam ko jaane diya - yah vaaky vyaakaranik nahi hai.

Figure 6: Zero-shot prompt P for grammaticality and response R by GPT-4.

6 Discussion and conclusion

In sum, this paper shows that drawing general conclusions as to the capabilities of LLMs with respect to under-resourced languages is tricky: Depending on the task and the language of the prompt, the quality of the responses varies. Our study overall mirrors results from previous work in that it is difficult to observe consistent behavior. The field has yet to find ways of systematically studying LLMs, but first approaches are being made, for instance Schulhoff et al. (2024), who offer the first systematic study of prompting techniques. However, given the multitude of tasks, benchmarks and languages, there is significant work yet to be done.

In this paper, we focused on V-V complex predicates in Urdu and Hindi, a complex linguistic phenomenon that is comparatively well understood from a theoretical perspective and for which we can rely on previous computational work. The novelty of this paper lies in exploring the capacity of the new generation of Large Language Models (LLMs) regarding their capabilities of detecting and analyzing the structure of complex predicates. For future work, it would be interesting to examine the LLM’s knowledge of constructions that have not been written on as extensively as complex predicates in Urdu/Hindi but are nevertheless found in the language.

Acknowledgments

This paper deals with intricacies of lexical semantics that have repercussions for morphology and syntax, an area that has fascinated Dick Crouch. Without

his continual work on integrating semantics and syntax to make intelligent NLP viable, we would not be where we are today.

Appendix A: Language variation for prompting

Zero-shot, minimal pair generation, Urdu Arabic

P: اپ ایک لسانی ہیں جو اپ ایک لسانی ہیں جو جنوبی ایشیائی زبان اردو میں مہارت کر رہے ہیں۔ اکثر لسانی رحمان پشیدہ پیش گوئیاں ہیں جو مریم بٹ (1995) کے مطابق اس وقت بنتی ہیں جب دو یا دو سے زیادہ پیش گوئی والے عناصر مشترکہ پیش گوئی کے رشتے میں داخل ہوتے ہیں۔ ہر پیش گوئی کا عنصر یک ایسا جملہ فراہم کریں جو ٹریل سنگل کوٹس میں جملے سے ملتا جلتا ہو "انجم نے خط لکھ ل" اور ایک پیچیدہ پیش گوئی بھی استعمال کریں۔ مونو کلوزل پیشنگوئی میں دلائل کا اضافہ کرتا ہے۔

Figure 7: One-shot prompt for the generation of a sentence with a complex predicate and the GPT-4 response. Both the prompt instructions and the input sentence are in Urdu Arabic script.

Zero-shot, minimal pair generation, Hindi Devanagari

P: आप दक्षिण एशियाई भाषा हिंदी में विशेषज्ञता रखने वाले भाषाविद् हैं। एक लगातार भाषाई घटना जटिल विधेय हैं, जो मिरियम बट (1995) के अनुसार तब बनते हैं जब दो या अधिक विधेय तत्व सह-विधेय के रिश्ते में प्रवेश करते हैं। प्रत्येक विधेय तत्व एक मोनोक्लॉसल विधेय में तर्क जोड़ता है। एक वाक्य प्रदान करें जो तीन एकल उद्धरण चिह्नों 'अंजुम ने खत लिख लिया' में वाक्य के समान हो और एक जटिल विधेय का भी उपयोग करें।

Figure 8: Zero-shot prompt *P* for grammaticality judgment in Hindi Devanagari by GPT-4.

Zero-shot, minimal pair generation, Hindi Roman

P: aap dakshin eshiyaae bhaasha hindee mein visheshagyata rakhane vaale bhaashaavid hain. ek lagaataar bhaashaee ghatana jatil vidhey hain, jo miriyam bat (1995) ke anusaar tab banate hain jab do ya adhik vidhey tatv sah-vidhey ke rishte mein pravesh karate hain. pratyek vidhey tatv ek monoklosal vidhey mein tark jodata hai. ek vaaky pradaan karen jo teen ekal uddharan chihnon '''anjum ne khat likh liya''' mein vaaky ke samaan ho aur ek jatil vidhey ka bhee upayog karen.'

Figure 9: Zero-shot prompt P for grammaticality judgment in Hindi Roman by GPT-4.

Appendix B: Remaining prompts

Zero-shot, grammaticality judgment, English

P: You are a linguist specialising in the South Asian language Urdu. Your task is to check whether the sentence in triple single quotes '''gH0rA doR bHAgA''' is grammatical or not. Respond with True for grammatical and False for ungrammatical.

R: False

Figure 10: Zero-shot prompt P for grammaticality judgment response R by GPT-4.

Zero-shot, POS tagging, English

P: You are a linguist specialising in the South Asian language Urdu. Your task is to identify the verbs in the sentence in triple single quotes ``chor mukaan se bahar kud nikla'''.

R: In the sentence ``chor mukaan se bahar kud nikla'''', the verbs are:

1. kud (jumped)
2. nikla (came out)

Figure 11: Zero-shot prompt *P* for POS tagging and response *R* by GPT-4.

References

- Achiam, Josh, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altmenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. *GPT-4 technical report*. <https://arxiv.org/abs/2303.08774>.
- Ahmed, Tafseer & Miriam Butt. 2011. Discovering semantic classes for Urdu N-V complex predicates. In Johan Bos & Stephen Pulman (eds.), *Proceedings of the ninth international conference on computational semantics (IWCS 2011)*. <https://aclanthology.org/W11-0132>.
- Ahmed, Tafseer, Miriam Butt, Annette Hautli & Sebastian Sulger. 2012. A reference dependency bank for analyzing complex predicates. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the eighth international conference on language resources and evaluation (LREC'12)*, 3145–3152. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/pdf/474_Paper.pdf.
- Anwar, Maaz, Riyaz Ahmad Bhat, Dipti Sharma, Ashwini Vaidya, Martha Palmer & Tafseer Ahmed Khan. 2016. A proposition bank of Urdu. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the tenth international conference on language resources and evaluation (LREC'16)*, 2379–2386. European Language Resources Association (ELRA). <https://aclanthology.org/L16-1377>.

- Baldwin, Timothy & Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkha & Fred J. Damerau (eds.), *Handbook of natural language processing (2nd ed.)* Chapman & Hall/CRC. DOI: <https://doi.org/10.1201/9781420085938>.
- Barančíková, Petra & Václava Kettnerová. 2017. ParaDi: Dictionary of paraphrases of Czech complex predicates with light verbs. In Stella Markantonatou, Carlos Ramisch, Agata Savary & Veronika Vincze (eds.), *Proceedings of the 13th workshop on multiword expressions (MWE 2017)*, 1–10. Association for Computational Linguistics. DOI: 10.18653/v1/W17-1701. <https://aclanthology.org/W17-1701>.
- Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever & Dario Amodei. 2020. *Language models are few-shot learners*. <https://arxiv.org/abs/2005.14165>.
- Butt, Miriam. 1995. *Complex predicates in Urdu*. CSLI Publications.
- Butt, Miriam. 1998. Constraining argument merger through aspect. In Erhard Hinrichs, Andreas Kathol & Tsuneko Nakazawa (eds.), *Syntax and semantics, complex predicates in nonderivational syntax*, vol. 30, 73–113. Academic.
- Butt, Miriam. 2010. The light verb jungle: Still hacking away. In Mengistu Amberber, Brett Baker & Mark Harvey (eds.), *Complex predicates: Cross-linguistic perspectives on event structure*, 48–78. Cambridge University Press.
- Butt, Miriam, Tina Bögel, Annette Hautli, Sebastian Sulger & Tafseer Ahmed. 2012. Identifying Urdu complex predication via bigram extraction. In Martin Kay & Christian Boitet (eds.), *Proceedings of COLING 2012*, 409–424. The COLING 2012 Organizing Committee. <https://aclanthology.org/C12-1026>.
- Butt, Miriam & Wilhelm Geuder. 2001. On the (semi)lexical status of light verbs. In Norbert Corver & Henk van Riemsdijk (eds.), *Semi-lexical categories: On the content of function words and the function of content words*, 323–370. Mouton de Gruyter.
- Butt, Miriam & Tracy Holloway King. 2006. Restriction for morphological valency alternations: The Urdu causative. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*, 235–258. CSLI Publications.
- Butt, Miriam, Tracy Holloway King & Gillian Catriona Ramchand. 2009. Complex predication: Who made the child pinch the elephant? In Linda Uyechi &

- Lian Hee Wee (eds.), *Reality exploration and discovery: Pattern interaction in language and life*, 231–256. CSLI Publications.
- Butt, Miriam & Gillian Catriona Ramchand. 2005. Complex aspectual structure in Hindi/Urdu. In Nomi Ertischik-Shir & Tova Rapoport (eds.), *The syntax of aspect*, 117–153. Oxford University Press.
- Cai, Zhenguang G., Xufeng Duan, David A. Haslett, Shuqi Wang & Martin J. Pickering. 2024. *Do large language models resemble humans in language use?* <https://arxiv.org/abs/2303.08014>.
- Cao, Yong, Li Zhou, Seolhwa Lee, Laura Cabello, Min Chen & Daniel Hershcovich. 2023. Assessing cross-cultural alignment between ChatGPT and human societies: An empirical study. In Sunipa Dev, Vinodkumar Prabhakaran, David Adelani, Dirk Hovy & Luciana Benotti (eds.), *Proceedings of the first workshop on cross-cultural considerations in NLP (C3NLP)*, 53–67. Association for Computational Linguistics. DOI: 10.18653/v1/2023.c3nlp-1.7. <https://aclanthology.org/2023.c3nlp-1.7>.
- Constant, Mathieu, Gülşen Eryiğit, Johanna Monti, Lonneke van der Plas, Carlos Ramisch, Michael Rosner & Amalia Todirascu. 2017. Multiword expression processing: A survey. *Computational Linguistics* 43(4). 837–892. <https://aclanthology.org/J17-4005>.
- Dentella, Vittoria, Elliot Murphy, Gary Marcus & Evelina Leivada. 2023. *Testing AI performance on less frequent aspects of language reveals insensitivity to underlying meaning*. <https://arxiv.org/abs/2302.12313>.
- Hautli-Janisz, Annette. 2013. Moving right along: Motion verb sequences in Urdu. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG13 conference*, 295–215. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2013/lfg13.html>.
- Hautli-Janisz, Annette. 2015. Urdu/Hindi complex predicates of motion and the manner/result complementarity. *Journal for South Asian Languages* 7.
- Hautli-Janisz, Annette & Erisa Bytyqi. 2024. The facts speak for themselves: Fallacy detection with GPT. In *Proceedings of OSSA13*.
- Hautli-Janisz, Annette, Tracy Holloway King & Gilian Catriona Ramchand. 2015. Encoding event structure in Urdu/Hindi VerbNet. In Eduard Hovy, Teruko Mitamura & Martha Palmer (eds.), *Proceedings of the 3rd workshop on EVENTS: Definition, detection, coreference, and representation*, 25–33. Association for Computational Linguistics. DOI: 10.3115/v1/W15-0804. <https://aclanthology.org/W15-0804>.
- Hendrickx, Iris, Amália Mendes, Sílvia Pereira, Anabela Gonçalves & Inês Duarte. 2010. Complex predicates annotation in a corpus of Portuguese. In Nianwen Xue & Massimo Poesio (eds.), *Proceedings of the fourth linguistic anno-*

- tation workshop*, 100–108. Association for Computational Linguistics. <https://aclanthology.org/W10-1812>.
- Herbold, Steffen, Annette Hautli-Janisz, Ute Heuer, Zlata Kikteva & Alexander Trautsch. 2023. A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports* 13(1). DOI: 10.1038/s41598-023-45644-9. <http://dx.doi.org/10.1038/s41598-023-45644-9>.
- Hook, Peter Edwin. 1974. *The compound verb in Hindi*. Center for South and Southeast Asian Studies. The University of Michigan.
- Hwang, Jena D., Archana Bhatia, Claire Bonial, Aous Mansouri, Ashwini Vaidya, Nianwen Xue & Martha Palmer. 2010. PropBank annotation of multilingual light verb constructions. In Nianwen Xue & Massimo Poesio (eds.), *Proceedings of the fourth linguistic annotation workshop*, 82–90. Association for Computational Linguistics. <https://aclanthology.org/W10-1810>.
- Jagfeld, Glorianna & Lonneke van der Plas. 2015. Towards a better semantic role labeling of complex predicates. In Diana Inkpen, Smaranda Muresan, Shibamouli Lahiri, Karen Mazidi & Alisa Zhila (eds.), *Proceedings of the 2015 conference of the North American chapter of the Association for Computational Linguistics (NAACL): student research workshop*, 33–39. Association for Computational Linguistics. DOI: 10.3115/v1/N15-2005. <https://aclanthology.org/N15-2005>.
- Kocoń, Jan, Igor Cichecki, Oliwier Kaszyca, Mateusz Kochanek, Dominika Szydło, Joanna Baran, Julita Bielaniewicz, Marcin Gruza, Arkadiusz Janz, Kamil Kancierz, Anna Kocoń, Bartłomiej Koptyra, Wiktoria Mieszczenko-Kowszewicz, Piotr Miłkowski, Marcin Oleksy, Maciej Piasecki, Łukasz Radliński, Konrad Wojtasik, Stanisław Woźniak & Przemysław Kazienko. 2023. ChatGPT: Jack of all trades, master of none. *Information Fusion* 99. DOI: <https://doi.org/10.1016/j.inffus.2023.101861>. <https://www.sciencedirect.com/science/article/pii/S156625352300177X>.
- Kojima, Takeshi, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo & Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. In *Advances in neural information processing systems*, vol. 35, 22199–22213.
- Lai, Viet, Nghia Ngo, Amir Pouran Ben Veyseh, Hieu Man, Franck Dernoncourt, Trung Bui & Thien Nguyen. 2023. ChatGPT beyond English: Towards a comprehensive evaluation of large language models in multilingual learning. In Houda Bouamor, Juan Pino & Kalika Bali (eds.), *Findings of the Association for Computational Linguistics: EMNLP 2023*, 13171–13189. Association for Computational Linguistics. DOI: 10.18653/v1/2023.findings-emnlp.878. <https://aclanthology.org/2023.findings-emnlp.878>.

- Mukerjee, Amitabha, Ankit Soni & Achla M. Raina. 2006. Detecting complex predicates in Hindi using POS projection across parallel corpora. In Begoña Villada Moirón, Aline Villavicencio, Diana McCarthy, Stefan Evert & Suzanne Stevenson (eds.), *Proceedings of the workshop on multiword expressions: Identifying and exploiting underlying properties*, 28–35. Association for Computational Linguistics. <https://aclanthology.org/W06-1205>.
- Open AI. 2024a. OpenAI models. <https://platform.openai.com/docs/models/gpt-3-5-turbo>.
- OpenAI. 2024b. *Prompt engineering*. Tech. rep. OpenAI. <https://platform.openai.com/docs/guides/prompt-engineering/prompt-engineering>.
- Pal, Santanu, Tanmoy Chakraborty & Sivaji Bandyopadhyay. 2011. Handling multiword expressions in phrase-based statistical machine translation. In *Proceedings of machine translation summit XIII: Papers*. <https://aclanthology.org/2011.mtsummit-papers.23>.
- Riemenschneider, Frederick & Anette Frank. 2023. Exploring large language models for classical philology. In Anna Rogers, Jordan Boyd-Graber & Naoaki Okazaki (eds.), *Proceedings of the 61st annual meeting of the Association for Computational Linguistics (ACL) (volume 1: long papers)*, 15181–15199. Association for Computational Linguistics. DOI: 10.18653/v1/2023.acl-long.846. <https://aclanthology.org/2023.acl-long.846>.
- Sanches Duran, Magali, Carlos Ramisch, Sandra Maria Aluísio & Aline Villavicencio. 2011. Identifying and analyzing Brazilian Portuguese complex predicates. In Valia Kordoni, Carlos Ramisch & Aline Villavicencio (eds.), *Proceedings of the workshop on multiword expressions: From parsing and generation to the real world*, 74–82. Association for Computational Linguistics. <https://aclanthology.org/W11-0812>.
- Schulhoff, Sander, Michael Ilie, Nishant Balepur, Konstantine Kahadze, Amanda Liu, Chenglei Si, Yinheng Li, Aayush Gupta, HyoJung Han, Sevien Schulhoff, Pranav Sandeep Dulepet, Saurav Vidyadhara, Dayeon Ki, Sweta Agrawal, Chau Pham, Gerson Kroiz, Feileen Li, Hudson Tao, Ashay Srivastava, Hevander Da Costa, Saloni Gupta, Megan L. Rogers, Inna Goncareenco, Giuseppe Sarli, Igor Galynker, Denis Peskoff, Marine Carpuat, Jules White, Shyamal Anadkat, Alexander Hoyle & Philip Resnik. 2024. *The prompt report: A systematic survey of prompting techniques*. <https://arxiv.org/abs/2406.06608>.
- Shi, Freda, Mirac Suzgun, Markus Freitag, Xuezhi Wang, Suraj Srivats, Soroush Vosoughi, Hyung Won Chung, Yi Tay, Sebastian Ruder & Denny Zhou. 2022. *Language models are multilingual chain-of-thought reasoners*. <https://arxiv.org/abs/2210.03057>.

11 *Revisiting the Light Verb Jungle: How Good Is GPT at Hacking Away?*

- Singh, Mona. 1991. The perfective paradox: Or how to eat your cake and have it too. In *Proceedings of the seventeenth annual meeting of the Berkeley Linguistics Society (BLS)*, 469–479.
- Sinha, R. Mahesh K. 2009. Mining complex predicates in Hindi using a parallel Hindi-English corpus. In Dimitra Anastasiou, Chikara Hashimoto, Preslav Nakov & Su Nam Kim (eds.), *Proceedings of the workshop on multiword expressions: Identification, interpretation, disambiguation and applications (MWE 2009)*, 40–46. Association for Computational Linguistics. <https://aclanthology.org/W09-2906>.
- Sulger, Sebastian & Ashwini Vaidya. 2014. Towards identifying Hindi/Urdu noun templates in support of a large-scale LFG grammar. In Christian Boitet & M.G. Abbas Malik (eds.), *Proceedings of the fifth workshop on south and south-east Asian natural language processing*, 1–10. Association for Computational Linguistics & Dublin City University. DOI: 10.3115/v1/W14-5501. <https://aclanthology.org/W14-5501>.
- Vaidya, Ashwini, Owen Rambow & Martha Palmer. 2019. Syntactic composition and selectional preferences in Hindi light verb constructions. *Linguistic Issues in Language Technology* 17(1). 1–30.

Chapter 12

Poking at GPT's NL Generalizations

Annie Zaenen^a, Alexis Cooper^b, Felix Zheng^b, Susan W. Brown^b & Martha Palmer^b

^aStanford University ^bUniversity of Colorado, Boulder

In this contribution we look at how far large language models (LLMs), and more specifically GPT-4 (Open AI 2023), can be trusted to make the right inferences when these inferences depend on lexical information of the type that VerbNet, FrameNet and WordNet try to encode. These inferences rely to a large degree on knowledge about the world. We are therefore exploring two different aspects of LLMs – the degree to which they have captured knowledge about language, and the degree to which they also capture the world knowledge reflected in that language knowledge. Specifically, we look at how well GPT-4 does with subcategorization alternations and with various aspects of the Caused Motion Construction (CMC, Goldberg 1995). Its performance with the CMC allows us to distinguish between behavior that is based on the integration of a moderately complex structural pattern and behavior that depends on the rich connotations of specific content words. Maybe surprisingly, GPT-4 does better on the latter than on the former.

1 Introduction

Symbolic language analysis systems tend to have a semi-modular approach, embodying the belief that the various subsystems have their own regularities which can be exploited in defining the range of possible utterances. The more recent systems such as LLMs give up modularity and the search for pre-defined subregularities. This does not mean that they might not find such subregularities; they just do not find it necessary to identify them explicitly as such. With enough data and sophisticated analysis methods, the patterns will reveal themselves without having been encoded separately.

Annie Zaenen, Alexis Cooper, Felix Zheng, Susan W. Brown & Martha Palmer. 2025. Poking at GPT's NL Generalizations. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 265–298. Konstanz: PubliKon.

In this paper we look at a type of subregularity that has historically proven to be difficult to codify in symbolic systems: subregularities that are centered on specific lexical items and more specifically subcategorization information. That such patterns exist has long been noticed and has led to many alternative valency lists in traditional grammars. It has also been hypothesized, at least since the work of Leila Gleitman (Gleitman 1990), that the knowledge of these patterns might play a role in language acquisition. For English they have been extensively described in Levin (1993), extended later in a more computationally accessible way in the work on VerbNet (explicitly) and FrameNet (implicitly) (Dang et al. 1998, Kipper et al. 2000, Schuler 2005, Kipper et al. 2006, 2008, Brown et al. 2022, Ruppenhofer et al. 2016). The claim embodied in these approaches is that, to a certain extent, these subcategorization frames are productive, so that it is impossible to list all instances. The other hallmark of these patterns is that the conditions of felicitous use are difficult to describe formally because they are at the intersection between syntax, semantics and world knowledge. With respect to these two claims, one can speculate that the generative aspect of LLMs could represent a significant advance in the direction of capturing the meaning of these constructions. In this paper we first probe GPT-4 to see whether it correctly detects a number of valency alternations and then concentrate on the caused motion/resultative construction to investigate whether it captures the real-world inferences of these constructions.

There is still a lot of mystery around the way very Large Language Models approximate meaning, especially with regard to GPT-4. We do know, however, that the training objective involves learning to accurately predict the next word in a sentence, and that the more data they can be trained on the better those predictions can be. GPT-4 is trained on truly vast amounts of data, resulting in values for 1.76 trillion parameters across 120 layers. We cannot fathom exactly how each word is represented, but we know the large number of layers and parameters allows it to store very precise information about the types of contexts in which a word appears in a numerical form, known as a vector representation or word embedding. These contexts are primarily syntactic and lexical contexts, allowing GPT-4 to place representations for closely related words near each other in an enormous multi-dimensional space, and to excel at paraphrasing. It can produce and recognize alternative wordings and structures for the same concepts very easily since their numerical representations are so similar. At a fundamental level it is using pattern matching to determine that this word in this context functions very similarly to the way another similar word would in the same context. From this perspective it truly “models” language knowledge (formal linguistic competence) extremely accurately. However, how well does it model the world?

How well does it interpret the mappings between the sentences it generates so readily and what is happening in the world (functional linguistic competence)? A classic AI representation given in Figure 1 of how natural language sentences are supposed to relate to the world shows a correspondence between individual sentences and their entailments or the facts that follow from the statement of the sentence. Does GPT-4 have access to the same entailments that a person would, upon processing the same sentence?

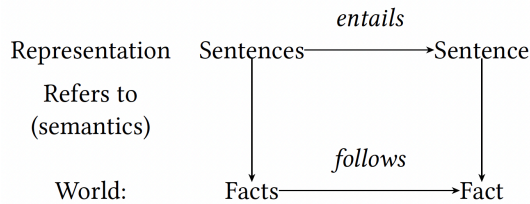


Figure 1: Classic AI representation of how natural language sentences relate to the world

In this paper we explore GPT-4’s ability to accurately identify grammatical syntactic variations for well-known verb classes, as well as its dexterity with respect to the corresponding typical entailments associated with those verb classes. We find, not surprisingly, that it handles straightforward syntactic variation extremely well, yet has more difficulty with entailments. This is especially true for long-distance constructions and unusual constructions. In general it seems the case that when both the linguistic construction is challenging and the real-world situation depicted is not run-of-the mill, the system does not get to the right interpretation (cf. example (14) and several of the type 4 CMC cases).

2 Valency alternations: grammaticality judgments

To systematically test GPT-4’s detection of predictable valency alternations, we referred to VerbNet. VerbNet is a computational lexicon of around 5,200 English verbs, organized primarily around Levin (1993)’s English verb classes (Dang et al. 1998, Kipper et al. 2000, Schuler 2005, Kipper et al. 2006, 2008, Brown et al. 2018, 2022). Verbs are clustered into VerbNet semantic classes primarily based on their syntactic behavior, as evidenced by their participation or non-participation in specific diathesis alternations, different subcategorization frames that can be meaning preserving or imply systematic changes in meaning (Bonial et al. 2011). Each VerbNet class contains, in addition to a core set of semantic roles and the

likely subcategorization frames for that class, a general-purpose 1st order predicate logic semantic representation that applies to all members of the class. The majority of the classes center around representations for basic semantic concepts such as change-of-state, change-of-location, or change-of-possession.

VerbNet has been shown to be effective in improving NLP results for tasks such as semantic role labeling, verb sense disambiguation and ontology mapping (Shi & Mihalcea 2005, Giuglea & Moschitti 2006, Loper et al. 2007, Brown et al. 2014). Often compared to the lexical resources FrameNet and PropBank, which also provide semantic roles, VerbNet differs from these primarily by providing semantic representations for each class that are interpretable in predicate logic. Both FrameNet and VerbNet group verbs semantically, although VerbNet takes into consideration the syntactic regularities of the verbs as well whereas FrameNet includes related nominals and adjectives. Both resources define semantic roles for these verb groupings, with VerbNet roles being fewer, more coarse-grained, and restricted to central participants in the events. For each class VerbNet also uses 1st order logical representations to show the general relationships between semantic roles for that class and to track any changes to the participants across the timeline of the event, with variations for different syntactic patterns. For example, the default representation for the VerbNet *manner_speaking-37.3* class explicitly specifies that all of the verbs in this class involve a transfer-of-information event, whereas the verbs in the *run-51.3.2-2-1* class involve a change-of-location event.

We assumed that LLMs are able to cope with most run-of-the-mill subcategorization information but tested whether GPT-4 could cope with standard valency alternations found in VerbNet. The results suggest that, at least in judging whether they are grammatical or not, they are up to the task.

We used GPT-4 0613, a snapshot of GPT-4 as of June 13th, 2023 that supports a context window of 8,192 tokens and consists of training data up to 3 September 2021. We selected several types of alternations to be tested. Certain verb classes are associated with each alternation, as shown in Table 1.

In total, we constructed 426 sentences across each of the six alternation types for a total of 17 subsets (i.e. combinations of alternation types, verb subclass, and sentence construction variations). Using the ChatCompletions API, we opened a new chat conversation for each prompt. The sentence was specified under the user content field and each one was introduced by a prompt. We ran a test probing grammaticality, using the following prompt: *For the given sentence, return whether a native English speaker would consider the sentence to be grammatical. If yes, only respond with "T"; if not, respond with "F" followed by a dash (no spaces), then a reason why it is not grammatical.* Because GPT-4 is not always consistent in its responses, we submitted each query three times.

Table 1: Selected alternation types and corresponding VerbNet classes. The example sentences showcase alternations that are possible or not usually accepted (indicated by *).

Verbs that alternate		Example Sentence	Alternation
Alternation Type & VerbNet Class			
Causative/Inchoative: break-45.1		Pat broke the vase.	The vase broke.
Causative: light-emission-43.1		Pat flashed the light.	The light flashed.
Understood body part: wink-40.3.1		Pat nodded his head.	Pat nodded.
Understood recip. obj.: marry-36.2 & meet-36.3		Pat and Kim met	Pat met Kim.
Dative: give-13.1 & instr_communication-37.4.1		Pat lent a tool to Kim.	Pat lent Kim a tool.
Benefactive: build-26.1		Pat baked a cake for Kim.	Pat baked Kim a cake.
Verbs that do not alternate			
Causative/Inchoative: cut-21.1 & carve-21.2		Pat chopped the vegetable.	*The vegetable chopped.
Causative/Inchoative: contribute-13.2		Pat submitted the draft to Kim.	*The draft submitted.
Understood body part: crane-40.3.2		Pat wrinkled his forehead.	*Pat wrinkled.
Understood recip. obj.: correspond-36.1		Pat and Kim bickered.	*Pat bickered Kim.
Dative: say-37.7 & manner_speaking-37.3		Pat mentioned a fact to Kim.	*Pat mentioned Kim a fact.
Benefactive: obtain-13.5.2		Pat acquired an item for Kim.	*Pat acquired Kim an item.

It is unclear what the meaning of grammaticality is for GPT-4. As far as we can tell, the system does not make a clear distinction between grammaticality and acceptability. The notion seems to be a broad one of acceptability. For example, we initially used the names Anne and Cathy in various examples, including with *marry* verbs, and GPT-4 responded by saying that marriage is usually between a man and a woman and did not answer the original question. To avoid this particular interference (and bias) from world knowledge, we replaced those names with ones that are traditionally considered gender-neutral. This example shows, however, that GPT-4 is not focusing solely on grammatical distinctions.

With this caveat, GPT-4 performed well when evaluating alternations. Out of 426 total sentences, 295 were grammatically valid according to Levin, and 131 were invalid. GPT-4 correctly classified 92.4% with a precision of 85.7% and recall of 96.3% for an F1 of 90.7%. The worst-performing subset was the non-alternating Understood Body Part category with *crane* verbs, and the best-performing subsets were the alternating Understood Reciprocal Object and Dative categories with *marry*, *meet*, and *give* verbs. All other subsets were close to the mean and median values.

For the sentences GPT-4 labeled ungrammatical, we found that GPT-4 was good at explaining why a sentence is not grammatical. For example, for the sentence “The fish speared”, GPT-4 responded with “The sentence is missing an object or a subject. If the fish is doing the action, an object is needed (“The fish speared the worm”). If something else is spearing the fish, a subject is needed (“The man speared the fish”).” We did not require an explanation from positive responses.

Generally, GPT-4 was more likely to assess a sentence as grammatical than Levin (1993). Often the disagreement seemed due to ambiguity in the meaning of the verb. For example, “Pat batted.” is classified by Levin as an ungrammatical alternation of “Pat batted his eyelashes”. GPT-4 evaluated it as grammatical, most likely because “Pat batted” can refer to batting a ball. Without more context, GPT-4 cannot know which sense of *batted* was meant. Several other cases of disagreement between GPT-4 and Levin (1993) can be accounted for either by word sense ambiguity or by GPT-4’s assessment that non-typical usages of a verb are ungrammatical. For example, GPT-4’s explanation for judging “Pat cubed the vegetable” as ungrammatical was “In English, we usually say “Pat chopped the vegetable into cubes.” The phrase “Pat cubed the vegetable” is not a commonly used or recognized phrase.”

On the whole, GPT-4 is relatively consistent in its responses, returning the same answer for all three prompts for 96.9% of all sentences. The subset with

the most disagreement (2 of 15 sentences tested) was the non-alternating Understood Body Part category with *crane* verbs constructed without objects (e.g. *‘‘Pat wrinkled’’); the majority of subsets had 100% agreement across each of the three prompts.

In our tests discussed below, the verb and its arguments occur in their canonical position (mainly Subj V Obj PrepObj). It has been pointed out that when the arguments are not in their canonical positions, specifically when their linear order is inverted, the results are much less good. We constructed a few examples to see whether GPT-4 shows a similar deterioration. The preliminary results show that simple questions seem to be assigned the correct grammaticality judgments but more complicated structures lead to problems.

We asked the same question as above, only probing for explanations when the sentence was found to be ungrammatical. We again repeated every probe three times to see whether GPT was consistent. As before, we used the ChatCompletions feature of the API on the June 13, 2023 snapshot of GPT-4. The model was run thrice, each time given a system prompt instructing it to determine the grammaticality of the given sentence (given as a user prompt) and explain its reasoning if the sentence was deemed ungrammatical. The prompts and full set of data collected are available in our GitHub¹ as Appendix A.

Sentences like:

- (1) I wonder what the heron speared.
- (2) I wonder what speared the fish.
- (3) What did you say the heron speared?

were unanimously judged to be grammatical, even though in (1) and (3) the linear order of the subject and the object are inverted.

A slightly more complicated example is also accepted:

- (4) I wonder what Louise thought the heron speared.

The longer distance subject extraction in (5), however, is judged ungrammatical although different reasons for ungrammaticality are given in each of the three attempts.

- (5) What did you say speared the fish?

¹<https://github.com/AACooper1/GPTGeneralizationsData>

The first judgment says the sentence is ungrammatical but then proposes as a correction the same sentence again: “Ungrammatical. The correct sentence should be “What did you say speared the fish?”” The second judgment would accept a version with *who* instead of *what* or the addition of a *with*, proposing the clearly ungrammatical: “What did you say speared the fish with?”. The third judgment again judges the input ungrammatical but proposes as a correct version the original sentence or a version with *who*: “A more grammatical sentence could be: “What did you say speared the fish?” or “Who did you say speared the fish?””

Some more complicated versions were accepted (“What did you say it was that speared the fish?” “What did you say you saw spear the fish?” “What did you say you saw spearing the fish?”). In these examples the subject-object linear order is respected.

Similar structures with object extraction, and hence an inverted order, confuse GPT. GPT-4’s reactions to (6a) are given in (6b-6d):

- (6) a. What did you say you saw the heron spear?
- b. Ungrammatical. The sentence should be, “Where did you say you saw the heron spear?”;
 - c. Ungrammatical. The correct sentence should be “Where did you say you saw the heron spear?” or “What did you say you saw the heron spearing?” The original sentence is missing either a noun to be the object of the verb “spear” or a preposition to indicate location;
 - d. Ungrammatical. Extra “the” in the sentence. The correct sentence would be “What did you say you saw heron spear?”

The clefted version seems to lead to similar confusions:

- (7) a. What was it you said you saw the heron spear?
- b. Ungrammatical. The sentence is missing a word for clarification. It should be “What was it you said you saw the heron spear with?”;
 - c. Ungrammatical. The sentence is missing a preposition separating the two actions. A correct version could be: “What was it you said you saw the heron spear?”
 - d. Grammatical.

These results suggest that GPT-4 has trouble with long-distance dependencies.

3 LLM and caused-motion construction inferences: overall impressions

The previous section dealt with argument structures that are syntactically complex but still canonical with respect to their verbs, tasks which natural language processing techniques prior to LLMs, such as syntactic parsing and semantic role labeling, tend to do well with. The inconsistency of GPT-4's interpretation of long-distance dependencies demonstrated in the previous section raises questions about the model's performance on other types of argument structure complexities. This section probes the model's ability to draw implications from argument structures that are not necessarily syntactically complex, but are not canonical with respect to their verbs—more specifically the caused motion construction (CMC) (see e.g. Goldberg 1995, Goldberg & Jackendoff 2004).

- (8) a. She threw the ball out of the park.
 b. I washed the chalk off the sidewalk.
 c. They laughed him off the stage.
 d. The market tilted the economy into recession.

While (8a) has a canonical argument structure for *throw* – the direct object is the theme and the argument following the preposition is the source – (8b) and (8c) do not for their respective verbs. In (8b), the argument structure is not typical of *wash*. Normally, the direct object would be the argument being washed (as in “I washed the sidewalk”), but here the direct object is the contaminant, while the theme – the sidewalk being washed – is in a prepositional phrase; and in (8c), *laugh*, typically intransitive, appears with a direct object. In these two latter examples, the argument structure is encoded in the construction rather than the verb itself.

Semantically, CMCs are interesting, as they are able to imbue a sense of motion not encoded in the vocabulary. Although only (8a) uses a verb inherently carrying caused motion in its internal semantics (i.e. “She causes the ball to move”), all four are interpreted as caused motion. *Wash*, for example, implies a caused *change of state* with no movement specified, but in (8b), the verb carries a sense of caused motion, while its semantics describe the manner in which the movement happened (“I caused the chalk to move off of the sidewalk by means of washing”). The verb *tilt* is a verb of spatial configuration normally taking, as its object, the inclined item (e.g. “He tilted the bottle”). In (8d) the motion described

is abstract rather than physical, as in “He tilted the liquid into his mouth and swallowed”. Whether the motion is physical or abstract, the semantics of the sentences are parallel: all four sentences have a causal argument responsible for the event, an argument in motion, and a path that specifies the initial, middle, or final location, state or condition of the argument in motion.

The reason these coerced usages proved challenging for previous natural language analysis systems is that those methods rely heavily on statistical patterns to capture the lexical aspects of the language. New and unusual usages of lexical items are problematic, because they are often infrequent in the data and the frequency of their usage is far outnumbered by that of more conventional verbs in the same semantic context. This is especially significant in discriminative learning methods employed by semantic classification tasks, because their predictive capability is dependent on the examples they have seen or experienced during the training stage. If examples for usages like (8b) are sparse (i.e. dwarfed in frequency) or lacking in the training data, the classifier cannot gather the necessary statistics about the verb. When *laugh* appears in a caused motion context, a discriminative approach is unlikely to correctly predict the meaning of this usage. It is a well-known problem: various studies showed that verbs of ambiguous meaning or verbs that do not have training examples contributed to classifier performance drop (Carreras & Màrquez 2005).

Hwang (2014) holds the premise that the challenge of generalizing to previously unseen CMC usages could be addressed by linking lexical semantics more directly with syntactic patterns, as theorized by Construction Grammar (Fillmore 1988, Goldberg 1995, Kay & Fillmore 1999, Michaelis 2004, Goldberg 2005). Construction Grammar is in agreement with the common assumption underlying Levin (1993) and VerbNet (Schuler 2005) that there is a relationship between the meaning of a word and its syntactic realization (Gleitman 1990). However, it goes a step further in proposing that particular syntactic structures can themselves carry meaning, in the same way a lexical item is thought to bear meaning. Meaning must be interpreted at both the lexical and constructional levels of sentences. This is especially important for sentences such as (8b) and (8c), in which the prevalent lexical semantics of the verb is at odds with the semantics of the sentence.

Thus, if the semantic interpretation available is strictly based on the expected semantics of the verb and its arguments, in some cases it will fail to include relevant information coming from the construction. Accurate interpretations for such sentences require that NLP classifiers first identify coerced usages in data. Furthermore, once the construction is identified and the semantic roles are properly assigned, the sentence also requires an accurate semantic interpretation with

appropriate representations. In sentence (8c), the conclusion is that the person left the room, and this was precipitated by the audience’s negative response. The *laughing* is the manner of the response. For (8d), the economy moved into a recession, and this was precipitated by the market *tilting*, presumably downward.

Hwang’s research (Hwang 2014) focused on capturing the effect of a CMC in descriptive features that could guide a symbolic semantic interpretation system to the correct in-context representation with VerbNet as the representation mechanism. Hwang’s proposal was to create a new class for the construction itself that contains a general semantic representation (using updated semantic predicates) of change of location, similar to the one for the “run” class. By combining this with the semantic representation for the manner-of-speaking class, the change of location and its instigation can both be captured.

Now we are curious as to how well GPT-4’s interpretation of CMCs might mirror Hwang’s suggestion. Is GPT-4 associating a specific semantic interpretation with similar CMCs? In the following two sections we investigate this possibility.

As far as their occurrence in texts is concerned, then, caused motion NP V NP PP cases can be classified in three subclasses:

- (9) a. the usual meaning of the verb combines with its arguments in the expected way, we will call this the *canonical combination*
- b. the verb is coerced into an unexpected frame to allow for a caused motion interpretation but this combination of the verb and its arguments is relatively common, so an LLM will most likely have been trained on instances of this combination, *coerced combination*
- c. the verb is coerced into an unusual frame and it is unlikely that the corpus the LLM is trained on will have seen a similar combination, *novel combination*

Grammaticality judgments for cases of the first two classes were part of the probe described in Section 2. We tested a few of those with respect to inferences, but as expected, GPT-4 did well on them. The full data set and prompts fed into the model are available in our GitHub² as Appendix B.

For instance, when given (10a), GPT-4 gave three times the correct answer to the questions: “Where was the ash before the blinking?” (10b) and “Where is it at the end?” (10c).

- (10) a. She blinked the volcanic ash off of her eyelashes.

²<https://github.com/AACooper1/GPTGeneralizationsData>

- b. The ash was on her eyelashes before the blinking.
- c. At the end, the ash is no longer on her eyelashes. It's likely in the air or on the ground / presumably in the air or on her cheeks (it is implied that the ash has been blinked away, likely falling off).

GPT-4 seems also to do well with respect to metaphorical expressions. It has been exposed to a lot of figurative/metaphoric language use, as can be seen in the following examples. For instance, the question *Did they touch him?* is answered adequately in the context of sentences like (11a) and (12a) as shown by (11b) to (11d) and (12b) to (12d):

- (11) a. They nodded him into the chair.
 - b. The text does not provide information on whether they touched him. Simply stating “they nodded him into the chair” suggests they motioned with their heads for him to sit, which does not require physical touch.
 - c. The phrase does not specify any physical contact, so it can be assumed that they did not touch him but indicated with their heads (by nodding) for him to sit in the chair.
 - d. The sentence does not provide information about whether they touched him. “They nodded him into the chair” generally suggests they used nonverbal communication (like a nod) to indicate he should sit, not physical touch.
- (12) a. They nodded him into the country.
 - b. No, they did not physically touch him. The phrase “nodded him into the country” is metaphorical, expressing that he was given approval or permission, possibly in a non-verbal manner such as with a nod, to enter the country.
 - c. No, they did not touch him. “Nodding” in this context means they gave their approval or consent, likely via a physical nod of the head, but it would not involve physical touch.
 - d. No, they did not touch him. “They nodded him into the country” is a metaphorical expression meaning others indicated approval or acceptance, possibly granting him entry into the country, without physically touching him.

However other examples show that not everything is understood the way a human would understand it as illustrated in (13).

- (13) a. They cold-shouldered her out of the pharmacy.
- b. Where is she now?
- After being cold-shouldered, she is now outside the pharmacy.
 - She is now out of the pharmacy.
 - She is now outside of the pharmacy.
- c. Where did she work?
- She worked at the pharmacy.
 - She worked in the pharmacy.
 - She worked at the pharmacy.

The answer to “Where did she work?” suggests that GPT-4 is sure that the pharmacy should be understood as her workplace rather than a physical location, but in the answer to “Where is she now?”, it seems to believe that she has been forced out of a physical location.

We then created some cases we hypothesized might be more difficult because they are uncommon and less plausible, such as:

- (14) The antiques dealer veneered the table into a higher price bracket.

Here we asked the system to put together linguistic information and world knowledge, asking: “Did the veneer make the table look better or worse?” GPT does not do this integration: three times we get the answer: “The text does not provide specific information on whether the veneer made the table look better or worse.”

4 A closer look at GPT-4’s interpretation of the grammar of CMC

The previous sections have shown that GPT-4 does well on a broad set of CMC examples. In this section we investigate which linguistic generalizations it might

draw on in providing answers. There is a lot of discussion about GPT and other LLMs mimicking the opinions of subgroups of users of English because of biases in the input data. We are interested in studying whether there are biases towards a subset of English. It is clear that such a bias must exist with respect to the lexicon as a whole. GPT-4 can only generalize over content words that are used frequently. We would like to know whether its input allows it to make the same generalizations adult native speakers would make over function words and grammatical constructions. The investigation below is a small step towards this goal.

The data we have investigated is not extensive enough to allow conclusions even with respect to CMC but does support hypothesis development. We presented GPT-4 with examples in which the subject and the object are pronouns, reducing the context to the minimum.³ For example, knowing the identity of the object being kicked in “She kicked the ball/wall under the shelf” allows the LLM to draw on the many examples it must have of balls moving and walls NOT moving to lead it to a CMC interpretation with *ball* as the object but a non-CMC interpretation with *wall* as an object. Just using *it* instead forces the LLM to rely more on structural patterns. We also limited ourselves to CMCs that describe physical events.

We start from the assumption that GPT-4 builds up its answers by combining high probability (usually appropriate) patterns in the input text together: although one of its advantages is that it does ‘everything everywhere all at once’ instead of proceeding through a pipeline, we hypothesize that its responses are still built out of parts. We assume that it has built up some probabilistic generalizations over patterns that occur frequently in its input and that it associates these patterns with possible patterns of continuations (answers). In the previous examples, it could appeal to a relatively rich context to construct appropriate responses. We are curious to see how much it can rely on those structural patterns in the impoverished context of this section. This does not mean it will always appeal (solely) to that knowledge in interpreting CMC examples but only that it can often get by by relying on this type of structural pattern recognition.

We have assumed above that interpreting a CMC implies being capable of constructing sentences that answer at the very least the following questions (DO = Direct Object):⁴

³Having the direct object being a pronoun also avoids interpretations where the NP Prep NP constellation forms one NP and the NP V NP Prep NP sequence could be a simple transitive construct.

⁴There is more to the meaning of CMCs as discussed in detail in Goldberg (1995) and related publications.

- (15) a. Did the DO move?
 b. Where is the DO at the beginning/end of the event?
 c. What is the cause of the movement?
 d. Who/what performed the action?

What are the minimal patterns in a CMC example that would allow a human to answer these questions? Interestingly enough, it depends on the semantic type of the verbs and prepositions the example contains. We can distinguish at least 8 types, some of which lead to an unambiguous interpretation and some of which allow a CMC and a non-CMC interpretation. We illustrate these types before looking at the GPT-4 responses.

All types that we are investigating have the same phrase structure:

NP/subj Verb NP/obj Prep NP/location.

The types differ in the kind of verb and/or preposition they contain. In what follows, we call a verb that implies that its DO changes location a *movement* verb; a verb that implies that the subject exerts some force against the DO without the latter necessarily moving a *force-exerting* verb; *locative alternation* verbs are verbs that have two subcategorization frames, one in which the location is expressed by the prepositional object and the DO expresses an object that is moved to that location or away from it, and another in which the location itself is the DO and there is a prepositional object whose nature varies depending on the preposition, as shown in “He stuffed toys into the suitcase. / He stuffed the suitcase with toys.” (see Levin (2006) for further discussion). *Coerced* verbs are verbs that are normally intransitive but that are used transitively within the CMC. The most famous example is “Fred/Frank sneezed the napkin off the table” (Goldberg 1995).

Prepositions are subdivided into *directional* prepositions which imply, when the event described is physical, that there is movement towards or away from the location expressed in the prepositional object and *neutral* prepositions that can be used to express such movement or describe a stationary position vis-à-vis of the prepositional object, e.g. an area.

Type 1a: movement verb and directional preposition. In this pattern, the meaning of the stretch of text always corresponds to a CMC. For example:

(16) He launched it into the field.

To answer our first question: *Did the DO move?* GPT-4 has only to have classified *launch* as a movement verb and to have analyzed the basic syntax of the verb-DO combination: the simpler sentence, “He launched it”, also implies that the DO moves. This pattern also entails the answers to the questions (15c) and (15d) (what is the cause of the movement and who/what performed the action). It is only to answer the question about the movement path that GPT-4 has to interpret the prepositional phrase. It has to know whether the prepositional object indicates the start point or the end point of the movement and it has to correctly interpret the trajectory. To fully interpret the trajectory, it has to go beyond purely linguistic knowledge and have some further information about what verbs like *launch* imply. The right answers to our questions for this type do not require a holistic interpretation of the sentence, as also pointed out in Weissweiler et al. (2024). These cases can be understood compositionally along the lines of Gawron (1985) and Pustejovsky (1991), who pointed out that such sentences involved putting together the interpretation of the Subject-V-Object part with the interpretation of the prepositional phrase.

Type 1b: movement verb and neutral preposition. This pattern can encode either a CMC or an event in which the subject causes the DO to move in an area described by the prepositional phrase. For example:

(17) They moved it over the table.

In one reading, they move it so that it ends up located over the table (and this fits the CMC); in the other, they move it around in the area over the table. In the first, *over* is the destination, whereas in the second, *over* is the area in which the movement takes place. In both cases, there is motion, but in the second, the pattern is not that of the caused motion construction in which the PP describes the trajectory.

Type 2a: force-exerting verb with directional preposition. Again, the CMC interpretation is forced by the preposition. For example:

(18) She kicked him through the street.

To answer the first question, for certain verbs GPT-4 can rely on statistical patterns: although they do not entail movement of the DO, they are often used in

contexts where movement occurs, or it could rely on a generalization about directional prepositions: when describing events in the physical world they imply change of location. For the causation question, GPT-4 has to have an interpretation of the verb as a movement verb or have an interpretation of the verb together with the preposition as indicating movement. The interpretation of the trajectory, again, depends on interpreting the prepositions and their relation to the verb.

Type 2b: force-exerting verb with neutral preposition. These are ambiguous along the same lines as those of type 1b. For example:

(19) She kicked him under the table.

Type 3a: locative alternation verb and directional preposition. The CMC reading is the only possibility. For example:

(20) He stuffed it into the suitcase.

Here GPT-4 cannot rely on the meaning of the verb alone as a sentence like “He stuffed it” does not have a movement interpretation. The movement reading only comes about by the combination of *stuff* with the directional preposition: *stuff into* implies movement.

Type 3b: locative alternation verb and neutral preposition. These are again ambiguous. However the ambiguity is of a different type from that in type 1b and 2b. For example:

(21) He washed it from the sidewalk.

Under the CMC interpretation, the direct object is the contaminant, while the theme – the sidewalk being washed – is in a prepositional phrase. However under the non-CMC interpretation to the sentence, *it* is the thing washed, and *the sidewalk* is the location from which it is done (e.g. *it* could be a car standing on the street and *from the sidewalk* is the location of the person doing the washing).

Type 4a: coerced verb and directional preposition. The CMC interpretation is the only possible one. For example:

(22) She sneezed it off the table.

Type 4b: coerced verb and neutral preposition. Again, the CMC interpretation is the only possible one. In this class, the subcategorization frame of the main verb forces the interpretation. For example:

(23) She sneezed it under the table.

These examples require a subtle interpretation of the subcategorization possibilities of the verbs involved. In their more common frame (intransitive and/or combined with *at*), these verbs do not have a DO, so they certainly do not imply movement of the DO. Here GPT-4 would require rather detailed statistical generalizations or could possibly rely on the interpretation of very specific patterns that have been encoded without being analyzed in their parts.

It is interesting that in the case where the verb is a movement verb, it is in fact not the verb that forces a CMC interpretation but only its combination with a directional preposition. If the preposition is neutral, there is also a non-CMC reading. It is only in type 4 that it is the verbal syntax that forces the CMC interpretation: even with a neutral preposition, coerced intransitive verbs lead to a CMC, a non-CMC interpretation would require e.g. the use of *glare at*, *snore at*, *sneeze at*. For instance, “She sneezed the napkin under the table.” only has a CMC interpretation and implies movement. “She sneezed at the napkin under the table.” does not imply movement.

4.1 Experiment

Each GPT-4 session consists of the presentation of a potential CMC sentence followed by five questions. For each sentence three sessions are run. (See results in the GitHub⁵ under Appendix C.)

In the analysis we try to determine whether GPT-4 recognizes that:

- (24) a. there is movement of the DO,
- b. the movement of the DO follows a trajectory described by the combination of the verb and the preposition, and
- c. it is the action described by the verb that causes this movement.

We see that overall, even in this semantically impoverished environment, GPT-4 does reasonably well, with the ‘best’ results being in the type 1a and 41 configurations.

⁵<https://github.com/AACooper1/GPTGeneralizationsData>

In type 1a, in all three runs, GPT-4 recognizes that there is movement of the DO, that the movement has a specific trajectory and that the action by the subject of the sentence is the cause of the movement. As said above, for GPT-4 to answer that the DO moved and that the movement is caused by the action of the subject, we only need to assume that GPT-4 recognizes that in the environment of *eject*, *steer*, *launch* and *move*, there is a very high probability that the DO of these verbs has moved and that the reason for this movement is the action of the subject, in other words, that GPT-4 sees these verbs as movement verbs. To answer the question “where is the *it* at the beginning and the end of the event”, however, GPT-4 has to have made generalizations over directional prepositions. *Into* indicates the end point of a movement, whereas *through* and *towards* indicate a trajectory and *out of* indicates a movement away from. GPT-4 gives clear and rather detailed answers about the movements that might be described by these sentences, not only getting the right result about the end point of the movement but also making reasonable inferences about the starting point. These inferences seem to rely on aspects of the meaning of the verb GPT-4 has encoded as highly likely; e.g. *launch* does not strictly entail the proximity of its agent to the object to be launched but GPT-4 assumes such proximity, stating that “It was with her before she launched it”. Presumably the majority of uses of *launch* in the training corpus allow for this inference and the use of this information seems to show that GPT-4 makes rather detailed calculations about the co-occurrence possibilities associated with these verbs and prepositions.

The results for type 1b complicate the picture somewhat. First, it seems that GPT-4 sees *across*, *down* and *over* as directional prepositions without considering other possibilities, and it answers the questions the same way as for type 1a. *Under*, however, is treated differently. Although the verb is *eject*, GPT-4 does not see that this sentence implies movement of the DO. In fact the answers given seem to indicate that GPT-4 does not have a coherent view on what this sentence means. For a human being, the sentence, “He ejected it under the porch”, can mean two things:

Reading 1: *He* and *it* are not under the porch at the beginning of the event and he ejects it so that it ends up under the porch.

Reading 2: *He* and *it* are under the porch from the beginning on and the whole event takes place under the porch.

Under both readings, there is movement, but under reading 1, the trajectory of the movement is specified whereas under reading 2 the trajectory is not specified. The answers GPT-4 provides are not consistent with either of these two

readings. The only part of the meaning that is correctly interpreted is that the DO is under the porch at the end of the event. This result is a bit odd: given that *eject* itself implies movement, the question about the DO moving should not have been problematic. In all three sessions, however, GPT-4 interprets the question as being about another movement. This raises a question about how the system puts together its interpretation of the sentence as a whole, or even whether it puts together such an interpretation. We will call interpretations in which the system ends up answering the questions in a way that does not allow us to conclude it has ended up with one of the two possible readings as *incoherent*; if it ends up with one of the readings, we call the interpretation *coherent*. Some readers have pointed out to us that to eject something under a location is most likely extremely rare. One would rather expect to eject something from under a location. Maybe GPT-4 rejects the sentence because the situation it describes is extremely improbable. This is, of course, worrisome: one important use of language is to describe unusual situations. However, we are getting evidence that humans also find the “eject” sentences confusing, so GPT-4’s confusion may be reflecting human confusion appropriately.

In type 2, we use verbs that do not strictly entail movement. Intuitively, however, the four verbs chosen differ greatly in how likely movement is. With *push* most people will think movement likely, with *hit* and *kick* movement is possible but not obvious, and *squeeze* is clearly not a movement verb. If we combine these preferences with the hypothesis we have made that GPT-4 seems to see *over* and *across* as directional prepositions, we expect that GPT-4 will interpret “He pushed it towards the enemy”, “He pushed it over the wall”, “She hit it out of the room”, “He hit it across the room” and “She kicked it through the street” as CMC, implying movement of the DO and a causal relation between the action indicated by the verb and the movement. This is the result we get.

The three other examples present problems. If GPT-4 had put together an interpretation of the CMC similar to that of humans, it would interpret “She squeezed it into the suitcase” as implying movement, on the basis of it having a preposition that implies directional change. This is not what we find: the sentence is deemed not to imply movement. The only aspects of the meaning that are decoded correctly are that at the end of the event; the DO is in the suitcase and that the subject of the sentence performs the action. “He squeezed it in the suitcase” allows for a coherent non-movement interpretation but that is also not exactly what we get: to the question of where the DO was at the beginning of the event, we get the answer that this is unknown or that it is outside of the suitcase, whereas at the end of the event it is inside the suitcase.

In “She kicked him under the table” both the verb and the preposition allow a non-movement interpretation. We suspect on the basis of example “He ejected it under the porch” (of type 1b) that GPT-4 prefers to see *under* as a non-directional preposition, and indeed we get a coherent interpretation along these lines.

Type 3 allows us to make further hypotheses about what GPT-4 pays attention to when answering the questions. The verbs in this category, *stuff*, *feed*, *wash* and *hose*, are clearly not verbs that imply a change of location of their object and regardless of the preposition. GPT-4 opts for non-movement in the case of *stuff* and *feed*. In the case of *wash* and *hose*, we get a recognition of movement in some sessions, but the causal relation between the movement and the action of the subject is not recognized most of the time. The only correct and coherent interpretation is given for “He hosed it out of the street”. It is not far-fetched to assume that GPT-4 sees *hose* as a movement verb. The prepositions influence the interpretation in as far as GPT-4 comes more or less to a non-CMC interpretation with *under* and *from* but the interpretations are not completely coherent. They fail to show the correct interpretations of the situations that we found in most of the cases of type 1 and some of the cases of type 2. This is interesting because it is the type 3 cases that should show the causal interpretation specific to the CMC: in this type the verb by itself does not imply movement so the causal interpretation has to be made on the basis of an interpretation of the combination of the verb and the preposition.

The performance of type 3 verbs would lead us to expect that GPT-4 would perform badly on the type 4 verbs because with those the verbs in the CMC are used in a way that is the furthest removed from their canonical use: the subcategorization frame is different and the verb on its own does not have a movement inference. However that is not the case. GPT-4 performs well on coerced verbs combined with directional prepositions, inferring start locations that are not explicitly mentioned and recognizing the causal relation between the action encoded in the verb and the movement of the DO when the preposition is neutral. So, the results are good with prepositions that we suspect from previous examples to be seen as being directional, but with prepositions that we know from previous examples to be interpreted as static (*under*), it sticks to the static interpretation and hence cannot come up with the right interpretation. Given the results, we hypothesize that *against* is also a preposition that is not seen as directional. We hypothesize that it has learned the CMC for these verbs as a special idiomatic construction. This impression is reinforced by responses such as “The phrase “laughed him out of the field” is idiomatic and denotes being ridiculed or dismissed, often causing a person to retreat from a situation”.

Although the data we have do not allow for a statistical interpretation, we summarize the results numerically: GPT-4, given 22 unambiguous CMC sentences, recognizes the causal relation involved in CMCs in 16 cases and the movement of the DO in 15 and provides 23 coherent interpretations out of 34 possible ones. The results point to several issues that need further study.

1. It might have been better to focus the questions more on the interpretation of the trajectory of the DO than on movement per se as it is movement along a trajectory that is central to the CMC, not just change of location.
2. It is clear that GPT-4 does not necessarily categorize the verbs and the prepositions at issue in the way humans do. For instance, lexical semantic investigations classify *push* and *hit* as verbs that encode an exertion of force but not necessarily movement. GPT-4 interprets them here as implying movement. Our example sentences combine the verbs and prepositions in Table 2 and give the movement-or-not interpretations indicated. On the basis of this distribution, we hypothesize that GPT-4 assigns a very high probability to a movement interpretation of *eject* (M), *steer* (M), *launch* (M), *move* (M), *push* (F), *hit* (F) and *kick* (F) and the prepositions *into*, *onto*, *through*, *toward*, *out of*, *down*, *off*, *across* and *over*. It would be interesting to see how stable these interpretations are for each verb and preposition. We could, for instance, ask how likely the movement interpretation of “X hit/pushed Y” is. Of course, humans might also have preferred interpretations for these verbs and the similarities and differences should be compared.
3. While *push* and *hit* are interpreted as movement verbs, *kick* and *squeeze* seem not to be. When *kick* is combined with a directional preposition, we get the expected CMC interpretation but with *squeeze* this is not the case. When we present a sentence with a neutral preposition, neither verb gives a CMC interpretation, but they also do not give a clear-cut non-CMC interpretation. For “She kicked him under the table”, we can interpret the answer as more or less correct on the non-CMC interpretation, but with *squeeze*, GPT-4 is completely inconsistent. In the answer to one question, it assumes movement but in the answer to another it does not. The interpretation is independent of whether the preposition is *in* or *into* (although the answer to the question of where it was before being squeezed in the suitcase refers to a state of being “before it was squeezed into (sic) the suitcase”). It is odd that even in the contexts in which the CMC is not taken

Table 2: GPT-4's tendency to interpret the combinations of verbs and prepositions as implying movement. (M: movement; F: force-exerting; L: locative alternation; C: coerced transitive)

Verb	into	through	tow-ards	out of	down	off	un-der	over	in	ac-ross	against	from	around	onto
Eject (M)	yes						no							
Steer (M)	yes							yes						
Launch (M)		yes								yes				
Move (M)			yes	yes										
Kick (F)	yes						no							
Push (F)		yes						yes						
Squeeze (F)	no								no					
Hit (F)				yes						yes				
Hose (L)			yes							no				
Wash (L)						no						?		
Feed (L)	no													
Stuff (L)	no						no							
Snore (C)	yes						no							
Laugh (C)			yes					yes						
Sneeze (C)				yes						yes				yes
Glare (C)										no				yes
Jeer (C)			yes											yes?

to be plausible, GPT-4 does not consistently propose simple locative interpretations to sentences such as “He squeezed it in the suitcase”. We need to investigate further how GPT-4 interprets “NP V NP prep NP” sentences more generally.

4. When the movement has to be inferred from the combination of a non-movement verb with a directional preposition, GPT-4 fails: We saw that in type 2, most of the correct answers could be accounted for by assuming that the GPT-4 class of movement verbs is bigger than the one that linguists assume, or rather that it is based on high probabilities rather than on entailments. With type 3 verbs GPT-4 has trouble. Here one must combine the meaning of the verb and that of the preposition to get the right result. GPT-4 cannot do this reliably. To make more sense of the results, we should confront GPT-4 with the locative alternation outside of the CMC construction (some work in that vein has been done in Wilson et al. (2023) on the BERT family of LLMs and the *spray/load* alternation).
5. At first blush, our experiment seems to show that GPT-4 captures the causal relation that is the hallmark of the CMC construction rather well. However a closer inspection of the results casts doubt on that. As discussed above, to correctly interpret type 1 examples one must combine the meaning of a simple transitive clause with a movement verb and the meaning of a prepositional phrase with a preposition that indicates or can indicate directed movement. Causing the movement is part of the meaning of the verb and does not have to be ascribed to the construction. Causing the movement to follow a particular trajectory is part of the meaning of the construction but we did not ask about that explicitly. The answers we have are rather ambiguous, but some seem to indicate that when GPT-4 decodes both the verb and the preposition as indicating movement, it goes further in putting together a picture of the event. For instance, answering “Why did it move?” for the sentence “She launched it towards the enemy”, we get “It moved because she launched it, applying force to send it towards the enemy.” This seems to go beyond recognizing that *launch* implies movement and that the movement is towards the enemy. However for combinations that do not include a verb that we hypothesize to be recognized as a movement verb by GPT-4 the results are not good. To recognize the causal relation involving the movement of the DO, one needs first to recognize the meaning of the combination of the verb and the preposition and GPT-4 does not seem to do that. This again can be tested further by studying how GPT-4 handles other cases of complex verbs.

6. We assume that for type 4 verbs, GPT-4 has developed a specific strategy. This hypothesis needs to be further investigated with more examples of different types of coerced verbs. It must be admitted that humans would also have problems with the interpretation of some of these sentences out of context and it remains to be seen under which circumstances both humans and GPT-4 generalize from specific verbs to a more general scheme such as ‘An intransitive verb that can also occur with *at*, can occur in the NP V NP dir-prep NP environment as a force-exerting verb causing the movement of its DO.’

Overall, this investigation leads to the hypothesis that GPT-4 has made some generalizations about movement verbs and their objects and about prepositions and their objects and that it has learned that some verbs occur in a special environment with a special interpretation, but it does not have a general interpretation of what linguists recognize as the CMC. More specifically it does not seem to combine a verb and a preposition to get to a movement interpretation when the verb on its own is not a movement verb. At this point it is not clear whether that is due to it concentrating on units that are too small or this being a relatively rare phenomenon, even in the vast amounts of training data GPT-4 has had access to. (Exponentially more data than any one very long-lived human would ever have had access to.)

5 GPT-4's interpretation of differing implications of CMCs

The previous section focused on the model's performance on the varying forms of caused motion constructions. This section focuses on its performance on the varying semantics of caused motion constructions. An interesting ambiguity in the interpretation of caused motion constructions is discussed in Iwata (2020). Iwata focuses on CMCs of the form *X beat the Y out of Z* as they appear in several corpora, and describes four different semantic categories elicited by the construction, detailed below with illustrating examples:

- (25) a. ‘Content coming out of container’
They beat the dust out of the rug.
- b. ‘Get rid of’
We beat these attitudes out of them.
- c. ‘Get by coercion’
They beat the confession out of them.

- d. ‘Physiological/emotional effect’
They beat the crap out of them.

We have done similar work to Iwata, but generalized beyond *beat* constructions. In our preliminary studies (Cooper 2024), we identified five categories, which we have dubbed *suppressive*, *extractive*, *transitional*, *preventative*, and *idiomatic*. These have some overlap with Iwata’s categories, but include sense distinctions not possible with *beat*. Our categories are exemplified in (26a-26e):

- (26) a. Suppressive: Trudeau’s government immediately *took the profit out of the long-term care industry*.
b. Extractive: Attempting to strengthen the nation, the USSR *squeezed the wealth out of Ukraine*.
c. Transitional: The by-law will *take the country out of the grips of drought*.
d. Preventative: Dostum managed to *keep the Taliban out of Kabul*.
e. Idiomatic: Constantine used the opportunity to *take the piss out of the pompous pricks* that populate the occult world.

(all adapted from the enTenTen Web Corpus)

In the suppressive interpretation, hypernymic to both the *get rid of* and *content coming out of container* categories from Iwata, the theme argument is an attribute which is removed from the source; it is an intangible and inalienable object and thus no longer exists at the end. In contrast, the theme in the extractive sense, like in Iwata’s *get by coercion*, does not have its destruction entailed. Rather, it is a more tangible item, alienably possessed by the source and the construction represents the transfer of possession to the theme.

The transitional sense of the construction demonstrated in (26c) is similar in that the theme is an independent object which exists after the event, but critically there is no possession involved. Instead, the agent leads the theme from one state to another, relying on the STATES ARE LOCATIONS conceptual metaphor.

The last two involve less literal forms of movement. The preventative sense is prevented motion rather than caused motion; rather than a source, the location argument is a goal. Finally, in the idiomatic sense, analogous to Iwata’s *Physiological/emotional effect* sense, motion is not invoked at all. Instead it is a construction of its own, with the caused motion form mostly being fossilized. The *take the piss out of* construction in (26e) is one example of this, but another very common idiomatic use of the caused motion construction is an emphatic sense, as in (27):

(27) They beat the absolute hell out of the American team.

By and large, GPT-4 does a good job of drawing the correct interpretations. As elsewhere, we tested the model's interpretation of these constructions across three instances with a common set of questions. These questions, shown in (28a-g), are focused on the states of arguments at different points in the event as well as the relationship between arguments:

- (28) a. Does (direct object) exist at the end of the event described in the sentence above?
- b. Does (subject) have (direct object) at the end of the event described in the sentence above?
- c. Does (locative argument) have (direct object) at the end of the event described in the sentence above?
- d. Describe (direct object)'s state at the beginning of the event described in the sentence above.
- e. Describe (indirect object)'s state at the beginning of the event described in the sentence above.
- f. Describe (direct object)'s state at the end of the event described in the sentence above.
- g. Describe (indirect object)'s state at the end of the event described in the sentence above.

We first tested the model on the five context-rich sentences in (26a–26e) and found that it performs in line with human expectations. Its responses to these are available in the GitHub⁶ as Appendix D.

Given its performance in these high-context environments, we tested it on some sentences containing less contextual information. Three examples from each category were fed to the model, the chat logs for which can be found in the GitHub⁷ as Appendix E. By and large, the model performed well on these too, with a few minor caveats where its interpretations seemed inconsistent, particularly when it came to distinguishing the suppressive sense from the extractive sense. Consider (29):

⁶<https://github.com/AACooper1/GPTGeneralizationsData>

⁷<https://github.com/AACooper1/GPTGeneralizationsData>

- (29) a. She mocked the fantasy out of him.
b. She blackmailed the fantasy out of him.

Typically, (29a) has a suppressive reading and (29b) has an extractive reading. The meaning of *the fantasy* is ambiguous and is only resolved by the sense of the construction evoked by the predicate. In (29a), *the fantasy* is an attribute, a synonym with *whimsy*; in (29b), *the fantasy* is a specific idea, making it compatible with the internal semantics of *blackmail*. GPT-4 tends to interpret *the fantasy* in the second sense when given (29a):

- (30) USER: Describe his state at the beginning of the event described in the sentence above.
ASSISTANT: Before the event described in the sentence, he was likely indulging in some form of fantasy or unrealistic belief, perhaps having a sense of hope regarding a particular situation or idea.
USER: Describe his state at the end of the event described in the sentence above.
ASSISTANT: He would likely be feeling embarrassed, disheartened or disappointed because his fantasy, something deeply personal to him, was made fun of. His confidence might also have been diminished.

However, it still gets a suppressive interpretation of the sentence, just regarding a specific fantasy as being diminished rather than a general sense of fantasy.

Relatedly, the model often had trouble with some of the implications of information transfer, namely that the actor transferring the information typically still retains that knowledge after the fact; see (31–32):

- (31) *She blackmailed the fantasy out of him.*
USER: Does the fantasy exist at the end of the event described in the sentence above?
ASSISTANT: The sentence suggests that the fantasy no longer exists within him after the event, as it was “blackmailed out of him”.
USER: Does he have the fantasy at the end of the event described in the sentence above?
ASSISTANT: The sentence suggests that he does not have the fantasy at the end of the event, as it was “blackmailed out of him.”
(N.B.: Only one of the iterations of the model had this issue, while the other two did not.)

(32) *She worked the information out of it.*

USER: Does it have the information at the end of the event described in the sentence above?

ASSISTANT: No, it no longer has the information at the end of the event as it has been worked out of "it" by the woman.

The model also sometimes had slightly odd interpretations of the preventative construction; see (33), in which GPT-4 seems to think that *the ideas* were hers, where the typical analysis would be that they were others' ideas which she wanted to shelter him from; or drew implications that are plausible, but not entailed, as in (34).

(33) *She kept those ideas out of him.*

USER: Describe those ideas' state at the beginning of the event described in the sentence above.

ASSISTANT: Before the event described in the sentence, the ideas were presumably known or present with her.

USER: Describe his state at the beginning of the event described in the sentence above.

ASSISTANT: Before the event described in the sentence, he was unaware of her ideas.

(N.B.: One iteration of the model interpreted the ideas as something she did not hold: "At the beginning of the event described in the sentence above, the status of those ideas is unspecified. They may have been considered, formed, or known, but they had not yet been introduced or shared with him.")

(34) *He barricaded them out of it.*

USER: Describe its state at the beginning of the event described in the sentence above.

ASSISTANT: The initial state is not specifically described in the sentence. However, before the event, the place was presumably accessible to "them."

These are minor issues, however. Overall, the model generally performs well. There were next to no observed issues with its responses to the idiomatic examples. These results also show that the model is adept at drawing the interpretation holistically and not solely from the verb or argument. It was given samples of different categories distinguished only by the direct object (*she worked the*

[*knot/information/hell*] out of it) and by the verb (*she [mocked/blackmailed] the fantasy out of him*) and still seemed to correctly capture the distinction, the slight deviations above notwithstanding.

6 Conclusion and future directions

The immense sets of language data that LLMs are trained on contain lexemes that are arranged in complex and varied grammatical patterns which convey a substantial amount of information about the world. In this paper we have probed GPT-4 in an attempt to tease apart its reliance on these various aspects of language. LLMs are often described as achieving human-like language generation by discovering patterns and regularities in the data, especially the patterns based on word co-occurrence frequencies. However, grammatical patterns are there to be discovered as well. We only have indirect evidence of what patterns in the data these models are discovering, drawn from the successes and failures in their output. This question is vast, and we have focused on some very specific patterns to begin to answer it.

First, we investigated GPT-4's ability to recognize common grammatical alternations, such as the dative alternation ("Pat lent a tool to Kim/Pat lent Kim a tool"). Only certain classes of verbs participate in each alternation. When verbs whose meanings conflict with the alternation are used in that alternation, we consider the resulting sentence as ungrammatical (e.g. *"Pat mentioned Kim a fact"). When presented with both grammatical and ungrammatical sentences for six different types of alternations, GPT-4 correctly categorized the grammaticality of about 90% of them. Some of the incorrect categorizations cannot be considered mistakes because of insufficient context in the queries, whereas others showed that GPT-4 considered some acceptable sentences as ungrammatical because of the rarity of the verb being used in that context. Overall, we judge its performance in this task as quite good.

Next, we looked closely at the caused motion construction because of the complex interplay of specific verbs, prepositions and nouns with the grammatical construction itself. This study led to the hypothesis that GPT-4 mainly relies on local patterns (e.g. subcategorization patterns) and does not integrate these into bigger structural units (e.g. constructions). What it seems to do well in most cases, although not all, is associate appropriate entailments with the patterns.

Finally, we presented 5 semantic categories of CMC and probed GPT-4's ability to correctly interpret them. Here again we see that GPT-4 draws on the connotations of content words to come to rather subtle interpretations.

We know that LLMs rely on what has occurred together frequently in their databases to interpret their prompts. That reliance gives them their power but also makes them susceptible to bias. GPT-4 excels in interpreting commonly occurring situations expressed in commonly occurring language. In that they may not be that different from people. However people can be corrected because they can be told that generalizations based on their previous experiences are not correct. One thing such corrections can rely on is a speaker's understanding of the grammar of their language, and the grounding of language in the real world. If the findings of this study hold up, this is not something that one can rely on when correcting LLMs. Their grammatical generalizations are not solid enough.

Acknowledgments

[Annie Zaenen] When Dick and I with many other colleagues worked on XLE, the PARC NLU system (Crouch et al. 2011), one aspect we tried to get our head around with little success was the combination of various lexical resources, mainly WordNet and VerbNet, into a component of the system that would help with the detection of entailments and contradictions (Condoravdi et al. 2003, Crouch & King 2004, Zaenen et al. 2010). The various resources used different terminologies, different granularities; they were incomplete and often contained errors. Putting them all together helped in some cases and hurt in others. While we are no longer involved in these rather frustrating experiments, it is interesting to see whether newer approaches do better.

We thank Joan Bresnan, the participants in the CLEAR (Computational Language and Education Research) meeting at the University of Colorado, October 2, 2024, and the audience of the Konstanz workshop LLMs and linguistic theory, October 14 2024 for comments on oral presentations of some of the material in this paper.

References

- Bonial, Claire, William Corvey, Martha Palmer, Volha V. Petukhova & Harry Bunt. 2011. A hierarchical unification of LIRICS and VerbNet semantic roles. In *2011 IEEE fifth international conference on semantic computing (IEEE)*, 483–489.
- Brown, Susan W., Julia Bonn, Ghazaleh Kazeminejad, Annie Zaenen, James Pustejovsky & Martha Palmer. 2022. Semantic representations for NLP using VerbNet and the generative lexicon. *Frontiers in artificial intelligence* 5.

- Brown, Susan W., Dmitriy Dligach & Martha Palmer. 2014. VerbNet class assignment as a WSD task. In *Computing meaning*, 203–216. Springer.
- Brown, Susan W., James Pustejovsky, Annie Zaenen & Martha Palmer. 2018. Integrating generative lexicon event structures into VerbNet. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*, 56–61.
- Carreras, Xavier & Lluís Màrquez. 2005. Introduction to the CoNLL-2005 shared task: Semantic role labeling. In *Proceedings of the ninth conference on computational natural language learning (CoNLL-2005)*, 152–164.
- Condoravdi, Cleo, Dick Crouch, Valeria De Paiva, Reihardt Stolle & Danny G. Bobrow. 2003. Entailment, intensionality and text understanding. In *Proceedings of the HLT-NAACL 2003 workshop on text meaning*, 38–45.
- Cooper, Alexis. 2024. Differing implications in caused motion constructions. Unpublished paper, University of Colorado Boulder.
- Crouch, Dick, Mary Dalrymple, Ronald Kaplan, Tracy Holloway King, John T. Maxwell III & Paula Newman. 2011. *XLE Documentation*. Palo Alto Research Center (PARC). <https://ling.sprachwiss.uni-konstanz.de/pages/xle/index.html>.
- Crouch, Dick & Tracy Holloway King. 2004. Unifying lexical resources. In *Proceedings of the interdisciplinary workshop on the identification and representation of verb features and verb classes*.
- Dang, Hao T., Karin Kipper, Martha Palmer & Joseph Rosenzweig. 1998. Investigating regular sense extensions based on intersective Levin classes. In *36th annual meeting of the Association for Computational Linguistics (ACL) and 17th international conference on computational linguistics*, vol. 1, 293–299.
- Fillmore, Charles J. 1988. The mechanisms of construction grammar. In Shelley Axmaker, Annie Jaissner & Helen Singmaster (eds.), *Proceedings of the Berkeley linguistics society (BLS)*, vol. 14, 35–55.
- Gawron, Jean Marc. 1985. Situations and prepositions. *Linguistics and Philosophy* 9(4). 427–476.
- Giuglea, Ana-Maria & Alessandro Moschitti. 2006. Semantic role labeling via FrameNet, VerbNet and PropBank. In *Proceedings of the 21st international conference on computational linguistics and the 44th annual meeting of the Association for Computational Linguistics (ACL)*, 929–936. Association for Computational Linguistics.
- Gleitman, Lila. 1990. Structural sources of verb learning. *Language Acquisition* 1(1). 1–63.
- Goldberg, Adele. 1995. *A Construction Grammar approach to argument structure*. The University of Chicago Press.

- Goldberg, Adele. 2005. *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Goldberg, Adele & Ray Jackendoff. 2004. The English resultative as a family of constructions. *Language* 80. 532–568. DOI: 10.1353/Language,2004.0129.
- Hwang, Jena D. 2014. *Identification and representation of caused motion constructions*. University of Colorado at Boulder. (Doctoral dissertation).
- Iwata, Seizi. 2020. *English resultatives: A force-recipient account*. John Benjamins Publishing Company.
- Kay, Paul & Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The what's X doing Y? construction. *Language* 75(1). 1–33.
- Kipper, Karin, Hao Trang Dang & Martha Palmer. 2000. Class-based construction of a verb lexicon. In *Proceedings of the seventeenth national conference on artificial intelligence and twelfth conference on innovative applications of artificial intelligence*, 691–696.
- Kipper, Karin, Anna Korhonen, Neville Ryant & Martha Palmer. 2006. Extending VerbNet with novel verb classes. In *LREC*, 1027–1032.
- Kipper, Karin, Anna Korhonen, Neville Ryant & Martha Palmer. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation* 42(1). 21–40.
- Levin, Beth. 1993. *English verb classes and alternations*. The University of Chicago Press.
- Levin, Beth. 2006. English object alternations: A unified account. Unpublished manuscript, Stanford University.
- Loper, Edward, Szu-Ting Yi & Martha Palmer. 2007. Combining lexical resources: Mapping between PropBank and VerbNet. In *Proceedings of the 7th international workshop on computational linguistics*.
- Michaelis, Laura A. 2004. Type shifting in Construction Grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15. 1–67.
- Open AI. 2023. *GPT-4 technical report Open AI*. <https://arxiv.org/abs/2303.08774>.
- Pustejovsky, James. 1991. The syntax of event structure. *Cognition* 41. 47–81.
- Ruppenhofer, Joseph, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, Collin F. Baker & Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice*. Institut für Deutsche Sprache, Bibliothek.
- Schuler, Karin K. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania.
- Shi, Lei & Rada Mihalcea. 2005. Putting pieces together: Combining FrameNet, VerbNet and WordNet for robust semantic parsing. In *International conference on intelligent text processing and computational linguistics*, 100–111. Springer.

- Weissweiler, Leonie, Abdullatif Köksal & Hinrich Schütze. 2024. *Hybrid human-LLM corpus construction and LLM evaluation for rare linguistic phenomena*. Submitted for publication. <https://arxiv.org/abs/2403.06965>.
- Wilson, Michael, Jackson Petty & Robert Frank. 2023. How abstract is linguistic generalization in large language models? Experiments with argument structure. In *TACL*, vol. 11, 1377–1395. <https://doi.org/10.1162/tacl%20a%2000608>.
- Zaenen, Annie, Cleo Condoravdi, Danny Bobrow & Raphael Hoffmann. 2010. Supporting rule-based representations with corpus-derived lexical information. In *Proceedings of the NAACL HLT 2010 first international workshop on formalisms and methodology for learning by reading*, 114–121.

Part IV

Applications

Chapter 13

Semantic Representation in Biomedical Text Mining

David Milward^a

^aIndependent Scholar

Computational semantics is concerned with mapping from natural language text to abstract semantic representations that can be used for inference. This paper describes the kinds of representation that have been used in academic and commercial tools in the field of biomedical text mining, including representation of entities via ontologies, and relationships via tabular data, JSON or RDF. It explores some of the operations that are available for manipulating these representations in spreadsheets, relational databases, NoSQL databases, and the semantic web.

1 Introduction

Information extraction uses Natural Language Processing (NLP) to take one or more sentences and fill out relevant information, e.g. within a template or a form. Interest in mining biomedical text for scientific research has increased both in the academic and commercial world as the number of scientific papers has expanded (see e.g. the trend in MEDLINE citations by year (MEDLINE 2024)). This increased volume of literature makes it difficult for researchers to keep abreast of the information they require, even within their own specialties, without the use of automated tools. Interest in mining clinical records for patient care and for medical research has also increased due to the improvement of the tools available and to the greater availability of data as hospitals have moved to electronic medical records.

Although information extraction has long been considered a key application area for computational semantics (Cooper et al. 1996), most systems that have

been used for biomedical information extraction do not use the traditional tools of computational semantics such as compositional semantics and logical inference. However, they do employ abstract semantic representations and some of the operations commonly used on these representations can be reformulated in terms of inference.

This paper describes several kinds of representation that are used as the output of biomedical text mining systems, including different variants of tabular data, JSON (JavaScript Object Notation (JSON 2024)), RDF (Resource Description Framework (RDF 2024)) and BEL (Biological Expression Language (BEL 2024)). It also looks at the kinds of use that are made of these representations within spreadsheets, databases or the semantic web. This includes, but is not limited to inference.

2 Introduction to biomedical text mining

Biomedical text mining includes the extraction of information from scientific articles, patents and drug labels for biological research and drug discovery, and the extraction from clinical records for medical research, patient care and selecting patients for clinical trials. An important application of biomedical text mining concerns understanding relationships between genes and diseases. Genes correspond to pieces of DNA that lead to production of particular proteins which improve or worsen disease (among other things) either directly or indirectly via interaction with other proteins.

Mining interactions between proteins was an early application of information extraction in biology (e.g. Thomas et al. 2000). Protein-protein interactions can be extracted without trying to understand the whole document, or even the whole sentence. For example, systems can focus on the fragment “the direct interaction of Tat and P-TEFb” containing the two proteins, Tat and P-TEFb, from within a much more complex sentence such as in (1).

- (1) The stimulatory effect involved **the direct interaction of Tat and P-TEFb** and was evident at the earliest step of TC assembly, the TBP-TATA box interaction.

From the fragment, a relationship can be populated in a template, or added to a table of protein-protein interactions such as that in Figure 1.

Systems can be configured or trained to recognize and identify protein names such as “Tat” or P-TEFb”, and to deal with different fragment constructions such

Entity	Relation	Entity
TAT	interact	P-TEFb
HSP105	associate	HSC70
CUL1	interact	SKP2

Figure 1: A table of protein-protein interactions extracted from scientific articles

as “TAT interacts with P-TEFb”, “TAF, which interacts with P-TEFb” or “the interaction of TAT with P-TEFb”.

Recognizing and identifying protein names is non-trivial. Although there are attempts at standardization such as the NIH’s International Protein Nomenclature Guidelines (NIH 2024), there can be a lot of variation in the way a single protein is expressed in practice, including alternative full names and alternative acronyms. There are also variations in how terms are written including different hyphenations and use of Roman or Greek letters e.g. Raf Alpha vs. Raf α or Raf 1 vs. Raf-1 vs. Raf I.

Competitions in biological or biological text mining such as BioCreative (BioCreative 2024) have included finding mentions of entities such as proteins (named entity extraction), identifying named entities (named entity identification) or finding properties of entities or the relationships between them. For example, Track 1 of BioCreative 2023 (BioCreative Track 1 2024) included extraction of biomedical relationships such as disease-gene, protein-protein, and chemical-chemical.

Although medical records often include structured fields, much useful information for both research and clinical care is within the unstructured text. An example application is selecting patients for clinical trials, where text mining can avoid expensive and slow review of records by hand. Medical text mining is now provided by the world’s largest companies e.g. Google has a Healthcare Natural Language API (*GoogleAPI* 2024) and Amazon provides Amazon Comprehend Medical (Amazon 2024a). There are competitions such as i2b2 and its successor, n2c2 (n2c2 2024). For example, the 2018 n2c2 task focused on clinical trial selection (Stubbs et al. 2019), and the 2022 n2c2 medication task (n2c2-T1 2024) looked for medication mentions, any change in medication (start/stop), as well as annotating negation, temporality and certainty.

In biomedical text mining, fragments of text may be provided with a relatively simple representation such as a table of results, similar to the protein-protein interaction example above, relating one biomedical entity to another, or to a value. The entries in each column will typically be more normalized than the original

text, so multiple ways of saying the same thing will be collapsed together (e.g. “carcinoma of the breast” as “breast cancer”, “fourteen milligrams” as “14 mg”, “body mass index” as “BMI”). This is often done using biomedical ontologies, which will be discussed in the next section.

Although systems often focus on fragments within a sentence, wider sentence or document context may be used as part of the representations that is built, or as a filter to ensure only information from relevant fragments is extracted (more details are provided in the section below on dealing with context). Although this can involve the use of NLP, it is also possible to exploit XML tags that are often used to annotate electronic documents. These can inform whether the fragment is within the Introduction, Discussion, Background etc. Tagged documents of this kind are often referred to as semi-structured text.

Standards have been developed for XML tagging of scientific articles and patents. The Journal Article Tag Suite (JATS) was introduced to provide standardization of XML formatting in journal publishing (National Center for Biotechnology Information 2024). The st.36 XML format (WIPO 2024) is recommended by the World Patent Office, WIPO, for patents, and this provides useful distinctions e.g. between information within the Background section or from a Claim. Tables within both JATS or st.36 documents may also be structured using XML, and this can be particularly useful for extracting numerical information which is often provided in tables rather than in free text.

The results of biomedical text mining may be used to feed machine learning algorithms or to populate spreadsheets, relational databases or the semantic web. The primary driver for choosing a representation for the output is typically the existing format of the information that the text mining results will be added to or merged with.

Understanding the kinds of representations used to support applications of text mining is important when considering the whole process. It is also important for interactive systems (e.g. Ontology Based Interactive Information Extraction (Milward et al. 2005)) where end-users can directly access the output. However, this does not mean that all biomedical text mining systems need to support a number of different representations. For systems designed primarily for automated workflows it is possible to use a standard semantic representation for the output as long as the information required downstream is not lost. Mapping to required output formats is then achieved by adding an extra translation step in the workflow that maps from the NLP output format into the required end-user format. An example of a standard output format for an NLP system (Amazon Comprehend Medical (Amazon 2024a)) is given in section 3.1.

This paper provides more background on semantic representation in general and on the importance of dealing with context. Different representations and their uses are then summarized. Inference is important, but other motivations for abstract semantic representations will also be discussed.

3 Background on semantic representation in biomedical text mining

3.1 Semantic representations of entities

Biology and medicine benefit from well-curated ontologies. These typically incorporate a thesaurus (i.e. capture synonymy) and a taxonomy (i.e. is-a relationships), and sometimes include more complex ontological relationships. Many text mining tasks involve mapping to the namespace of a particular ontology (e.g. Mädche et al. (1999)), with the use of a semantic ID or a preferred term (a human-readable concept name which is uniquely mapped to a specific ontology ID). If there is existing structured data using the same namespace, then it is possible to aggregate text-mined information with the pre-existing information using the shared identifiers.

Examples of common biomedical ontologies for diseases or adverse reactions include MedDRA (MedDRA 2024), SNOMED CT (*SnoMed* 2024) and MeSH (MeSH 2024). Biomedical ontologies for drugs or chemicals include RxNorm (*RxNorm* 2024) and ChEBI (ChEBI 2024). Many ontologies are readily available from portal web sites such as BioPortal (BioPortal 2024). Some require licensing, particularly for commercial use.

There is no single or consistent set of ontologies used by everyone. Different communities such as medical practitioners, medical researchers or biopharma researchers use different ontologies. Even within a particular research field some researchers may prefer the hierarchy of one ontology to another, or may use a particular ontology since it has a particular strength in their research area e.g. particularly good cancer coverage. Some standardization does occur due to reporting requirements e.g. the FDA (US Food and Drug Administration) requires the use of MedDRA for adverse events, and the WHO (World Health Organization) specifies ICD-11 (ICD 2024) for clinical coding of diseases.

Another issue with ontologies is that they change over time as scientific knowledge changes. In text-mined results, it is useful to include the version or date stamp from any ontologies used.

For example, Amazon Comprehend Medical (Amazon 2024a) can provide entities with a SNOMED CT code and a description (the Preferred Term) along with

the original text, and the version of the SNOMED CT ontology used. A fragment of example output (Amazon 2024b) is provided below. This is provided in JSON format which is an open standard data exchange format providing hierarchical structure. Many programming languages such as Python provide support to read and parse JSON data structures.

```
{
  "Entities": [
    {
      "Category": "MEDICAL_CONDITION",
      "BeginOffset": 39,
      "EndOffset": 59,
      "Text": "oropharyngeal lesion",
      "SNOMEDCTConcepts": [
        {
          "Code": "33431000119109",
          "Description": "Lesion of oropharynx (disorder)"
        }
      ]
    }
  ],
  "SNOMEDCTDetails": {
    "Edition": "US",
    "VersionDate": "20200901",
    "Language": "en"
  }
}
```

Figure 2: A fragment of JSON output from Amazon Comprehend Medical

It is also useful to include the original natural language text e.g. “oropharyngeal lesion” as in the Amazon Comprehend Medical output in figure 2, since this may need to be reinterpreted if the ontology information goes out of date. For example, the original text might have been quite precise, but the best ontology concept at the time might have been more generic. After several ontology iterations, a better-matching, more specific concept might become available.

Ontologies can be used for known entities. However, there are cases which require a representation for new entities. For example, when mining patent documents it is useful to be able to represent the compounds that have been newly defined within the patent in a way that allows further processing.

Name-to-structure algorithms such as OPSIN (Lowe et al. 2011) can take textual descriptions of novel, newly named chemicals and map them to a representation

of the chemical structure in InChI (International Chemical Identifier) (Heller et al. 2013) or SMILES (Simplified Molecular-Input Line-Entry System) (Weininger 1988) formats. There are also dictionaries which map existing drug names to their corresponding SMILES or InChI structures. For example, the SMILES representations for the following compounds are as follows:

- (2) a. Nitrobenzene c1c(N(=O)=O)cccc1
- b. 2,2-Dimethylbutane CC(C)(C)CC

Once chemicals are represented by their structures, this allows filtering based on structure. For example, chemicals can be found which are similar in structure to a chemical of interest (chemical similarity search) or have a particularly interesting substructure (chemical substructure search).

Finally, we should consider numerical entities. Similar to novel chemicals, many numerical entities cannot be mapped to a finite namespace. However, numerical entities can be standardized in various ways. For example, dates or times can be mapped to formats defined by the International Organization for Standardization (ISO), fractions can be mapped to decimals (e.g. $\frac{1}{2}$ to 0.5), units can be mapped to symbols (e.g. milligram to mg) or to powers of units defined by the International System of Units (SI) e.g. 10⁻³g. More complex units such as g/kg can be normalized to mg/g or represented using UCUM (Unified Code for Units of Measure) (Schadow et al. 1999). To allow comparisons between measurements (comprising a value and a unit) or to allow filtering by range, these can be converted to the same unit where this is possible e.g. 4g mapped to 4000mg, or even 5 feet to 152.4cm.

3.2 Dealing with context

Can the output of a text mining system be regarded as a semantic representation if it does not encode context from the document, or even context from within the same sentence?

Just as biomedical text mining does not typically try to interpret the whole sentence, it also does not try to interpret the whole context. However, it may extract or filter based on relevant parts of the context. Example contextual information includes:

1. Information about the document
 - a) What kind of document e.g. a patent, journal article, patient record, laboratory report

- b) What is the document about e.g. a patent class code using the CPC standard (Cooperative Patent Classification) (*Cooperative Patent Classification 2024*)
 - c) Where was it published e.g. Nature
 - d) Who was the author
 - e) How old is it e.g. patent priority date for establishing prior art
2. Information concerning the position within the document
- a) What section does the sentence occur in e.g. Background, Conclusion, Methods, Patent Claims
 - b) Does the sentence occur at the start or end of the document
3. Information from other sentences or from within the sentence
- a) Conditions e.g. inside or outside an organism (in vitro vs in vivo)
 - b) Tense e.g. a current or past condition
 - c) Species e.g. human vs. monkey vs. rat
4. Information from the sentence itself
- a) Stated vs. suggested e.g. there was or may be a relationship between A and B
 - b) Negated or not

One approach is to add contextual information of interest to the representation e.g. as an extra column in a table. The other approach is to filter to relevant context as part of the text mining strategy. In this case, similar factors may be used such as the patent code, or the status of the assertion, but the fragment is only analysed if contextual parameters are satisfied e.g. that the document is from Nature and refers to an in vitro experiment. In this approach, the semantic representation typically only includes a representation of the semantics of the fragment rather than its context.

Structural context is particularly important when mining certain kinds of documents. For example, when mining a patent, something in the Claims section has a very different status to something in the Background section. This information can be obtained by parsing the XML structure of the document when available.

Just as rule-based and statistical systems can be used to interpret fragments, they can also be used for context. For example, statistical systems can be used to

filter to particular contexts by the system being trained on relevant vs. irrelevant example contexts. They can also be used to populate extra columns by being trained on specific aspects of the context. Similarly, rule-based systems do not typically try to understand the whole context, but look for relevant wording to distinguish between particular aspects. When used to filter to specific contexts, a rule-based system may explicitly code the required context as a set of constraints within a declarative query language (e.g. the Extraction and Search Language (EASL) of Linguamatics I2E (Cormack et al. 2015), as used, e.g. by Trivedi et al. (2020)).

It is not always clear at the start of a text mining task which parts of context will be relevant. Negation can be a good indicator of what needs to be included or filtered out by the text mining strategy. Consider if two apparently contradictory statements are extracted:

- (3) a. A does not interact with B
- b. A interacts with B

This may be a genuine contradictory finding by two separate authors, but it can also be a difference due to context, e.g. one applies *in vitro*, the other *in vivo*. In fact, negative findings in a specific context are often interesting only because the positive is usually true. In these cases, the strategy may need to be changed to either include extra context in the representation or to filter out contexts which are not relevant for the task.

Finally, context may also be relevant for establishing the meaning of the fragment itself. It may be required for disambiguation e.g. to ensure RAF is mapped to its protein meaning in a biological text vs. its meaning of Royal Air Force in a medical text referring to RAF hospitals. It may also be required for co-reference e.g. to ensure that “it” or “the protein” is linked to a prior mention of “RAF”.

3.3 Adding constraints

Context can be regarded as one example of adding a constraint on text mining output. Another example is the ontologies themselves. Just as there are two approaches to dealing with context, i.e. add context to the semantic representation to allow post-filtering or pre-filter within the text mining strategy, there are two approaches to dealing with ontology filtering. Ontology-based interactive information extraction (Milward et al. 2005) incorporates ontologies within the text mining strategy so that output is pre-filtered to particular classes of interest, whether this is a particular protein or disease, or a protein family or class of

diseases. For example, refining to the cancer class provides all results referring to cancer or any type of cancer such as “breast carcinoma” or “leukemia”, and the ways they might be referred to such as “cancer of the breast” or “leukaemia”. The alternative approach is to output everything and then later use inferences based on ontological structure to constraint the results to those most relevant. RDF (RDF 2024) is particularly well suited to supporting ontological inference.

4 Semantic representation in biomedical text mining

4.1 Tabular formats for semantic representation

Tabular data is a convenient format for loading into spreadsheets or databases. Tabular data can also be used as input to packages for statistical analysis and machine learning. Examples include open source libraries such as Pandas for Python (*Pandas* 2024), and commercial packages such as SAS (*SAS Institute Inc.* 2024). An example use case is the analysis of trends in heart failure using text mining to feed a SAS analysis (Ambrosy et al. 2021).

Spreadsheets or databases can input data in their own proprietary formats, e.g. XML for Excel, but also support import of tabular data in standard formats such as comma separated value (CSV) or tab separated value (TSV) files, with the option to treat the first line as a header. The table in Figure 1 represented as a tab separated plain text file including the column headers would look as in Figure 3.

Entity	Relation	Entity
TAT	interact	P-TEFb
HSP105	associate	HSC70
CUL1	interact	SKP2

Figure 3: A tab separated file of protein interactions

A typical table may include columns containing the preferred terms or the node identifiers of concepts from biomedical ontologies or numerical values. Relationships ideally are normalized to preferred terms or node identifiers of a relationship ontology or, if this is not available, may be normalized to some extent by being mapped to the infinitive form of the verb (e.g. “phosphorylated” to “phosphorylate”).

More complex relationship information or contextual information can be added as extra columns. It is typical in spreadsheets to have relatively large numbers of columns. For example, patient information such as Weight and Body Mass

Index (BMI) might be expressed in tabular format as in (4), with all the properties of the patient in different columns.

(4)

Patient ID	BMI	Weight (kg)
0001	30	90
0002	25	70
0003	33	100

For databases there is more opportunity to join information back together so smaller tables can be used. For example, there might be a separate table for each patient property:

(5)

a.	Patient ID	BMI	b.	Patient ID	Weight (kg)
	0001	30		0001	90
	0002	15		0002	40
	0003	33		0003	100

The split into smaller tables can better reflect an underlying semantics or fit more directly with the output of text mining.

4.2 Spreadsheets

Spreadsheet formats such as Microsoft Excel are a common way of passing around information, and consequently, provide a useful output for text mining, especially for one-off research tasks.

What can we do with information in a spreadsheet that we could not have done with the original sentences? Firstly, we can sort the information, choosing particular columns of semantic entities for priority sorting. This makes it easier to find information. It also gathers together information about a concept, which would otherwise have been kept separate, and can thereby provide new insights. Secondly, the spreadsheet information can be used to create graphs allowing trends to be visualized, e.g. a patient's body mass index (BMI) over several years or the size of a patient's tumor over several weeks. Thirdly, the information can be filtered. For example, if a clinical trial wants to select patients based on a set of criteria, the spreadsheet can be filtered to the relevant patients, e.g. ones at a particular hospital with a particular condition. Formulae can be used for numerical criteria to filter on one column, e.g. Body Mass Index is less than 17.5, or filter on multiple columns, e.g. to select patients where a more recent value is greater than an earlier one.

These capabilities show that even a basic tabular format provides a level of semantic representation that abstracts from the original text and allows useful operations at the semantic level. Spreadsheet filtering capabilities can be regarded as a type of inference.

A test suite entry in the style of FraCaS (Cooper et al. 1996) could be:

(6) Patients eligible for the trial must have a BMI less than 17.5.

George Smith has a BMI of 25.

Is George Smith eligible for the trial? [No]

This is similar to inference over subintervals (Cooper et al. 1996)

(7) (3.252) Since 1992 ITEL has been in Birmingham.

It is now 1996.

Was ITEL in Birmingham in 1993? [Yes]

In summary, tabular formats as used in spreadsheets provide a simple to understand semantic representation for the output of text mining. Example capabilities provided by spreadsheets for information manipulation include graphs of trends, filtering of data via numerical ranges, and sorting.

4.3 Relational databases

Pharmaceutical companies and healthcare organizations store large amounts of information as tables within relational databases. This typically includes proprietary information but also information downloaded from open-source or commercial databases.

Text-mined results may be used to add to this, either by creating extra columns to existing tables, adding to existing tables, or generating new separate tables of results, often as part of automated workflows. Relational databases provide ways to manipulate information using the SQL query language. In addition to allowing sorting and filtering, SQL allows joining of two or more different tables through common identifiers in particular columns. This enables the discovery of previously unknown information by finding new relationships constructed from information found in two different documents, or information found in a document with information from a structured data source.

For example, if there is a row in a protein-protein relationship table relating EGFR to activation of MAP3K1, and a row in another table relating MAP3K1 to breast cancer, we may infer the possibility of an indirect relationship between EGFR and breast cancer.

(8) Given

EGFR \rightarrow MAP3K1MAP3K1 \rightarrow Breast Cancer

There is a potential indirect relationship

EGFR \rightarrow Breast Cancer

Similar indirect relationships can be used to repurpose an existing drug for the treatment of a new disease (drug repurposing). The drug being repurposed could already be on the market or it might have been found to be safe but not as effective as existing drugs. In both cases there will be interest in seeing if the same drug can be used for other diseases, since much of the expensive safety testing can be reused.

In a typical scenario for drug repurposing, there may be a table of relationships between the drugs and the proteins they interact with, e.g. Drug X with Protein Y, and another table extracted from the literature which associates proteins with diseases, e.g. Protein Y with Disease Z. The join between the two tables on the Protein column then suggests Drug X as a potential treatment for Disease Z i.e.

(9) Given

Drug X \rightarrow Protein YProtein Y \rightarrow Disease Z

There is a potential indirect relationship

Drug X \rightarrow Disease Z

Another application of database joins is for finding the mechanism underlying a side effect. If a clinical trial shows a side effect, it is critical to discover the mechanism. The relationships here can be similar to drug repurposing, but looking at the causes of a specific side effect rather than treating a disease. There may be a table from the literature relating the drug to all the proteins it effects, e.g. Drug X to Protein Y, and another table from the literature relating any protein to the specific side effect found in the clinical trial, e.g. Protein Y causes Side Effect Z. The join then provides a potential indirect relationship between Drug X and Side Effect Z which can be investigated as a possible explanation.

(10) Given

Drug X \rightarrow Side Effect Z

And given

Drug X \rightarrow Protein YProtein Y \rightarrow Side Effect ZDrug X \rightarrow Protein Y \rightarrow Side Effect Zis a potential mechanism for Drug X \rightarrow Side Effect Z

The examples above have involved a single join, i.e. $A \rightarrow B$ and $B \rightarrow C$ giving $A \rightarrow C$. However, two joins can be used to create longer distance indirect relationships e.g. $A \rightarrow B$, $B \rightarrow C$ and $C \rightarrow D$ giving $A \rightarrow D$.

A test suite entry in the style of FraCaS (Cooper et al. 1996) for one of these indirect relationships could be:

- (11) EGFR causes increased MAP3K1 expression.
Increased MAP3K1 expression is associated with breast cancer.
Is EGFR associated with breast cancer? [Yes]

Providing transitive relationships by database joins is similar to the transitive temporal inferences concerning temporal subordinate clauses (Cooper et al. 1996).

- (12) (3.262) Smith left after Jones left.
Jones left after Anderson left.
Did Smith leave after Anderson left? [Yes]

In summary, tabular data as used in databases provides a simple semantic representation of the output of text mining with more flexibility to split into separate, smaller tables. Using a database provides tools for access control and data integrity, and for adding or joining to existing structured data. This is in addition to typical capabilities for graphing trends, filtering or sorting.

4.4 RDF (Resource Description Framework)

RDF (RDF 2024) is one of the foundational components of the semantic web. An RDF statement consists of a triple of a node for the subject, an arc for the relationship and a node for the object. The nodes and relationships are either literal values or specified by a URI (Uniform Resource Identifier). The triples are stored in a triplestore and queried using the SPARQL query language.

In biomedical applications the URIs are often based on standard ontologies. For example, “<http://bio2rdf.org/go:0004396>” provides a URI for the gene ontology ID, 0004396 (*GeneOntology* 2024). For each ontology concept you can normally find multiple URIs so it is useful to use something similar to bio2rdf (Belleau et al. 2008) links to provide consistency.

RDF semantic triples combined with the SPARQL query language provide a convenient mechanism for capturing indirect relationships. For example, it is possible to find an indirect relationship between A and C by traversing the graph

of triples from A to B and from B to C. This covers examples such as (8) relating EGFR to breast cancer via MAP3K1.

RDF has some limitations in expressing more complex knowledge. It is similar to only having transitive verbs, without a direct way to express ditransitive verbs such as “John gave Mary a book”. It is possible to use a reified form of RDF where relationships are expressed as events, with triples provided for each attribute of the event, but this may limit the ability for efficient inference.

RDF can also provide inferences by combining text-mined results with knowledge of ontology structure. The ontology hierarchy is converted into RDF, e.g. using SKOS, Simple Knowledge Organization System (SKOS 2024). Once everything is in RDF, it is possible to traverse from an ontology node to its child, and then from the child via a text-mined relationship arc.

For example, the following inference would be possible:

- (13) Leukemia is-a cancer.
George Smith has leukemia.
Therefore, George Smith has cancer.

This kind of inference may be useful if we wanted to select cancer patients for priority treatment, ensuring that patients with leukemia get priority.

A test suite entry in the style of FraCaS (Cooper et al. 1996) could be:

- (14) Priority treatment is given to cancer patients.
Leukemia is a type of cancer.
Do leukemia patients get priority treatment? [Yes]

This is similar to the monotonicity inferences provided in FraCaS D16 such as:

- (15) (3.17) An Irishman won the Nobel prize for literature.
Did an Irishman win a Nobel prize? [Yes]

In summary, RDF allows easy integration with open-source semantic web data, including ontological information. The SPARQL query language allows for information manipulation including indirect relationships and inference using ontology structure. There are also tools available for graph visualization which can help with the understanding of interaction pathways.

4.5 NoSQL databases

NoSQL approaches provide more flexibility in data format than relational databases. There are a wide variety of NoSQL approaches including document databases, key-value databases, wide column databases and graph databases. The focus here is on document databases and graph databases as they are the most different from relational databases (section 4.3) in the way they structure information.

4.5.1 Document-oriented databases

Document-oriented databases such as MongoDB (*MongoDB* 2024) store data in documents similar to JSON data structures. These provide fields and values, but, unlike tables in relational databases, also allow for nested hierarchical structures. For example, information about an individual can be stored as in Figure 4.

```
{
  "id": 12345,
  "name": "John Smith",
  "age": 42
  "hobbies": {
    "indoor": [
      "Chess"
    ],
    "outdoor": [
      "Basketball", "Tennis"
    ]
  }
}
```

Figure 4: Document-oriented database example entry

Document-oriented databases include many of the capabilities available for relational databases, including the ability to sort by field or perform range filtering, along with free text searching. The lookup function in MongoDB is similar to a database join and so allows for derivation of indirect relationships.

In summary, document-oriented NoSQL databases allow for more natural hierarchical semantic representations while providing the benefits of a database for access control, as well as capabilities for data visualization and querying.

4.5.2 Graph databases

A graph database uses nodes, edges, and properties instead of tables or documents to represent and store data. Graph databases such as Neo4j (Neo4j 2024) can be used similarly to RDF. For example, they can represent interaction networks to help discover the mechanism behind a side effect or to combine ontology relationships with relationships extracted from text. For the example in (8) this would involve using the Cypher query language in Neo4j to find a path between the node “EGFR” and the node “breast cancer” by traversing the graph from the node “EGFR” to the node “MAP3K1”, and then between “MAP3K1” and “breast cancer”.

Unlike database joins where it is necessary to specify the number of joins involved in an indirect relationship, a graph database allows discovery of paths of arbitrary length between two concepts using shortest path algorithms, e.g. Dijkstra Source-Target Shortest Path (Needham & Hodler 2024).

In summary, a graph database allows the representation of relationships as a network helping to clarify mechanisms, e.g. interaction pathways. They allow discovery of indirect relationships of arbitrary distance, and allow inference using ontology structure. They support access control as well as standard operations such as range filtering and sorting.

4.6 BEL (Biological Expression Language)

BEL (BEL 2024) is not a well-known format, but it is useful to include here as an example of a domain-specific language designed to allow specific inferences and to provide more readable semantic representations. This extra readability is useful when annotating corpora or checking results from text mining. The BEL representation can be automatically converted to other formats such as reified RDF for further processing.

BEL is designed to capture causal, correlative, and associative relationships between biological entities along with the experimental/biological context in which they were observed, as well as the provenance of the publication from which the relation was reported.

BEL can capture statements which are not naturally expressed as triples involving two entities and a relationship. For example, the following sentence can be translated into BEL format to capture a relationship between an entity and a process (simplified from an example provided by *Selventa Knowledge* 2024).

(16) Azaserine reversed the stimulatory effect of glucose on c-Jun.

a(SCHEM:Azaserine) decreases (a(CHEBI:glucose) increases
tscript(p(RGD:Jun)))

Translating back you might get “Azaserine decreases the increase of c-Jun activity by glucose”. *p* stands for protein family, *a* for abundance, *tscript* for transcriptional activity. The named entities are normalized to the preferred terms from ontologies, e.g. glucose from ChEBI.

In summary, BEL provides the ability to more naturally represent complex relationships of interest in systems biology. This helps readability of the semantic representation which is useful when annotating or checking results. It can be used as an intermediate representation which is then translated into another format such as RDF.

5 Conclusion

Although much of biomedical text mining may look far removed from classical computational semantics, this paper has shown that some similar notions of representation and inference apply. Biomedical representations have a particularly strong focus on the semantics of entities. These typically use ontology identifiers, but it is also possible to use representations based on chemical structure to capture newly patented chemicals. Relatively simple representations for relationships or properties provide strong capabilities for sorting, filtering or joining which would not be available when using free text directly. Even though some representations do not provide a rich context, this is often substituted by contextual filtering within the text mining strategy. Similarly, ontologies filtering within text mining strategies may be used as a substitute for ontology inference.

Acknowledgments

I am very grateful to Ian Lewin and Sneha Bhat for their comments on earlier drafts of this paper. I am also appreciative of the useful comments of the editors and an anonymous reviewer.

References

Amazon. 2024a. *Amazon Comprehend Medical*. On-line documentation. <https://aws.amazon.com/comprehend/medical/>.

- Amazon. 2024b. *Amazon Comprehend Medical SNOMED CT linking*. On-line documentation. <https://docs.aws.amazon.com/comprehend-medical/latest/dev/ontology-linking-snomed.html>.
- Ambrosy, Andrew P., Rishi V. Parikh, Sue Hee Sung, Anand Narayanan, Rajeev Masson, Phuong-Quang Lam, Kevin Kheder, Alan Iwahashi, Alexander B. Hardwick, Jesse K. Fitzpatrick, Harshith R. Avula, Van N. Selby, Xian Shen, Navneet Sanghera, Joaquim Cristino & Alan S. Go. 2021. A natural language processing-based approach for identifying hospitalizations for worsening heart failure within an integrated health care delivery system. *JAMA Network Open*. November 1, 2021. DOI: 10.1001/jamanetworkopen.2021.35152.
- BEL. 2024. *Biological expression language*. Online documentation. <https://biological-expression-language.github.io/>.
- Belleau, François, Marc-Alexandre Nolin, Nicole Tourigny, Philippe Rigault & Jean Morissette. 2008. Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. *Journal of Biomedical Informatics* 41(5). 706–716. DOI: 10.1016/j.jbi.2008.03.004..
- BioCreative. 2024. *Biocreative Workshop*. On-line workshop materials. <https://biocreative.bioinformatics.udel.edu/>.
- BioCreative Track 1. 2024. *Track 1: BioRED (Biomedical Relation Extraction Dataset) Track*. On-line task materials. <https://biocreative.bioinformatics.udel.edu/tasks/biocreative-viii/track-1/>.
- BioPortal. 2024. *BioPortal: Ontologies*. On-line resource. <https://bioportal.bioontology.org/ontologies>.
- ChEBI. 2024. *(chemical entities of biological interest)*. Online resource. <https://www.ebi.ac.uk/chebi/aboutChebiForward.do>.
- Cooper, Robin, Dick Crouch, Jan van Eijck, Chris Fox, Josef van Genabith, Jan Jaspars, Hans Kamp, David Milward, Manfred Pinkal, Massimo Poesio & Steve Pulman. 1996. *Using the framework: FraCaS deliverable D16*. Tech. rep. <https://gu-clasp.github.io/multifracas/D16.pdf>.
- Cooperative Patent Classification*. 2024. On-line documentation. <https://www.cooperativepatentclassification.org/>.
- Cormack, James, Chinmoy Nath, David Milward, Kalpana Raja & Siddhartha Jonnalagadda. 2015. Agile text mining for the 2014 i2b2/UTHealth cardiac risk factors challenge. *Journal of Biomedical Informatics* 15. DOI: 10.1016/j.jbi.2015.06.030.
- GeneOntology*. 2024. Online resource. <https://www.geneontology.org/>.
- GoogleAPI*. 2024. On-line documentation. <https://cloud.google.com/healthcare-api/docs/concepts/nlp>.

- Heller, Stephen, Alan McNaught, Stephen Stein, Dmitrii Tchekhovskoi & Igor Pletnev. 2013. InChI – the worldwide chemical structure identifier standard. *Journal of Cheminformatics* 5(1). PMC 3599061. DOI: 10.1186/1758-2946-5-7.
- ICD. 2024. *Icd-11 (international classification of diseases 11th revision)*. <https://icd.who.int/en>.
- JSON. 2024. *Introducing JSON*. On-line documentation. <https://www.json.org/json-en.html>.
- Lowe, Daniel M., Peter T. Corbett, Peter Murray-Rust & Robert C. Glen. 2011. Chemical name to structure: OPSIN, an open source solution. *Journal of Chemical Information and Modeling* 51(3). 739–53.
- Mädche, Alexander, Steffen Staab & Rudi Studer. 1999. Ontology-oriented information extraction and integration. In *Workshop on language technologies in information and knowledge management. 7th conference on computational linguistics of the German society for language technologies*, 7–8.
- MedDRA. 2024. *Medication dictionary for regulatory activities*. <https://www.meddra.org/>.
- MEDLINE. 2024. *Medline Citation Counts by Year of Publication*. https://www.nlm.nih.gov/bsd/medline_cit_counts_yr_pub.html.
- MeSH. 2024. *MeSH (medical subject headings)*. Online thesaurus. <https://www.ncbi.nlm.nih.gov/mesh>.
- Milward, David, Marcus Bjärelund, William Hayes, Michelle Maxwell, Lisa Oberg, Nick Tilford, James Thomas, Roger Hale, Sylvia Knight & Barnes Julie. 2005. Ontology-based interactive information extraction from scientific abstracts. *Computational Functional Genomics* 6. 67–71.
- MongoDB. 2024. Online documentation. <https://www.mongodb.com/>.
- n2c2. 2024. *National NLP Clinical Challenges (n2c2)*. On-line conference materials. <https://n2c2.dbmi.hms.harvard.edu/>.
- n2c2-T1. 2024. *National NLP Clinical Challenges (n2c2): Track 1 Contextualized Medication Event Extraction CMED*. On-line conference materials. <https://n2c2.dbmi.hms.harvard.edu/2022-track-1>.
- National Center for Biotechnology Information. 2024. *Journal Publishing Tag Set*. <https://jats.nlm.nih.gov/publishing/>.
- Needham, Mark & Amy E. Hodler. 2024. *Graph algorithms in Neo4j: Shortest path*. <https://neo4j.com/blog/graph-algorithms-neo4j-shortest-path/>.
- Neo4j. 2024. <https://neo4j.com/>.
- NIH. 2024. *International Protein Nomenclature Guidelines*. On-line documentation. https://www.ncbi.nlm.nih.gov/genbank/internatprot_nomenguide/.
- Pandas. 2024. On-line documentation. <https://pandas.pydata.org/>.

- RDF. 2024. *RDF (resource description framework)*. Online documentation. https://en.wikipedia.org/wiki/Resource_Description_Framework.
- RxNorm. 2024. Online documentation. <https://www.nlm.nih.gov/research/umls/rxnorm/index.html>.
- SAS Institute Inc. 2024. On-line documentation. <https://www.sas.com/>.
- Schadow, Gunther, Clement J. McDonald, Jeffrey G. Suico, Ulrich Föhring & Thomas Tolxdorff. 1999. Units of measure in clinical information systems. *Journal of American Medical Informatics Association* 6(2). PMID: 10094068; PMCID: PMC61354, 151–62. DOI: 10.1136/jamia.1999.0060151.
- Selventa Knowledge. 2024. On-line documentation. https://github.com/cthoht/selventaknowledge/blob/master/selventa_knowledge/small_corpus.bel.
- SKOS. 2024. *SKOS Simple Knowledge Organization System Reference*. On-line documentation. <https://www.w3.org/TR/skos-reference/>.
- SnoMed. 2024. Online documentation. <https://www.snomed.org/about-us>.
- Stubbs, Amber, Michele Filannino, Ergin Soysal, Samuel Henry & Özlem Uzuner. 2019. Cohort selection for clinical trials: n2c2 2018 shared task track 1. *Journal of the American Medical Informatics Association* 26(11). PMID: 31562516; PMCID: PMC6798568, 1163–1171. DOI: 10.1093/jamia/ocz163.
- Thomas, James, David Milward, Christos Ouzounis, Stephen Pulman & Mark Carroll. 2000. Automatic extraction of protein interaction from scientific abstracts. In *Pacific symposium on biocomputing (PSB)*, 541–552.
- Trivedi, Sapna, Roger Gildersleeve, Sandra Franco, Andrew S. Kanter & Afzal Chaudhry. 2020. Evaluation of a concept mapping task using named entity recognition and normalization in unstructured clinical text. *Journal of Healthcare Informatics Research* 4. 395–410.
- Weininger, David. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* 28(1). 31–36. DOI: 10.1021/ci00057a005..
- WIPO. 2024. *Recommendation for the Processing of Patent Information Using XML (Extensible Markup Language)*. <https://www.wipo.int/export/sites/www/standards/en/pdf/03-36-01.pdf>.

Chapter 14

An Information Theoretic Evaluation of AI Cooperation

Marisa Ferrara Boston^a

^aReins AI

Large Pre-Trained Language Models (LLMs) (Vaswani et al. 2017) and multi-agent systems (Chase 2022) are changing how humans and machines cooperate in organizations. This paper proposes a method to measure cooperation using principles from Information Theory (Shannon 1948), organizational behavior (March & Simon 1958), and Cooperative AI (Dafoe et al. 2021). The paper proposes a methodology to measure entropy on human-machine and multi-agent communication, and then validates the approach on the Enron Corpus (Klimt & Yang 2004). The results demonstrate that entropy can measure communication complexity during healthy organizational growth and during crises. Building on established natural language processing benchmarks (Condoravdi et al. 2003, Bowman et al. 2015, Kalouli et al. 2020), this work aims to advance the development of AI systems capable of effective human collaboration. By offering a new way to evaluate cooperative behavior, this method contributes to the increasingly important field of human-machine teaming and multi-agent systems.

1 Introduction

Historically, humans have used automation and artificial intelligence technologies (*machines*) as supervised tools, with humans commanding the technologies and controlling or reviewing the output. Yet the evolution from expert systems (Hayes-Roth 1983) to knowledge management (Alavi & Leidner 2001) to cognitive automation (Ferrucci et al. 2010) and Generative AI (Vaswani et al. 2017) has led to the possibility of machines acting as collaborators, peers, and even creators of information (Seeber et al. 2020). LLMs demonstrate remarkable fluency and

Marisa Ferrara Boston. 2025. An Information Theoretic Evaluation of AI Cooperation. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 323–332. Konstanz: PubliKon.

robustness to human interactions across a variety of use cases, making machine agency not only feasible, but also likely (Malone 2018, Butler et al. 2023).

Despite their fluency and robustness, LLMs demonstrate uncooperative behavior such as confabulations (Ji et al. 2022), objective-forgetting (Wang, Yang, et al. 2023), and unethical outputs (Cox 2022). Scholars worry that the increasing output of LLMs might suppress the diversity and dynamism of human knowledge, amplifying the hegemonic worldviews present in their limited training corpora (Bender et al. 2021). This paper offers a quantifiable metric for organizational health that can be sensitive to these issues. Specifically, knowledge sharing is not only a key tenet of human-machine teaming and multi-agent systems (see Table 1), it is also a critical component of organizational health demonstrating better decisions and the growth of organizational knowledge (Simon 1971, Malone 2018). The proposed metric, entropy, can serve as a suitable metric of knowledge¹ sharing. This study outlines a straight-forward method to evaluate entropy on any set of textual communications, from multi-agent system logs to human-machine chats to emails. The method is validated on the Enron Corpus, with plans outlined to expand this work into a larger information-theoretic benchmark for Cooperative AI.

Table 1: The principles of teaming intelligence (Johnson & Vera 2019).

Principle	Definition
Observability	Mutually sharing knowledge about past and present state to maintain common ground.
Predictability	Being expressive of one’s intent, the ability to receive information about the intentions of others, and the ability to predict future states.
Directability	Ability to follow explicit task and role assignments and subtle suggestions, guidance, and warnings.

¹Although information and knowledge are not synonymous, this paper aligns with organizational behavior theory in considering them similarly, as both contribute to the flow of meaningful content within organizations (March & Simon 1958).

2 Background

2.1 Organizational behavior

While this work targets cooperative behavior in machines, organizational science has historically emphasized the importance of effective human cooperation for organizational health. This field sets particularly useful precedents for Cooperative AI, outlining communication patterns that support decision-making and retain organizational knowledge (March & Simon 1958, Simon 1971). The performance measures that organizational science targets tend to be focused on metrics like efficiency, output quantity, or employee satisfaction because it is difficult to measure the level of information-sharing among team members. This work provides a solution by providing a measure that correlates information-sharing with performance. The metric could also be used to better understand effective retention and dissemination of institutional knowledge.

2.2 Information Theory

Information Theory (Shannon 1948) is seeing a resurgence in today's technical landscape, proving a potentially useful technique for tuning LLMs toward more efficient communication. Beyond efficiency, Shannon's work has provided a theoretical foundation for effective communication in linguistics, cognitive science, and psychology. This paper brings Shannon's theory to human-machine teaming, specifically toward measuring the amount of cooperation based on the complexity and amount of communication. Rather than targeting single-turn interactions, entropy determines the overall complexity across many interactions and agents. Entropy can therefore be used not only as a measure of cooperative behavior, but even a harbinger of organizational difficulty.

2.3 Cooperative AI

Human-only teaming relies on cooperative behavior that allows groups to share knowledge, understand goals and tasks, predict future states, and communicate status. Cooperative AI targets the tuning of LLMs toward similar behavior so they can interact with human teammates and each other in ways that promote organizational health while accomplishing objectives (Dafoe et al. 2021). Several evaluations of LLM coordination (Agashe et al. 2023) and planning (Wang, Cai, et al. 2023) highlight current deficiencies, particularly when expanded beyond game-play. Despite these issues, enterprises are investing in multi-agent systems to update their workflows and team with their employees (CBInsights 2024). The

focus of this research is to find a metric that can be used to benchmark machines' abilities to interface more naturally with a team by optimizing informativeness of communications. Machines could potentially facilitate even human-human communication when optimized toward this behavior (Malone 2018).

3 Entropy as a measure of cooperation

One of the key tenets of cooperation is the sharing of information, but not all communication is equally informative. Entropy (Shannon 1948) is a measure of the unpredictability of unique words in a vocabulary. When considered under the guise of human-machine teaming and even multi-agent systems, higher entropy should correlate to more diverse communications, indicating higher cooperation between entities.

This study uses vocabulary size as the information unit for calculating entropy, with the vocabulary defined by novel words within communication logs between any number of human and machine agents. These textual communication logs not only provide a corpus for measuring entropy following standard natural language processing techniques (Jurafsky & Martin 2009), but they also reflect tangible interactions between agents for measuring cooperation. Shannon's calculation of entropy H is the sum of the probability of each unique word p_i in a vocabulary multiplied by the log of that word's probability (Equation 1).

$$H = - \sum_i p_i \log_2(p_i) \quad (1)$$

Entropy is a useful metric because it can be calculated over any number of communication logs with diverse vocabulary sizes. Entropy can also be measured over time, allowing for the identification of periods of higher and lower levels of cooperation across tasks. While other information theoretic calculations like mutual information may also be translated to cooperative behavior, they often require experimentation with multiple variables difficult to encode within communication logs. Entropy is also simpler to calculate than state distribution metrics like KL-Divergence, where state encoding can be impractical or even impossible for human teaming. Current multi-agent framework logs may make this easier, which is a potential direction of future work.

4 Case study: Enron Corpus evaluation

This research uses the Enron Corpus (Klimt & Yang 2004) to test entropy’s ability to measure cooperative information sharing. While the Enron Corpus is typically used for email classification, it also provides insight for organizational behavior and communication research (Diesner et al. 2005). Because the corpus spans the years before, during, and after the crisis in October 2001, its 600,000 emails demonstrate patterns from healthy—even mundane—organizational communications to crisis communications.

Figure 1 shows the steps taken to process the messages from the Enron corpus so that vocabulary size and entropy can be calculated by month. The monthly time increment allows for a substantial enough vocabulary size to have meaningful entropy calculations while being fine-grained enough to cover critical organizational moments (like the crisis). Pre-processing stripped emails to their subjects and messages, and collated them by month. Vocabulary size and entropy were then calculated for each month.²

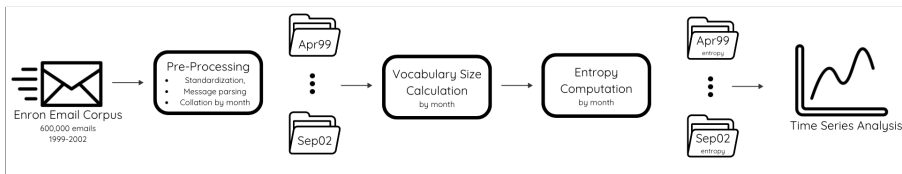


Figure 1: Methodology to calculate entropy on the Enron corpus.

Figure 2 shows entropy and vocabulary size results by month. The first month of the corpus is an outlier (largely because the email messages at this time reflected early email behavior in the organization). Following this first month, vocabulary size and entropy gradually increased, indicating healthy organizational communication and increasing complexity of operations.

In the months leading to October 2001, there is a decline in entropy leading to the sharp change at the point of the crisis. Following the crisis, there is a marked change in vocabulary sizes but relatively stable (and low entropy) until the eventual end. This timing maps to the public disclosure of accounting irregularities. One would predict the organization shifted toward more controlled, less diverse communication during this time, potentially even limiting certain types of information over email.

²Full code is available at <https://github.com/maReins/enronEntropy>.

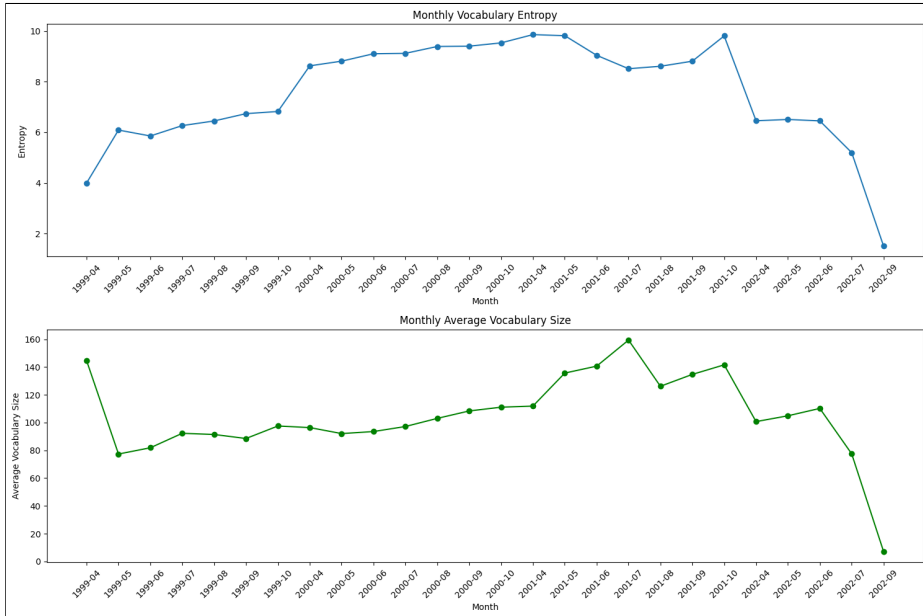


Figure 2: Time series of entropy and vocabulary size by month over the Enron Corpus. The Enron Crisis occurred in October 2001.

These findings demonstrate how entropy and vocabulary size metrics can capture significant shifts in organizational communication patterns, particularly during periods of crisis. The alignment of these changes with known events in Enron’s timeline suggests that these metrics could potentially serve as early indicators of organizational stress or change. These results indicate that calculating entropy over communication logs between agents – human and machine – can inform whether the addition of machine agents is leading to more knowledge sharing, operational complexity, and overall health.

5 Impact and limitations

This paper provides a straight-forward implementation of entropy to measure cooperative behavior from the communication logs of human-machine teams and multi-agent systems. Because of the conversational fluency of LLMs, these communication logs are not only readily available but also hold key insight into cooperative behavior based on any interaction.

The paper then validates this hypothesis against the Enron Corpus, showing that changes in entropy across communication logs within the company map

to the organizational growth before the crisis and the restricted communication afterwards. The findings indicate that entropy can provide insight into organizational issues before they arise and is sensitive to information control or even information attrition that many worry LLMs may bring to organizations through increased use (Bender et al. 2021).

While these results are promising, additional work is needed to determine entropy thresholds predictive of good and poor cooperation. Current research is evaluating entropy's ability to differentiate very good and very poor information-sharing in simulated multi-agent systems. These simulated environments would make it easier to assess other information theoretic metrics, like Mutual Information (Shannon 1948), to create a more cohesive evaluation of knowledge sharing.

Another thread of research is building out evaluations to assess the other tenets of teaming, predictability and directability (Johnson & Vera 2019). Multi-agent frameworks are stateful, and building out metrics based on the KL-Divergence (Kullback & Leibler 1951) between actual and predicted state is feasible. It is more difficult to create state representations for the human teammates, although the machine agent's internal representation of its human teammate may be sufficient.

This work fits within a larger goal of creating a Cooperative AI benchmark. A benchmark would provide principled and scalable evaluation techniques to guide progress. While LLM benchmarks for performance (Liang et al. 2023) and hallucinations (Hong et al. 2024) exist, translating strictly numeric evaluations into guidance toward complex human behavior remains difficult. Even benchmarks that attempt similar goals like AI Safety tend to focus on reducing negative behavior rather than promoting good behavior (Vidgen et al. 2024). This work provides a counterweight to these important efforts, focusing on the key role of communication and knowledge sharing within any human organization, with the hope of translating it into a principled and robust metric over any set of human and machine communications.

6 Conclusion

If LLMs are to provide value to organizations, they should augment knowledge discovery, creation, and dissemination to solve real problems. Information theory provides a solid theoretical background, and entropy offers a simple and practical metric to both assess organizational health and determine whether the addition of LLMs increases information flow across human and machine team members. Future research will explore developing a comprehensive Cooperative AI benchmark using a full range of information-theoretic measures.

Acknowledgments

Once, in despair as a junior member of his team, I asked Dick how he could bear to let go of our work when the organization shifted direction. “It’s not lost,” he said. “You can tinker with it in the background, and then one day it will come back in a new form.” This is my attempt to tinker with a few of those nearly-lost projects, in the spirit of the wisest tinkerer I know.

References

- Agashe, Saaket, Yue Fan, Anthony Reyna & Xin Eric Wang. 2023. *LLM-coordination: Evaluating and analyzing multi-agent coordination abilities in large language models*. <https://arxiv.org/abs/2310.03903>.
- Alavi, Maryam & Dorothy Leidner. 2001. Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly* 1.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major & Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 610–623.
- Bowman, Samuel R., Gabor Angeli, Christopher Potts & Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Conference on empirical methods in natural language processing (EMNLP)*, 632–642.
- Butler, Jenna, Sonya Jaffe, Nancy Baym, Mary Czerwinski, Shamsi Iqbal, Kate Nowak, Sean Rintel, Abigail Sellen, Mihaela Vorvoreanu, Brent Hecht & Jaime Teevan. 2023. *Microsoft new future of work report 2023*. Tech. rep.
- CBInsights. 2024. *Future of the workforce: How AI agents will transform enterprise workflows*. Tech. rep. <https://www.cbinsights.com/research/report/future-workforce-ai-agents/>.
- Chase, Harrison. 2022. *Langchain*. Online code and documentation. <https://github.com/langchain-ai/langchain>.
- Condoravdi, Cleo, Dick Crouch, Valeria De Paiva, Reihardt Stolle & Danny G. Bobrow. 2003. Entailment, intensionality and text understanding. In *Proceedings of the HLT-NAACL 2003 workshop on text meaning*, 38–45.
- Cox, Andrew. 2022. The ethics of AI for information professionals: Eight scenarios. *Journal of the Australian Library and Information Association* 71. 201–214.

- Dafoe, Allan, Edward Hughes, Yoram Bachrach, Tantum Collins, Kevin R. McKee, Joel Z. Leibo, Kate Larson & Thore Graepel. 2021. Open problems in cooperative AI. In *NEURIPS 2021 (Cooperative AI Workshop)*. arXiv. <https://arxiv.org/abs/2012.08630>.
- Diesner, Jana, Terrill L. Frantz & Kathleen M. Carley. 2005. Communication networks from the Enron email corpus: “It’s always about the people. Enron is no different”. *Computational and Mathematical Organizational Theory* 11(3). 201–228.
- Ferrucci, David A., Eric W. Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John M. Prager, Nico Schlaefer & Chris Welty. 2010. Building Watson: An overview of the DeepQA project. *AI Magazine* 31. 59–79.
- Hayes-Roth, Frederick. 1983. *Building expert systems*. Addison-Wesley.
- Hong, Giwon, Aryo Pradipta Gema, Rohit Saxena, Xiaotang Du, Ping Nie, Yu Zhao, Laura Perez-Beltrachini, Max Ryabinin, Xuanli He, Clementine Fourier & Pasquale Minervini. 2024. *The hallucinations leaderboard—an open effort to measure hallucinations in Large Language Models*. <https://arxiv.org/abs/2404.05904>.
- Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Delong Chen, Wenliang Dai, Andrea Madotto & Pascale Fung. 2022. Survey of hallucination in natural language generation. *ACM Computing Surveys* 55. 1–38.
- Johnson, Matthew Thomas & Alonso H. Vera. 2019. No AI is an island: The case for teaming intelligence. *AI Magazine* 40. 16–28.
- Jurafsky, Daniel & James H. Martin. 2009. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Pearson Prentice Hall.
- Kalouli, Aikaterini-Lida, Dick Crouch & Valeria de Paiva. 2020. Hy-NLI: A hybrid system for natural language inference. In *International conference on computational linguistics*, 5235–5249.
- Klimt, Bryan & Yiming Yang. 2004. The Enron corpus: A new dataset for email classification research. In *Machine learning: ECML 2004*, vol. 3201/2004, 217–226. Springer.
- Kullback, S. & R. A. Leibler. 1951. On information and sufficiency. *The Annals of Mathematical Statistics* 22(1). 79–86.
- Liang, Percy, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, Benjamin Newman, Binhang Yuan, Bobby Yan, Ce Zhang, Christian Cosgrove, Christopher D. Manning, Christopher Ré, Diana Acosta-Navas, Drew A. Hud-

- son, Eric Zelikman, Esin Durmus, Faisal Ladhak, Frieda Rong, Hongyu Ren, Huaxiu Yao, Jue Wang, Keshav Santhanam, Laurel Orr, Lucia Zheng, Mert Yuksekgonul, Mirac Suzgun, Nathan Kim, Neel Guha, Niladri Chatterji, Omar Khattab, Peter Henderson, Qian Huang, Ryan Chi, Sang Michael Xie, Shibani Santurkar, Surya Ganguli, Tatsunori Hashimoto, Thomas Icard, Tianyi Zhang, Vishrav Chaudhary, William Wang, Xuechen Li, Yifan Mai, Yuhui Zhang & Yuta Koreeda. 2023. Holistic evaluation of language models. *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=iO4LZibEqW>.
- Malone, Thomas W. 2018. *Superminds: The surprising power of people and computers thinking together*. Little, Brown & Company.
- March, J. G. & Herbert A. Simon. 1958. *Organizations*. Wiley.
- Seeber, Isabella, Eva Bittner, Robert O. Briggs, Triparna de Vreede, Gert-Jan de Vreede, Aaron Elkins, Ronald Maier, Alexander B. Merz, Sarah Oeste-Reiß, Nils Randrup, Gerhard Schwabe & Matthias Söllner. 2020. Machines as teammates: A research agenda on AI in team collaboration. *Information & Management* 57(2). 103–174.
- Shannon, Claude Elwood. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27. 379–423.
- Simon, Herbert A. 1971. Designing organizations for an information rich world. In Martin Greenberger (ed.), *Computers, communications, and the public interest*, 37–72.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser & Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st international conference on neural information processing systems*, 6000–6010. Curran Associates Inc.
- Vidgen, Bertie et al. 2024. *Introducing v0.5 of the AI safety benchmark from ML-Commons*. <https://arxiv.org/abs/2404.12241>.
- Wang, Zhenyi, Enneng Yang, Li Shen & Heng Huang. 2023. *A comprehensive survey of forgetting in deep learning beyond continual learning*. <https://arxiv.org/abs/2307.09218>.
- Wang, Zihao, Shaofei Cai, Anji Liu, Xiaojian Ma & Yitao Liang. 2023. *Describe, explain, plan and select: Interactive planning with large language models enables open-world multi-task agents*. <https://arxiv.org/abs/2302.01560>.

Chapter 15

Multi-modal Search: To Embed or not to Embed

Tracy Holloway King^a

^aAdobe Inc.

This is an exciting time for search and recommendations. Deep learning and embedding-based algorithms have provided game changing new techniques for finding relevant content, while users' needs for multi-modal content is ever increasing. This is especially the case for search over images, video, and audio. In multi-modal and cross-modal search (e.g. text to image/video/audio), users struggle to form keyword queries that represent their needs and information retrieval systems struggle to match keywords against non-textual content. This paper considers when to use off-the-shelf vs. in-house models, what types of queries do not work well with embedding-based search, and how to address latency, product, and business requirements. Examples are based on the techniques we are using at Adobe to improve multi-modal search in Adobe Stock (an image, video, and audio marketplace).

1 Introduction

Over the past 25 years, search and information retrieval have been an active area of applied research, with a strong natural language processing (NLP) component (see Manning et al. (2008)). Although there were major search applications and research before then, the advent of large-scale web search drove research in this area. The past few years have seen a shift in search capabilities with the introduction of embedding-based search techniques, especially ones that are feasible in real-world applications. These embeddings are a mathematical representation of data (e.g. text, image) in a lower-dimensional space, where each asset is converted into a vector that captures its key characteristics and relationships with

other assets. More recently, generative AI has seen broader adoption for question answering for search queries that could be interpreted as looking for an answer or specific information as opposed to looking for a specific document (e.g. the query *insert carriage return in excel* when issued on Google). This has led to claims that search is dead. In contrast, I would argue that search is needed more than ever, including as a component for question answering systems, ensuring that the information they provide is correct, validatable and up to date. In addition, there are still many situations in which users need to find specific documents and assets (e.g. personal photos) and are not looking for question answers.

Not only is search not dead, it is not solved. Shah et al. (2023) detail how advances in search, question answering, and dialog have only begun to lay the groundwork for new, mission- or task-driven experiences (e.g. finding an after-school activity for your child, putting together a financial report, applying for a passport). These experiences require a deep understanding of what the user wants to accomplish, feedback from the user at key decision points, and a strong grounding in up-to-date public and private data. Currently these missions are accomplished by the user breaking down the task and conducting a series of searches. However, search can evolve to help the user focus on the broader goal and then with synthesizing and interpreting information found.

Many of the machine learning (ML) modeling approaches used for state-of-the-art question answering in search experiences can be leveraged directly for search. In particular, the class of embedding-based techniques, which have been around for a decade but have only recently been broadly used, provides a way to make major improvements to search. These approaches do not solve all the issues with search, but when combined with classic search techniques, they can alleviate certain issues with finding relevant assets and ranking them appropriately.

In this paper, section 2 outlines some image-based search use cases to understand why search is an unsolved research and applied science area. Section 3 then examines how non-embedding-based search is done using inverted indices. With this base, first section 4 discusses how embeddings can be used in pure embedding-based search and then section 5 discusses how to combine classic and embedding-based techniques to get the best of both worlds. This still leaves several classes of queries that cannot be handled by either approach, but moves us significantly closer to high-quality search over a broad class of asset types.

2 Image search use cases

People often think of search as finding text documents which match a text query, e.g. using a keyword query in web search to find web pages that match the intent of the query. However, many products rely on search over other types of assets, including private asset collections which often have less textual data associated with them (e.g. you probably did not manually provide tags for the photos on your camera roll). This section describes some search-related products and features at Adobe which will be used as examples throughout the paper. Each of these has a different set of attributes which in turn requires a slightly different approach to providing relevant search results.

Adobe Stock is a marketplace where contributors can sell photos, illustrations, music, and short videos. Users search on Adobe Stock for content to use in presentations, advertising materials, etc. Most of these searches are simple keyword searches. These describe the objects depicted in the image (e.g. *banana white background, elderly patient and nurse*) or abstract concepts that they want representations of (e.g. *happiness, resilience*). The asset set is large (over 400M assets). This poses a challenge for search using text: either the query terms match contributor-provided or ML-produced text and tags or multi-modal techniques need to be introduced.

Adobe Stock also has image-based search, including the ability to search for particular image attributes. The query can be just an image or an image plus text. This provides users with a way to locate similar images in situations where natural language is hard to use to describe what they want (Figure 1).

Similar to search for images on Adobe Stock, Adobe also provides search for user's personal photos in Lightroom. Although the overall number of Lightroom photos is large (over 20 billion images), each user has a much smaller number of photos to search over. While Adobe Stock has contributor provided captions and tags, Lightroom users rarely tag their photos and so an auto-tagger model is used to enable keyword search for commonly depicted objects and concept (e.g. dog, landscape). In addition, many of the Lightroom searches are for dates (e.g. *aug 2019*) and so mapping from keywords to date-related metadata is key for success.

Another largely image-related search is in Adobe Express. Adobe Express allows users to create customized cards, marketing materials, and social media posts. These can be based on templates which, similar to Adobe Stock, have extensive metadata such as designer-created tags and titles. Unlike Stock, the number of templates is limited (hundreds of thousands for Express vs. hundreds of millions for Stock). However, since users customize these for their specific purpose, the goal is to find inspirational, easy-to-remix templates for the user's intent. These

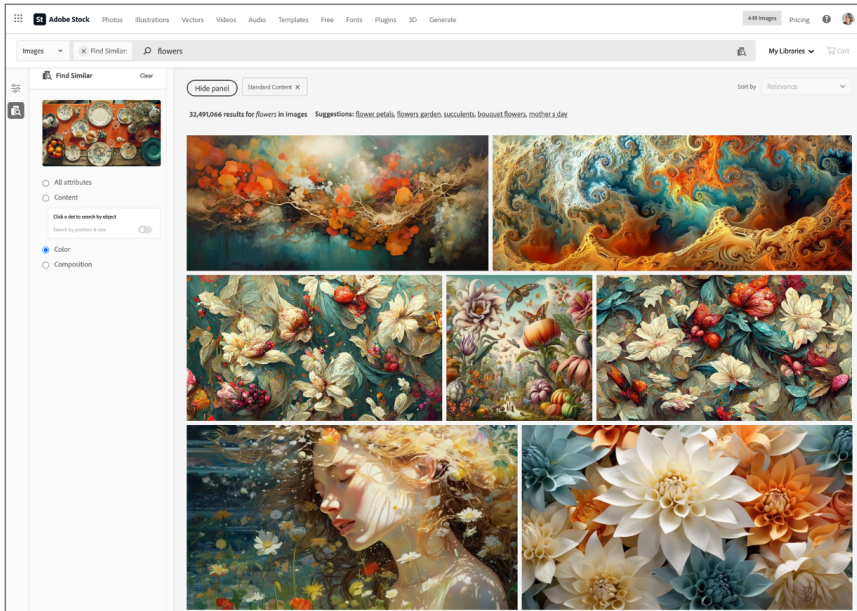


Figure 1: Adobe Stock color similarity search combined with keyword search. The color reference (query) image is on the upper left and color similarity is selected. The keyword *flowers* searches for flower images even though the reference image is of dishes on a table.

intentions are expressed as queries and can be general (e.g. *birthday*) or specific (e.g. *mary's 1 birthday party purple unicorn*). Express users also search for design elements like icons and backgrounds to customize the templates and for help using Express. As such, the search system has to determine the type of content needed and then to find relevant content within that (see Table 1 for example Adobe Express queries).

In the remainder of this paper I focus on Adobe Stock image search since it highlights the benefits and challenges of classic approaches to search and newer embedding-based approaches.

3 Classic approaches

In order to understand how the newer embedding-based search works, it is important to first understand how classic keyword search works.

Table 1: Example Adobe Express queries. Queries can be looking for templates, design elements, tools, or help.

Template	Express Query Types	
	Design Element	Tools & Help
collage	circle	remove background
halloween	mail icon	add a page
save the date dance party	dark blue background	how do you make a
we will be closed for thanksgiving and black friday	with white clouds	photo collage

3.1 Classic search indices

Classic search engines are based on an inverted index (see Manning et al. (2008) for an introduction to information retrieval techniques).¹ An inverted index is similar to the index at the end of a book, hence the name. A book index allows the reader to look up a word or phrase and see what pages are associated with it. In search, these pages are assets² (e.g. web pages in web search, images in Adobe Stock, help pages and videos in Adobe HelpX).

To create a search index for a set of assets, first the system identifies all the words associated with each asset. Then it builds an index of the words and associates them with an identifier for relevant assets. Table 2 shows four simple one-sentence assets and the inverted index that is created from them.

When the user issues a query, the query is broken into words.³ All of the assets associated with each word are found by looking up the word in the index. The asset lists for each query word are compared. The assets that appear on all the lists are returned. This set of assets is referred to as the match set or recall set. These assets are then ranked so that the most relevant one is first, then the next more relevant, etc. Four sample queries and the assets returned for them are shown in Table 3.

The one-word query *dog* returns all the assets with the word *dog* in them. For the ranking, here we ranked the two assets with the word *dog* as the subject higher

¹Brin & Page (1998) provide background on Google large-scale web search. In this paper we are concerned with scaling only with how embeddings affect it (sections 4.2 and 5.2).

²In search these are traditionally called documents, but since the term document is strongly associated with text, in this paper the term asset is used instead.

³Here we ignore the complex issues of tokenization and normalization. See Manning et al. (2008) for an overview.

Table 2: Sample inverted index based on 4 assets, each with only one phrase. The word *a* is not indexed since it is a closed-class “stop” word.

Assets		Inverted Index	
Asset Id	Text	Word	Asset Ids
1	Happy dog	dog	1, 3, 4
2	Happy girl	happy	1, 2
3	Girl chasing a dog	girl	2, 3
4	Dog chasing a ball	ball	4
		chasing	3, 4

Table 3: Sample queries and the assets they return for the assets and inverted index in Table 2

Query	Initial assets	Final assets
dog	1, 3, 4	1, 4, 3
dog chasing	dog: 1, 3, 4; chasing: 3, 4	4, 3
dog cat	dog: 1, 3, 4; cat: —	—
girl ball	girl: 2, 3; ball: 4	—

than the one with it as an object. For the query *dog chasing*, there are two assets which contains both words and so those two are returned. We ranked the asset which matches the word order in the query higher.⁴ In contrast, the query *dog cat* does not return any results because there is no asset that contains the word *cat*. Similarly, the query *girl ball* also returns no results. In this case, there are two assets which contain the word *girl* and one which contains the word *ball*, but no assets that contain both. Many search engines return “partial matches” where not all the query words are matched if there are no assets that match all the words. For example, for the query *girl ball*, assets 2, 3, and 4 would be returned, often with an indication that no exact matches to the query were found. This is not shown in Table 3.

⁴There are several techniques for efficiently capturing information about multiword expressions (e.g. *post office*, *hot dog*) and word adjacency (e.g. *dog chasing* in Table 3) in inverted indices. These techniques helped bridge the gap between the initial inverted indices described in this section and the embedding-based approaches in section 4. These ngram- and phrase-based techniques are not discussed here: search over images benefits less from these techniques due to the lack of coherent text, i.e. even if image captions are available they are often short and composed of keywords, not phrases. See Manning et al. (2008) on standard techniques and Wang et al. (2012) for an example of ngram selection for web search.

3.2 Classic search over images

Classic search depends on text-based matching as described above. How does classic search over images (and other non-textual assets such as music, sound effects, and video) work? There are two types of textual data that can be associated with these assets.⁵

The first is information provided from metadata and by the creator of the image. Images can have metadata such as the date of creation, location (e.g. embedded lat-long coordinates) and the file type, where the file type can be key in determining whether an image is, for example, a vector graphic or some other type of image. Creators often provide captions or tags describing the content of the image (e.g. *girl in blue shirt and shorts chasing a golden retriever on a beach in Maine*). For professional photos, such as those on Adobe Stock, images containing people have model release forms which provide basic information about the people in the photograph. Non-PII information⁶ in these release forms can be valuable in providing metadata for use in search. The textual information (e.g. captions and tags) is used directly as keywords in the search index. The metadata can either be searched by converting keywords in user queries into metadata for matching or by converting the metadata into text and inserting that in the index.

The second type of information used in search for images is created by models used to analyze the images. Auto-taggers associate textual tags or sometimes inter-lingual concepts such as WordNet senses (Fellbaum 1998) or Babelscape concepts (Navigli et al. 2021) with an image. Auto-taggers usually provide a confidence score or probability with each tag and the higher confidence tags for each image can be indexed and searched over. Auto-captioning models work similarly but provide captions instead of isolated tags. The downside of auto-taggers and auto-captioning is that the models have errors and are dependent on the original training vocabulary and image styles. However, they can be easily applied to all the images to be indexed, even if the image does not have a creator-provided set of tags or caption, and they provide a uniform set of vocabulary to search over.

4 Leveraging embeddings

Instead of relying solely on an inverted index of words, it is possible to search using embeddings. These embeddings are a mathematical representation of data

⁵This paper focuses on images, but the basic principles apply to audio and video, although the complexity of the ML models needed for each of these varies.

⁶PII = Personally Identifiable Information, which is securely stored and not used for model training or public asset search.

(e.g. text, image) in a lower-dimensional space: each asset is converted into a vector (i.e. a list of numbers) that captures its key characteristics and relationships with other assets. This is sometimes referred to as semantic search since it depends on the more abstract embedding representation of an asset instead of the words in the asset.⁷

Embeddings are vectors, i.e. an ordered list of numbers representing a position in a multidimensional abstract mathematical space (see Jurafsky & Martin (2024) for a good introduction from an NLP approach). In embedding-based search, each asset is converted into an embedding by a machine-learned model.⁸ At query time, the query is converted into an embedding in the same mathematical space. The query embedding is compared with each asset embedding, resulting in a score. Commonly this score is the cosine distance or dot product between the two embeddings. The higher the score, the more similar the embeddings are and hence the more relevant the asset.

The remainder of this section discusses some of the advantages and disadvantages of embedding-based search. Section 5 then discusses approaches to leveraging the strengths of embedding-based search while avoiding the pitfalls.

4.1 Pure embedding-based search

Pure embedding-based search uses only the embeddings of the assets and the query to find relevant assets. Intuitively this is simple to do since it only involves selecting an ML model to create the embeddings and a scoring function, e.g. cosine. Then for each query, the score is calculated and the assets are ranked by that score.

Embedding-based search was initially used for a single modality: text-to-text, image-to-image, audio-to-audio, etc. Since the query and the assets are of the same type, the same ML model can be used to create the embeddings. An example of image-to-image search is shown in Figure 2. The first image of the results is the query image itself since it is a perfect match. The remaining images are progressively more different and hence less relevant than the query image. Figure

⁷Throughout our working together, Dick demonstrated an unfailing patience in explaining the complexities of semantics and reasoning. This paper does not leverage the deep semantic search and question answering we worked on at PARC (Bobrow et al. 2007) and Microsoft. I refer to the approaches in this paper as “embedding-based search” instead of “semantic search” to avoid any potential confusion.

⁸This paper does not discuss how to train such models. There is extensive literature on semantic search and on image and text processing which discuss the ML models used for embeddings (e.g. Radford et al. (2021) and Wadekar et al. (2024) and papers cited by and citing that work).

1 in the introduction demonstrated the use of specialized embeddings, namely color embeddings (and in the figure, also a keyword constraint).

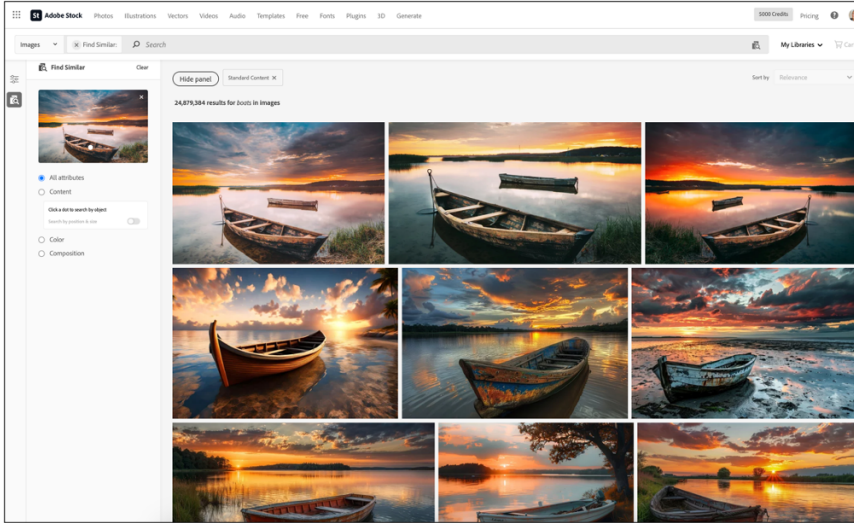


Figure 2: Adobe Stock image similarity search. The reference (query) image is on the upper left. The first result is the query image itself, which is a perfect match. The resulting images are progressively further away visually from the reference image.

Embedding-based search can also be used in a cross-modal (usually referred to as multi-modal) setting where the query and the assets are in different modalities. For search, these multi-modal embeddings are generally text-to-X (Radford et al. 2021, Jia et al. 2021), allowing textual queries for search over non-textual assets such as images, video, music, and sound effects. One of the most influential approaches to multi-modal embeddings is CLIP (contrastive language-image pretraining (Radford et al. 2021)). The CLIP approach takes a text model and an image model and aligns the embeddings into the same mathematical space. Training data for creating CLIP models generally comes from image-caption pairs but can be supplemented with query-clicked image pairs for search use cases. The CLIP approach can be extended to include multi-modal asset information, e.g. an image and its textual caption, to improve accuracy (Arora et al. 2024, Liu et al. 2021). Other modality pairings are possible such as video-music multi-modal embeddings, for example, to find recommended music for short videos.

Similar to the single modality embeddings used in text-to-text or image-to-image search, multi-modal embeddings are extremely useful in search: they enable textual queries to search over images without having to tag or caption the

images, whether manually or with an auto-tagger. This provides a robustness to the vocabulary mismatches which occur when the queries use different words (i.e. vocabulary) than those used to tag or caption the images (e.g. when an image is labeled with *bridal gown* but the query is for *wedding dress*, which contain no common words even though they mean the same thing).⁹ The multi-modal embeddings also cover, albeit extremely abstractly, a broader range of concepts than are usually represented by image tags (often only 10–50 tags per image) and captions (often <10 words). The embedding-based approaches provide a robustness to the exact matching of terms, something which applies also to text-to-text embeddings.

Another advantage to embedding-based search is that the query will always match something. That is, unless a threshold is set on the scoring, there is a score for every query-asset pair and there is always a highest score, even though this highest score may not reflect a relevant result. This allows for a form of null and low search result set recovery, retrieving results for every query. For example, in Table 3, two of the queries (*dog cat* and *girl ball*) had no results; text embedding-based search would have returned the most similar assets, even though they are incomplete matches.

4.2 Challenges with embedding search

There are several challenges with embedding-based search. Two of the challenges are common to all search relevance features: determining which embedding model to use and how many different embedding models to use. Section 5.1 discusses selecting embeddings, including whether to use a pretrained, off-the-shelf model or a custom one. Whether to use multiple embeddings (e.g. more than one image-text model, each of which may have particular strengths and weaknesses) is largely a matter of how to combine the signals and to keep the memory usage and latency at acceptable levels.

There are two embedding-specific challenges. Computing scores for millions of assets is slow. The latency can be improved with more powerful, but more expensive machines, and with custom vector databases. In addition, vector databases usually have to be combined with more conventional search engines in order to provide additional features (e.g. filters, refinements), enable keyword search for certain query types (see below) and for hybridizing with additional ranking

⁹Indexing of concepts in addition to words or of a list of synonyms can also alleviate the issues of synonyms such as *wedding dress/bridal gown*. Those approaches are easier to debug and to control for domain-specific vocabulary, but they often miss synonym pairs, especially for multi-word expressions.

features (section 5). Latency can also be improved by using sparse embeddings and quantization methods (section 5.2), which are less accurate but much faster.

From a relevancy perspective one of the most serious drawbacks of embeddings is that there is no good/bad, relevant/irrelevant cut-off. Thresholding with embedding scoring is difficult since different queries have different ranges for what is relevant. This makes it difficult to detect out-of-domain queries, for which highly irrelevant results will be returned, i.e. the top scoring results are always returned but in this case they are nonsense. This is also a problem if the search engine includes deterministic sorts like price low to high (see Tsagkias et al. (2020) on deterministic sorts in e-commerce search), where there is no way to restrict the result set to just the relevant items when sorting on dimensions like price, recency or review ratings. Incorporating the embeddings into a broader recall and ranking scheme can alleviate these issues (section 5).

Using only embeddings for recall and ranking has specific challenges. Most classic ranking algorithms leverage aggregated user behavior data and other popularity signals to improve the ranking and ensure that high quality assets appear at the top of search (Manning et al. 2008, Buttcher et al. 2016, Sorokina & Cantú-Paz 2016). However, many embeddings rely only on content signals, e.g. the text or the image, and are not sensitive to what constitutes a quality or popular asset (section 5.3). Using just embeddings can also result in a “Pictionary”¹⁰ effect where the most canonical asset for the query is shown first because it is the mathematically closest, centroid for the meaning in the query. This is particularly noticeable in text-image models with shorter queries, where the query concept maps to images that are canonical and stereotypical for that concept. Related to this issue is the lack of diversity in the top search results for pure embedding-based search. This arises because the assets are ranked by embedding score and visually similar assets have similar scores. An example is shown in Figure 3 for an image-based query on Adobe Stock.

Certain types of textual queries are particularly difficult for embeddings to handle. Terms which distribute similarly but with different meanings are difficult for embeddings to discriminate. These include numbers, years and model numbers, which are common in text search and e-commerce search. People names can also cause difficulty, often returning results with people names of same gender or ethnicity or for extremely famous people names returning associated people.

¹⁰Pictionary is a game where one teammate is given a word and has to draw an image representing that word for their teammate to guess. So, if the word is *cheese*, you might draw a wedge of Swiss cheese, perhaps with a mouse next to it, since that is the canonical, children’s book representation of cheese.

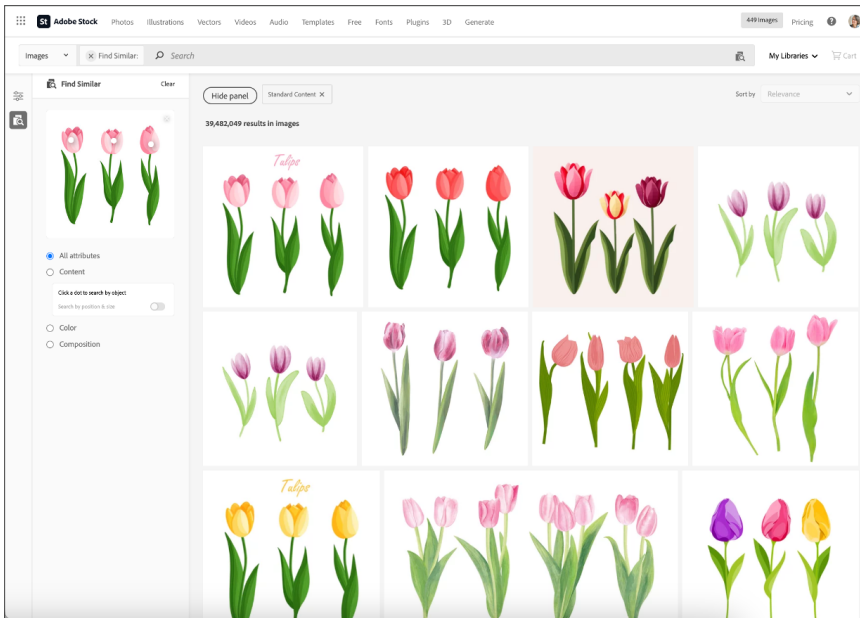


Figure 3: Adobe Stock image similarity search. The reference (query) image is on the upper left. The results are extremely similar, which fails to show the diversity of tulip images available, even among tulip illustrations on white backgrounds.

The same holds for location names. This can be helpful in finding related references, but can be frustrating in high precision applications.

Finally, negation (e.g. *snacks without nuts*), counterfactuals and information determined by syntactic structure (e.g. *dresses over \$100*) and not just co-occurrence can be difficult for embedding-based search (King 2023). These types of queries are also problematic for classic keyword search (section 3; Kumar et al. (2023), Sun et al. (2016), Wang et al. (2014, 2015)) and unfortunately, the move to embedding-based search does not solve the problem. These are usually addressed with domain-specific solutions which target the specific classes which are key for that domain. For example, for price-related queries in e-commerce, the price phrases (e.g. *over \$100*) are converted into constraints on the metadata, namely the price field, and only the non-price query terms (e.g. *dresses*) are used for retrieval. Negation can be treated similarly by first searching for the non-negated terms (e.g. *snacks*) and then removing all results which contain the negated term (e.g. *nuts*). This is complex to implement, slow, and error prone (e.g. a snack with almonds has nuts in it but may not contain the term *nut* and hence erroneously

be returned for *snacks without nuts*).¹¹ As a result, negation, counterfactuals, and syntactic constructions remain an area of active research.

5 Towards solutions: Hybrid systems and embedding improvements

By combining embedding-based techniques with more traditional search techniques, it is possible to build search systems which leverage the strengths of both approaches while alleviating their weaknesses. Some areas, such as handling negation, remain a challenge and will require further innovations. This section reviews three areas for improving embedding-based search: customizing embeddings for the specific domain, mitigating the high latencies involved in embedding-based search, and incorporating embeddings into more traditional search ranking techniques.

5.1 Customized embeddings

There are many off-the-shelf models available for creating embeddings, including the multi-modal image-text embeddings used in text query search over images (e.g. the CLIP models and variants of those). These can serve as a solid baseline for multi-modal embedding-based search. However, they are usually general purpose models and may not be ideal for a given use case.

To create a domain-specific embedding model, either an existing model can be fine-tuned or the model can be trained from scratch. Regardless of which approach is chosen, data will be needed for the fine-tuning or training and for an evaluation set to determine whether the model is accurate enough. In the case of search, past user search behavior (i.e. queries and clicks) can be used in conjunction with manual evaluation. For example, for Adobe Stock and Express image search (section 2), the off-the-shelf CLIP model lacked several capabilities including handling short text (i.e. queries), Adobe-specific vocabulary and asset types, and multiple languages.¹²

Two main types of training data were collected to enable this. The first was caption-asset pairs from Stock images and Express templates. These provide

¹¹For this particular query, converting it to *nut-free snacks* provides much better results since that phrase is frequently used in product titles and descriptions. However, this is not a generalizable solution.

¹²At the time the initial model was trained, multilingual CLIP models were not available for industry use.

longer text in the Adobe vocabulary and high quality images compared to those often found in publicly available image datasets. The second was aggregated query-to-clicked-asset pairs. These provided shorter text since queries are generally much shorter (e.g. 1–3 words) than captions. They also provide vocabulary of Adobe’s users, e.g. people searching for Adobe Stock images or Express templates. Both types of data were largely in English but also contained high-quality non-English instances. This Adobe-specific data was supplemented by industry-standard datasets and by machine translated versions of the English data.

The evaluation data for this Adobe-specific image-text model was compiled in two ways for rapid iteration while training the model. A set of the query-click data was held out for evaluation: when the query was issued, better models ranked the clicked assets higher than assets that were not clicked for that query. Similarly, a set of the caption-image data was held out: when the caption was used as a query, better models ranked the image the caption came from higher than other images. These automatic, quantitative evaluations allowed for rapid model iteration. However, they could not replace a final manual evaluation of results for a stratified sample query set to see absolute quality and to understand where the model did not perform well. This manual step is common to search relevance modeling in general and not just embedding-based search (Sanderson & Clough 2013).¹³

5.2 Match set and latency

Calculating the embedding score for millions of assets can be prohibitively slow and reducing the latency by using more powerful machines, including GPUs, can significantly increase the cost.¹⁴ There are several ways to enable the advantages

¹³A similar approach is used for single modality search with embeddings. Kumar et al. (2023) discuss training a multi-lingual text embedding model for use in searching over Adobe help documentation. The primary issue for this model was the specific language usage in Adobe products (e.g. *rush* and *acrobat* are product names in this application, not common English words). There are two interesting areas around help-related search. The first is supporting the wide number of languages (>25) that the documentation is available in, including determining the language of the query (Tambi et al. 2020). The second is the domain-specific vocabulary used in the documentation and the fact that users who need help documentation often do not know what Adobe calls specific tools and actions. This means that there is a vocabulary mismatch between how users refer to what they need and how the help documents refer to how the products work. For example, Photoshop has a tool called *magic wand* which automatically selects an object or area of the image. This is something users often want to do but a novice user might query this as *select object* or *tool to select area* and would not think to search for *magic wand*.

¹⁴The goal is to reduce cost, whether by using less expensive GPUs or by running GPU-trained models on CPUs, which maintaining high accuracy and low latency.

of the broad coverage of embeddings while lowering latency and maintaining high accuracy.

The first set of approaches uses embeddings only when most needed and relies on classic keyword search techniques the majority of the time. For example, embedding-based search can be reserved for null and low recovery, especially for longer, more natural language like queries. That is, the search is first done with classic keyword techniques, perhaps with embeddings used in a second pass ranker to improve relevancy (section 5.3). This provides for low latency for the majority of searches, especially if caching is enabled for frequent head queries. Only when there are few results, with the exact number dependent on the application but usually less than one page worth, are embeddings used for creating the match set. Having the keyword index available also enables accurate search for numbers and proper names (section 4.2) and so serves a dual purpose in accuracy for those queries and lower latency for other queries. Even when embedding search is necessary, e.g. when searching for images using an image as the query, additional constraints can be encouraged from the user such as high level keywords (e.g. in Figure 1 where the keyword *flower* constrains the match set size for the image color embeddings to work over) and application of filters (e.g. photo vs. illustration, aspect ratio, date of capture).

The second set of approaches aims to make embedding-based search faster. Vector search databases are optimized for this purpose. These include systems like FAISS, OpenSearch, Vespa, and Pinecone, which were created especially for this purpose. These specialized search systems provide both exact scoring over dense embeddings and faster but approximate search and scoring methods.

The approximate methods reduce accuracy in that the list of assets whose embeddings are closest to the query embedding is not exact. Depending on the method, either certain assets may be missing from the set altogether or the scoring and hence ordering may be slightly off. However, if the asset set size is large enough and the application is tolerant of a few missed assets, then these approximate methods allow for rapid search over large numbers of asset embeddings. One such approach is to create clusters from the embeddings and select a centroid embedding to represent each cluster. When a query is run, it is first compared to the relatively small number of centroid embeddings. The top n are selected and the query is compared only to the embeddings in those clusters.

Embeddings can also be made sparser by bucketizing or quantizing them (Kong et al. 2023). PQ (product quantization) codes were one of the first, and highly effective, ways of creating embeddings that can be retrieved at low latency (Jégou et al. 2011). These early approaches are continually being improved

upon. For example, Huggingface Matryoshka (Kusupati et al. 2022) creates embeddings where the first part of the embedding can be truncated to create embeddings which are shorter, and hence fast to compute scores with. Once a smaller set of embeddings is selected (e.g. by using the top n embeddings from the scoring of truncated embeddings), then the entire embedding can be used to get more accurate scores for a final ranking. Embeddings can also be sparsified so that they can be used similar to keywords. An example of this is shown in Table 4.

Table 4: Dense and Sparse representations of embeddings with sample scoring for sparse embeddings. Dense embeddings are shown with 2048 dimensions. Sparse embeddings have more dimensions (here 8192) but most of the dimensions have no values.

Dense Embedding Example			Sparse Embedding Example				
Dim.	Img. 1	Img. 2	Dim.	Query = Img. 1	Img. 2	Img. 3	Img. 4
1	0.11	1.23	1	—	1.12	—	—
2	1.21	0.42	2	—	—	0.81	—
3	0.15	0.53	3	1.16	0.83	—	0.64
4	0.22	2.25	4	—	—	1.83	—
...			...				
2048	2.17	0.64	8192	0.13	—	—	0.01

Sparse Embedding Matching and Scoring
Matches for query image 1: image 2: 1 dimension (dimension 3) image 3: 0 dimensions image 4: 2 dimensions (dimension 3 & 8192)
Scoring for ranking: image 2: $1.16 * 0.83 = 0.96$ image 4: $1.16 * 0.64 + 0.13 * 0.01 = 0.74$

The dense embeddings have values for every dimension, 2048 dimensions in table 4. The sparse version of the embeddings, which is derived from the dense version, has many more dimensions (8192 in table 4) but most of them have no values. For a given query, only assets which match at least n dimensions are returned. In the example, n is set to 1 and so image 2 matches dimension 3 and image 4 matches dimensions 3 and 8192, while image 3 is not matched. The scoring for ranking is the sum of the matched dimension values weighted by the score of that dimension for the query. This sparse encoding for matching and

scoring is extremely fast, but, as with all sparsification techniques, comes at the cost of lower accuracy and with the need to select and implement the sparsification model.

5.3 Ranking with embeddings

Search ranking uses a variety of features and machine learning algorithms (Manning et al. 2008, Sorokina & Cantú-Paz 2016), including transformer-based approaches (Pradeep et al. 2023). The features roughly divide into two classes: ones that capture core relevance (does the asset match the query, e.g. if the query is *dog on beach*, is the image of a beach and is there a dog on it) and ones that capture quality. The quality features are often derived via past user behavior such as clicks and purchases aggregated across all users, both in general and for specific queries, and ones from other signals such as the producer (e.g. web pages by Wikipedia are considered more reliable than ones from lesser known sources). Embeddings can provide strong signals for core relevance (section 4), but are often less useful in capturing quality, especially quality as reflected by aggregated past user behavior.

The goal of ranking with embeddings is to leverage the strength of the behavioral signals with the strength of the embedding semantics, thereby avoiding issues with too much similarity (Figure 3) while improving core relevance.

One simple approach is to use two rounds of ranking. The first round uses a traditional ML ranker. The second round re-orders the top n results using the embedding score. The value of n depends on the application but is likely between 100 and 10,000. This allows for low latency since only n embedding scores have to be calculated and can avoid having overly similar results since the first round ranker provides enough diversity. However, this approach only takes one type of embedding into account.

Another approach is to use the embedding scores as a feature in a traditional ML ranker, where it supplements the other core relevance scores such as BM25 (Robertson & Zaragoza 2009). Since calculating the dense score may not be feasible over large match sets (section 5.2), this embedding-feature ranker can also be a second round ranker. This approach can improve over simply having the embedding score be the second round ranker since the relative weight of the embedding score vs. the other scores is learned in the ML training of the ranker instead of being arbitrarily set at 100%. This approach also allows multiple embedding types to be used in the ranker, each with their own learned weight.

Ultimately, the behavioral signals themselves may be incorporated into embeddings (e.g. Chen et al. (2023)), which in turn opens up more ways to use

embeddings for ranking, including capturing more fine-grained user intent and interaction styles. This will also benefit the quality of transformer-based rankers (Pradeep et al. 2023). One downside of moving to transformer-based rankers is that the scores and impact of different features are much less interpretable than with an SVM (Joachims 2002) or GBDT (Zheng et al. 2007) based ranker. This means that it can be more difficult to determine why issues arise (e.g. why does a seemingly irrelevant asset appear high in the search results) and hence harder to determine what to try to improve the ranker.

This section discussed ways to leverage embeddings in image rich search applications like Adobe Stock. As with search throughout the decades, the addition of embedding-based techniques still requires combining matching and ranking techniques based on the specific product and feature and will continue to evolve as the technical capabilities evolve.

6 Conclusion

This is an exciting time to work on search, especially on search over non-textual assets such as images, video or audio. Embeddings are unlocking new capabilities, abstracting away from the exact words in textual assets and the limited textual data such as captions and tags on non-textual assets. However, they are not a magic elixir. Even with embeddings customized for the domain in question (e.g. Adobe Stock image search), they work best when combined with other signals such as popularity and traditional keyword relevance signals, and for certain types of queries, such as model numbers, dates and person names, non-embedding-based approaches, including mapping to metadata fields like image creation dates, is necessary. Some areas remain unsolved for both classic and embedding-based techniques, in particular the handling of queries with negation, counterfactuals, and syntactic constructions. These remain an area of active research and solving them is crucial for more complex question answering and task-driven applications.

Contrary to some recent claims, search is not dead or even dying. However, it is evolving. Not only are the techniques used in search evolving, as discussed in this paper, but the user's interaction with search is changing. In some cases, such as search for images on a phone camera roll, search is simply getting better, allowing the user to describe in words which photo they are looking for. In other cases, search is working behind the scenes to provide the information that fuels the answers at the top of search pages or for within-asset question answering. This is referred to as retrieval augmented generation or RAG (Lewis et al. 2020) since

the search results are used to generate the answer to the user's question. Even with question answering directly on search pages, links to relevant assets within the answer allows the user to quickly confirm the answer or to dive more deeply into the topic. For images, search and generation are being combined in other ways, such as showing Adobe Stock image results on image generation pages for inspiration, for use as reference images for refining and controlling generation, or for direct use. All of these advances are needed to enable mission-based search (Shah et al. 2023) so that the search experience no longer requires the user to break their task into individual queries and then assemble the resulting data into a coherent whole. Each of these new user experiences will require improvement to search relevance and performance, including to embedding-based search techniques.

Acknowledgments

I am pleased to contribute this paper in honor of Dick Crouch. Dick and I worked closely together at PARC's Natural Language Theory and Technology group, at Powerset and at Microsoft Bing. His ability to simultaneously understand broad architectural concerns and detailed empirical issues and to leverage this deep understanding to create state-of-the-art computational linguistic systems continues to amaze me. I hope someday to achieve even a fraction of his skill.

I would like to thank Chirag Shah, Jon Degenhardt, Aikaterini-Lida Kalouli, Annie Zaenen and the audiences at Etsy, Madhive, Rakuten and the Aristotle University of Thessaloniki for their comments.

References

- Arora, Abhishek, Xinmei Yang, Shao-Yu Jheng & Melissa Dell. 2024. *Linking representations with multimodal contrastive learning*. <https://arxiv.org/abs/2304.03464>.
- Bobrow, Daniel, Dick Crouch, Tracy Holloway King, Cleo Condoravdi, Lauri Karttunen, Rowan Nairn, Valeria de Paiva & Annie Zaenen. 2007. Precision-focused textual inference. In *Proceedings of the ACL-PASCAL workshop on textual entailment and paraphrasing*, 16–21.
- Brin, Sergey & Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks* 30. 107–117. <http://www-db.stanford.edu/~backrub/google.html>.

- Buttcher, Stefan, Charles L. A. Clarke & Gordon V. Cormack. 2016. *Information retrieval: Implementing and evaluating search engines*. The MIT Press.
- Chen, Haonan, Zhicheng Dou, Qiannan Zhu, Xiaochen Zuo & Ji-Rong Wen. 2023. Integrating representation and interaction for context-aware document ranking. *ACM Transactions on Information Systems* 41. 1–23.
- Fellbaum, Christiane (ed.). 1998. *WordNet: An electronic lexical database*. The MIT Press.
- Jégou, Hervé, Matthijs Douze & Cordelia Schmid. 2011. Product quantization for nearest neighbor search. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33. 117–128.
- Jia, Chao, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc V. Le, Yunhsuan Sung, Zhen Li & Tom Duerig. 2021. Scaling up visual and vision-language representation learning with noisy text supervision. In *Proceedings of the 38th international conference on machine learning PMLR*.
- Joachims, Thorsten. 2002. Optimizing search engines using clickthrough data. In *Proceedings of the ACM conference on knowledge discovery and data mining*.
- Jurafsky, Daniel & James H. Martin. 2024. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models*. 3rd edn. Online manuscript released August 20, 2024; print 2nd edition 2009. <https://web.stanford.edu/~jurafsky/slp3/>.
- King, Tracy Holloway. 2023. White roses, red backgrounds: Bringing structured representations to search. In Roussanka Loukanova, Peter LeFanu Lumsdaine & Reinhard Muskens (eds.), *Logic and algorithms in computational linguistics 2021 (LACompLing2021)*, 191–215. Springer. DOI: 10.1007/978-3-031-21780-7_9.
- Kong, Weize, Jeffrey M. Dudek, Cheng Li, Mingyang Zhang & Mike Bendersky. 2023. SparseEmbed: Learning sparse lexical representations with contextual embeddings for retrieval. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval (SIGIR '23)*, 2399–2403.
- Kumar, Jayant, Ashok Gupta, Zhaoyu Lu, Andrei Stefan & Tracy Holloway King. 2023. Multi-lingual semantic search for domain-specific applications: Adobe Photoshop and Illustrator help search. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval (SIGIR '23)*, 3225–3229. Association for Computing Machinery. DOI: 10.1145/3539618.3591826.
- Kusupati, Aditya, Gantavya Bhatt, Aniket Rege, Matthew Wallingford, Aditya Sinha, Vivek Ramanujan, William Howard-Snyder, Kaifeng Chen, Sham Kakade, Prateek Jain & Ali Farhadi. 2022. Matryoshka representation learning. In *Advances in neural information processing systems*, 30233–30249. Cur-

- ran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2022/file/c32319f4868da7613d78af9993100e42-Paper-Conference.pdf.
- Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel & Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in neural information processing systems*, vol. 33, 9459–9474. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf.
- Liu, Huidong, Shaoyuan Xu, Jinmiao Fu, Yang Liu, Ning Xie, Chien-Chih Wang, Bryan Wang & Yi Sun. 2021. *CMA-CLIP: Cross-modality attention CLIP for image-text classification*. <https://arxiv.org/abs/2112.03562>.
- Manning, Christopher D., Prabhakar Raghavan & Hinrich Schütze. 2008. *Introduction to information retrieval*. Cambridge University Press.
- Navigli, Roberto, Michele Bevilacqua, Simone Conia, Dario Montagnini & Francesco Cecconi. 2021. Ten years of BabelNet: A survey. In *Proceedings of IJCAI 2021*, 4559–4567.
- Pradeep, Ronak, Kai Hui, Jai Gupta, Honglei Zhuang, Adam Lelkes, Jimmy Lin, Don Metzler & Vinh Tran. 2023. Understanding generative retrieval at scale. In *EMNLP 2023*, 1305–1321. <https://aclanthology.org/2023.emnlp-main.83/>.
- Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger & Ilya Sutskever. 2021. Learning transferable visual models from natural language supervision. In Marina Meila & Tong Zhang (eds.), *Proceedings of the 38th international conference on machine learning*, vol. 139, 8748–8763. PMLR. <https://proceedings.mlr.press/v139/radford21a.html>.
- Robertson, Stephen & Hugo Zaragoza. 2009. *The probabilistic relevance framework: BM25 and beyond*. NOW Publishers, Inc.
- Sanderson, Paul & Mark Clough. 2013. Evaluating the performance of information retrieval systems using test collections. *IR Information Research* 18.
- Shah, Chirag, Ryen W. White, Paul Thomas, Bhaskar Mitra, Shawon Sarkar & Nicholas Belkin. 2023. Taking search to task. In *Proceedings of CHIIR’23*, 1–13. ACM.
- Sorokina, Daria & Erick Cantú-Paz. 2016. Amazon search: The joy of ranking products. In Raffaele Perego, Fabrizio Sebastiani, Javed A. Aslam, Ian Ruthven & Justin Zobel (eds.), *Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval, SIGIR 2016*, 459–460. ACM.

- Sun, Xiangyan, Haixun Wang, Yanghua Xiao & Zhongyuan Wang. 2016. Syntactic parsing of web queries. In *Proceedings of the 2016 conference on empirical methods in natural language processing (EMNLLP)*, 1787–1796.
- Tambi, Ritiz, Ajinkya Kale & Tracy Holloway King. 2020. Search query language identification using weak labeling. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the twelfth language resources and evaluation conference (LREC)*, 3520–3527. European Language Resources Association. <https://aclanthology.org/2020.lrec-1.432>.
- Tsagkias, Manos, Tracy Holloway King, Surya Kallumadi, Vanessa Murdock & Maarten de Rijke. 2020. Challenges and research opportunities in eCommerce search and recommendations. *ACM SIGIR Forum* 54. 1–23.
- Wadkar, Shakti N., Abhishek Chaurasia, Aman Chadha & Eugenio Culurciello. 2024. *The evolution of multimodal model architectures*. <https://arxiv.org/abs/2405.17927>.
- Wang, Chen, Keping Bi, Yunhua Hu, Hang Li & Guihong Cao. 2012. Extracting search-focused key n-grams for relevance ranking in web search. In *WSDM '12: Proceedings of the fifth ACM international conference on web search and data mining*, 343–352.
- Wang, Zhongyuan, Haixun Wang & Zhirui Hu. 2014. Head, modifier, and constraint detection in short texts. In *Proceedings of the international conference on data engineering*, 280–291.
- Wang, Zhongyuan, Kejun Zhao, Haixun Wang, Xiaofeng Meng & Ji-Rong Wen. 2015. Query understanding through knowledge-based conceptualization. In *Proceedings of IJCAI*, 3264–3270.
- Zheng, Zhaohui, Hongyuan Zha, Tong Zhang, Olivier Chapelle, Keke Chen & Gordon Sun. 2007. A general boosting method and its application to learning ranking functions for web search. In *Proceedings of NIPS*.

Index

- %stem variable, 33
- accessibility relation, 178
- Achiam, Josh, 238
- adjective, 41, 212–214, 224, 226, 227, 229, 234, 268
- adjectivization, 224
- adposition, 81
- adverb, 213
- Agashe, Saaket, 325
- agent, 94, 100, 246, 324
 - oblique, 39
 - optional, 39
- aggregation, 196
- agreement, 28, 36, 61, 242
 - gender, 70
 - inflection, 58
 - number, 70
- agreement attraction, 239
- Ahmed, Tafseer, iii, 238, 240
- Ahn, Byron, 4, 211, 228
- Ahn, David, iii
- AI
 - benchmark, 329
 - collaboration, 5
 - cooperative, 323–329
 - ethics, 324
 - explainable, 165
 - generative, 323, 334
 - safety, 329
 - statistical, 133
 - systems, 323
- Alavi, Maryam, 323
- Albert, Réka, 141
- algebra
 - coproduct, 103, 104
 - count, 102–104
- algebraic correctness criterion, 85, 101, 102
- algebraic expression, 101
- Allen, James F., 177
- allomorph, 58
- Alon, Uri, 141, 148, 149
- Alzheimer's, 133
- Amanaki, Eirini, 188
- Amazon, 1, 21, 303–306
- ambiguity, 2, 26, 27, 31–32, 35, 37, 39–42, 45–47, 85, 195, 196, 222, 270, 274, 289, 292
 - compositional, 45
 - quantifier, 31, 42
 - scope, 31
 - semantic, 32
 - spurious, 26, 27, 43–47
 - syntax, 67
- Ambrosy, Andrew P., 310
- anaphora, 111–133
- Ancient Greek, 239
- Anderson, Stephen Robert, 35
- Andrews, Avery D., 2, 35, 39, 46, 81, 88, 89
- Andrews, Peter B., 166

Index

- annotation, 138, 140–141, 188, 189,
 - 196, 317, 318
 - disagreement, 194
 - functional, 37
 - graph-based, 141
 - multi-label, 194
 - variability, 196
- antonym, 122
- Anwar, Maaz, 240
- application, 1, 4, 137, 302, 333
 - Adobe Stock, 333–351
 - biomedical, 302
 - education, 133
- applicative, 3, 94, 211–234
 - out, 211–234
- Ardila, Frederico, 102
- argument, 82–86
 - external, 89, 90, 94, 96, 100
 - internal, 89
 - optional, 39, 46
- argument structure, 211–234
 - non-canonical, 273–277
- argumentation, 187–204
- Aristotle, 197, 198
- Arora, Abhishek, 341
- Asher, Nicholas, 113, 117, 118, 120,
 - 138–140
- Asudeh, Ash, iii, 2, 15, 26, 30, 32, 37,
 - 39, 40, 43, 44, 46, 47, 49,
 - 56–59, 61–64, 67, 69, 70,
 - 72–74, 82, 93, 100, 105
- attachment point, 93
- attribute, 122
 - interpretable, 89
- attribute-value pair, 12
- auxiliary, 61
- axiom link, 84, 86, 87, 91, 93, 94, 98,
 - 101
 - principle, 92
- Babelscape, 339
- Bach, Emmon W., 22
- Bachmair, Leo, 166
- Baldwin, Timothy, 237
- Barabási, Albert-László, 141, 148
- Barančíková, Petra, 240
- Barsalou, Lawrence W., 231
- Barzel, Baruch, 141
- Barzilay, Regina, 137
- Bashan, Amir, 141
- Battiston, Stefano, 141
- BDU, *see* discourse unit, basic
- Beber, Moritz E., 148
- Beesley, Kenneth, 15
- BEL, 302, 317–318
- Belleau, François, 314
- Bender, Emily M., iii, 324, 329
- Bengali, 240
- Benzmüller, Christoph, iii, 165, 166,
 - 168, 171
- Bernardy, Jean-Philippe, 188
- Bernódusson, Helgi, 83
- Berwick, Robert, 79
- Bhat, Sneha, 318
- bialgebra, 102, 104
- bidirectional node connection, 144
- BioPortal, 305
- BioCreative, 303
- BioCreative Track 1, 303
- Bjorkman, Bronwyn, 61
- Blackburn, Patrick, 165
- blocking, 55–74
 - semantic, 69–74
- BM25, 349
- Bobrow, Daniel, 340
- Bonial, Claire, 267
- Borer, Hagit, 59
- Borschev, Vladimir, 87
- Boston, Marisa, 5

- bounding function, 14, 19
 Bowman, Samuel R., 323
 Boyer, Lorie A., 148
 Brazilian Portuguese, 240
 Breitholtz, Ellen, 197–201, 205
 Bresnan, Joan, 9, 10, 25, 28, 34–36,
 56, 57, 66, 119, 214, 295
 Brin, Sergey, 337
 Brown, Chad E., 180, 183
 Brown, Susan W., 4, 266–268
 Brown, Tom B., 246, 253
 Bullmore, Ed, 148
 Bunt, Harry, 42
 Burnett, Heather, 190, 197
 Butler, Jenna, 324
 Butt, Miriam, iii, 4, 15, 26, 36, 49,
 119, 238, 240–243, 247
 Buttcher, Stefan, 343
 Bytyqi, Erisa, 4, 239
 Bögel, Tina, 15, 49, 56
 Börjars, Kersti, 25, 29

 c-structure, *see*
 constituent-structure
 C58, 138, 156
 Cabitza, Federico, 194
 Cai, Shaofei, 325
 Cai, Zhenguang G., 239
 calculus
 lambda, 31, 32, 101, 102, 200
 predicate, 189
 sequent, 84
 superposition, 166
 calque, 227
 Cambridge University, 1
 Cantú-Paz, Erick, 343, 349
 Cao, Yong, 238, 250
 Cappelen, Herman, 129
 captioning, 196

 Carreras, Xavier, 274
 Cases, Ignacio, 192–194
 categorical grammar, 47
 categorizer formative, 88
 category, 65
 syntactic, 90
 causation, 281, 283, 285, 286, 288,
 290
 causative, 70–72, 247
 caused motion construction, 4,
 265–295
 extractive, 290, 291
 idiomatic, 290, 293
 preventative, 290, 293
 suppressive, 290, 291
 transitional, 290
 CBInsights, 325
 CDU, *see* discourse unit, complex
 CFG, *see* context free grammar
 change of location, 268, 275, 281
 change of possession, 268
 change of state, 268
 Chase, Harrison, 323
 chat, 324
 Chatzichrisafis, Nikos, 48
 Chatzikyriakidis, Stergios, iii, 188,
 205
 ChEBI, 305
 Chegg, 1
 Chen, Haonan, 349
 Chen, Tongfei, 194
 Chomsky, Noam, 56, 57, 59, 79–81
 Choudhury, Monojit, 141
 Church, Alonzo, 166
 Clark, Eve, 227
 class partition
 disjoint, 170
 exhaustive, 170
 classification, 274, 308, 327

Index

- classifier, 203, 204, 274
clause, 114, 116
 event, 123
Clauset, Aaron, 141
cleft, 114, 272
CLIP model, 341, 345
Clough, Mark, 346
co-description, 13, 32, 34, 91
coassociativity, 104
code switching, 227
coercion, 218, 219, 227, 274, 275, 279,
 281, 282, 285
cognitive disability, 133
cognitive science, 325
coherence, 29, 38, 39, 46, 199
commensurability condition, 218
common sense, 166, 172
communication, written, 112
commutativity, 26, 27, 40–47
comparative, 3, 211–234
 comparand, 216, 221, 233
 comparandum, 216, 221, 223,
 233
 degree, 215, 219–221, 228
 event, 215–219, 228
 modal, 221–234
 modal degree, 215
 nominal, 218
competition, 231–233
completeness, 29, 38, 39, 46
complex predicate, 4, 231, 237–257
 aspectual, 240–241, 252
 dataset, 243–244
 motion, 242–243, 252
 permissive, 241–242, 252
complexity, 12, 165
composition, 14
 extended head, 18
compositionality, 30, 61, 115, 280
computational linguistics, 3, 15,
 25–48, 188, 194
computational semantics, 302
conditionality, 38
Condoravdi, Cleo, 3, 295, 323
confabulation, 324
conjugation, 61
conjunction, 73, 113, 213
 simplification, 71
conjunctive, 113
connective, 202
conservation, 9–21
 of information, 12–16
 principle, 1, 11–16, 19, 21
consistency, 28, 199
Constant, Mathieu, 237
constituency, 27
constituent, 9
 discontinuous, 16, 18, 19, 115
constituent, discontinuous, 10
constituent-structure, 2, 9–21,
 26–29, 33, 37, 38, 44, 56–58,
 63, 65, 67, 70
 category, 67
 node, 65
 optional node, 59
 rule, 57, 65
 terminal node, 58, 63, 64, 67
constraint
 indecisive, 67, 69
 most informative, 57, 66, 67,
 69–73
 mutual, 13
 violation, 67
construction, 4, 123, 265–295
 double infinitive, 10
 double object, 83
 long distance, 267
 proposed, 123

- uncommon, 294
- Construction Grammar, 82–85, 274
- context, 85, 213, 307–309
 - dependent, 220
- context free grammar, 15, 27, 36
- contradiction, 189, 198
- control
 - construction, 242
 - obligatory, 99
- Cooper, Alexis, 4, 290
- Cooper, Robin, iii, 3, 188, 195,
 - 200–203, 301, 312, 314, 315
- cooperative behavior, 325
- coordination, 80, 93, 118
- coproduct, 80
- Cormack, James, 309
- corpus, 140–141, 197
 - complex predicate, 243–244
 - creation, 196–197
 - discourse, 138
 - human-computer
 - collaboration, 323–329
- correspondence, 65
 - function, 28
 - inverse, 13
- cosine similarity, 340
- counterfactual, 344
- Cox, Andrew, 324
- Creswell, Cassandre, 139
- Cristea, Dan, 117
- Crouch, Dick, 1, 2, 16, 21, 25, 49, 72,
 - 74, 105, 119, 188, 193, 205,
 - 255, 295, 330, 351
- Crouch, Richard, 5, 15, 16, 32, 33, 41,
 - 93, 101, 183, 234
- cue phrase, 114
- Curry, Haskell Brooks, 31
- Curry-Howard isomorphism, 31, 32,
 - 200
- Czech, 240
- D’Souza, Raissa M., 141
- Dafoe, Allan, 323, 325
- Dalrymple, Mary, iii, 15, 16, 23, 26,
 - 29, 30, 32, 37, 53, 56, 59, 62,
 - 77, 81, 82, 89, 99, 107, 134
- Dang, Hao T., 266, 267
- Dask Development Team, 155
- data aggregation, 194
- date, 177
- De Groote, Philippe, 44, 83–85, 101
- de Paiva, Valeria, 49
- decidability, 10, 16, 19, 20, 195
- declension, 60
- deduction, 92, 99
 - intuitionistic, 101
- deep neural network, 138
- definite, 34, 35, 37
- Degenhardt, Jon, 351
- deixis, 112, 177
- denotation, 113
- Dentella, Vittoria, 239
- dependency
 - crossed, 141
 - long distance, 46, 272
- derivation, 92, 93
- description by analysis, 13, 32–34,
 - 91
- determiner, 35, 37, 64, 113, 214
- deterministic sort, 343
- dialog, 196, 334
- diathesis, 267
- Diesner, Jana, 327
- directability, 324
- disambiguation, 239
 - sense, 268
- discourse, 2
 - antimotif, 157

Index

- coherence, 151
- context, 113
- definition, 112
- directed network, 142–144
- embedded context, 127–132
- motif, 3
- repair, 124
- segment, 141
- segmentation, 139
- speaker index, 129, 131
- syntax, 114
- turn, 99
- discourse constituent, 116, 117
- discourse constituent unit, 118, 120–125
- Discourse GraphBank, 138–141, 155, 157
 - metric, 146
 - network motif, 152–155
- Discourse Lexicalized Tree
 - Adjoining Grammar, 139
- discourse node
 - coordination, 118, 122–123, 126
 - n-ary, 118, 123–125, 130
 - operator, 118
 - subordination, 118, 121–122, 126, 128, 129
- discourse operator, 119
- discourse parse tree, 119–120
 - independent, 128, 129
 - pointer, 128
- discourse parsing, 112, 115–126
 - rule, 120–125
- discourse relation, 138–140
 - asymmetrical, 140, 149
 - continuation, 140
 - coordination, 140
 - directed, 140, 142
 - elaboration, 139, 152
 - example, 143
 - explanation, 139
 - importance, 140
 - narration, 139, 140
 - result, 139
 - subordination, 140
 - symmetrical, 140, 144, 149
 - temporal sequence, 140
- discourse representation, 120, 137–157
- Discourse Representation Theory, 32, 115, 120
- discourse structure, 111–133, 137–157
 - coherence, 156
 - network motif, 148–152
 - shift, 113
- Discourse Structures Theory, 113, 118
- discourse unit, 138–140, 142
 - basic, 113–115
 - complex, 139, 149
 - elementary, 139–140, 142, 149–151
 - nucleus, 140
 - satellite, 140
- discriminative learning, 274
- disjoint union, 92
- disjunction, 35, 37, 38, 65, 73
- distinctness condition, 19, 20
- Distributed Morphology, 2, 4, 55–59, 61, 65, 66, 69, 70, 74, 228
- DM, *see* Distributed Morphology
- Doetjes, Jenny, 218
- dogwhistle, 3, 201–202
- Dowle, Frances, 49
- DRT, *see* Discourse Representation Theory
- Ducrot, Oswald, 199

- Dutch, 10
- E, 171
- Eckert, Penelope, 196
- EDU, *see* discourse unit, elementary
- efficiency, 325
- elsewhere condition, 35, 61
- email, 324, 327
- embedding, 5, 266, 333–351
 - image-text, 341–342, 345–346
- emotional state, 192
- emptiness problem, 10, 19
- English, 4, 35, 42, 55–74, 79–101,
 - 111–133, 187–204, 211–234,
 - 237–257, 265–295
- Enron corpus, 5, 324, 327–328
- entailment, 69–71, 73, 188, 267, 283,
 - 284, 288, 294
 - lexical, 72
 - proof, 72
 - syntactic, 72
- enthymeme, 197–202
- entity, 5, 83
 - numerical, 307
 - relation, 302, 303, 310, 312, 313,
 - 315, 317
 - rigidity, 180
- entity normalization, 303
- entropy, 5, 324–329
- equality, 9, 27, 34
 - quantifier free, 9
- equation
 - constraining, 27, 34–38, 40, 61
 - defining, 27, 34–37, 40
- evaluation, 323–329, 346
- event, 3, 213
 - structure, 239, 252
- Everdell, Michael, 59
- Exo-Skeletal Syntax, 59
- expert systems, 323
- exponence, 57, 63–74
 - competition, 74
 - morphophonology, 56
 - morphosyntax, 56
- expression, unseen, 274
- extra-sentential, 123
- extraction
 - long distance, 271
 - object, 272
 - subject, 271
- f-structure, *see* functional-structure
- Falk, Yehuda, 29
- fallacy identification, 239
- feature
 - category, 90
 - null, 36
- Fellbaum, Christiane, 339
- Ferrucci, David A., 323
- Feynmann diagram, 80
- figurative language, 276
- Fillmore, Charles J., 82, 84, 274
- filtering, 5, 309, 312, 316, 318, 342,
 - 347
- Findlay, Jamie Y., 41, 43, 45, 47, 49,
 - 101
- finite-state technology, 15
- finite-state transducer, 15
- Fitting, Melvin, 165
- focus, 115
- Foissy, Loïc, 80, 102
- formal computation, 1
- formal language, 26, 194
- formal model, 48
- formula, mathematical, 194
- FraCaS, 4, 188, 193, 205, 301, 312, 314,
 - 315
- FrameNet, 266, 268

Index

- Frank, Anette, 32, 239
French, 199, 224, 226, 227
fugui, 63, 66
function, 200
 Φ , 64–66
 Φ , 104
 anchor, 17, 18, 20
 annotation, 10, 16, 27, 28, 33
 bridging, 64
 correspondence, 56
 inverse, 28
 matching, 64
 operator, 115
 term, 113
 unit, 113
function application, 27
functional description, 37
functional precedence, 18
functional-structure, 2, 9, 12, 14, 15,
 17–21, 26–29, 33, 35, 36, 38,
 44, 46, 47, 56–58, 63, 65–67,
 70, 79, 89, 90
 input, 11
 minimal, 65, 67
future, 177

Gabbay, Dov M., 172
game theory, 199
Garrette, Dan, 32
Gawron, Jean Marc, 280
Gazdar, Gerald, 60
gender neutral, 270
generation, 12, 16, 20, 46, 71
 minimal pair, 252–253
generation problem, 11, 19
generic, 121
genre meta structure, 124
German, 47, 224
gerund, 114

gesture, 112, 113
Geuder, Wilhelm, 241
Gibson, Edward, 138, 139
Ginzburg, Jonathan, 200
Giorgolo, Gianluca, 39, 40, 74
Girard, Jean-Yves, 30
Giuglea, Ana-Maria, 268
Gleitman, Lila, 266, 274
Glue semantics, 2, 15, 25–48, 56, 63,
 67, 79–105
 correspondence architecture,
 45
 derivation compilation, 44
 first order, 81, 93
 overview, 30
 proof, 67
 propositional, 81, 93
 System F, 81, 93
Glue Semantics Workbench, 49
Goldberg, Adele, 265, 273, 274, 278,
 279
Goldblatt, Robert, 165
Gotham, Matthew, 32, 41, 47, 81, 85,
 99
Gowers, Tim, 103
GPT, 4, 237–257, 265–295
 consistency, 271
grammar
 arbitrary, 10
 context free, 9, 10
 parallel, 26
 rule, 16
 unification, 36
grammar engineering, 25–48
grammatical function, 9, 27, 39, 121
grammatical mapping problem, 13
grammaticality, 243, 294
 judgment, 238, 244, 246,
 249–250, 254, 267–272

- vs. acceptability, 270
- graph, 141
 - directed, 141, 144
 - dynamic, 84
 - multilink, 141
 - self-loop, 141, 149
 - undirected, 144
 - unweighted, 141
- graph database, 317
- graph edge, discourse relation, 141
- graph node, utterance, 141
- greeting, 124
- GRIPES, 201
- Grosz, Barbara J., 113, 117, 118, 138
- Gupta, Kuhika, 137
- Gupta, Pranshu, 4
- Gupta, Vineet, 42
- Gurevych, Iryna, 138

- Hagberg, Aric, 154
- Halle, Morris, 56
- hallucination, 329
- Halvorsen, Per-Kristian, 32, 91
- Harley, Heidi, 59, 88, 89
- Harvey, Carol, 227
- Haug, Dag Trygve Truslew, 10, 32, 39, 41, 43, 47
- Haugen, Jason D., 58
- Hautli-Janisz, Annette, iii, 4, 238–240, 243
- Hayes-Roth, Frederick, 323
- head, 82
 - extended, 18, 20
- Head-driven Phrase Structure Grammar, 32
- Heim, Irene, 81, 113
- Heller, Stephen, 307
- Henderson, Robert, 201
- Hendrickx, Iris, 240

- Henkin, Leon, 166
- Hepple, Mark, 44
- Herbold, Steffen, 239
- Hindi, 4, 237–257
- Hirschberg, Julia, 113
- Hobbs, Jerry, 138
- Hodler, Amy E., 317
- Holler, Anke, 117
- Hong, Giwon, 329
- Hook, Peter Edwin, 241
- Hopf algebra of trees, 2, 79–105
 - overview, 102–104
- Hornstein, Norbert, 99
- Howard, William A, 31
- HPSG, *see* Head-driven Phrase Structure Grammar
- human-computer collaboration, 5, 323–329
- Hwang, Jena D., 240, 274, 275
- Hymes, Dell H., 114
- hyponym, 218

- ICD, 305
- Icelandic, 83
- idiom, 81, 88
- idiosyncratic case, 81
- image auto-caption, 339
- image auto-tagger, 339
- implementation, 2, 15, 113, 133, 137, 203
- implication, 101
 - elimination, 93
 - linear, 30, 42
- implicative, 232
- inchoative, 70–72
- inconsistent, 66
- indefinite, 34, 35
- indexical, 114

Index

- inference, 1, 4, 165–183, 187–204,
 231–234, 310, 315, 317
 chain, 167
 defeasible, 197, 198
 logical, 193, 302
 natural language, 188
 rule, 200
 temporal, 314
- inferential closure, 166
- inflection, 88
- information extraction, 301
- information retrieval, 333–351
- information sharing, 325
- information theory, 325, 329
- information-structure, 39, 56, 70, 121
- instantiation, 72, 81, 92–94, 96, 97
- interjective, 113
- interpretation, higher order,
 182–183
- interrogative, 81, 96–100
 echo question, 99
 in situ, 98–100
 question answer, 123
- interruption, 113, 121, 127–129, 132
- intransitive, 279, 282
- inverted index, 337–338, 348
- Irmen, Lisa, 117
- island constraint, 119
- Iwata, Seizi, 289
- Iyer, Ranjani, 234
- Jackendoff, Ray, 46, 89, 273
- Jagfeld, Glorianna, 240
- Japanese, 98, 133
- Ji, Ziwei, 324
- Jia, Chao, 341
- Jiang, Congcong, 137
- Jiang, Lanlan, 137
- Joachims, Thorsten, 350
- Johnson, Mark, 10
- Johnson, Matthew Thomas, 324, 329
- joining, 5
- Joshi, Aravind, 137
- JSON, 302, 306, 316
- Jurafsky, Daniel, 36, 326, 340
- Jégou, Hervé, 347
- Kalouli, Aikaterini-Lida, 323, 351
- Kamp, Hans, 113, 115, 120
- Kaplan context, 120
- Kaplan, Ronald M., iii, 1, 9–19, 23,
 25, 27, 32, 34–36, 48, 53, 56,
 65, 77, 91, 107, 119, 134
- Karttunen, Lauri, 15, 188, 193
- Kay, Martin, 14
- Kay, Paul, 82, 84, 274
- Kennedy, Chris, 218
- Kettnerová, Václava, 240
- Khoo, Justin, 201
- Kim, Su Nam, 237
- King, Tracy Holloway, 5, 15, 16, 33,
 34, 37, 49, 247, 295, 344
- Kiparsky, Paul, 3, 35, 61, 226
- Kipper, Karin, 266, 267
- KL divergence, 326, 329
- Klein, Ewan, 32
- Klimt, Bryan, 323, 327
- Klosik, David F., 148
- Knott, Alistair, 138
- knowledge management, 323
- knowledge sharing, 328, 329
- Kocoń, Jan, 238
- Kogkalidis, Konstantinos, 194
- Kojima, Takeshi, 244
- Kokkonidis, Miltiadis, 32, 81
- Kolmogorov rule, 202
- Kong, Weize, 347
- Korean, 133

- Kornai, Andras, 105
 Koskenniemi, Kimmo, 15
 Kosmidis, Kosmas, 3, 138, 139, 149, 156
 Kotowski, Sven, 211, 223, 224, 231
 Kratzer, Angelika, 81
 Kripke, Saul A., 165, 178, 180, 183
 Kuhn, Jonas, 39, 46
 Kullback, S., 329
 Kumar, Jayant, 344, 346
 Kuppevelt, Jan Van, 138
 Kusupati, Aditya, 348
 Kühnlein, Peter, 117
- L-TAG, *see* Lexical Tree Adjoining Grammar
 Labov, William, 123, 190
 Lacasa, Lucas, 141
 Lai, Alice, 137
 Lai, Viet, 250
 Lakoff, George, 93
 Lamarche, Francois, 84
 lambda, 85
 abstract, 85
 Lamping, John, 42
 language
 formal, 34
 under-resourced, 255
 language acquisition, 266
 Lapointe, Steven, 57
 LaPorte, Joseph, 180
 Larsson, Staffan, 203
 Lascarides, Alex, 113, 118, 138, 139
 Latin, 10, 19
 LDM, *see* Linguistic Discourse Model
 Leibiz, Gottfried, 178
 Leibler, R. A., 329
 Leidner, Dorothy, 323
- Lev, Iddo, 41, 42, 44, 45
 Levelt, Willem J. M., 71
 Levin, Beth, iii, 266, 267, 270, 274, 279
 Lewin, Ian, 318
 Lewis, David K., 178
 Lewis, Patrick, 350
 lexical entry, 16, 31, 33, 39, 91
 Lexical Functional Grammar, 2,
 9–21, 25–48, 56, 79–105, 119
 correspondence architecture, 2,
 13–15, 27, 36, 45, 46, 56, 74
 optionality, 26
 theory-practice, 34–40
 lexical inference, 265–295
 lexical insertion, 65
 lexical item, 65, 96, 101
 shared, 96
 lexical resource, 240
 lexical rule, 65
 lexical semantics, 4, 121, 238, 244,
 265–295
 Lexical Tree Adjoining Grammar,
 32
 Lexical-Realizational Functional
 Grammar, 55–74
 lexicalism, 2
 strong, 57
 lexicon, 27, 58
 morphology, 88
 semantic, 88
 LFG, *see* Lexical Functional
 Grammar
 Liang, Percy, 329
 Lin, Tianyang, 138
 linear logic, 15, 30, 31, 43–45, 71
 assumption, 31
 complex antecedent, 45
 proof, 30, 31, 39, 41–44

Index

- representation, 30
- linear order, 27, 83, 101
- linearization, 80
- Linguistic Discourse Model, 2, 111–133, 139
- link
 - dynamic, 87
 - implicit, 84
 - many-to-one, 90
- LISP, 168
- list, 122
- listicon, 65
- Litman, Diane, 113
- Liu, Huidong, 341
- Liu, Jiangming, 137
- LLM, 5, 237–257, 265–295, 323, 325, 328, 329
 - BERT, 288
 - bias, 278
 - incoherent reading, 284
 - inconsistent, 286, 291
 - laziness, 248
 - probing, 138
 - syntactic generalization, 277–289
 - training, 266
- logic, 30, 105, 187, 189
 - complexity, 165
 - connective, 123, 124
 - deontic, 172, 175–177
 - first order, 166, 169, 172, 203, 268
 - formula, 13
 - higher order, 3, 165, 166, 171, 172
 - intensional, 194
 - intuitionistic implicational, 84
 - multi-modal, 3
 - non-commutative, 47
 - quantifier free, 27, 34
 - temporal, 177–178
- logical form, 69, 187, 194
- long distance binding, 120
- Loper, Edward, 268
- Lowe, Daniel M., 306
- LRFG, 2, *see* Lexical-Realizational Functional Grammar
- Lyons, John, 177
- Ma, Yuhao, 137
- MacCartney, Bill, 194
- machine learning, 197, 203, 205
- macro, 59, 61, 62, 74
 - bang, 61, 64
 - list, 61
 - root, 59, 61, 64
- main clause, 119
- Malone, Thomas W., 324, 326
- Mann, William C., 118, 139
- manner, 273
- Manning, Christopher D., 194, 333, 337, 338, 343, 349
- mapping, many-to-one, 58
- Marantz, Alec, 56, 58, 59, 83, 94
- March, J. G., 323–325
- Marculli, Matilde, 79–82, 93, 100–102, 104
- Marcu, Daniel, 113, 117, 118
- Marr, Carsten, 148
- Marsh, William, 22
- Martin, James H., 36, 326, 340
- mathematics, 105, 187
- Matsson, Arild, 203
- Maxwell III, John T., 23, 53, 77, 107, 134
- McCready, Elin, 201
- McIntyre, Andrew, 231
- McNamara, Paul, 175
- meaning, 14, 45

- uncommon, 274, 277, 294
- meaning constructor, 30, 31, 33, 43, 46, 63, 65, 67, 85, 91
 - obligatory, 73
 - optional, 73
- meaning representation, 30
- measure function, 220
- measure phrase, 214–216, 225, 234
- MedDRA, 305
- MEDLINE, 301
- Mehler, Alexander, 141
- Melchin, Paul B., 56, 64, 70
- Mendelsohn, Richard L., 165
- mental state, 191, 194
- Merchant, Jason, 58
- MeSH, 305
- meta language, 26
- meta-variable, 28
- metarule, 60
- metric, 249, 325
 - accuracy, 249, 252
 - assortative degree, 145, 146
 - assortative mixing, 145, 146
 - BLEU, 252
 - clustering, 144, 145
 - connectivity, 146
 - degree centrality, 144
 - disassortative mixing, 145, 146
 - edge fraction, 144
 - F1, 249
 - inter-annotator agreement, 140
 - kappa, 140
 - network size, 144
- Meßmer, Moritz, 49
- Michaelis, Laura A., 274
- Microsoft, 1, 21, 351
- middle, 230
- Middle English, 224
- Mihalcea, Rada, 268
- Milo, Ron, 141, 148
- Miltsakaki, Eleni, 114, 125
- Milward, David, 4, 304, 309
- minimal pair, 4
- Minimalism, 2, 32, 55, 56, 59, 69, 79–105
 - external merge, 80, 88–97
 - internal merge, 80, 92, 93, 95–100
 - merge, 59, 80–82, 88–101
- Mocho, 133
- modal
 - necessity, 172, 174, 175
 - obligation, 175, 176, 183
 - permission, 175, 176
 - possibility, 172, 174, 175
- modal logic
 - axiom, 172–175
- modal operator, 178
- modality, 121
 - irrealis, 121
 - realis, 121
- model training, 190
- modifier, 42, 45
 - skeleton, 42
- modus ponens, 30
- monotonicity, 2, 26, 27, 46, 315
- Montague, Richard, 83, 85, 194, 195
- Moot, Richard, 47, 101
- morpheme, 11
 - tense, 70, 72
- morphology, 14, 15, 35, 88, 190, 234
 - exponence, 55–74
 - marked, 61
 - realizational, 56
 - unmarked, 61
- morphology-prosody-phonology
 - path, 56, 58

Index

- morphology-semantics interface, 55–74
- morphology-syntax interface, 15
- morphosemantics, 2
- morphosyntax, 55
- Moschitti, Alessandro, 268
- Moss, Lawrence S., 194, 195
- motion verb
 - directional, 243
 - manner, 243
- MPPP, *see* morphology-prosody-phonology path
- Mukerjee, Amitabha, 239
- Mukherjee, Animesh, 141
- multi-agent system, 324, 325, 328, 329
- multi-modality, 333–351
- multi-word expression, 81, 88, 90, 96, 237, 240, 338
- Muskens, Reinhard, 195
- mutual information, 326, 329
- MWE, *see* multi-word expression
- Mycock, Louise, 15, 98
- Màrquez, Lluís, 274
- Mädche, Alexander, 305

- n2c2, 303
- n2c2-T1, 303
- Nadathur, Prerna, 223
- Nagano, Akiko, 231
- named entity recognition, 303, 306, 318
- Narasimhan, Karthik, 137
- narrative, 122, 123
- National Center for Biotechnology Information, 304
- Natural Language Inference, 188
- Natural Language Processing, 133, 304, 326

- natural reasoning, 187–204
- Navigli, Roberto, 339
- NBD, *see* nonbranching dominance condition
- Needham, Mark, 317
- negation, 192, 193, 344
- negative results, 309
- NER, *see* named entity recognition
- network, 317
 - connectivity, 141–155
 - edge, 138
 - metric, 144–146
 - subgraph, 148–152
- network analysis, 3, 137–157
- network antimotif, 149, 154
 - common cause, 156
 - linear chain, 156
- network motif, 137–157
 - common cause, 148, 152, 154
 - cycle, 148
 - Discourse GraphBank, 152–155
 - dual cause, 148, 151–152, 154, 156, 157
 - feed-forward loop, 148, 149, 151, 154, 156, 157
 - full connectivity, 151
 - linear chain, 148–151, 154
- Newman, M. E. J., 141
- Newman, Mark, 141, 158
- NIH, 303
- Nikolaeva, Irina, 56
- Niles, Ian, 166
- Nishino, Tetsuro, 10
- NLP, 137–157, 188, 194, 256
 - model, 239
 - task, 238
- Noble, Bill, 199
- node, 101
 - empty, 38

- terminal, 58, 63–65, 67, 81, 92–94
- nominal
 - quantified, 85
- nominalization, 114, 224
- non-branching dominance, 9–21
 - recursive, 10
- nonce word, 212
- nonreferential, 215
- normalization, 318
- normative force, 176
- NoSQL database, 4, 316
- noun, 35, 37, 113, 212, 213, 218, 224, 226, 227, 229, 234
 - proper, 82
 - unmarked, 34, 35
- Nuance, 1, 21
- numeral, 171
- object, 33, 85, 89, 96, 121, 126, 211–234, 278–280, 290, 291
 - cognate, 227
 - indirect, 89, 291
- observability, 324
- off-line parsability, 10
- Old English, 224, 226
- ontology, 165–183, 268, 304–306, 310, 315, 317, 318
 - relation, 166, 168, 169, 176
- Open AI, 238, 265
- operator, 122
 - alethic, 172–175
 - coproduct, 93
 - disjoint union, 80
 - doxastic, 175
 - epistemic, 174–175
 - form set, 100
 - logical, 113, 126
 - modal, 173
 - most specific common denominator, 122
- Q, 96
- tense, 63
- Optimality Theory, 37
- optionality, 26, 27, 37–40, 45–47
- order of application, 83
- ordering, partial, 43
- organizational behavior, 327
 - theory, 324
- organizational health, 324
- Otten, Jens, 165
- overgeneration, 25–48
- p-diagram, 56
- Pacuit, Eric, 178
- Page, Lawrence, 337
- Pal, Santanu, 240
- Palmer, Martha, iii, 4
- Panini, 61
- paradox, 172
- PARC, 1, 21, 74, 295, 351
- parenthetical, 118, 121, 122
- ParGram project, 26, 36
- parsing, 20, 46, 71, 119, 273
- Parsons, Terence, 43
- part of speech, 4, 140, 240
- part-of-speech tagging, 238, 244, 246, 250
- Partee, Barbara H., 87
- particle, 89
- passive, 213, 215
- past, 177
- Patejuk, Agnieszka, 43
- path, 274
- Paulson, Lawrence C., 165, 166
- Pease, Adam, 3, 166, 168, 171
- Penn Discourse Treebank, 125
- Perrier, Guy, 101

Index

- persona, 196, 199, 202
- Peters, Stanley, 12
- phonological form, 69
- phonological string, 56
- phonology, 14
- phrase structure
 - annotation, 58
 - rule, 27
- Piñón, Chris, 234
- pied piping, 97
- pointer, 89, 90
- Polanyi, Livia, 2, 111, 117, 119, 121, 123, 125, 127, 133, 138, 139
- polarity, 101
- Pollard, Carl, 71
- polysemy, 215
- possession, alienable, 290
- possible world, 178, 180
- Powerset, 1, 21, 351
- Pradeep, Ronak, 349, 350
- pragmatics, 40, 139, 228
- PRED FN, 72, 73
- predicate
 - resultative, 231
 - variable, 170, 182
 - with arguments, 82
- predictability, 324
- prefixation, 211–234
 - productive, 212
- premise, 30, 33, 101, 102, 188
 - atomic, 30
 - complex, 30
 - modifier, 41–45, 47
 - skeleton, 42, 43
- preposition, 89, 213–215, 280, 285, 286, 288, 289, 294
 - argument, 89, 90
 - directional, 279, 281–286, 288, 289
 - neutral, 279–282
- Prior, Arthur Norman, 177
- probability, 197
 - theory, 202
- process, 213
- product quantization, 347
- production, 71
- projection architecture, 46
- Prolog, 168
- prompt, 4, 238, 244–247, 268
 - chain of thought, 239
 - language, 237–257
 - one shot, 246–247, 253
 - parameters, 247
 - zero shot, 238, 245–246, 249, 250, 253
- pronoun, 132, 278
- pronunciation, 190
- proof, 93
 - normal form, 42, 44
 - tree, 30
- proofnet, 44, 84, 101–102
 - polarity, 83
- PropBank, 240, 268
- proper anchoring, 9–21
- proposition, 83
 - atomic, 92
- prosodic-structure, 56
- prosody, 14, 15, 39
 - prosodic hierarchy, 15
 - prosodic inversion, 58
 - prosodic word, 58
- prosody-syntax interface, 15
- Przepiórkowski, Adam, 43
- Prüst, Hub, 122
- psychology, 325
- Pustejovsky, James, 280
- Python, 154, 155, 203
- PYTTR, 203

- probilistic, 203
- quantification, 99
 - linear universal, 81
 - universal, 198
- quantifier, 41, 42, 44, 45, 47, 85, 98,
 - 99, 120, 168
 - existential, 39
 - raising, 81, 85
 - scope, 67, 81, 195
- question answering, 1, 165, 334, 351
- quoting, 3, 127, 129–132
- R, 155
- Rad, Siavash Rafiee, 40
- Radford, Alec, 340, 341
- Ramchand, Gillian Catriona, 58, 241
- Ranta, Arne, 195
- Raths, Thomas, 165
- RDF, 302, 310, 314–315, 317
- realization, 57
 - problem, 10, 11
- reasoning, 2
 - locality, 197
 - multi-modal, 165–183
 - natural, 3
 - probabilistic variational, 193
- recognition problem, 9, 19
- Recognizing Textual Entailment,
 - 188
- record type, 200
- recoverability of deletion, 12
- recursion, 16–20
- reduplication, 15
- reentrancy, 79
- regression testing, 48
- Reimers, Nils, 138
- relation
 - Γ , 13, 14
 - coherence, 140–141
 - conservative, 13–16, 19
 - inverse, 17
 - parameterization, 180
 - post-syntax, 14–16
 - pre-syntax, 14, 16
 - regular, 15
 - rigidity, 180
 - temporal sequence, 142
 - undirected, 144
- relational database, 304, 312–314,
 - 316
- Relevance Theory, 199
- resource sensitivity, 33
 - hypothesis, 12
- resultative, 228, 231–233
- Retoré, Christian, 47
- Reyle, Uwe, 120
- rhetoric, 113, 197
 - rhetorical relation, 118
- Rhetorical Structure Theory, 113,
 - 118, 139, 140, 142, 151
 - Discourse Treebank, 141
- Riemenschneider, Frederick, 239
- right edge, 117, 119, 120, 122, 132
 - constraint, 3, 117, 132, 133
- Ritchie, R. W., 12
- Roach, Kelly, 10
- Robertson, Stephen, 349
- Rohde, Hannah, 138
- root, 59, 64, 89, 90
 - individuation, 59
- Rothwell, William, 227
- row variable, 170–171
- rule-based system, 308
- Ruppenhofer, Joseph, 266
- Sacks, Harvey, 124
- Safran-Naveh, Gila, 22

Index

- Sag, Ivan A., 71
saliency, 189
Sanchez Duran, Magali, 240
Sanderson, Paul, 346
Saraswat, Vijay, 36
Sassen, Claudia, 117
Savitch, Walter J., 22
Scha, Remko, 111, 127, 138
Schadow, Gunther, 307
Schegloff, Emanuel A., 124
Schiffrin, Deborah, 114
Schlangen, David, 204
Schuler, Karin K., 266, 267, 274
Schulhoff, Sander, 244, 255
Schulz, Stephan, 166
scope, 79, 81, 96, 98, 99
 quantifier, 82, 99
 restriction, 81
Scott, Dana, 178
search, 5, 333–351
 diversity, 343
 hybrid, 345–350
 image-to-image, 335
 keyword, 5, 335, 337–339, 342, 347
 latency, 342, 346–349
 metadata, 339
 mission-based, 334
 multimodal, 5
 null and low, 342, 347
 numerals, 343
 semantic, 1, 340
 user behavior data, 343, 345, 349
 vector, 5, 347
 web, 333, 335
search index, 337–339, 347, 348
search ranking, 343, 349–350
Seeber, Isabella, 323
Segmented Discourse
 Representation Theory, 139, 140
selectional restriction, 229
semantic
 form, 29, 59
 role, 274
 type, 87–88
semantic assembly, 88–96
semantic lexicon entry, 88–91, 93, 94, 96–100
semantic projection, 33
semantic representation, 301–318
semantic role labeling, 268, 273
semantic triple, 314
semantic web, 4, 302, 304
semantic-structure, 26, 33, 56
semantics, 15, 25–48, 57, 79–105, 115, 118, 120, 123, 126, 165–183, 187–204, 228, 234, 266, 273
 assembly point, 90–93
 compositional, 79, 82
 computational, 25–48, 188
 context, 111–133
 dynamic, 113
 event, 43, 44, 70
 lexical, 115
 lexicon, 88
 model theoretic, 194
 Montague, 79, 81, 83, 85, 194, 195
 natural language, 81
 possible worlds, 178
 relation, 139
 relational, 178
 theoretical, 188
 type theory, 83, 84
Serbo-Croatian, 37
set

- multiset, 80
- union of, 63
- set theory, 94
- SFMP, *see* structure-function-meaning path
- Shah, Chirag, iii, 334, 351
- Shakespeare, 227
- Shannon, Claude Elwood, 323, 325, 326, 329
- Shen-Orr, Shai S., 148
- Shi, Freda, 239
- Shi, Lei, 268
- Siddiqi, Daniel, 15, 56, 58, 59, 63, 64, 69, 73, 75
- Sidner, Candace L., 113, 117, 118, 138
- Sigma, 171
- Silué, Songfólo Lacina, 49
- similarity threshold, 342, 343
- Simon, Herbert A., 323–325
- Singh, Mona, 241
- Sinha, R. Mahesh K., 239
- Sinnott-Armstrong, Walter, 172
- situation, 200
- SKOS, 315
- SNLI, 188
- social framing, 114
- social meaning, 196
- sociolinguistics, 196, 204
 - variation, 190–193
- Song, Chaoming, 141
- Sorokina, Daria, 343, 349
- sorting, 5
- spanning, 58, 63
- SPARQL, 314, 315
- speaker, message, 71
- speech event, 114, 119, 120, 124
- spellout, 234
- Sperber, Daniel, 199
- Sporns, Olaf, 148
- spreadsheet data, 311–312
- SQL, 312
- SRI, 1
- SRI Cambridge, 1
- Stabler, Edward P., 80
- STAC, 138, 156
- state, 213
- state of affairs, 122
- stateful system, 329
- statistical patterns, 274
- Steedman, Mark, 129
- storytelling, 125
- string anchor, 17, 18, 20
- Strogatz, Steven H., 141
- structure-function-meaning path,
 - 56
- structured data, 303, 308, 312
- Structured Discourse
 - Representation Theory, 118
- Stubbs, Amber, 303
- Stump, Gregory T., 55
- sub-event, 231
- subcategorization, 4, 29, 266
- subject, 28, 33, 89, 94, 96, 100, 121, 122, 126, 228–230, 291
 - locative, 230
- subset principle, 66
- subsistence, 12
- subsumption, 66, 67
- Suda, Martin, 183
- Sulger, Sebastian, 238
- summarization, 133
- SUMO, 3, 165–183
 - axiom, 166, 168–170
 - modal attribute, 173
- Sun, Xiangyan, 344
- SUO-KIF, 168, 170, 171, 178
- suppletion, 70
- Sutcliffe, Geoff, 168, 171

Index

- Sutton, Peter, 205
Svenonius, Peter, 58
syllogism, 198
symbolic analysis, 265
symbolic system, 275
synonym, 122
syntactic derivation
 proper anchoring, 17–20
syntax, 9–21, 39, 40, 42, 46, 55, 58,
 79–105, 114, 118–120, 132,
 228, 230, 234, 244, 266, 268,
 273
 structure, 140, 274
 theoretical-computational,
 25–48
 variation, 267
syntax-semantics interface, 26,
 32–34, 47, 81

table joining, 311–314, 316, 318
tabular data, 4, 302, 310–311
TAG, *see* Tree Adjoining Grammar
Talman, Aarne, 188
Tambi, Ritiz, 346
Tantos, Alexandros, 3, 138, 139, 149,
 156
taxonomy, 305
template, 40, 59
tense, 63, 70, 88
 past, 36
tensor, 103, 104
Tetreault, Joel, 137
text mining, 301–318
 biomedical, 4, 301–318
theorem prover, 3, 165–183
thesaurus, 305
theta-role, 80, 213, 228
Thomas, James, 302
Thompson, Sandra A., 118, 139

tokenization, 337
topic, 115, 125
topos, 3, 197–202
TPTP, 168, 170, 171
TQM, 181
Trac, Steven, 168
trace, 82, 100
trajectory, 280–283, 286, 288
Transcriptional Regulatory
 Network, 149
Transformational Grammar, 12, 93,
 104
transformer model, 138
translation, 4, 238, 240, 252
tree, 90–94, 101–103
 binary branching, 42, 59, 80,
 100
 forest, 103, 104
 multiple branching, 42
 n-ary branching, 80, 93, 100
 rooted, 84
 unordered, 80, 93, 103, 104
Tree Adjoining Grammar, 101, 129
tree description language, 90
Treinish, Matthew, 154
triple, 315, 317
Trivedi, Sapna, 309
Troelstra, Anne S., 81, 101
Tsagkias, Manos, 343
type guard, 169–170, 182
type restriction, 169
Type Theory with Records, 200,
 202, 203
 probabilistic, 202
typed first order form, 181–183
 with arithmetic, 171
Typed First order form with
 Arithmetic, 181, 182

- typed higher order form, 171, 179,
181–183
- undecidability, 11
- underspecification, 60, 61
- ungrammaticality, 213
- Universität Konstanz, iii, 295
- University of Colorado, 295
- Urdu, 4, 237–257
- v-structure, *see*
vocabulary-structure
- Vaidya, Ashwini, 238
- valency alternation, 267–272
- value
default, 37
- Vampire, 171
- Van Benthem, Johan, 194
- Van De Putte, Frederik, 175
- Van den Berg, Martin H., iii, 2
- Van den Berg, Martin H., 113, 122
- Van der Plas, Lonneke, 240
- Van der Sandt, Rob A., 120
- Van Eijck, Jan, 120
- Van Genabith, Josef, 32, 41, 101
- Van Noord, Rik, 137
- variable arity relation, 170
- variable, list, 170
- variation, 202, 204
- Vaswani, Ashish, 138, 323
- vector
representation, 266
space, 102–104
- vector search
database, 347
- Vera, Alonso H., 324, 329
- verb, 31, 113, 114, 211–234
agreement, 58, 70, 72
alternation, 288
dative, 294
locative, 279, 281
- argument, 43, 89
argument as modifier, 43, 44
auxiliary, 35
change of location, 286
exertion of force, 286
force exerting, 279–281, 289
implicative, 192, 194
information transfer, 292
intransitive, 212, 213, 226
locative, 224
movement, 279, 280, 282, 284,
286, 288, 289
participle, 35, 36
passive, 35, 39
predicative, 213
transitive, 33, 39, 96, 212, 213,
226, 234, 278, 288
weather, 229, 230
- VerbNet, 266–268, 274, 275, 295
Body Part, 270
Dative, 270
Manner of Speaking, 275
Understood Reciprocal Object,
270
- VI, *see* vocabulary item
- Vidgen, Bertie, 329
- Vieu, Laure, 140
- visualization, 311
- vocabulary, 57, 62, 63, 65, 67, 69, 74
lookup, 56, 63–69
size, 326–328
- vocabulary item, 57, 58, 62–67,
69–73
- vocabulary-structure, 56
- vocative, 113
- Wadekar, Shakti N., 340

Index

- Waletsky, Joshua, 123
Walker, Marilyn A., 199
Wang, Bang, 137
Wang, Chen, 338
Wang, Zhenyi, 324
Wang, Zhongyuan, 344
Wang, Zihao, 325
Warlpiri, 37
Wasow, Thomas, 12
Watts, Duncan J., 141
Webber, Bonnie, 117, 137
Wedekind, Jürgen, 2, 9–12, 16–19, 21,
32, 34, 36
Weininger, David, 307
Weissweiler, Leonie, iii, 280
Welsh, 34
Wilson, Deirdre, 199
Wilson, Michael, 288
WIPO, 304
witness condition, 202–204
Wolf, Florian, 138, 139
Wood, Jim, 94
word, 9, 11, 141
 function, 278, 338
 stop, 338
word order, 42, 47, 58, 244, 246
 canonical, 271
 permutation, 238
 variation, 246, 250–251
WordNet, 295, 339
world knowledge, 166, 174, 177,
187–204, 219, 266, 270, 277,
294
Wu, Changxing, 137
Xiang, Wei, 137
XLE, 1, 15, 25, 33–39, 74, 119, 295
 transfer component, 15
 transfer system, 33
XML, 304, 308, 310
Yang, Enneng, 324
Yang, Yiming, 323, 327
z-score, 149, 152, 154, 156
 heatmap, 154
Zaenen, Annie, 4, 23, 53, 77, 101, 105,
107, 134, 188, 193, 295, 351
Zaragoza, Hugo, 349
Zhang, Yingxue, 137
Zheng, Feliz, 4
Zheng, Zhaohui, 350
Zymla, Mark-Matthias, iii, 2

Bibliography

- Achiam, Josh, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. *GPT-4 technical report*. <https://arxiv.org/abs/2303.08774>.
- Afantenos, Stergos, Nicholas Asher, Farah Benamara, Myriam Bras, Cécile Fabre, Mai Ho-dac, Anne L. Draoulec, Philippe Muller, Marie-Paule Péry-Woodley & Laurent Prévot. 2012. An empirical resource for discovering cognitive principles of discourse organisation: The ANNODIS corpus. In *Proceedings of the eighth international conference on language resources and evaluation (LREC'12)*, 2727–2734. <https://aclanthology.org/L12-1498/>.
- Afantenos, Stergos D. & Nicholas Asher. 2010. Testing SDRT's right frontier. In Chu-Ren Huang & Daniel Jurafsky (eds.), *Proceedings of the COLING 2010, 23rd international conference on computational linguistics*, 1–9. Tsinghua University Press.
- Agashe, Saaket, Yue Fan, Anthony Reyna & Xin Eric Wang. 2023. *LLM-coordination: Evaluating and analyzing multi-agent coordination abilities in large language models*. <https://arxiv.org/abs/2310.03903>.
- Ahmed, Tafseer & Miriam Butt. 2011. Discovering semantic classes for Urdu N-V complex predicates. In Johan Bos & Stephen Pulman (eds.), *Proceedings of the ninth international conference on computational semantics (IWCS 2011)*. <https://aclanthology.org/W11-0132>.
- Ahmed, Tafseer, Miriam Butt, Annette Hautli & Sebastian Sulger. 2012. A reference dependency bank for analyzing complex predicates. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the eighth international conference on language resources and evaluation (LREC'12)*, 3145–3152. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/pdf/474_Paper.pdf.
- Ahn, Byron. 2022. Mapping *out*-argument structure. *Syntax* 25. 417–465.
- Alavi, Maryam & Dorothy Leidner. 2001. Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly* 1.

Bibliography

- Allen, James F. 1984. Towards a general theory of action and time. *Artificial Intelligence* 23. 123–154. <https://api.semanticscholar.org/CorpusID:22263624>.
- Alon, Uri. 2019. *An introduction to systems biology: Design principles of biological circuits*. CRC Press.
- Aloni, Maria, Harald Bastiaanse, Tikitu de Jager & Katrin Schulz (eds.). 2010. *Logic, language, and meaning: Selected papers from the seventeenth Amsterdam colloquium*. Springer.
- Amanaki, Eirini, Jean-Philippe Bernardy, Stergios Chatzikyriakidis, Robin Cooper, Simon Dobnik, Aram Karimi, Adam Ek, Eirini Chrysovalantou Gianikouri, Vasiliki Katsouli, Ilias Kolokousis, Eirini Chrysovalantou Mamatzaki, Dimitrios Papadakis, Olga Petrova, Erofilis Psaltaki, Charikleia Soupiona, Efrosyni Skoulataki & Christina Stefanidou. 2022. Fine-grained entailment: Resources for Greek NLI and precise entailment. In Jonne Sälevä & Constantine Lignos (eds.), *Proceedings of the workshop on dataset creation for lower-resourced languages within the 13th language resources and evaluation conference (LREC)*, 44–52. European Language Resources Association. <https://aclanthology.org/2022.dclrl-1.6>.
- Amazon. 2024a. *Amazon Comprehend Medical*. On-line documentation. <https://aws.amazon.com/comprehend/medical/>.
- Amazon. 2024b. *Amazon Comprehend Medical SNOMED CT linking*. On-line documentation. <https://docs.aws.amazon.com/comprehend-medical/latest/dev/ontology-linking-snomed.html>.
- Ambrosy, Andrew P., Rishi V. Parikh, Sue Hee Sung, Anand Narayanan, Rajeev Masson, Phuong-Quang Lam, Kevin Kheder, Alan Iwahashi, Alexander B. Hardwick, Jesse K. Fitzpatrick, Harshith R. Avula, Van N. Selby, Xian Shen, Navneet Sanghera, Joaquim Cristino & Alan S. Go. 2021. A natural language processing-based approach for identifying hospitalizations for worsening heart failure within an integrated health care delivery system. *JAMA Network Open*. November 1, 2021. DOI: 10.1001/jamanetworkopen.2021.35152.
- Anderson, Stephen Robert. 1969. *West Scandinavian vowel systems and the ordering of phonological rules*. Massachusetts Institute of Technology. (Doctoral dissertation).
- Andrews, Avery. 2025. Glue Semantics for Hopf-Algebraic Minimalism. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 79–108. Konstanz: PubliKon.
- Andrews, Avery D. 2016. *Propositional glue semantics is fully variable-free*. <https://semanticsarchive.net/Archive/mY4MTI2N/gluevarfree.pdf>.

- Andrews, Avery D. 1990. Unification and morphological blocking. *Natural Language & Linguistic Theory* 8(4). 507–557.
- Andrews, Avery D. 2007. Generating the input in OT-LFG. In Jane Grimshaw, Jane Simpson, Tracy Holloway King, Joan Maling, Christopher D. Manning & Annie Zaenen (eds.), *Architectures, rules, and preferences: A festschrift for Joan Bresnan*, 319–340. CSLI Publications.
- Andrews, Avery D. 2008. The role of PRED in LFG+glue. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG08 conference*, 46–76. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2008/lfg08.html>.
- Andrews, Avery D. 2010. Propositional glue and the correspondence architecture of LFG. *Linguistics and Philosophy* 33. 141–170.
- Andrews, Avery D. 2019. A one-level analysis for Icelandic Quirky Case. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG19 conference*, 24–27. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2019/index.shtml>.
- Andrews, Peter B. 1972. General models and extensionality. *Journal of Symbolic Logic* 37. 395–397.
- Andrews, Peter B., Matthew Bishop, Sunil Issar, Dan Nesmith, Frank Pfenning & Hongwei Xi. 1996. TPS: A theorem-proving system for classical type theory. *Journal of Automated Reasoning* 16(3). 321–353.
- Anwar, Maaz, Riyaz Ahmad Bhat, Dipti Sharma, Ashwini Vaidya, Martha Palmer & Tafseer Ahmed Khan. 2016. A proposition bank of Urdu. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the tenth international conference on language resources and evaluation (LREC'16)*, 2379–2386. European Language Resources Association (ELRA). <https://aclanthology.org/L16-1377>.
- Ardila, Frederico. 2012. *Hopf algebras*. <https://fardila.com/Clase/Hopf/lectures.html>.
- Aristotle. 2007. *On rhetoric: A theory of civic discourse*. 2nd edn. Translated with introduction, notes, and appendices by George A. Kennedy. Oxford University Press.
- Arka, I Wayan, Ash Asudeh & Tracy Holloway King (eds.). 2021. *Modular design of grammar: Linguistics on the edge*. Oxford University Press.
- Arora, Abhishek, Xinmei Yang, Shao-Yu Jheng & Melissa Dell. 2024. *Linking representations with multimodal contrastive learning*. <https://arxiv.org/abs/2304.03464>.
- Asher, Nicholas. 1993. *Reference to abstract objects in discourse*. Springer.

Bibliography

- Asher, Nicholas. 2008. Troubles on the right frontier. In Anton Benz & Peter Kühnlein (eds.), *Constraints in discourse*, vol. 2 (Pragmatics & Beyond New Series), 29–52. John Benjamins Publishing Company. DOI: 10.1075/pbns.172.02ash. <https://doi.org/10.1075/pbns.172.02ash>.
- Asher, Nicholas, Julie Hunter & Kate Thompson. 2020. Modelling structures for situated discourse. *Dialogue Discourse* 11. 89–121.
- Asher, Nicholas & Alex Lascarides. 1998a. Bridging. *Journal of Semantics* 15. 83–113. <https://doi.org/10.1093/jos/15.1.83>.
- Asher, Nicholas & Alex Lascarides. 1998b. The semantics and pragmatics of presupposition. *Journal of Semantics* 15. 239–300. <https://doi.org/10.1093/jos/15.3.239>.
- Asher, Nicholas & Alex Lascarides. 2003. *Logics of conversation*. Cambridge University Press.
- Asher, Nicholas, Antoine Venant, Philippe Muller & Stergos Afantenos. 2011. Complex discourse units and their semantics. In Nicholas Asher (ed.), *CID 2011—constraints in discourse*. Agay.
- Asher, Nicholas & Laure Vieu. 2005. Subordinating and coordinating discourse relations. *Lingua* 115. 591–610.
- Asher, Nicolas, Julie Hunter, Mathieu Morey, Benamara Farah & Stergos Afantenos. 2016. Discourse structure and dialogue acts in multiparty dialogue: The STAC corpus. In *Proceedings of the tenth international conference on language resources and evaluation (LREC 2016)*, 2721–2727.
- Asudeh, Ash. 2004. *Resumption as resource management*. Stanford University. (Doctoral dissertation).
- Asudeh, Ash. 2005. Control and semantic resource sensitivity. *Journal of Linguistics* 41(3). 465–511.
- Asudeh, Ash. 2006. Direct compositionality and the architecture of LFG. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*, 363–387. CSLI Publications.
- Asudeh, Ash. 2012. *The logic of pronominal resumption* (Oxford Studies in Theoretical Linguistics). Oxford University Press. DOI: 10.1093/acprof:oso/9780199206421.001.0001. <https://doi.org/10.1093/acprof:oso/9780199206421.001.0001>.
- Asudeh, Ash. 2021. The unrealized and the unheard. In I Wayan Arka, Ash Asudeh & Tracy Holloway King (eds.), *Modular design of grammar: Linguistics on the edge*, 22–42. Oxford University Press.
- Asudeh, Ash. 2022. Glue semantics. *Annual Review of Linguistics* 8. 321–341.

- Asudeh, Ash. 2023. Glue semantics. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 651–697. Language Science Press. DOI: 10.5281/zenodo.10185964. <https://doi.org/10.5281/zenodo.10185964>.
- Asudeh, Ash. 2025. Morphological Exponence and Morphosemantics. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 55–78. Konstanz: PubliKon.
- Asudeh, Ash, Bronwyn Bjorkman, Frances Dowle, Neil Myler, Daniel Siddiqi & Lisa Sullivan. 2024. Fusional morphology, metasyncretism, and secondary exponence: A morphemic, realizational approach to Latin declension. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG24 conference*, 44–68. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/3>.
- Asudeh, Ash, Tina Bögel & Daniel Siddiqi. 2023. Modelling exponents. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG23 conference*, 23–44. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/2>.
- Asudeh, Ash & Richard Crouch. 2002a. Coordination and parallelism in glue semantics: Integrating discourse cohesion and the element constraint. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG02 conference*, 19–39. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2002/lfg02.html>.
- Asudeh, Ash & Richard Crouch. 2002b. Derivational parallelism and ellipsis parallelism. In Line Mikkelsen & Christopher Potts (eds.), *WCCFL 21 proceedings*, 1–14. Cascadilla Press.
- Asudeh, Ash & Richard Crouch. 2002c. Glue semantics for HPSG. In *Proceedings of the 8th international HPSG conference*, 1–19. CSLI Publications.
- Asudeh, Ash, Mary Dalrymple & Ida Toivonen. 2008. Constructions with lexical integrity: Templates as the lexicon-syntax interface. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG08 conference*, 68–88. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2008/lfg08.html>.
- Asudeh, Ash, Mary Dalrymple & Ida Toivonen. 2013. Constructions with lexical integrity. *Journal of Language Modelling* 1(1). 1–54.
- Asudeh, Ash & Gianluca Giorgolo. 2012. Flexible composition for optional and derived arguments. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG12 conference*, 64–84. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2012/lfg12.html>.

Bibliography

- Asudeh, Ash & Gianluca Giorgolo. 2020. *Enriched meanings: Natural language semantics with category theory*, vol. 13. Oxford University Press.
- Asudeh, Ash, Gianluca Giorgolo & Ida Toivonen. 2014. Meaning and valency. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG14 conference*, 68–88. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2014/lfg14.html>.
- Asudeh, Ash, Paul B. Melchin & Daniel Siddiqi. 2024. Constraints all the way down: DM in a representational model of grammar. In *Proceedings of the 39th West Coast Conference on Formal Linguistics (WCCFL)*, 468–476. Cascadilla Press. <http://www.lingref.com/cpp/wccfl/39/index.html>.
- Asudeh, Ash & Siavash Rafiee Rad. 2023. Persian perception verbs. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG23 conference*, 45–65. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/2>.
- Asudeh, Ash & Daniel Siddiqi. 2022. Realizational morphosemantics in L_RFG . In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG22 conference*, 20–40. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/1>.
- Asudeh, Ash & Daniel Siddiqi. 2023. Morphology in LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 855–901. Language Science Press. DOI: 10.5281/zenodo.10185976. <https://doi.org/10.5281/zenodo.10185976>.
- Asudeh, Ash & Daniel Siddiqi (eds.). 2026. *Constraint-based morphosyntax: The L_RFG approach*. Forthcoming. With contributions by Oleg Belyaev, Bronwyn Bjorkman, Tina Bögel, Frances Dowle, Michael Everdell, Paul B. Melchin, Neil Myler, Will Oxford, and Lisa Sullivan. Cambridge University Press.
- Bachmair, Leo, Harald Ganzinger, Christopher Lynch & Wayne Snyder. 1992. Basic paramodulation and superposition. In *Proceedings of the 11th international conference on automated deduction*, vol. 607 (Lecture Notes in Computer Science), 462–476. Springer-Verlag. DOI: 10.1007/3-540-55602-8_185.
- Baldrige, Jason, Nicholas Asher & Julie Hunter. 2007. Annotation for and robust parsing of discourse structure on unrestricted texts. *Zeitschrift für Sprachwissenschaft* 26. 213–239. <https://doi.org/10.1515/ZFS.2007.018>.
- Baldwin, Timothy & Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkha & Fred J. Damerau (eds.), *Handbook of natural language processing (2nd ed.)* Chapman & Hall/CRC. DOI: <https://doi.org/10.1201/9781420085938>.
- Barabási, Albert-László. 2013. Network science. *Philosophical Transactions of the Royal Society A – Mathematical, Physical and Engineering Sciences* 371. 20120375.

- Barabási, Albert-László & Réka Albert. 1999. Emergence of scaling in random networks. *Science* 286. 509–512.
- Barabási, Albert-László, Natali Gulbahce & Joseph Loscalzo. 2011. Network medicine: A network-based approach to human disease. *Nature Reviews Genetics* 12. 56–68.
- Barabási, Albert-László & Márton Pósfai. 2016. *Network science*. Cambridge University Press.
- Barančíková, Petra & Václava Kettnerová. 2017. ParaDi: Dictionary of paraphrases of Czech complex predicates with light verbs. In Stella Markantonatou, Carlos Ramisch, Agata Savary & Veronika Vincze (eds.), *Proceedings of the 13th workshop on multiword expressions (MWE 2017)*, 1–10. Association for Computational Linguistics. DOI: 10.18653/v1/W17-1701. <https://aclanthology.org/W17-1701>.
- Barsalou, Lawrence W. 1982. Context-independent and context-dependent information in concepts. *Memory & Cognition* 10(1). 82–93.
- Barzel, Baruch & Albert-László Barabási. 2013. Universality in network dynamics. *Nature Physics* 9. 673–681. <https://doi.org/10.1038/nphys2741>.
- Bashan, Amir, Ronny P. Bartsch, Jan W. Kantelhardt, Shlomo Havlin & Plamen C. Ivanov. 2012. Network physiology reveals relations between network topology and physiological function. *Nature Communications* 3. <https://doi.org/10.1038/ncomms1705>.
- Battiston, Stefano, J. Doyne Farmer, Andreas Flache, Diego Garlaschelli, Andrew G. Haldane, Hans Heesterbeek, Cars Hommes, Carlo Jaeger, Robert May & Marten Scheffer. 2016. Complexity theory and financial regulation. *Science* 351. 818–819. <https://doi.org/10.1126/science.aad0299>.
- Beber, Moritz E., Christoph Fretter, Shubham Jain, Nikolaus Sonnenschein, Matthias Müller-Hannemann & Marc-Thorsten Hütt. 2012. Artefacts in statistical analyses of network motifs: General framework and application to metabolic networks. *Journal of the Royal Society Interface* 9. 3426–3435.
- Beesley, Kenneth & Lauri Karttunen. 2003. *Finite-state morphology*. CSLI Publications.
- BEL. 2024. *Biological expression language*. Online documentation. <https://biological-expression-language.github.io/>.
- Belleau, François, Marc-Alexandre Nolin, Nicole Tourigny, Philippe Rigault & Jean Morissette. 2008. Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. *Journal of Biomedical Informatics* 41(5). 706–716. DOI: 10.1016/j.jbi.2008.03.004..

Bibliography

- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major & Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 610–623.
- Benzmüller, Christoph & Lawrence C. Paulson. 2013. Quantified multimodal logics in simple type theory. *Logica Universalis* 7(1). 7–20. DOI: 10.1007/s11787-012-0052-y.
- Benzmüller, Christoph & Adam Pease. 2010. Progress in automating higher-order ontology reasoning. In Boris Konev, Renate Schmidt & Stephan Schulz (eds.), *Workshop on practical aspects of automated reasoning (PAAR-2010)*. CEUR Workshop Proceedings.
- Benzmüller, Christoph & Adam Pease. 2012. Higher-order aspects and context in SUMO. In Jos Lehmann, Ivan José Varzinczak & Alan Bundy (eds.), *Special issue on reasoning with context in the semantic web*, vol. 12-13. Science, Services & Agents on the World Wide Web.
- Benzmüller, Christoph, Florian Rabe & Geoff Sutcliffe. 2008. THF0 – the core of the TPTP language for higher-order logic. In Alessandro Armando, Peter Baumgartner & Gilles Dowek (eds.), *Automated reasoning*, 491–506. Springer.
- Bernardy, Jean-Philippe & Stergios Chatzikyriakidis. 2020. Improving the precision of natural textual entailment problem datasets. In *Proceedings of the twelfth language resources and evaluation conference (LREC)*, 6835–6840.
- Bernódusson, Helgi. 1982. *Ópersunólegar Setningar [impersonal sentences]*. University of Iceland. (MA thesis).
- BioCreative. 2024. *Biocreative Workshop*. On-line workshop materials. <https://biocreative.bioinformatics.udel.edu/>.
- BioCreative Track 1. 2024. *Track 1: BioRED (Biomedical Relation Extraction Dataset) Track*. On-line task materials. <https://biocreative.bioinformatics.udel.edu/tasks/biocreative-viii/track-1/>.
- BioPortal. 2024. *BioPortal: Ontologies*. On-line resource. <https://bioportal.bioontology.org/ontologies>.
- Bjørner, Nikolaj & Andrei Voronkov (eds.). 2012. *Logic for Programming, Artificial Intelligence, and Reasoning - 18th International Conference LPAR-18 Proceedings*. Vol. 7180. (LNCS). Springer. <http://dx.doi.org/10.1007/978-3-642-28717-6>.
- Blackburn, Patrick, Maarten de Rijke & Yde Venema. 2001. *Modal logic* (Cambridge Tracts in Theoretical Computer Science 53). Cambridge University Press.
- Bloomfield, Leonard. 1925. On the sound-system of central Algonquian. *Language* 1(4). 130–156.

- Bobrow, Daniel, Dick Crouch, Tracy Holloway King, Cleo Condoravdi, Lauri Karttunen, Rowan Nairn, Valeria de Paiva & Annie Zaenen. 2007. Precision-focused textual inference. In *Proceedings of the ACL-PASCAL workshop on textual entailment and paraphrasing*, 16–21.
- Bögel, Tina. 2015. *The syntax-prosody interface in Lexical Functional Grammar*. University of Konstanz. (Doctoral dissertation).
- Bögel, Tina. 2023. Prosody and its interfaces. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 779–821. Language Science Press. DOI: 10.5281/zenodo.10185970. <https://doi.org/10.5281/zenodo.10185970>.
- Bögel, Tina, Miriam Butt, Ronald M. Kaplan, Tracy Holloway King & John T. Maxwell III. 2009. Prosodic phonology in LFG: A new proposal. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG09 conference*, 146–166. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2009/lfg09.html>.
- Bögel, Tina, Miriam Butt, Ronald M. Kaplan, Tracy Holloway King & John T. Maxwell III. 2010. Second position and the prosody-syntax interface. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG10 conference*, 106–126. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2010/lfg10.html>.
- Bolt, Joe, Bob Coecke, Fabrizio Genovese, Martha Lewis, Dan Marsden & Robin Piedeleu. 2019. Interacting conceptual spaces I: Grammatical composition of concepts. In Mauri Kaipainen, Frank Zenker, Antti Hautamäki & Peter Gärdenfors (eds.), *Conceptual spaces: Elaborations and applications*, 151–181. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-12800-5_9.
- Bonial, Claire, William Corvey, Martha Palmer, Volha V. Petukhova & Harry Bunt. 2011. A hierarchical unification of LIRICS and VerbNet semantic roles. In *2011 IEEE fifth international conference on semantic computing (IEEE)*, 483–489.
- Borer, Hagit. 2005a. *Structuring sense, vol. I: In name only*. Oxford University Press.
- Borer, Hagit. 2005b. *Structuring sense, vol. II: The normal course of events*. Oxford University Press.
- Borer, Hagit. 2005c. *Structuring sense, vol. III: Taking form*. Oxford University Press.
- Börjars, Kersti. 2020. Lexical-Functional Grammar: An overview. *Annual Review of Linguistics* 6. 155–172.
- Börjars, Kersti, Rachel Nordlinger & Louisa Sadler. 2019. *Lexical-Functional Grammar: An introduction*. Cambridge University Press.

Bibliography

- Boston, Marisa Ferrara. 2025. An Information Theoretic Evaluation of AI Cooperation. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 323–332. Konstanz: PubliKon.
- Bowman, Samuel R., Gabor Angeli, Christopher Potts & Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Conference on empirical methods in natural language processing (EMNLP)*, 632–642.
- Boyer, Lorie A., Tong Ihn Lee, Megan F. Cole, Sarah E. Johnstone, Stuart S. Levine, Jacob P. Zucker, Matthew G. Guenther, Roshan M. Kumar, Heather L. Murray, Richard G. Jenner, David K. Gifford, Douglas A. Melton, Rudolf Jaenisch & Richard A. Young. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122. 947–956.
- Breitholtz, Ellen. 2020. *Enthymemes and topoi in dialogue: The use of common sense reasoning in conversation*, vol. 41 (Current Research in the Semantics/Pragmatics Interface). Brill.
- Breitholtz, Ellen & Robin Cooper. 2021. Dogwhistles as inferences in interaction. In *Proceedings of the reasoning and interaction conference (ReInAct 2021)*, 40–46. ACL. <https://aclanthology.org/2021.reinact-1.6.pdf>.
- Bresnan, Joan. 1980. Polyadicity. In Teun Hoekstra, Harry van der Hulst & Michael Moortgat (eds.), *Lexical grammar*, 97–121. Revised as Joan Bresnan, ed., 1982. *The Mental Representation of Grammatical Relations*, Chapter 3, 149–172. MIT Press. Foris.
- Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-functional syntax*. 2nd edn. Wiley-Blackwell.
- Bresnan, Joan, Ronald M. Kaplan, Stanley Peters & Annie Zaenen. 1982. Cross-serial dependencies in Dutch. *Linguistic Inquiry* 13(4). 613–635. <https://www.jstor.org/stable/4178298>. Reprinted in Savitch, Bach, Marsh & Safran-Naveh (1987: 286-319).
- Brin, Sergey & Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks* 30. 107–117. <http://www-db.stanford.edu/~backrub/google.html>.
- Brown, Chad E., Adam Pease & Josef Urban. 2023. Translating SUMO-K to higher-order set theory. In *International symposium on frontiers of combining systems*. <https://api.semanticscholar.org/CorpusID:258685309>.
- Brown, Susan W., Julia Bonn, James Gung, Annie Zaenen, James Pustejovsky & Martha Palmer. 2019. VerbNet representations: Subevent semantics for transfer verbs. In *Proceedings of the first international workshop on designing meaning representations*, 154–163.

- Brown, Susan W., Julia Bonn, Ghazaleh Kazeminejad, Annie Zaenen, James Pustejovsky & Martha Palmer. 2022. Semantic representations for NLP using VerbNet and the generative lexicon. *Frontiers in artificial intelligence* 5.
- Brown, Susan W., Dmitriy Dligach & Martha Palmer. 2014. VerbNet class assignment as a WSD task. In *Computing meaning*, 203–216. Springer.
- Brown, Susan W., James Pustejovsky, Annie Zaenen & Martha Palmer. 2018. Integrating generative lexicon event structures into VerbNet. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*, 56–61.
- Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever & Dario Amodei. 2020. *Language models are few-shot learners*. <https://arxiv.org/abs/2005.14165>.
- Bullmore, Ed & Olaf Sporns. 2009. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* 10. 186–198.
- Bunt, Harry. 2008. Semantic underspecification: Which technique for what purpose? In *Computing meaning*, 55–85. Springer.
- Burnett, Heather. 2019. Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy* 42(5). 419–450.
- Butler, Jenna, Sonya Jaffe, Nancy Baym, Mary Czerwinski, Shamsi Iqbal, Kate Nowak, Sean Rintel, Abigail Sellen, Mihaela Vorvoreanu, Brent Hecht & Jaime Teevan. 2023. *Microsoft new future of work report 2023*. Tech. rep.
- Butt, Miriam. 1995. *Complex predicates in Urdu*. CSLI Publications.
- Butt, Miriam. 1998. Constraining argument merger through aspect. In Erhard Hinrichs, Andreas Kathol & Tsuneko Nakazawa (eds.), *Syntax and semantics, complex predicates in nonderivational syntax*, vol. 30, 73–113. Academic.
- Butt, Miriam. 2010. The light verb jungle: Still hacking away. In Mengistu Amberber, Brett Baker & Mark Harvey (eds.), *Complex predicates: Cross-linguistic perspectives on event structure*, 48–78. Cambridge University Press.
- Butt, Miriam, Tina Bögel, Annette Hautli, Sebastian Sulger & Tafseer Ahmed. 2012. Identifying Urdu complex predication via bigram extraction. In Martin Kay & Christian Boitet (eds.), *Proceedings of COLING 2012*, 409–424. The COLING 2012 Organizing Committee. <https://aclanthology.org/C12-1026>.

Bibliography

- Butt, Miriam, Mary Dalrymple & Tracy Holloway King (eds.). 2006. *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*. CSLI Publications.
- Butt, Miriam, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi & Christian Rohrer. 2002. The parallel grammar project. In *Proceedings of the 2002 workshop on grammar engineering and evaluation*, vol. 15, 1–7. Association for Computational Linguistics.
- Butt, Miriam, Jamie Y. Findlay & Ida Toivonen (eds.). 2021. *Proceedings of the LFG21 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2021/lfg21.html>.
- Butt, Miriam, Jamie Y. Findlay & Ida Toivonen (eds.). 2022. *Proceedings of the LFG22 Conference*. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/1>.
- Butt, Miriam, Jamie Y. Findlay & Ida Toivonen (eds.). 2023. *Proceedings of the LFG23 Conference*. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/2>.
- Butt, Miriam, Jamie Y. Findlay & Ida Toivonen (eds.). 2024. *Proceedings of the LFG24 Conference*. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/3>.
- Butt, Miriam & Wilhelm Geuder. 2001. On the (semi)lexical status of light verbs. In Norbert Corver & Henk van Riemsdijk (eds.), *Semi-lexical categories: On the content of function words and the function of content words*, 323–370. Mouton de Gruyter.
- Butt, Miriam & Tracy Holloway King. 1998a. Interfacing phonology with LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG98 conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG1998/index.html>.
- Butt, Miriam & Tracy Holloway King (eds.). 1998b. *Proceedings of the LFG98 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG1998/index.html>.
- Butt, Miriam & Tracy Holloway King (eds.). 2001. *Proceedings of the LFG01 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2001/lfg01.html>.
- Butt, Miriam & Tracy Holloway King (eds.). 2002. *Proceedings of the LFG02 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2002/lfg02.html>.
- Butt, Miriam & Tracy Holloway King (eds.). 2004. *Proceedings of the LFG04 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2004/lfg04.html>.

- Butt, Miriam & Tracy Holloway King (eds.). 2006a. *Proceedings of the LFG06 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2006/lfg06.html>.
- Butt, Miriam & Tracy Holloway King. 2006b. Restriction for morphological valency alternations: The Urdu causative. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*, 235–258. CSLI Publications.
- Butt, Miriam & Tracy Holloway King (eds.). 2008. *Proceedings of the LFG08 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2008/lfg08.html>.
- Butt, Miriam & Tracy Holloway King (eds.). 2009. *Proceedings of the LFG09 Conference*. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2009/lfg09.html>.
- Butt, Miriam & Tracy Holloway King (eds.). 2010. *Proceedings of the LFG10 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2010/lfg10.html>.
- Butt, Miriam & Tracy Holloway King (eds.). 2011. *Proceedings of the LFG11 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2011/lfg11.html>.
- Butt, Miriam & Tracy Holloway King (eds.). 2012. *Proceedings of the LFG12 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2012/lfg12.html>.
- Butt, Miriam & Tracy Holloway King (eds.). 2013. *Proceedings of the LFG13 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2013/lfg13.html>.
- Butt, Miriam & Tracy Holloway King (eds.). 2014. *Proceedings of the LFG14 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2014/lfg14.html>.
- Butt, Miriam & Tracy Holloway King (eds.). 2018. *Proceedings of the LFG18 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2018/index.shtml>.
- Butt, Miriam & Tracy Holloway King (eds.). 2019. *Proceedings of the LFG19 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2019/index.shtml>.
- Butt, Miriam, Tracy Holloway King, María-Eugenia Niño & Frédérique Segond. 1999. *A grammar writer's cookbook*. CSLI Publications.
- Butt, Miriam, Tracy Holloway King & Gillian Catriona Ramchand. 2009. Complex predication: Who made the child pinch the elephant? In Linda Uyechi &

Bibliography

- Lian Hee Wee (eds.), *Reality exploration and discovery: Pattern interaction in language and life*, 231–256. CSLI Publications.
- Butt, Miriam & Gillian Catriona Ramchand. 2005. Complex aspectual structure in Hindi/Urdu. In Nomi Ertischik-Shir & Tova Rapoport (eds.), *The syntax of aspect*, 117–153. Oxford University Press.
- Butt, Miriam & Ida Toivonen (eds.). 2020. *Proceedings of the LFG20 Conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2020/index.shtml>.
- Buttcher, Stefan, Charles L. A. Clarke & Gordon V. Cormack. 2016. *Information retrieval: Implementing and evaluating search engines*. The MIT Press.
- Cabitzza, Federico, Andrea Campagner & Valerio Basile. 2023. Toward a perspectivist turn in ground truthing for predictive computing. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, 6860–6868. DOI: 10.1609/aaai.v37i6.25840. <https://ojs.aaai.org/index.php/AAAI/article/view/25840>.
- Cai, Zhenguang G., Xufeng Duan, David A. Haslett, Shuqi Wang & Martin J. Pickering. 2024. *Do large language models resemble humans in language use?* <https://arxiv.org/abs/2303.08014>.
- Camilleri, Maris & Louisa Sadler. 2016. Relativisation in Maltese. *Transactions of the Philological Society* 114(1). 117–145. DOI: 10.1111/1467-968X.12070. <https://doi.org/10.1111/1467-968X.12070>.
- Cao, Yong, Li Zhou, Seolhwa Lee, Laura Cabello, Min Chen & Daniel Hershcovich. 2023. Assessing cross-cultural alignment between ChatGPT and human societies: An empirical study. In Sunipa Dev, Vinodkumar Prabhakaran, David Adelani, Dirk Hovy & Luciana Benotti (eds.), *Proceedings of the first workshop on cross-cultural considerations in NLP (C3NLP)*, 53–67. Association for Computational Linguistics. DOI: 10.18653/v1/2023.c3nlp-1.7. <https://aclanthology.org/2023.c3nlp-1.7>.
- Cappelen, Herman, Ernest Lepore & Matthew McKeever. 2023. Quotation. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford encyclopedia of philosophy*, Summer 2023. Metaphysics Research Lab, Stanford University.
- Carlson, Lynn, Daniel Marcu & Mary Ellen Okurowski. 2002. RST discourse treebank, LDC2002T07. In Linguistic Data Consortium. University of Pennsylvania.
- Carreras, Xavier & Lluís Màrquez. 2005. Introduction to the CoNLL-2005 shared task: Semantic role labeling. In *Proceedings of the ninth conference on computational natural language learning (CoNLL-2005)*, 152–164.
- Cases, Ignacio, Clemens Rosenbaum, Matthew Riemer, Atticus Geiger, Tim Klinger, Alex Tamkin, Olivia Li, Sandhini Agarwal, Joshua D. Greene, Daniel Jurafsky, Christopher Potts & Lauri Karttunen. 2019. Recursive routing net-

- works: Learning to compose modules for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies (ACL-HLT), volume 1 (long and short papers)*, 3631–3648. Association for Computational Linguistics. DOI: 10.18653/v1/N19-1365. <https://aclanthology.org/N19-1365>.
- CBInsights. 2024. *Future of the workforce: How AI agents will transform enterprise workflows*. Tech. rep. <https://www.cbinsights.com/research/report/future-workforce-ai-agents/>.
- Chase, Harrison. 2022. *Langchain*. Online code and documentation. <https://github.com/langchain-ai/langchain>.
- Chatzichrisafis, Nikos, Dick Crouch, Tracy Holloway King, Rowan Nairn, Manny Rayner & Marianne Santaholma. 2007. Regression testing for grammar-based systems. In *Proceedings of the grammar engineering across frameworks (GEAF07) workshop*, 128–143. CSLI Publications.
- ChEBI. 2024. (*chemical entities of biological interest*). Online resource. <https://www.ebi.ac.uk/chebi/aboutChebiForward.do>.
- Chen, Haonan, Zhicheng Dou, Qiannan Zhu, Xiaochen Zuo & Ji-Rong Wen. 2023. Integrating representation and interaction for context-aware document ranking. *ACM Transactions on Information Systems* 41. 1–23.
- Chen, Tongfei, Zhengping Jiang, Adam Poliak, Keisuke Sakaguchi & Benjamin Van Durme. 2020. Uncertain natural language inference. In Daniel Jurafsky, Joyce Chai, Natalie Schluter & Joel Tetreault (eds.), *Proceedings of the 58th annual meeting of the Association for Computational Linguistics (ACL)*, 8772–8779. Association for Computational Linguistics. DOI: 10.18653/v1/2020.acl-main.774. <https://aclanthology.org/2020.acl-main.774>.
- Chomsky, Noam. 1957. *Syntactic structures*. Mouton.
- Chomsky, Noam. 1970. Remarks on nominalization. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in English transformational grammar*, 184–221. Ginn & Company.
- Chomsky, Noam. 1995. *The Minimalist Program*. MIT Press.
- Chomsky, Noam, T. Daniel Seely, Robert C. Berwick, Sandiway Fong, M.A.C. Huybrechts, Hisatsugu Kitihara, Andrew McInnerney & Yushi Sugimoto. 2023. *Merge and the strong Minimalist thesis*. Cambridge University Press.
- Choudhury, Monojit & Animesh Mukherjee. 2009. The structure and dynamics of linguistic networks. In *Dynamics on and of complex networks*, 145–166. Springer.
- Church, Alonzo. 1940. A formulation of the simple theory of types. *Journal of Symbolic Logic* 5. 56–68.

Bibliography

- Clarke, Edmund M. & Andrei Voronkov (eds.). 2010. *Logic for programming, artificial intelligence, and reasoning (LAPAR) – revised selected papers*, vol. 6355 (Lecture Notes in Computer Science). Springer. <http://dx.doi.org/10.1007/978-3-642-17511-4>.
- Clauset, Aaron, Christopher Moore & M.E.J. Newman. 2008. Hierarchical structure and the prediction of missing links in networks. *Nature* 453. 98–101. <https://doi.org/10.1038/nature06830>.
- Comrie, Bernard. 1981. *Language universals and linguistic typology*. Basil Blackwell.
- Condoravdi, Cleo, Dick Crouch, Valeria De Paiva, Reihardt Stolle & Danny G. Bobrow. 2003. Entailment, intensionality and text understanding. In *Proceedings of the HLT-NAACL 2003 workshop on text meaning*, 38–45.
- Condoravdi, Cleo & Tracy Holloway King (eds.). 2019. *Tokens of meaning: Papers in honor of Lauri Karttunen*. CSLI Publications.
- Condoravdi, Cleo & Paul Kiparsky. 2025. Out-Applicatives are Concealed Comparatives. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 211–235. Konstanz: PubliKon.
- Constant, Mathieu, Gülşen Eryiğit, Johanna Monti, Lonneke van der Plas, Carlos Ramisch, Michael Rosner & Amalia Todirascu. 2017. Multiword expression processing: A survey. *Computational Linguistics* 43(4). 837–892. <https://aclanthology.org/J17-4005>.
- Cooper, Alexis. 2024. Differing implications in caused motion constructions. Unpublished paper, University of Colorado Boulder.
- Cooper, Robin. 2019. Types as learnable cognitive resources in PyTTR. In Cleo Condoravdi & Tracy Holloway King (eds.), *Tokens of meaning: Papers in honor of Lauri Karttunen*. CSLI Publications.
- Cooper, Robin. 2023. *From perception to communication: A theory of types for action and meaning*. Open access. Oxford University Press. <https://global.oup.com/academic/product/from-perception-to-communication-9780192871312>.
- Cooper, Robin. 2025. Variation in Natural Reasoning. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 187–208. Konstanz: PubliKon.
- Cooper, Robin, Dick Crouch, Jan van Eijck, Chris Fox, Josef van Genabith, Jan Jaspars, Hans Kamp, David Milward, Manfred Pinkal, Massimo Poesio & Steve Pulman. 1996. *Using the framework: FraCaS deliverable D16*. Tech. rep. <https://gu-clasp.github.io/multifracas/D16.pdf>.

- Cooper, Robin, Simon Dobnik, Shalom Lappin & Staffan Larsson. 2015. Probabilistic type theory and natural language semantics. *Linguistic Issues in Language Technology* 10(4). 1–45. <https://journals.colorado.edu/index.php/lilt/article/view/1357/1199>.
- Cooper, Robin & Jonathan Ginzburg. 2015. Type theory with records for natural language semantics. In Shalom Lappin & Chris Fox (eds.), *The handbook of contemporary semantic theory*, 2nd edn., 375–407. Wiley-Blackwell.
- Cooper, Robin & Ruth Kempson (eds.). 2008. *Language in flux: Dialogue coordination, language variation, change and evolution*, vol. 1 (Communication, Mind and Language). College Publications.
- Cooper, Robin & Aarne Ranta. 2008. Natural languages as collections of resources. In Robin Cooper & Ruth Kempson (eds.), *Language in flux: Dialogue coordination, language variation, change and evolution*, vol. 1 (Communication, Mind and Language), 109–120. College Publications.
- Cooperative Patent Classification*. 2024. On-line documentation. <https://www.cooperativepatentclassification.org/>.
- Cormack, James, Chinmoy Nath, David Milward, Kalpana Raja & Siddhartha Jonnalagadda. 2015. Agile text mining for the 2014 i2b2/UTHealth cardiac risk factors challenge. *Journal of Biomedical Informatics* 15. DOI: 10.1016/j.jbi.2015.06.030.
- Cox, Andrew. 2022. The ethics of AI for information professionals: Eight scenarios. *Journal of the Australian Library and Information Association* 71. 201–214.
- Creswell, Cassandre, Katherine Forbes, Eleni Miltsakaki, Rashmi Prasad, Aravind K. Joshi & Bonnie Webber. 2003. Penn discourse treebank: Building a large scale annotated corpus encoding DLTAG-based discourse structure and discourse relations. In *Proceedings of the Penn Discourse Treebank*.
- Cristea, Dan. 2005. The right frontier constraint holds unconditionally. In *Proceedings of the multidisciplinary approaches to discourse 2005 (MAD'05)*.
- Crouch, Dick, Mary Dalrymple, Ronald Kaplan, Tracy Holloway King, John T. Maxwell III & Paula Newman. 2011. *XLE Documentation*. Palo Alto Research Center (PARC). <https://ling.sprachwiss.uni-konstanz.de/pages/xle/index.html>.
- Crouch, Dick & Tracy Holloway King. 2004. Unifying lexical resources. In *Proceedings of the interdisciplinary workshop on the identification and representation of verb features and verb classes*.
- Crouch, Richard. 2005. Packed rewriting for mapping semantics to KR. In *Proceedings of the sixth international workshop on computational semantics (IWCS-6)*, 103–114.

Bibliography

- Crouch, Richard. 2006. Packed rewriting for mapping text to semantics and KR. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: variations on a theme by Ronald M. Kaplan*, 389–416.
- Crouch, Richard & Tracy Holloway King. 2006. Semantics via f-structure rewriting. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG06 conference*, 145–165. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2006/lfg06.html>.
- Crouch, Richard & Josef van Genabith. 1999. Context change, underspecification, and the structure of glue language derivations. In Mary Dalrymple (ed.), *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*, 117–189. The MIT Press.
- Crouch, Richard & Josef van Genabith. 2000. *Linear logic for linguists*. <http://www.coli.uni-saarland.de/courses/logical-grammar/contents/crouch-genabith.pdf>.
- Curry, Haskell Brooks, Robert Feys, William Craig, J. Roger Hindley & Jonathan P. Seldin. 1958. *Combinatory logic*, vol. 1. North-Holland Amsterdam.
- D'Souza, Raissa M., Jesus Gómez-Gardeñes, Jan Nagler & Alex Arenas. 2019. Explosive phenomena in complex networks. *Advances in Physics* 68. 123–223. <https://doi.org/10.1080/00018732.2019.1650450>.
- Dafoe, Allan, Edward Hughes, Yoram Bachrach, Tantum Collins, Kevin R. McKee, Joel Z. Leibo, Kate Larson & Thore Graepel. 2021. Open problems in cooperative AI. In *NEURIPS 2021 (Cooperative AI Workshop)*. arXiv. <https://arxiv.org/abs/2012.08630>.
- Dalrymple, Mary (ed.). 1999. *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*. The MIT Press.
- Dalrymple, Mary. 2001. *Lexical Functional Grammar*, vol. 34 (Syntax and Semantics). Academic Press.
- Dalrymple, Mary (ed.). 2024. *Handbook of Lexical Functional Grammar*. Language Science Press.
- Dalrymple, Mary, Vineet Gupta, John Lamping & Vijay Saraswat. 1999. Relating resource-based semantics to categorial semantics. In Mary Dalrymple (ed.), *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*, 261–280. The MIT Press.
- Dalrymple, Mary, Ronald M. Kaplan & Tracy Holloway King. 2004. Linguistic generalizations over descriptions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG04 conference*, 199–208. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2004/lfg04.html>.
- Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell III & Annie Zaenen (eds.). 1995. *Formal issues in Lexical-Functional Grammar*. CSLI Publications.

- Dalrymple, Mary, John Lamping, Fernando Pereira & Vijay Saraswat. 1997. Quantification, anaphora and intensionality. *Logic, Language and Information* 6. Appears with some modifications in Dalrymple (1999), pp. 39–90, 219–273.
- Dalrymple, Mary, John Lamping & Vijay Saraswat. 1993. LFG semantics via constraints. In *Proceedings of the sixth conference of the European chapter of the Association for Computational Linguistics (EACL) (EACL '93)*, 97–105. Association for Computational Linguistics. DOI: 10.3115/976744.976757. <https://doi.org/10.3115/976744.976757>.
- Dalrymple, Mary, John J. Lowe & Louise Mycock. 2019. *The Oxford reference guide to Lexical Functional Grammar*. Oxford University Press.
- Dalrymple, Mary & Louise Mycock. 2011. The prosody-semantics interface. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 conference*, 173–193. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2011/lfg11.html>.
- Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge University Press.
- Dang, Hao T., Karin Kipper, Martha Palmer & Joseph Rosenzweig. 1998. Investigating regular sense extensions based on intersective Levin classes. In *36th annual meeting of the Association for Computational Linguistics (ACL) and 17th international conference on computational linguistics*, vol. 1, 293–299.
- Dask Development Team. 2016. *Dask: Library for Dynamic Task Scheduling*. Software package. <https://docs.dask.org/en/stable/>.
- Davidson, Donald. 1979. Quotation. *Theory and Decision* 11(1). 27–40. DOI: 10.1007/bf00126690. <https://doi.org/10.1007/bf00126690>.
- de Groote, Philippe. 1999. An algebraic correctness criterion for intuitionistic multiplicative proof-nets. *Theoretical Computer Science* 224(1-2). 115–134.
- Dentella, Vittoria, Elliot Murphy, Gary Marcus & Evelina Leivada. 2023. *Testing AI performance on less frequent aspects of language reveals insensitivity to underlying meaning*. <https://arxiv.org/abs/2302.12313>.
- Diesner, Jana, Terrill L. Frantz & Kathleen M. Carley. 2005. Communication networks from the Enron email corpus: “It’s always about the people. Enron is no different”. *Computational and Mathematical Organizational Theory* 11(3). 201–228.
- Doetjes, Jenny. 2010. Incommensurability. In *Logic, language and meaning: 17th Amsterdam colloquium revised selected papers*, 254–263. Springer.
- Ducrot, Oswald. 1988. Topoi et formes topiques. *Bulletin d’études de la linguistique française* 22. 1–14.
- Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in the study of variation. *Annual Review of Anthropology* 41. 87–100.

Bibliography

- Egg, Markus & Gisela Redeker. 2010. How complex is discourse structure? In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *Proceedings of the seventh international conference on language resources and evaluation (LREC'10)*. European Language Resources Association (ELRA).
- Everdell, Michael, Paul B. Melchin, Ash Asudeh & Daniel Siddiqi. 2021. Beyond c-structure and f-structure: On the argument-adjunct distinction in O'dam. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG21 conference*, 125–145. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2021/lfg21.html>.
- Falk, Yehuda. 2011. *Lexical-Functional Grammar*. Oxford University Press.
- Fellbaum, Christiane (ed.). 1998. *WordNet: An electronic lexical database*. The MIT Press.
- Ferrucci, David A., Eric W. Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John M. Prager, Nico Schlaefter & Chris Welty. 2010. Building Watson: An overview of the DeepQA project. *AI Magazine* 31. 59–79.
- Fillmore, Charles J. 1988. The mechanisms of construction grammar. In Shelley Axmaker, Annie Jaissner & Helen Singmaster (eds.), *Proceedings of the Berkeley linguistics society (BLS)*, vol. 14, 35–55.
- Findlay, Jamie Y. 2019. *Multiword expressions and the lexicon*. Oxford University. (Doctoral dissertation).
- Findlay, Jamie Y. 2021. Meaning in LFG. In I Wayan Arka, Ash Asudeh & Tracy Holloway King (eds.), *Modular design of grammar: Linguistics on the edge*, 340–374. Oxford University Press.
- Findlay, Jamie Y. & Dag Trygve Truslew Haug. 2022. Managing scope ambiguities in glue via multistage proving. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG22 conference*, 144–163. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/1>.
- Fisher, Cynthia, Lila Gleitman & Henry Gleitman. 1991. On the semantic content of subcategorization frames. *Cognitive Psychology* 23. 331–392.
- Fitting, Melvin & Richard L. Mendelsohn. 1998. *First-order modal logic*. Kluwer. DOI: doi:10.1007/978-94-011-5292-1.
- Foissy, Loïc. 2013. *An introduction to the Hopf algebras of trees*. <https://www2.mathematik.hu-berlin.de/~kreimer/wp-content/uploads/Foissy.pdf>.
- Frank, Anette & Josef van Genabith. 2001. Linear logic-based semantics construction for LTAG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG01 conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2001/lfg01.html>.

- Gabbay, Dov M., John F. Horty, Xavier Parent, Ron van der Meyden & Leendert van der Torre. 2013. *Handbook of deontic logic and normative systems*. College Publications. <https://api.semanticscholar.org/CorpusID:60312009>.
- Garrette, Dan & Ewan Klein. 2009. An extensible toolkit for computational semantics. In *Proceedings of the eighth international conference on computational semantics*, 116–127.
- Gawron, Jean Marc. 1985. Situations and prepositions. *Linguistics and Philosophy* 9(4). 427–476.
- Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum & Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Harvard University Press.
- GeneOntology. 2024. Online resource. <https://www.geneontology.org/>.
- Girard, Jean-Yves. 1995. Linear logic: Its syntax and semantics. *London Mathematical Society Lecture Note Series*. 1–42.
- Giuglea, Ana-Maria & Alessandro Moschitti. 2006. Semantic role labeling via FrameNet, VerbNet and PropBank. In *Proceedings of the 21st international conference on computational linguistics and the 44th annual meeting of the Association for Computational Linguistics (ACL)*, 929–936. Association for Computational Linguistics.
- Gleitman, Lila. 1990. Structural sources of verb learning. *Language Acquisition* 1(1). 1–63.
- Goldberg, Adele. 1995. *A Construction Grammar approach to argument structure*. The University of Chicago Press.
- Goldberg, Adele. 2005. *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Goldberg, Adele & Ray Jackendoff. 2004. The English resultative as a family of constructions. *Language* 80. 532–568. DOI: 10.1353/Language,2004.0129.
- Goldblatt, Robert. 2003. Mathematical modal logic: A view of its evolution. *Journal of Applied Logic* 1(5). 309–392. DOI: [https://doi.org/10.1016/S1570-8683\(03\)00008-9](https://doi.org/10.1016/S1570-8683(03)00008-9). <https://www.sciencedirect.com/science/article/pii/S1570868303000089>.
- GoogleAPI. 2024. On-line documentation. <https://cloud.google.com/healthcare-api/docs/concepts/nlp>.
- Gotham, Matthew. 2015. Towards glue semantics for Minimalist syntax. *Cambridge Occasional Papers in Linguistics* 8. 56–83.
- Gotham, Matthew. 2018. Making logical form type-logical: Glue semantics for Minimalist syntax. *Linguistics and Philosophy* 41(5). 511–556.
- Gotham, Matthew. 2019. Constraining scope ambiguity in LFG+Glue. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG19 conference*. CSLI

Bibliography

- Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2019/index.shtml>.
- Gotham, Matthew. 2021. Approaches to scope islands in LFG+glue. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG21 conference*, 146–166. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2021/lfg21.html>.
- Gotham, Matthew & Dag Trygve Truslew Haug. 2018. Glue semantics for universal dependencies. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG18 conference*. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2018/index.shtml>.
- Gowers, Tim. 2024. *How to lose your fear of tensor products*. Undated webpage accessed in 2024. <https://www.dpmms.cam.ac.uk/~wtg10/tensors3.html>.
- Grosz, Barbara J. & Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3). 175–204. <https://aclanthology.org/J86-3001>.
- Guidi, Barbara, Andrea Michienzi & Laura Ricci. 2020. A graph-based socioeconomic analysis of Steemit. *IEEE Transactions on Computational Social Systems* 8. 365–376.
- Gupta, Kuhika, Joseph Ripberger, Andrew Fox, Hank Jenkins-Smith & Carol Silva. 2022. Discourse network analysis of nuclear narratives. In *Narratives and the policy process: Applications of the narrative policy framework*. Pressbooks.
- Gupta, Vineet & John Lamping. 1998. Efficient linear logic meaning assembly. In *Proceedings of the 17th international conference on computational linguistics – volume 1*, 464–470.
- Hagberg, Aric, Pieter Swart & Daniel S. Chult. 2008. *Exploring network structure, dynamics, and function using NetworkX*. Tech. rep.
- Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In Samuel Jay Keyser & Ken Hale (eds.), *The view from Building 20*, 111–176. MIT Press.
- Halvorsen, Per-Kristian & Ronald M. Kaplan. 1988. Projections and semantic description in Lexical-Functional Grammar. In *Proceedings of the international conference on fifth generation computer systems*, 1116–1122. Also published in Dalrymple, Kaplan, Maxwell & Zaenen (1995), pp. 279–292. Institute for New Generation Computer Technology.
- Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40(3–4). 225–276.
- Haug, Dag Trygve Truslew. 2008. Tense and aspect for glue semantics: The case of participial XADJs. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings*

- of the LFG08 conference, 291–311. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2008/lfg08.html>.
- Haug, Dag Trygve Truslew. 2017. Syntactic discontinuities in Latin. In Victoria Rosén & Koenraad De Smedt (eds.), *The very model of a modern linguist: In honor of Helge Dyvik*, 75–96. Bergen Language & Linguistic Studies.
- Haugen, Jason D. & Daniel Siddiqi. 2016. Towards a restricted realizational theory: Multimorphemic monolistemicity, portmanteaux, and post-linearization spanning. In Daniel Siddiqi & Heidi Harley (eds.), *Morphological metatheory*, 343–386. John Benjamins.
- Hautli-Janisz, Annette. 2013. Moving right along: Motion verb sequences in Urdu. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG13 conference*, 295–215. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2013/lfg13.html>.
- Hautli-Janisz, Annette. 2015. Urdu/Hindi complex predicates of motion and the manner/result complementarity. *Journal for South Asian Languages* 7.
- Hautli-Janisz, Annette & Erisa Bytyqi. 2024. The facts speak for themselves: Fallacy detection with GPT. In *Proceedings of OSSA13*.
- Hautli-Janisz, Annette, Erisa Bytyqi, Pranshu Gupta & Miriam Butt. 2025. Revisiting the Light Verb Jungle: How Good Is GPT at Hacking Away? In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 237–263. Konstanz: PubliKon.
- Hautli-Janisz, Annette, Tracy Holloway King & Gilian Catriona Ramchand. 2015. Encoding event structure in Urdu/Hindi VerbNet. In Eduard Hovy, Teruko Mitamura & Martha Palmer (eds.), *Proceedings of the 3rd workshop on EVENTS: Definition, detection, coreference, and representation*, 25–33. Association for Computational Linguistics. DOI: 10.3115/v1/W15-0804. <https://aclanthology.org/W15-0804>.
- Hayes-Roth, Frederick. 1983. *Building expert systems*. Addison-Wesley.
- Heim, Irene. 1988. *The semantics of definite and indefinite noun phrases* (Outstanding Dissertations in Linguistics). Garland Publishing. <http://semanticsarchive.net/Archive/Tk0ZmYyY/dissertation.pdf>.
- Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Blackwell.
- Heller, Stephen, Alan McNaught, Stephen Stein, Dmitrii Tchekhovskoi & Igor Pletnev. 2013. InChI – the worldwide chemical structure identifier standard. *Journal of Cheminformatics* 5(1). PMC 3599061. DOI: 10.1186/1758-2946-5-7.

Bibliography

- Henderson, Robert & Elin McCready. 2024. *Signaling without Saying: The Semantics and Pragmatics of Dogwhistles* (Oxford Studies in Semantics and Pragmatics 17). Oxford: Oxford University Press.
- Hendrickx, Iris, Amália Mendes, Sílvia Pereira, Anabela Gonçalves & Inês Duarte. 2010. Complex predicates annotation in a corpus of Portuguese. In Nianwen Xue & Massimo Poesio (eds.), *Proceedings of the fourth linguistic annotation workshop*, 100–108. Association for Computational Linguistics. <https://aclanthology.org/W10-1812>.
- Henkin, Leon. 1950. Completeness in the theory of types. *Journal of Symbolic Logic* 15. 81–91.
- Hepple, Mark. 1996. A compilation-chart method for linear categorial deduction. In *Proceedings of the 16th conference on computational linguistics — volume 1*, 537–542.
- Herbold, Steffen, Annette Hautli-Janisz, Ute Heuer, Zlata Kikteva & Alexander Trautsch. 2023. A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports* 13(1). DOI: 10.1038/s41598-023-45644-9. <http://dx.doi.org/10.1038/s41598-023-45644-9>.
- Hobbs, Jerry. 1985. *On the coherence and structure of discourse*. Tech. rep. CSLI Research report number 85–37.
- Holler, Anke & Lisa Irmen. 2007. Empirically assessing effects of the right frontier constraint. In António Branco (ed.), *Anaphora: Analysis, algorithms and applications*, 15–27. Springer.
- Hong, Giwon, Aryo Pradipta Gema, Rohit Saxena, Xiaotang Du, Ping Nie, Yu Zhao, Laura Perez-Beltrachini, Max Ryabinin, Xuanli He, Clementine Fourier & Pasquale Minervini. 2024. *The hallucinations leaderboard—an open effort to measure hallucinations in Large Language Models*. <https://arxiv.org/abs/2404.05904>.
- Hook, Peter Edwin. 1974. *The compound verb in Hindi*. Center for South and South-east Asian Studies. The University of Michigan.
- Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30. 69–96.
- Howard, William A. 1980. The formulae-as-types notion of construction. In J.P. Seldin & J.R. Hindley (eds.), *To H. B. Curry: Essays on combinatory logic, lambda calculus, and formalism*, vol. 44, 479–490.
- Howes, Christine, Stergios Chatzikiyriakidis, Adam Ek & Vidya Somashekarappa (eds.). 2020. *Proceedings of the Probability and Meaning Conference (PaM 2020)*. Association for Computational Linguistics. <https://aclanthology.org/2020.pam-1.0>.

- Hunter, Julie, Nicholas Asher & Alex Lascarides. 2015. Integrating nonlinguistic events into discourse structure. In *Proceedings of the 11th international conference on computational semantics*, 184–194.
- Hunter, Julie, Nicholas Asher & Alex Lascarides. 2018. A formal semantics for situated conversation. *Semantics and Pragmatics* 11. Article 10, 1–59.
- Hwang, Jena D. 2014. *Identification and representation of caused motion constructions*. University of Colorado at Boulder. (Doctoral dissertation).
- Hwang, Jena D., Archana Bhatia, Claire Bonial, Aous Mansouri, Ashwini Vaidya, Nianwen Xue & Martha Palmer. 2010. PropBank annotation of multilingual light verb constructions. In Nianwen Xue & Massimo Poesio (eds.), *Proceedings of the fourth linguistic annotation workshop*, 82–90. Association for Computational Linguistics. <https://aclanthology.org/W10-1810>.
- Hymes, Dell H. 1974. *Foundations in sociolinguistics: An ethnographic approach*. University of Pennsylvania Press. <https://archive.org/details/foundationsinsoc0000hyme>.
- ICD. 2024. *Icd-11 (international classification of diseases 11th revision)*. <https://icd.who.int/en>.
- Iwata, Seizi. 2020. *English resultatives: A force-recipient account*. John Benjamins Publishing Company.
- Jackendoff, Ray. 1973. The base rules for prepositional phrases. In Stephen R. Anderson & Paul Kiparsky (eds.), *A festschrift for Morris Halle*, 345–357. Holt, Rinehard & Winston.
- Jackendoff, Ray. 2010. The parallel architecture and its place in cognitive science. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 583–605. Oxford University Press.
- Jacobs, Roderick A. & Peter S. Rosenbaum (eds.). 1970. *Readings in English Transformational Grammar*. Ginn & Company.
- Jagfeld, Glorianna & Lonneke van der Plas. 2015. Towards a better semantic role labeling of complex predicates. In Diana Inkpen, Smaranda Muresan, Shibamouli Lahiri, Karen Mazidi & Alisa Zhila (eds.), *Proceedings of the 2015 conference of the North American chapter of the Association for Computational Linguistics (NAACL): student research workshop*, 33–39. Association for Computational Linguistics. DOI: 10.3115/v1/N15-2005. <https://aclanthology.org/N15-2005>.
- Jégou, Hervé, Matthijs Douze & Cordelia Schmid. 2011. Product quantization for nearest neighbor search. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33. 117–128.

Bibliography

- Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Delong Chen, Wenliang Dai, Andrea Madotto & Pascale Fung. 2022. Survey of hallucination in natural language generation. *ACM Computing Surveys* 55. 1–38.
- Jia, Chao, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc V. Le, Yunhsuan Sung, Zhen Li & Tom Duerig. 2021. Scaling up visual and vision-language representation learning with noisy text supervision. In *Proceedings of the 38th international conference on machine learning PMLR*.
- Jiang, Congcong, Tiejun Qian & Bing Liu. 2022. Knowledge distillation for discourse relation analysis. In *Proceedings of the companion proceedings of the web conference 2022*, 210–214.
- Jiang, Lanlan, Shengjun Yuan & Jun Li. 2021. A discourse coherence analysis method combining sentence embedding and dimension grid. *Complexity*. 6654925. <https://doi.org/10.1155/2021/6654925>.
- Joachims, Thorsten. 2002. Optimizing search engines using clickthrough data. In *Proceedings of the ACM conference on knowledge discovery and data mining*.
- Johnson, Mark. 1986. The LFG treatment of discontinuity and the double infinitive construction in Dutch. In *Proceedings of the fifth west coast conference on formal linguistics*, 102–118. CSLI Publications.
- Johnson, Matthew Thomas & Alonso H. Vera. 2019. No AI is an island: The case for teaming intelligence. *AI Magazine* 40. 16–28.
- JSON. 2024. *Introducing JSON*. On-line documentation. <https://www.json.org/json-en.html>.
- Jurafsky, Daniel & James H. Martin. 2009. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Pearson Prentice Hall.
- Jurafsky, Daniel & James H. Martin. 2024. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models*. 3rd edn. Online manuscript released August 20, 2024; print 2nd edition 2009. <https://web.stanford.edu/~jurafsky/slp3/>.
- Kalouli, Aikaterini-Lida, Dick Crouch & Valeria de Paiva. 2020. Hy-NLI: A hybrid system for natural language inference. In *International conference on computational linguistics*, 5235–5249.
- Kamp, Hans. 1981. A theory of truth and semantic representation. In Jeroen A. G. Groenendijk, Theo M. V. Janssen & Martin B. J. Stokhof (eds.), *Formal methods in the study of language*, vol. 1, 277–322. Mathematisch Centrum.
- Kamp, Hans & Uwe Reyle. 1993. *From discourse to logic: Introduction to model theoretic semantics of natural language, formal logic and discourse representation theory*. Kluwer Academic Publishers.

- Kaplan, Ronald M. 1987. Three seductions of computational psycholinguistics. In Peter Whitelock, Mary McGee Wood, Harold L. Somers, Rod Johnson & Paul Bennett (eds.), *Linguistic theory and computer applications*, 149–188. Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 337–367). Academic Press.
- Kaplan, Ronald M. 1989. The formal architecture of Lexical-Functional Grammar. *Journal of Information Science and Engineering* 5(4). Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 1–27), 305–322.
- Kaplan, Ronald M. 2005. A method for tokenizing text. In Antti Arppe, Lauri Carlson, Krister Lindén, Jussi Piitulainen, Mickael Suominen and Martti Vainio, Hanna Westerlund & Anssi Yli-Jyrä (eds.), *Inquiries into words, constraints and contexts: Festschrift for Kimmo Koskenniemi on his 60th birthday*. 57–66. CSLI Publications. <https://web.stanford.edu/group/cslipublications/cslipublications/koskenniemi-festschrift/>.
- Kaplan, Ronald M. 2025. Conservation as a Principle of Grammar. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 9–24. Konstanz: PubliKon.
- Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 173–281. The MIT Press. Reprinted in Dalrymple, Kaplan, Maxwell & Zaenen (1995: 29–130).
- Kaplan, Ronald M. & Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics* 20. 331–478.
- Kaplan, Ronald M. & Jürgen Wedekind. 1993. Restriction and correspondence-based translation. In *Proceedings of the sixth conference of the European chapter of the Association for Computational Linguistics (EACL)*, 193–202.
- Kaplan, Ronald M. & Jürgen Wedekind. 2023. Formal and computational properties of LFG. In Mary Dalrymple (ed.), *Handbook of Lexical Functional Grammar*, 1035–1082. Language Science Press. DOI: 10.5281/zenodo.10185982. <https://doi.org/10.5281/zenodo.10185982>.
- Karttunen, Lauri, Ronald M. Kaplan & Annie Zaenen. 1992. Two level morphology with composition. In *Proceedings of COLING 92*, vol. I, 141–148.
- Kay, Paul & Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The what’s X doing Y? construction. *Language* 75(1). 1–33.
- Kennedy, Chris. 1997. Comparison and polar opposition. In *Semantics and linguistic theory*, vol. 7, 240–257.
- Khoo, Justin. 2017. Code words in political discourse. *Philosophical Topics* 45(2). 33–64.

Bibliography

- King, Tracy Holloway. 2016. Theoretical linguistics and grammar engineering as mutually constraining disciplines. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King & Stefan Müller (eds.), *Proceedings of the joint 2016 conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, 339–359. CSLI Publications. DOI: 10.21248/hpsg.2016.18.
- King, Tracy Holloway. 2023. White roses, red backgrounds: Bringing structured representations to search. In Roussanka Loukanova, Peter LeFanu Lumsdaine & Reinhard Muskens (eds.), *Logic and algorithms in computational linguistics 2021 (LACompLing2021)*, 191–215. Springer. DOI: 10.1007/978-3-031-21780-7_9.
- King, Tracy Holloway. 2025. Multi-modal Search: To Embed or not to Embed. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 333–354. Konstanz: PubliKon.
- Kiparsky, Paul. 1973. “Elsewhere” in *phonology*. Indiana University Linguistics Club.
- Kiparsky, Paul. 1997. Remarks on denominal verbs. In Alex Alsina, Joan Bresnan & Peter Sells (eds.), *Argument structure*. CSLI.
- Kipper, Karin, Hao Trang Dang & Martha Palmer. 2000. Class-based construction of a verb lexicon. In *Proceedings of the seventeenth national conference on artificial intelligence and twelfth conference on innovative applications of artificial intelligence*, 691–696.
- Kipper, Karin, Anna Korhonen, Neville Ryant & Martha Palmer. 2006. Extending VerbNet with novel verb classes. In *LREC*, 1027–1032.
- Kipper, Karin, Anna Korhonen, Neville Ryant & Martha Palmer. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation* 42(1). 21–40.
- Kitsak, Maksim, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley & Hernán A. Makse. 2010. Identification of influential spreaders in complex networks. *Nature Physics* 6. 888–893. <https://doi.org/10.1038/nphys1746>.
- Klimt, Bryan & Yiming Yang. 2004. The Enron corpus: A new dataset for email classification research. In *Machine learning: ECML 2004*, vol. 3201/2004, 217–226. Springer.
- Klosik, David F., Stefan Bornholdt & Marc-Thorsten Hütt. 2014. Motif-based success scores in coauthorship networks are highly sensitive to author name disambiguation. *Physical Review E* 90. 032811.
- Knott, Alistair. 2007. Coherence in natural language: Data structures and applications. *Computational Linguistics* 33. 591–595.

- Kocoń, Jan, Igor Cichecki, Oliwier Kaszyca, Mateusz Kochanek, Dominika Szydło, Joanna Baran, Julita Bielaniewicz, Marcin Gruza, Arkadiusz Janz, Kamil Kanclerz, Anna Kocoń, Bartłomiej Koptyra, Wiktoria Mieszczenko-Kowszewicz, Piotr Milkowski, Marcin Oleksy, Maciej Piasecki, Łukasz Radliński, Konrad Wojtasik, Stanisław Woźniak & Przemysław Kazienko. 2023. ChatGPT: Jack of all trades, master of none. *Information Fusion* 99. DOI: <https://doi.org/10.1016/j.inffus.2023.101861>. <https://www.sciencedirect.com/science/article/pii/S156625352300177X>.
- Kogkalidis, Konstantinos, Stergios Chatzikyriakidis, Eirini Giannikouri, Vasiliki Katsouli, Christina Klironomou, Christina Koula, Dimitris Papadakis, Thelka Pasparaki, Erofilia Psaltaki, Efthymia Sakellariou & Charikleia Soupiona. 2024. OYXOY: A modern NLP test suite for Modern Greek. In Yvette Graham & Matthew Purver (eds.), *Findings of the Association for Computational Linguistics: EACL 2024*, 311–322. Association for Computational Linguistics. <https://aclanthology.org/2024.findings-eacl.21>.
- Kojima, Takeshi, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo & Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. In *Advances in neural information processing systems*, vol. 35, 22199–22213.
- Kokkonidis, Miltiadis. 2006. A glue/ λ -DRT treatment of resumptive pronouns. In *Proceedings of the eleventh ESSLI student session*, 51–63.
- Kokkonidis, Miltiadis. 2008. First-order glue. *Journal of Logic, Language and Information* 17. 43–68.
- Kong, Weize, Jeffrey M. Dudek, Cheng Li, Mingyang Zhang & Mike Bendersky. 2023. SparseEmbed: Learning sparse lexical representations with contextual embeddings for retrieval. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval (SIGIR '23)*, 2399–2403.
- Koskenniemi, Kimmo. 1983. *Two-level morphology: A general computational model for word-form recognition and production*. Tech. rep. 11. Department of General Linguistics, University of Helsinki.
- Kotowski, Sven. 2021. The semantics of English *out*-prefixation: A corpus-based investigation. *English Language and Linguistics* 25(1). 61–89.
- Kotowski, Sven. 2023. Modeling locative prefix semantics. A formal account of the English verbal prefix *out*. *Morphology* 33(2). 115–152.
- Kripke, Saul A. 1959. A completeness theorem in modal logic. *Journal of Symbolic Logic* 24(1). 1–14. DOI: [10.2307/2964568](https://doi.org/10.2307/2964568).
- Kripke, Saul A. 1980. *Naming and necessity*. Harvard University Press. <https://books.google.com/books?id=9vvAlOBfq0kC>.

Bibliography

- Kuhn, Jonas. 1998. *Resource sensitivity in the syntax-semantics interface and the German split NP construction*. Universität Stuttgart, Fakultät Philosophie.
- Kullback, S. & R. A. Leibler. 1951. On information and sufficiency. *The Annals of Mathematical Statistics* 22(1). 79–86.
- Kumar, Jayant, Ashok Gupta, Zhaoyu Lu, Andrei Stefan & Tracy Holloway King. 2023. Multi-lingual semantic search for domain-specific applications: Adobe Photoshop and Illustrator help search. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval (SIGIR '23)*, 3225–3229. Association for Computing Machinery. DOI: 10.1145/3539618.3591826.
- Kuppevelt, Jan Van. 1995. Main structure and side structure in discourse. *Linguistics* 33. 809–833.
- Kusupati, Aditya, Gantavya Bhatt, Aniket Rege, Matthew Wallingford, Aditya Sinha, Vivek Ramanujan, William Howard-Snyder, Kaifeng Chen, Sham Kakade, Prateek Jain & Ali Farhadi. 2022. Matryoshka representation learning. In *Advances in neural information processing systems*, 30233–30249. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2022/file/c32319f4868da7613d78af9993100e42-Paper-Conference.pdf.
- Labov, William. 1972. The transformation of experience in narrative syntax. In William Labov (ed.), *Language in the inner city*, 352–389. University of Pennsylvania.
- Labov, William. 2012. *Dialect diversity in America: The politics of language change*. University of Virginia Press.
- Labov, William & Joshua Waletzky. 1965. Narrative analysis: Oral versions of personal experience. In *Essays on the verbal and the visual arts*, 3–38. University of Washington Press.
- Lacasa, Lucas, Bartolo Luque, Fernando Ballesteros, Jordi Luque & Juan Carlos Nuño. 2008. From time series to complex networks: The visibility graph. In *Proceedings of the national academy of sciences of the United States of America*, vol. 105, 4972–4975. <https://doi.org/10.1073/pnas.0709247105>.
- Lahiri, Aditi (ed.). 2000. *Analogy, leveling, markedness: Principles of change in phonology and morphology* (Trends in Linguistics 127). Mouton de Gruyter.
- Lai, Alice & Joel Tetreault. 2018. Discourse coherence in the wild: A dataset, evaluation and methods. In *Proceedings of the 19th annual SIGdial meeting on discourse and dialogue*, 214–223. <https://doi.org/10.18653/v1/W18-5023>.
- Lai, Viet, Nghia Ngo, Amir Pouran Ben Veyseh, Hieu Man, Franck Dernoncourt, Trung Bui & Thien Nguyen. 2023. ChatGPT beyond English: Towards a comprehensive evaluation of large language models in multilingual learning. In Houda Bouamor, Juan Pino & Kalika Bali (eds.), *Findings of the Association*

- for *Computational Linguistics: EMNLP 2023*, 13171–13189. Association for Computational Linguistics. DOI: 10.18653/v1/2023.findings-emnlp.878. <https://aclanthology.org/2023.findings-emnlp.878>.
- Lakoff, George. 1970. Global rules. *Language* 46. 627–639.
- Lamarche, Francois. 1994. *Proof nets for intuitionistic linear logic 1: Essential nets*. Technical Report, Imperial College London.
- Lapointe, Steven. 1980. *A theory of grammatical agreement*. University of Massachusetts, Amherst. (Doctoral dissertation).
- LaPorte, Joseph. 2022. Rigid designators. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford encyclopedia of philosophy*, Winter 2022. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2022/entries/rigid-designators/>.
- Lappin, Shalom & Chris Fox (eds.). 2015. *The handbook of contemporary semantic theory*. 2nd edn. Wiley-Blackwell.
- Larsson, Staffan. 2020. Discrete and probabilistic classifier-based semantics. In *Proceedings of the probability and meaning conference (PaM 2020)*, 62–68.
- Lascarides, Alex, Ann A. Copestake & Ted Briscoe. 1996. Ambiguity and coherence. *Journal of Semantics* 13. 41–65. <https://doi.org/10.1093/jos/13.1.41>.
- Lev, Iddo. 2007. *Packed computation of exact meaning representations*. Stanford University. (Doctoral dissertation).
- Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. MIT Press.
- Levin, Beth. 1993. *English verb classes and alternations*. The University of Chicago Press.
- Levin, Beth. 2006. English object alternations: A unified account. Unpublished manuscript, Stanford University.
- Lewis, David K. 1986. *On the plurality of worlds*. Wiley-Blackwell.
- Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel & Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in neural information processing systems*, vol. 33, 9459–9474. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf.
- Li, Bai, Zining Zhu, Guillaume Thomas, Frank Rudzicz & Yang Xu. 2022. Neural reality of argument structure constructions. In *Proceedings of the 60th annual meeting of the Association for Computational Linguistics (ACL) (volume 1: long papers)*, 7410–7423. Association for Computational Linguistics.

Bibliography

- Liang, Percy, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, Benjamin Newman, Binhang Yuan, Bobby Yan, Ce Zhang, Christian Cosgrove, Christopher D. Manning, Christopher Ré, Diana Acosta-Navas, Drew A. Hudson, Eric Zelikman, Esin Durmus, Faisal Ladhak, Frieda Rong, Hongyu Ren, Huaxiu Yao, Jue Wang, Keshav Santhanam, Laurel Orr, Lucia Zheng, Mert Yuksekgonul, Mirac Suzgun, Nathan Kim, Neel Guha, Niladri Chatterji, Omar Khattab, Peter Henderson, Qian Huang, Ryan Chi, Sang Michael Xie, Shibani Santurkar, Surya Ganguli, Tatsunori Hashimoto, Thomas Icard, Tianyi Zhang, Vishrav Chaudhary, William Wang, Xuechen Li, Yifan Mai, Yuhui Zhang & Yuta Koreeda. 2023. Holistic evaluation of language models. *Transactions on Machine Learning Research*. <https://openreview.net/forum?id=iO4LZibEqW>.
- Lin, Tianyang, Yuxin Wang, Xiangyang Liu & Xipeng Qiu. 2021. *A survey of transformers*. <https://arxiv.org/abs/2106.04554>.
- Litman, Diane & Julia Hirschberg. 1990. Disambiguating cue phrases in text and speech. In *COLING 1990 volume 2: Papers presented to the 13th international conference on computational linguistics*. <https://aclanthology.org/C90-2044>.
- Liu, Huidong, Shaoyuan Xu, Jinmiao Fu, Yang Liu, Ning Xie, Chien-Chih Wang, Bryan Wang & Yi Sun. 2021. *CMA-CLIP: Cross-modality attention CLIP for image-text classification*. <https://arxiv.org/abs/2112.03562>.
- Liu, Jiangming, Shay B. Cohen & Mirella Lapata. 2019. Discourse representation structure parsing with recurrent neural networks and the transformer model. In *Proceedings of the IWCS shared task on semantic parsing*, 429–439. <https://doi.org/10.18653/v1/W19-1203>.
- Loper, Edward, Szu-Ting Yi & Martha Palmer. 2007. Combining lexical resources: Mapping between PropBank and VerbNet. In *Proceedings of the 7th international workshop on computational linguistics*.
- Lowe, Daniel M., Peter T. Corbett, Peter Murray-Rust & Robert C. Glen. 2011. Chemical name to structure: OPSIN, an open source solution. *Journal of Chemical Information and Modeling* 51(3). 739–53.
- Lyons, John. 1977. Deixis, space and time. In *Semantics*, 636–724. Cambridge University Press.
- Ma, Yuhao, Jian Zhu & Jie Liu. 2022. Enhanced semantic representation learning for implicit discourse relation classification. *Applied Intelligence* 52. 7700–7712.
- MacCartney, Bill. 2009. *Natural language inference*. Stanford University. (Doctoral dissertation).
- MacCartney, Bill & Christopher D. Manning. 2007. Natural logic for textual inference. In *Proceedings of the workshop on textual entailment and paraphrasing*, 193–200. Association for Computational Linguistics.

- Mädche, Alexander, Steffen Staab & Rudi Studer. 1999. Ontology-oriented information extraction and integration. In *Workshop on language technologies in information and knowledge management. 7th conference on computational linguistics of the German society for language technologies*, 7–8.
- Malone, Thomas W. 2018. *Superminds: The surprising power of people and computers thinking together*. Little, Brown & Company.
- Mann, William C. & Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text – Interdisciplinary Journal for the Study of Discourse* 8(3). 243–281. DOI: doi:10.1515/text.1.1988.8.3.243. <https://doi.org/10.1515/text.1.1988.8.3.243>.
- Manning, Christopher D., Prabhakar Raghavan & Hinrich Schütze. 2008. *Introduction to information retrieval*. Cambridge University Press.
- Marantz, Alec. 1984. *On the nature of grammatical relations*. MIT Press.
- Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4(2). 201–225.
- March, J. G. & Herbert A. Simon. 1958. *Organizations*. Wiley.
- Marcolli, Matilde. 2023a. *A mathematical model of syntactic merge*. Lecture series at MIT. https://www.youtube.com/watch?v=7Gtpc_E1Tfw.
- Marcolli, Matilde. 2023b. *A mathematical model of syntactic merge*. Lecture at University of Utrecht, June 25, 2023. <https://www.youtube.com/watch?v=JPdjRaU4JNU&t=2027s>.
- Marcolli, Matilde. 2024. *Mathematical models of generative linguistics (Ma191c spring 2024)*. CalTech course materials. <https://www.its.caltech.edu/~matilde/LinguisticsMa191Spring2024.html>.
- Marcolli, Matilde, Robert Berwick & Noam Chomsky. 2023. *Old and new Minimalism: A Hopf algebra comparison*. <https://arxiv.org/abs/2306.10270>.
- Marcolli, Matilde, Noam Chomsky & Robert Berwick. 2023a. *Mathematical structure of syntactic merge*. <https://arxiv.org/abs/2305.18278>.
- Marcolli, Matilde, Noam Chomsky & Robert Berwick. 2023b. *Syntax-semantics interface: An algebraic model*. <https://lingbuzz.net/lingbuzz/007696>.
- Marcolli, Matilde, Noam Chomsky & Robert Berwick. 2024. *Mathematical structure of syntactic merge*. Unpublished manuscript, CalTech. <https://www.its.caltech.edu/~matilde/MergeMCB-MITPress-LI.pdf>.
- Marcu, Daniel. 2000. *The theory and practice of discourse parsing and summarization*. MIT Press.
- Marr, Carsten, Fabian J. Theis, Larry S. Liebovitch & Marc-Thorsten Hütt. 2010. Patterns of subnet usage reveal distinct scales of regulation in the transcrip-

Bibliography

- tional regulatory network of escherichia coli. *PLoS Computational Biology* 6. e1000836.
- Matsson, Arild. 2018. *Implementing perceptual semantics in type theory with records (TTR)*. Masters in Language Technology. University of Gothenburg. (MA thesis).
- Matsusaka, Youichi. 2013. *A theory of direct discourse: Its semantics and pragmatics*. UCLA. (Doctoral dissertation).
- Maxwell, John T., III & Ronald M. Kaplan. 1989. An overview of disjunctive constraint satisfaction. In Masaru Tomita (ed.), *Proceedings of the first international workshop on parsing technologies*, 18–27. Carnegie Mellon University. <https://aclanthology.org/W89-0203>.
- McIntyre, Andrew. 2003. Preverbs, argument linking and verb semantics: Germanic prefixes and particles. In Geert Booij & Jaap Van Marle (eds.), *Yearbook of morphology 2003*, 119–144. Springer.
- McNamara, Paul & Frederik Van De Putte. 2022. Deontic logic. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford encyclopedia of philosophy*, Fall 2022. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2022/entries/logic-deontic/>.
- MedDRA. 2024. *Medication dictionary for regulatory activities*. <https://www.meddra.org/>.
- MEDLINE. 2024. *Medline Citation Counts by Year of Publication*. https://www.nlm.nih.gov/bsd/medline_cit_counts_yr_pub.html.
- Mehler, Alexander. 2008. Large text networks as an object of corpus linguistic studies. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook of the science of language and society*, 328–382. De Gruyter.
- Mehler, Alexander, Andy Lücking, Sven Banisch, Philippe Blanchard & Barbara Frank-Job. 2016. *Towards a theoretical framework for analyzing complex linguistic networks*. Springer.
- Melchin, Paul B., Ash Asudeh & Dan Siddiqi. 2020. Ojibwe agreement in Lexical-Realizational Functional Grammar. In Miriam Butt & Ida Toivonen (eds.), *Proceedings of the LFG20 conference*, 268–288. CSLI Publications. <https://typo.uni-konstanz.de/lfg-proceedings/LFGprocCSLI/LFG2020/index.shtml>.
- Merchant, Jason. 2015. How much context is enough? Two cases of span-conditioned stem allomorphy. *Linguistic Inquiry* 46(2). 273–303.
- MeSH. 2024. *MeSH (medical subject headings)*. Online thesaurus. <https://www.ncbi.nlm.nih.gov/mesh>.
- Michaelis, Laura A. 2004. Type shifting in Construction Grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15. 1–67.

- Milo, Ron & Uri Alon. 2002. Network motifs: Simple building blocks of complex networks. *Science* 298. 824–827. <https://doi.org/10.1126/science.298.5594.824>.
- Miltsakaki, Eleni, Aravind Joshi, Rashmi Prasad & Bonnie Webber. 2004. Annotating discourse connectives and their arguments. In *Proceedings of the workshop frontiers in corpus annotation at HLT-NAACL 2004*, 9–16. Association for Computational Linguistics. <https://aclanthology.org/W04-2703>.
- Milward, David. 2025. Semantic Representation in Biomedical Text Mining. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 301–321. Konstanz: PubliKon.
- Milward, David, Marcus Bjärelund, William Hayes, Michelle Maxwell, Lisa Oberg, Nick Tilford, James Thomas, Roger Hale, Sylvia Knight & Barnes Julie. 2005. Ontology-based interactive information extraction from scientific abstracts. *Computational Functional Genomics* 6. 67–71.
- MongoDB. 2024. Online documentation. <https://www.mongodb.com/>.
- Montague, Richard. 1974. *Formal philosophy: Selected papers of Richard Montague*. Edited and with an introduction by Richmond H. Thomason. Yale University Press.
- Moot, Richard. 2002. *Proof-nets for linguistic analysis*. University of Utecht. (Doctoral dissertation).
- Moot, Richard & Christian Retoré. 2012. *The logic of categorial grammars: A deductive account of natural language syntax and semantics* (Lecture Notes in Computer Science volume 6850). Springer.
- Moss, Lawrence S. 2010. Natural logic and semantics. In Maria Aloni, Harald Bastiaanse, Tikitou de Jager & Katrin Schulz (eds.), *Logic, language, and meaning: Selected papers from the seventeenth Amsterdam colloquium*, 84–93. Springer.
- Moss, Lawrence S. 2015. Natural logic. In Shalom Lappin & Chris Fox (eds.), *The handbook of contemporary semantic theory*, 2nd edn. Wiley-Blackwell.
- Mukerjee, Amitabha, Ankit Soni & Achla M. Raina. 2006. Detecting complex predicates in Hindi using POS projection across parallel corpora. In Begoña Villada Moirón, Aline Villavicencio, Diana McCarthy, Stefan Evert & Suzanne Stevenson (eds.), *Proceedings of the workshop on multiword expressions: Identifying and exploiting underlying properties*, 28–35. Association for Computational Linguistics. <https://aclanthology.org/W06-1205>.
- Muskens, Reinhard. 2010. An analytic tableau system for natural logic. In Maria Aloni, Harald Bastiaanse, Tikitou de Jager & Katrin Schulz (eds.), *Logic, language, and meaning: Selected papers from the seventeenth Amsterdam colloquium*, 104–113. Springer.

Bibliography

- Mycock, Louise. 2006. *The typology of constituent questions: A Lexical-Functional Grammar analysis of 'WH'-questions*. University of Manchester. (Doctoral dissertation).
- n2c2. 2024. *National NLP Clinical Challenges (n2c2)*. On-line conference materials. <https://n2c2.dbmi.hms.harvard.edu/>.
- n2c2-T1. 2024. *National NLP Clinical Challenges (n2c2): Track 1 Contextualized Medication Event Extraction CMED*. On-line conference materials. <https://n2c2.dbmi.hms.harvard.edu/2022-track-1>.
- Nadathur, Prerna. 2019. *Causality, aspect, and modality in actuality inferences*. Stanford University. (Doctoral dissertation).
- Nagano, Akiko. 2011. The right-headedness of morphology and the status and development of category-determining prefixes in English. *English Language and Linguistics* 15(01). 61–83.
- Narasimhan, Karthik & Regina Barzilay. 2015. Machine comprehension with discourse relations. In *Proceedings of the 53rd annual meeting of the Association for Computational Linguistics (ACL) and the 7th international joint conference on natural language processing (volume 1: long papers)*, 1253–1262. <https://doi.org/10.3115/v1/P15-1121>.
- National Center for Biotechnology Information. 2024. *Journal Publishing Tag Set*. <https://jats.nlm.nih.gov/publishing/>.
- Navigli, Roberto, Michele Bevilacqua, Simone Conia, Dario Montagnini & Francesco Cecconi. 2021. Ten years of BabelNet: A survey. In *Proceedings of IJCAI 2021*, 4559–4567.
- Needham, Mark & Amy E. Hodler. 2024. *Graph algorithms in Neo4j: Shortest path*. <https://neo4j.com/blog/graph-algorithms-neo4j-shortest-path/>.
- Neo4j. 2024. <https://neo4j.com/>.
- Newman, M. E. J. 2003. Mixing patterns in networks. *Physical Review E* 67. 026126.
- Newman, M. E. J., Steven H. Strogatz & Duncan J. Watts. 2001. Random graphs with arbitrary degree distributions and their applications. *Physical Review E* 64. 026118. <https://doi.org/10.1103/PhysRevE.64.026118>.
- Newman, Mark. 2010. *Networks: An introduction*. Oxford University Press.
- Newman, Mark, Albert-László Barabási & Duncan J. Watts. 2006. *The structure and dynamics of networks*. Princeton University Press.
- NIH. 2024. *International Protein Nomenclature Guidelines*. On-line documentation. https://www.ncbi.nlm.nih.gov/genbank/internatprot_nomenguide/.
- Niles, Ian & Adam Pease. 2001. Toward a standard upper ontology. In Chris Welty & Barry Smith (eds.), *Proceedings of the 2nd international conference on formal ontology in information systems (FOIS-2001)*, 2–9.

- Nishino, Tetsuro. 1991. Mathematical analysis of Lexical-Functional Grammars—complexity, parsability, and learnability. *Language Research* 27(1). 867–915.
- Noble, Bill, Ellen Breitholtz & Robin Cooper. 2020. Personae under uncertainty: The case of topoi. In Christine Howes, Stergios Chatzikyriakidis, Adam Ek & Vidya Somashekarappa (eds.), *Proceedings of the probability and meaning conference (PaM 2020)*, 8–16. Association for Computational Linguistics. <https://aclanthology.org/2020.pam-1.2>.
- Nordhoff, Sebastian. 2018. *BibTeX generator*. <https://www.langsci-press-gug.org/doc2tex/doc2bib>.
- Open AI. 2023. *GPT-4 technical report Open AI*. <https://arxiv.org/abs/2303.08774>.
- Open AI. 2024a. OpenAI models. <https://platform.openai.com/docs/models/gpt-3-5-turbo>.
- OpenAI. 2024b. *Prompt engineering*. Tech. rep. OpenAI. <https://platform.openai.com/docs/guides/prompt-engineering/prompt-engineering>.
- Pacuit, Eric. 2017. *Neighborhood semantics for modal logic*. Springer.
- Pal, Santanu, Tanmoy Chakraborty & Sivaji Bandyopadhyay. 2011. Handling multiword expressions in phrase-based statistical machine translation. In *Proceedings of machine translation summit XIII: Papers*. <https://aclanthology.org/2011.mtsummit-papers.23>.
- Pandas. 2024. On-line documentation. <https://pandas.pydata.org/>.
- Parsons, Terence. 1990. *Events in the semantics of English*. MIT Press.
- Partee, Barbara H. 2009. Do we need two basic types? *Snippets: Special issue for Manfred Krifka* 20. 37–41.
- Partee, Barbara H. & Vladimir Borschev. 2004. Genitives, types, and sorts. In Jiyung Kim, Yury A. Lander & Barbara H. Partee (eds.), *Possessives and beyond: Semantics and syntax*, 29–43. UMass GSLA. <http://people.umass.edu/partee/Research.htm>.
- Paulson, Lawrence C. 1986. Natural deduction as higher-order resolution. *The Journal of Logic Programming* 3(3). 237–258. DOI: [https://doi.org/10.1016/0743-1066\(86\)90015-4](https://doi.org/10.1016/0743-1066(86)90015-4). <https://www.sciencedirect.com/science/article/pii/0743106686900154>.
- Pease, Adam. 2009. *SUO-KIF reference manual*. Retrieved 20 June 2020. <https://github.com/ontologyportal/sigmakee/blob%20/master/suo-kif.pdf>.
- Pease, Adam. 2011. *Ontology: A practical guide*. Articulate Software Press.
- Pease, Adam. 2023. *Converting the suggested upper merged ontology to typed first-order form*. <https://arxiv.org/abs/2303.04148v1>.
- Pease, Adam. 2025. Modal and Higher Order Logical Reasoning with SUMO. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zae-

Bibliography

- nen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 165–186. Konstanz: PubliKon.
- Pease, Adam & Stephan Schulz. 2022. Contradiction detection and repair in a large theory. In *The international FLAIRS conference proceedings*.
- Perrier, Guy. 1999. Labelled proof-nets for the syntax and semantics of natural languages. *Logic Journal of the IGPL* 7. 629–654.
- Peters, Stanley & R. W. Ritchie. 1973. On the generative power of transformational grammars. *Information Sciences* 6. 49–83. DOI: 10.1016/0020-0255(73)90027-3. [https://doi.org/10.1016/0020-0255\(73\)90027-3](https://doi.org/10.1016/0020-0255(73)90027-3).
- Polanyi, Livia. 1985. A theory of discourse structure and discourse coherence. In Paul D. Kroeber, William H. Eilfort & Karen L. Peterson (eds.), *Papers from the general session at the 21st meeting of the Chicago Linguistics Society*. Chicago Linguistic Society.
- Polanyi, Livia. 1988. A formal model of discourse structure. *Journal of Pragmatics* 12. 601–638.
- Polanyi, Livia. 1989. *Telling the American story: A structural and cultural analysis of conversational storytelling*. MIT Press.
- Polanyi, Livia. 1996. *The linguistic structure of discourse*. Tech. rep. CSLI Publications.
- Polanyi, Livia, Chirs Culy, Martin van den Berg, Lorenzo Thione & David Ahn. 2004. A rule based approach to discourse parsing. In *Proceedings of the 5th SIGdial workshop on discourse and dialogue at HLT-NAACL*, 108–117.
- Polanyi, Livia & Remko Scha. 1983. On the recursive structure of discourse. In Konrad Ehlich & Henk C. van Riemsdijk (eds.), *Connectedness in sentence, discourse and text*, 141–178. Tilburg University.
- Polanyi, Livia & Remko Scha. 1984. A syntactic approach to discourse semantics. In *10th international conference on computational linguistics and 22nd annual meeting of the Association for Computational Linguistics (ACL)*, 413–419. Association for Computational Linguistics. DOI: 10.3115/980491.980576. <https://aclanthology.org/P84-1085>.
- Polanyi, Livia, Martin van den Berg & David Ahn. 2003. Discourse structure and sentential information structure: An initial proposal. *The Journal of Logic, Language and Information* 12(3). 337–350.
- Polanyi, Livia & Martin H. van den Berg. 2025. The Linguistic Discourse Model at 40. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 111–136. Konstanz: PubliKon.

- Polanyi, Livia, Martin Henk van den Berg, Gian Lorenzo Thione, Christopher Culy & David Ahn. 2004a. Hybrid text summarization: Combining external relevance measures with structural analysis. In *ACL2004 workshop text summarization branches out*, 51–55.
- Polanyi, Livia, Martin Henk van den Berg, Gian Lorenzo Thione, Christopher Culy & David Ahn. 2004b. Sentential structure and discourse parsing. In *ACL2004 workshop on discourse annotation*, 51–55. <https://aclanthology.org/W04-1009/>.
- Pollard, Carl & Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. The University of Chicago Press & CSLI Publications.
- Pradeep, Ronak, Kai Hui, Jai Gupta, Honglei Zhuang, Adam Lelkes, Jimmy Lin, Don Metzler & Vinh Tran. 2023. Understanding generative retrieval at scale. In *EMNLP 2023*, 1305–1321. <https://aclanthology.org/2023.emnlp-main.83/>.
- Prasad, Rashmi, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi & Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the sixth international conference on language resources and evaluation (LREC'08)*.
- Prévot, Laurent & Laure Vieu. 2008. The moving right frontier. In Anton Benz & Peter Kühnlein (eds.), *Constraints in discourse, pragmatics and beyond*, 53–66. John Benjamins Publishing Company.
- Prior, Arthur Norman. 1957. *Time and modality: Being the John Locke lectures for 1955–6 delivered in the University of Oxford*. Oxford University Press.
- Przepiórkowski, Adam & Agnieszka Patejuk. 2023. Filling gaps with glue. In Miriam Butt, Jamie Y. Findlay & Ida Toivonen (eds.), *Proceedings of the LFG23 conference*, 223–240. PubliKon. <https://lfg-proceedings.org/lfg/index.php/main/issue/view/2>.
- Pustejovsky, James. 1991. The syntax of event structure. *Cognition* 41. 47–81.
- Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger & Ilya Sutskever. 2021. Learning transferable visual models from natural language supervision. In Marina Meila & Tong Zhang (eds.), *Proceedings of the 38th international conference on machine learning*, vol. 139, 8748–8763. PMLR. <https://proceedings.mlr.press/v139/radford21a.html>.
- Ramchand, Gillian Catriona. 2008. *Verb meaning and the lexicon: A first phase syntax*. Cambridge University Press.
- Raths, Thomas & Jens Otten. 2012. The QMLTP problem library for first-order modal logics. In Bernhard Gramlich, Dale Miller & Uli Sattler (eds.), *International joint conference on automated reasoning (IJCAR)*, 454–461. Springer.

Bibliography

- RDF. 2024. *RDF (resource description framework)*. Online documentation. https://en.wikipedia.org/wiki/Resource_Description_Framework.
- Reimers, Nils & Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, 3982–3992. <https://doi.org/10.18653/v1/D19-1410>.
- Riemschneider, Frederick & Anette Frank. 2023. Exploring large language models for classical philology. In Anna Rogers, Jordan Boyd-Graber & Naoaki Okazaki (eds.), *Proceedings of the 61st annual meeting of the Association for Computational Linguistics (ACL) (volume 1: long papers)*, 15181–15199. Association for Computational Linguistics. DOI: 10.18653/v1/2023.acl-long.846. <https://aclanthology.org/2023.acl-long.846>.
- Roach, Kelly. 1983. LFG languages over a one-letter alphabet. Unpublished manuscript, Xerox Palo Alto Research Center.
- Robertson, Stephen & Hugo Zaragoza. 2009. *The probabilistic relevance framework: BM25 and beyond*. NOW Publishers, Inc.
- Rohde, Hannah, Alexander Johnson, Nathan Schneider & Bonnie Webber. 2018. Discourse coherence: Concurrent explicit and implicit relations. In *Proceedings of the 56th annual meeting of the Association for Computational Linguistics (ACL) (volume 1: long papers)*, 2257–2267. <https://doi.org/10.18653/v1/P18-1210>.
- Rothwell, William. 2000. Aspects of lexical and morphosyntactical mixing in the languages of medieval England. In D. A. Trotter (ed.), *Multilingualism in later medieval Britain*. D. S. Brewer.
- Ruppenhofer, Joseph, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, Collin F. Baker & Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice*. Institut für Deutsche Sprache, Bibliothek.
- RxNorm. 2024. Online documentation. <https://www.nlm.nih.gov/research/umls/rxnorm/index.html>.
- Sanches Duran, Magali, Carlos Ramisch, Sandra Maria Aluísio & Aline Villavicencio. 2011. Identifying and analyzing Brazilian Portuguese complex predicates. In Valia Kordoni, Carlos Ramisch & Aline Villavicencio (eds.), *Proceedings of the workshop on multiword expressions: From parsing and generation to the real world*, 74–82. Association for Computational Linguistics. <https://aclanthology.org/W11-0812>.
- Sanderson, Paul & Mark Clough. 2013. Evaluating the performance of information retrieval systems using test collections. *IR Information Research* 18.

- Saraswat, Vijay. 1999. LFG as concurrent constraint programming. In Mary Dalrymple (ed.), *Semantics and syntax in Lexical Functional Grammar: The resource logic approach*, 281–318. The MIT Press.
- SAS Institute Inc. 2024. On-line documentation. <https://www.sas.com/>.
- Sassen, Claudia & Peter Kühnlein. 2005. The right frontier constraint as conditional. In Alexander F. Gelbukh (ed.), *Computational linguistics and intelligent text processing international conference (CICLing) proceedings*, vol. 3406 (Lecture Notes in Computer Science), 222–225. Springer. DOI: 10.1007/978-3-540-30586-6_23. https://doi.org/10.1007/978-3-540-30586-6_23.
- Savitch, Walter J., Emmon W. Bach, William Marsh & Gila Safran-Naveh (eds.). 1987. *The formal complexity of natural language* (Studies in Linguistics and Philosophy). Springer. DOI: 10.1007/978-94-009-3401-6. <https://doi.org/10.1007/978-94-009-3401-6>.
- Schadow, Gunther, Clement J. McDonald, Jeffrey G. Suico, Ulrich Föhring & Thomas Tolxdorff. 1999. Units of measure in clinical information systems. *Journal of American Medical Informatics Association* 6(2). PMID: 10094068; PMCID: PMC61354, 151–62. DOI: 10.1136/jamia.1999.0060151.
- Schegloff, Emanuel A. & Harvey Sacks. 1973. Opening up closings. *Semiotica* 8(4). 289–327. DOI: doi:10.1515/semi.1973.8.4.289. <https://doi.org/10.1515/semi.1973.8.4.289>.
- Schiffrin, Deborah. 1987. *Discourse markers* (Studies in Interactional Sociolinguistics). Cambridge University Press.
- Schlangen, David, Sina Zarriess & Casey Kennington. 2015. *Resolving references to objects in photographs using the words-as-classifiers model*. <https://arxiv.org/abs/1510.02125>.
- Schuler, Karin K. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania.
- Schulhoff, Sander, Michael Ilie, Nishant Balepur, Konstantine Kahadze, Amanda Liu, Chenglei Si, Yinheng Li, Aayush Gupta, HyoJung Han, Sevien Schulhoff, Pranav Sandeep Dulepet, Saurav Vidyadhara, Dayeon Ki, Sweta Agrawal, Chau Pham, Gerson Kroiz, Feileen Li, Hudson Tao, Ashay Srivastava, Hevander Da Costa, Saloni Gupta, Megan L. Rogers, Inna Goncarencu, Giuseppe Sarli, Igor Galynker, Denis Peskoff, Marine Carpuat, Jules White, Shyamal Anadkat, Alexander Hoyle & Philip Resnik. 2024. *The prompt report: A systematic survey of prompting techniques*. <https://arxiv.org/abs/2406.06608>.
- Schulz, Stephan, Geoff Sutcliffe, Josef Urban & Adam Pease. 2017. Detecting inconsistencies in large first-order knowledge bases. In *Proceedings of CADE 26*, 310–325. Springer.

Bibliography

- Scott, Dana. 1970. Advice on modal logic. In Karel Lambert (ed.), *Philosophical problems in logic: Some recent developments*, 143–173. Springer.
- Seeber, Isabella, Eva Bittner, Robert O. Briggs, Triparna de Vreede, Gert-Jan de Vreede, Aaron Elkins, Ronald Maier, Alexander B. Merz, Sarah Oeste-Reiß, Nils Randrup, Gerhard Schwabe & Matthias Söllner. 2020. Machines as teammates: A research agenda on AI in team collaboration. *Information & Management* 57(2). 103–174.
- Selventa Knowledge. 2024. On-line documentation. https://github.com/cthoyst/selventaknowledge/blob/master/selventa_knowledge/small_corpus.bel.
- Shah, Chirag, Ryen W. White, Paul Thomas, Bhaskar Mitra, Shawon Sarkar & Nicholas Belkin. 2023. Taking search to task. In *Proceedings of CHIIR'23*, 1–13. ACM.
- Shannon, Claude Elwood. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27. 379–423.
- Shen-Orr, Shai S., Ron Milo, Shmoolik Mangan & Uri Alon. 2002. Network motifs in the transcriptional regulation network of escherichia coli. *Nature Genetics* 31. 64–68.
- Shi, Freda, Mirac Suzgun, Markus Freitag, Xuezhi Wang, Suraj Srivats, Soroush Vosoughi, Hyung Won Chung, Yi Tay, Sebastian Ruder & Denny Zhou. 2022. *Language models are multilingual chain-of-thought reasoners*. <https://arxiv.org/abs/2210.03057>.
- Shi, Lei & Rada Mihalcea. 2005. Putting pieces together: Combining FrameNet, VerbNet and WordNet for robust semantic parsing. In *International conference on intelligent text processing and computational linguistics*, 100–111. Springer.
- Siddiqi, Daniel & Heidi Harley (eds.). 2016. *Morphological metatheory*. John Benjamins.
- Simon, Herbert A. 1971. Designing organizations for an information rich world. In Martin Greenberger (ed.), *Computers, communications, and the public interest*, 37–72.
- Singh, Mona. 1991. The perfective paradox: Or how to eat your cake and have it too. In *Proceedings of the seventeenth annual meeting of the Berkeley Linguistics Society (BLS)*, 469–479.
- Sinha, R. Mahesh K. 2009. Mining complex predicates in Hindi using a parallel Hindi-English corpus. In Dimitra Anastasiou, Chikara Hashimoto, Preslav Nakov & Su Nam Kim (eds.), *Proceedings of the workshop on multiword expressions: Identification, interpretation, disambiguation and applications (MWE 2009)*, 40–46. Association for Computational Linguistics. <https://aclanthology.org/W09-2906>.

- Sinnott-Armstrong, Walter. 1985. A solution to Forrester's paradox of gentle murder. *Journal of Philosophy* 82(3). 162–168.
- SKOS. 2024. *SKOS Simple Knowledge Organization System Reference*. On-line documentation. <https://www.w3.org/TR/skos-reference/>.
- SnoMed. 2024. Online documentation. <https://www.snomed.org/about-us>.
- Song, Chaoming, Shlomo Havlin & Hernán A. Makse. 2005. Self-similarity of complex networks. *Nature* 433. 392–395. <https://doi.org/10.1038/nature03248>.
- Song, Chaoming, Shlomo Havlin & Hernán A. Makse. 2006. Origins of fractality in the growth of complex networks. *Nature Physics* 2. 275–281. <https://doi.org/10.1038/nphys266>.
- Sorokina, Daria & Erick Cantú-Paz. 2016. Amazon search: The joy of ranking products. In Raffaele Perego, Fabrizio Sebastiani, Javed A. Aslam, Ian Ruthven & Justin Zobel (eds.), *Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval, SIGIR 2016*, 459–460. ACM.
- Sperber, Daniel & Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Blackwell.
- Stabler, Edward P. 2010. Computational perspectives on Minimalism. In Cedric Boeckx (ed.), *Oxford handbook of linguistic minimalism*, 616–641. Oxford University Press.
- Stubbs, Amber, Michele Filannino, Ergin Soysal, Samuel Henry & Özlem Uzuner. 2019. Cohort selection for clinical trials: n2c2 2018 shared task track 1. *Journal of the American Medical Informatics Association* 26(11). PMID: 31562516; PMCID: PMC6798568, 1163–1171. DOI: 10.1093/jamia/ocz163.
- Stump, Gregory T. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge University Press.
- Sulger, Sebastian & Ashwini Vaidya. 2014. Towards identifying Hindi/Urdu noun templates in support of a large-scale LFG grammar. In Christian Boitet & M.G. Abbas Malik (eds.), *Proceedings of the fifth workshop on south and south-east Asian natural language processing*, 1–10. Association for Computational Linguistics & Dublin City University. DOI: 10.3115/v1/W14-5501. <https://aclanthology.org/W14-5501>.
- Sun, Xiangyan, Haixun Wang, Yanghua Xiao & Zhongyuan Wang. 2016. Syntactic parsing of web queries. In *Proceedings of the 2016 conference on empirical methods in natural language processing (EMNLLP)*, 1787–1796.
- Sutcliffe, Geoff. 2010. The TPTP world — infrastructure for automated reasoning. In Edmund M. Clarke & Andrei Voronkov (eds.), *LPAR*, vol. 6355 (Lecture Notes in Computer Science), 1–12. Springer. DOI: http://dx.doi.org/10.1007/978-3-642-17511-4_1. <http://dx.doi.org/10.1007/978-3-642-17511-4>.

Bibliography

- Sutcliffe, Geoff, Stephan Schulz, Koen Claessen & Peter Baumgartner. 2012. The TPTP typed first-order form with arithmetic. In Nikolaj Bjørner & Andrei Voronkov (eds.), *LPAR*, vol. 7180 (LNCS), 406–419. Springer. http://dx.doi.org/10.1007/978-3-642-28717-6_32.
- Svenonius, Peter. 2016. Spans and words. In Daniel Siddiqi & Heidi Harley (eds.), *Morphological metatheory*, 201–222. John Benjamins.
- Talman, Aarne & Stergios Chatzikyriakidis. 2019. Testing the generalization power of neural network models across NLI benchmarks. In Tal Linzen, Grzegorz Chrupała, Yonatan Belinkov & Dieuwke Hupkes (eds.), *Proceedings of the 2019 ACL workshop BlackboxNLP: Analyzing and interpreting neural networks for NLP*, 85–94. Association for Computational Linguistics. DOI: 10.18653/v1/W19-4810. <https://aclanthology.org/W19-4810>.
- Tambi, Ritiz, Ajinkya Kale & Tracy Holloway King. 2020. Search query language identification using weak labeling. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the twelfth language resources and evaluation conference (LREC)*, 3520–3527. European Language Resources Association. <https://aclanthology.org/2020.lrec-1.432>.
- Tantos, Alexandros, Giorgos Chatziioannidis, Katerina Lykou, Meropi Papatheohari, Antonia Samara & Kostas Vlachos. 2017. Corpus C58 and a data-analytic approach to the interface between intra- and inter-sentential linguistic information. In *Proceedings of the proceedings of 12th international conference on Greek linguistics*, 961–976.
- Tantos, Alexandros & Kosmas Kosmidis. 2023. From discourse relations to network edges: A network theory approach to discourse analysis. *Applied Sciences* 13(12). 6902.
- Tantos, Alexandros & Kosmas Kosmidis. 2025. From Network Theory to Linguistic Insights. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 137–163. Konstanz: PubliKon.
- Thomas, James, David Milward, Christos Ouzounis, Stephen Pulman & Mark Carroll. 2000. Automatic extraction of protein interaction from scientific abstracts. In *Pacific symposium on biocomputing (PSB)*, 541–552.
- Trac, Steven, Geoff Sutcliffe & Adam Pease. 2008. Integration of the TPTPWorld into SigmaKEE. In *Proceedings of IJCAR '08 workshop on practical aspects of automated reasoning (PAAR-2008)*. CEUR Workshop Proceedings.

- Treinisch, Matthew, Ivan Carvalho, Georgios Tsilimigkounakis & Nahum Sá. 2021. *Rustworkx: A high-performance graph library for Python*. <https://arxiv.org/abs/2110.15221>.
- Trivedi, Sapna, Roger Gildersleeve, Sandra Franco, Andrew S. Kanter & Afzal Chaudhry. 2020. Evaluation of a concept mapping task using named entity recognition and normalization in unstructured clinical text. *Journal of Healthcare Informatics Research* 4. 395–410.
- Troelstra, Anne S. 1992. *Lectures on linear logic*. CSLI Publications.
- Tsagkias, Manos, Tracy Holloway King, Surya Kallumadi, Vanessa Murdock & Maarten de Rijke. 2020. Challenges and research opportunities in eCommerce search and recommendations. *ACM SIGIR Forum* 54. 1–23.
- Vaidya, Ashwini, Owen Rambow & Martha Palmer. 2019. Syntactic composition and selectional preferences in Hindi light verb constructions. *Linguistic Issues in Language Technology* 17(1). 1–30.
- van Benthem, Johan. 1986. *Essays in logical semantics*. Springer. 240 pp. https://www.ebook.de/de/product/7614100/johan_van_benthem_johan_van_benthem_essays_in_logical_semantics.html.
- van Eijck, Jan & Hans Kamp. 2011. Discourse representation in context. In Johan van Benthem & Alice ter Meulen (eds.), *Handbook of logic and language*, 181–252. Elsevier. DOI: 10.1016/B978-0-444-53726-3.00003-7. <https://doi.org/10.1016/B978-0-444-53726-3.00003-7>.
- van Genabith, Josef & Richard Crouch. 1997. How to glue a donkey to an f-structure or porting a dynamic meaning representation language into LFG’s linear logic based glue language semantics. In *International workshop for computational semantics*, 52–65.
- van Noord, Rik, Lasha Abzianidze, Antonio Toral & Johan Bos. 2018. Exploring neural methods for parsing discourse representation structures. *Transactions of the Association of Computational Linguistics* 6. 619–633. https://doi.org/10.1162/tacl_a_00241.
- van den Berg, Martin H. 1996. *Some aspects of the internal structure of discourse – the dynamics of nominal anaphora*. Institute for Logic Language & Computation, University of Amsterdam. (Doctoral dissertation).
- van den Berg, Martin H. & Hub Prüst. 1990. Common denominators and default unification. In Ton van der Woude & Wietske Sijtsma (eds.), *Computational linguistics in the Netherlands: Papers from the first CLIN meeting*. University of Utrecht.
- van der Sandt, Rob A. 1992. Presupposition projection as anaphora resolution. *Journal of Semantics* 9(4). 333–377. DOI: <https://doi.org/10.1093/jos/9.4.333>. <https://academic.oup.com/jos/article-pdf/9/4/333/9836990/333.pdf>.

Bibliography

- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser & Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st international conference on neural information processing systems*, 6000–6010. Curran Associates Inc.
- Vidgen, Bertie et al. 2024. *Introducing v0.5 of the AI safety benchmark from ML-Commons*. <https://arxiv.org/abs/2404.12241>.
- Wadekar, Shakti N., Abhishek Chaurasia, Aman Chadha & Eugenio Culurciello. 2024. *The evolution of multimodal model architectures*. <https://arxiv.org/abs/2405.17927>.
- Walker, Marilyn A. 1996. Limited attention and discourse structure. *Computational Linguistics* 22(2). 255–264.
- Wang, Chen, Keping Bi, Yunhua Hu, Hang Li & Guihong Cao. 2012. Extracting search-focused key n-grams for relevance ranking in web search. In *WSDM '12: Proceedings of the fifth ACM international conference on web search and data mining*, 343–352.
- Wang, Zhenyi, Enneng Yang, Li Shen & Heng Huang. 2023. *A comprehensive survey of forgetting in deep learning beyond continual learning*. <https://arxiv.org/abs/2307.09218>.
- Wang, Zhongyuan, Haixun Wang & Zhirui Hu. 2014. Head, modifier, and constraint detection in short texts. In *Proceedings of the international conference on data engineering*, 280–291.
- Wang, Zhongyuan, Kejun Zhao, Haixun Wang, Xiaofeng Meng & Ji-Rong Wen. 2015. Query understanding through knowledge-based conceptualization. In *Proceedings of IJCAI*, 3264–3270.
- Wang, Zihao, Shaofei Cai, Anji Liu, Xiaojuan Ma & Yitao Liang. 2023. *Describe, explain, plan and select: Interactive planning with large language models enables open-world multi-task agents*. <https://arxiv.org/abs/2302.01560>.
- Wasow, Thomas. 1978. On constraining the class of transformational languages. *Synthese* 39(1). 81–104. <http://www.jstor.org/stable/20115307> (17 September, 2024).
- Watts, Duncan J. & Steven H. Strogatz. 1998. Collective dynamics of ‘small-world’ networks. *Nature* 393. 440–442.
- Webber, Bonnie. 1988. *Discourse deixis and discourse processing*. University of Pennsylvania.
- Webber, Bonnie. 2006. Accounting for discourse relations: Constituency and dependency. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*. CSLI Publications.

- Webber, Bonnie & Aravind Joshi. 2012. Discourse structure and computation: Past, present and future. In *Proceedings of the ACL-2012 special workshop on rediscovering 50 years of discoveries*, 42–54.
- Wedekind, Jürgen. 2014. On the universal generation problem for unification grammars. *Computational Linguistics* 40(3). 533–538.
- Wedekind, Jürgen & Ronald M. Kaplan. 1993. Type-driven semantic interpretation of f-structures. In *EACL '93: Proceedings of the sixth conference of the European chapter of the Association for Computational Linguistics*, 404–411. DOI: 10.3115/976744.976791.
- Weininger, David. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* 28(1). 31–36. DOI: 10.1021/ci00057a005..
- Weissweiler, Leonie, Abdullatif Köksal & Hinrich Schütze. 2024. *Hybrid human-LLM corpus construction and LLM evaluation for rare linguistic phenomena*. Submitted for publication. <https://arxiv.org/abs/2403.06965>.
- Whitelock, Peter, Mary McGee Wood, Harold L. Somers, Rod Johnson & Paul Bennett (eds.). 1987. *Linguistic theory and computer applications*. Academic Press.
- Wilson, Michael, Jackson Petty & Robert Frank. 2023. How abstract is linguistic generalization in large language models? Experiments with argument structure. In *TACL*, vol. 11, 1377–1395. <https://doi.org/10.1162/tacl%20a%2000608>.
- WIPO. 2024. *Recommendation for the Processing of Patent Information Using XML (Extensible Markup Language)*. <https://www.wipo.int/export/sites/www/standards/en/pdf/03-36-01.pdf>.
- Wolf, Florian & Edward Gibson. 2005. Representing discourse coherence: A corpus-based study. *Computational Linguistics* 31. 249–287. <https://doi.org/10.1162/0891201054223977>.
- Wood, Jim & Alec Marantz. 2017. The interpretation of external arguments. In Roberta D'Alessandro (ed.), *The verbal domain*, 255–278. Oxford University Press.
- Wu, Changxing, Liuwen Cao, Yubin Ge, Yang Liu, Min Zhang & Jinsong Su. 2022. A label dependence-aware sequence generation model for multi-level implicit discourse relation recognition. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, 11486–11494.
- Xiang, Wei & Bang Wang. 2023. A survey of implicit discourse relation recognition. *ACM Computing Surveys* 55. 1–34.
- Zaenen, Annie. 2025. Introduction. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 1–5. Konstanz: PubliKon.

- Zaenen, Annie, Cleo Condoravdi, Danny Bobrow & Raphael Hoffmann. 2010. Supporting rule-based representations with corpus-derived lexical information. In *Proceedings of the NAACL HLT 2010 first international workshop on formalisms and methodology for learning by reading*, 114–121.
- Zaenen, Annie, Alexis Cooper, Felix Zheng, Susan W. Brown & Martha Palmer. 2025. Poking at GPT’s NL Generalizations. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 265–298. Konstanz: PubliKon.
- Zaenen, Annie, Lauri Karttunen & Richard Crouch. 2005. Local textual inference: Can it be defined or circumscribed? In *Proceedings of the ACL workshop on empirical modeling of semantic equivalence and entailment*, 31–36. Association for Computational Linguistics.
- Zhang, Yingxue, Fandong Meng, Peng Li, Ping Jian & Jie Zhou. 2021. Context tracking network: Graph-based context modeling for implicit discourse relation recognition. In *Proceedings of the 2021 conference of the north american chapter of the Association for Computational Linguistics: Human language technologies (ACL-HLT)*, 1592–1599.
- Zheng, Zhaohui, Hongyuan Zha, Tong Zhang, Olivier Chapelle, Keke Chen & Gordon Sun. 2007. A general boosting method and its application to learning ranking functions for web search. In *Proceedings of NIPS*.
- Zymła, Mark-Matthias. 2025. Computational Challenges from Theoretical Semantic Modeling in LFG. In Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman & Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 25–54. Konstanz: PubliKon.

Semantics at the Crossroads

This Festschrift in honour of Richard (Dick) Crouch contains 14 articles which reflect the variety of Dick's interests and contributions to theoretical linguistics and natural language processing. Throughout his career, Dick has had a constant interest in the interaction between language and reasoning preferring insights embodied in implementations to mere speculations.

The papers in this volume range across topics such as: constraints to be put on computational systems so they generate and analyze only possible human languages; proposals to analyze discourse; linking language to theorem provers; accounts of lexical constructions in theoretical semantics and by LLMs; applications in biomedical research and multi-modal search.