Coordination of Theoretical and Computational Linguistics

Adam Przepiórkowski
ICS Polish Academy of Sciences
and University of Warsaw
adamp@ipipan.waw.pl

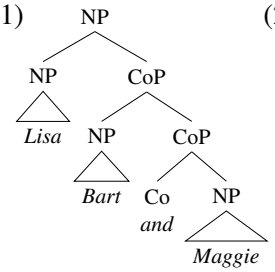
Agnieszka Patejuk
ICS Polish Academy of Sciences
aep@ipipan.waw.pl

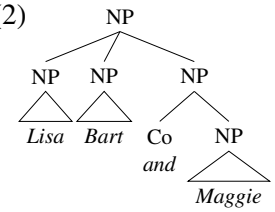
Abstract

The aim of this paper is to present a case study of a fruitful and, hopefully, inspiring interaction between formal and computational linguistics. A variety of NLP tools and resources have been used in linguistic investigations of the symmetry of coordination, leading to novel theoretical arguments. The converse impact of theoretical results on NLP work has been successful only in some cases.

1 Introduction

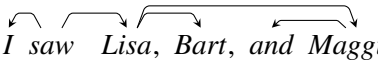
Coordination, as in *Lisa, Bart, and Maggie*, is a controversial theoretical linguistic phenomenon, with no agreement on its structure and properties. The two most common structures assumed in constituency syntax are those in (1)–(2), with variants of the binary structure in (1) almost universally assumed in Chomskyan linguistics, and variants of the flat structure in (2) universally adopted in LFG (Bresnan 1982, Dalrymple et al. 2019; see Patejuk 2023) and HPSG (Pollard and Sag 1987, 1994; see Abeillé and Chaves 2024).

(1)


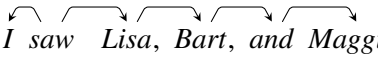
(2)


Similar disagreement is observed in theoretical dependency linguistics, see (4)–(6), and – consequently – in dependency corpora, where the current annotation standard, Universal Dependencies (UD), assumes (3), with Enhanced Dependencies (Schuster and Manning 2016) adding elements of (6), Surface-syntactic Universal Dependencies (Gerdes et al. 2018, 2021) adopts a variant of (4), and original Prague Dependency Treebanks (Hajič et al. 2006) assume (5).

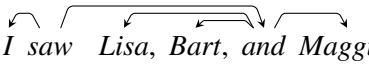
(3) Bouquet/Stanford (de Marneffe et al. 2021):



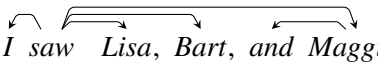
(4) Chain/Moscow (Mel'čuk 1988, 2009):



(5) &-headed/Prague (Sgall et al. 1986):



(6) Multi-headed/London (Hudson 1984, 1990):



This variety of structures reflects the lack of agreement regarding the fundamental issue of the symmetry of coordination: are all conjuncts syntactically equal? On some approaches, e.g., (1) and (3)–(4), the first conjunct is (closest to) the head of the coordinate structure and so it largely determines the properties of coordination. On other approaches, e.g., (2) and (5)–(6), all conjuncts determine such properties to the same extent.

This issue is related to another bone of contention: do conjuncts have to be alike, or can unlike categories or different grammatical functions be coordinated? Assuming that unlike category coordination is possible, as in the attested (7) (from the English Web 2015 corpus[1]), asymmetric approaches predict that the whole coordinate structure is an NP, while on symmetric approaches it has features of both NP and CP; see Figure 1.

(7) *I understand* [[$_{\text{NP}}$ *those concerns*] *and* [$_{\text{CP}}$ *that they are sincerely held*]].

For decades, these issues have been discussed on the basis of a handful of – usually constructed –

[1] http://www.sketchengine.eu (Jakubíček et al. 2013)

Proceedings of the Second Workshop on the Bridges and Gaps between Formal and Computational Linguistics (BriGap-2), pages 26–34

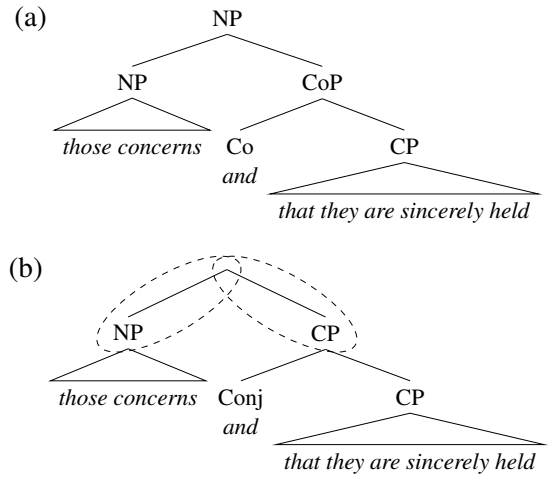


Figure 1: The structure of the coordination in (7) on the asymmetric approach of Munn 1993 (in (a)) and on the symmetric approach of Neeleman et al. 2023 (in (b))

examples and have remained unresolved. This paper shows that employing computational linguistic tools and resources makes it possible to construct novel theoretical arguments and that, conversely, awareness of theoretical issues may sometimes influence such tools and resources.

2 Morphosyntactic Corpora

The most basic use of NLP technologies that a theoretical linguist can make is the use of annotated corpora, often created with NLP applications in mind and/or annotated with NLP tools. There is surprisingly little awareness of morphosyntactically annotated corpora among generative syntacticians and of the power that lies in their associated query languages. Learning a given tagset and a given query language takes time, and tagsets and query languages may differ considerably even for a single language,[2] but doing so is well worth the effort.

For example, for many decades generative linguists believed, following a remark in Chomsky 1957, that only the same syntactic categories can be coordinated, a belief that was elevated to the status of a universal law (Williams 1981) and defended against some constructed counterexamples (e.g., in Sag et al. 1985) as recently as in 2020 (Bruening and Al Khalaf 2020). In particular, Sag et al.'s (1985) examples were claimed to involve the coordination of "supercategories" Predicate (in (8)) and Modifiers (in (9)), with no similar coordination of different categories possible in true argument positions.

(8) *Pat is* [[$_{\text{NP}}$ *a Republican*] *and* [$_{\text{AP}}$ *proud of it*]].

(9) *We walked* [[$_{\text{ADVP}}$ *slowly*] *and* [$_{\text{PP}}$ *with great care*]].

This long-held myth was refuted in Patejuk and Przepiórkowski 2023 – a paper in a prominent generative journal – on the basis of a few dozen attested examples; the effectiveness of this rebuttal was admitted by an erstwhile advocate of the refuted view, Bruening (2025): "[Patejuk and Przepiórkowski (2023)] are correct, and there is no requirement that conjuncts match in syntactic category".[3]

As coordination of unlikes is textually very rare, this would not be possible without an advanced use of morphosyntactically annotated corpora. In this case, the English Web 2015 corpus was queried with queries of varying complexity, e.g., (10) used to find examples such as (11), or (12) to find examples such as (13). Note that such queries require not only the basic knowledge of the query language and the relatively standard Penn Treebank (PTB; Marcus et al. 1993) tagset, but also the knowledge of regular expressions, not universally mastered by theoretical linguists.[4]

(10)
```
[lemma="with"] [lemma="respect|dignity"]
[tag="CC"] [tag="RB"]
```

(11) *. . . not all of us treat our animals* [[$_{\text{PP}}$ *with respect*] *and* [$_{\text{ADVP}}$ *humanly*]]*!*

(12)
```
[lemma="teach" & tag="VV.*"]
[tag="N.*|P.*|JJ.*|DT|CD.*"]{1,5} "that"
[tag="N.*|P.*|JJ.*|DT|CD.*"] []{1,5}
[word=","]? [tag="CC"] [tag="TO"]
```

(13) *You teach me* [[$_{\text{CP}}$ *that hard work pays off*] *and* [$_{\text{INFP}}$ *to never give up on a goal*]].

Among the claims of likeness of conjuncts is the claim that, in languages with rich nominal morphology, only the same grammatical cases can be coordinated (Weisser 2020). Advanced queries applied to the National Corpus of Polish (Przepiórkowski et al. 2011, 2012) and the Turkish Web 2012 corpus (Baisa and Suchomel 2012) helped to show that both kinds of coordination of unlikes – unlike categories and unlike cases – are readily found

[2] For example, those of The Corpus of Contemporary American English (COCA; Davies 2008–2025) are very different from those of enTenTen English corpora made available via SketchEngine (see fn. 1).

[3] See also Przepiórkowski and Patejuk 2025.

[4] While "`[tag="N.*|P.*|JJ.*|DT|CD.*"]{1,5}`" in (12) is a very poor regular definition of a nominal phrase, it returns a reasonable number of true positives.

in Polish (Przepiórkowski 2022) and in Turkish (Şenşekerci and Przepiórkowski 2024). Again, the intimate knowledge of relevant tagsets and query languages was crucial – for example, the awareness of the special feature of the Poliqarp search engine (Janus and Przepiórkowski 2007) of the Polish corpus, which makes it possible to use variables to specify that a given token must have a different case value than some other token.[5]

3 Valency Dictionaries

A resource that could be useful in an investigation of unlike category coordination is a valency dictionary, i.e., a lexicon containing information about arguments of verbs (and possibly predicates of other categories). Such a lexicon could encode information that a given argument of a given verb, e.g., the object of *understand* (cf. (7) above), could be realized as, say, an NP or a CP, which would make it worth checking whether this argument can be realized as NP and CP coordinated. Unfortunately, neither traditional valency dictionaries, nor machine-readable lexicons such as VerbNet (Kipper et al. 2006), contain such information.

Fortunately, however, the development of the largest and most detailed Polish valency dictionary, *Walenty* (Przepiórkowski et al. 2014, 2017), was informed by this issue, which resulted in the following unique feature. In this human- and machine-readable lexicon, arguments are described as sets of categories, often singleton sets, signaling that a given argument must bear a specific category, e.g., an NP. However, when there is corpus evidence that a given argument may be realized by a number of categories and, importantly, by coordinations of such unlike categories, this argument is described as a set of these categories.

For example, a valency schema for *zapowiedzieć* 'announce' contains information about the following two arguments (where ... indicates that more elements of the set are specified in the dictionary but omitted here for clarity), among others:

(14) a. `subj{np(str)}`
b. `obj{np(str);cp(int);cp(że);...}`

(14a) specifies one argument as a structural (i.e., normally nominative) NP subject, while (14b) specifies another argument as an object which may be realized as a structural (normally accusative) NP, an interrogative CP, a CP with the complementizer *że* 'that', etc. This information is supported with example (15) from the National Corpus of Polish.

(15) *Pan prezydent zapowiedział* [[$_{NP}$ *swój* Mr. president announced self's *patronat...*] *oraz* [$_{CP}$ *że takiej ustawy na* patronage and that such bill for *pewno nie podpisze*]].
certain not sign
'The president announced [[$_{NP}$ his patronage...] and [$_{CP}$ that he will definitely not sign such a bill]].'

The developed valency dictionary has been used in subsequent theoretical publications (e.g., Przepiórkowski 2022), as a rich source of examples of coordination of unlikes in Polish.

4 Implemented Grammars

As this valency dictionary is used by a number of grammar-based syntactic parsers of Polish (Patejuk 2015, Woliński 2015), these tools are able to parse sentences containing unlike category coordination, for example, the parse in Figure 2 produced by the *Świgra 2* parser (Woliński 2019).

Moreover, one of these parsers, the LFG parser *POLFIE* (Patejuk and Przepiórkowski 2012c,a, 2017b), has in turn been extensively used to verify a number of theoretical linguistic proposals, including analyses of agreement, predication, negation, numeral phrases, so-called reflexive markers, coordination of different grammatical functions, gapping, etc. (Patejuk and Przepiórkowski 2012b, 2014, 2015, 2017a, 2018, Przepiórkowski and Patejuk 2012, 2015, 2023).

Hence, in cases described in §§3–4, theoretical considerations fruitfully influenced the development of NLP resources (valency dictionary) and tools (syntactic parsers), which in turn helped in exemplifying and verifying theoretical analyses.

5 Syntactic Corpora

The possibility of theoretical influence on NLP tools and resources depends crucially on the stage of development of these tools and resources. For example, the representation of coordination in UD has various problems, e.g., it does not distinguish flat coordinations from certain nested coordinations. This problem was discussed – and solutions were proposed – in Przepiórkowski and Patejuk 2019,

[5] https://nkjp.pl/poliqarp/help/ense3.html/#x4-90003.4

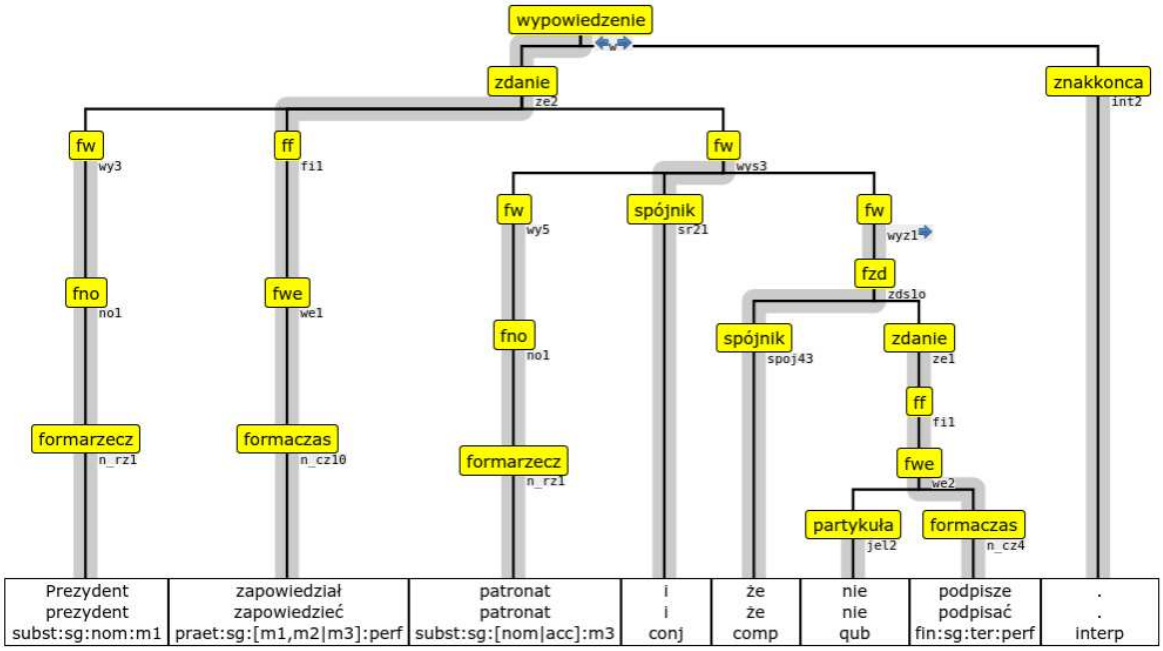


Figure 2: Parse of a much simplified version of example (15) produced by *Świgra 2*

but they have not been adopted by the UD community, probably because at this stage UD had already been employed in a number of corpora and was perceived as stable. Also other proposals, stemming from considerations of head-final languages (Kanayama et al. 2018), from the coordination of unlike grammatical functions (Patejuk and Przepiórkowski 2019), as in [*What and when*] *to eat to stay healthy*, and from the incompatibility of coordination with UD's core/non-core distinction (Przepiórkowski and Patejuk 2018), have not been adopted in UD.

Nevertheless, syntactic corpora – including UD corpora – were the basis for a novel theoretical argument against asymmetric approaches to coordination, described below.

6 Trained Parsers

Dependency Length Minimization (DLM) is a robustly demonstrated tendency for natural languages to strive for maximally local – shortest possible – dependencies.[6] Przepiórkowski and Woźniak 2023 argue that, given DLM, the distribution of lengths in binary coordinations in $\text{PTB}_{\&}$ – a version of PTB with enhanced annotation of coordination (Ficler and Goldberg 2016) – is compatible with symmetric approaches to coordination, but not with asymmetric approaches.

Specifically, they show that when the length difference between the two conjuncts increases, the tendency for the shorter constituent to be the initial conjunct also increases, but only when the governor is on the left (as in *I saw Bart and Lisa*) or absent (as in *Bart came and Lisa left*), and not when the governor is on the right (as in *Bart and Lisa laughed*). This tendency was observed whether the length of conjuncts was measured in the number of words, syllables, or characters; see Figure 3 in Appendix A. On various assumptions about the exact nature of DLM this observation is compatible with symmetric dependency representations such as (5)–(6) (and, by extension to constituency representations, (2)), but on no reasonable assumptions about DLM is it compatible with asymmetric representations such as (3)–(4) (and (1)).

This argument was reproduced on the basis of a variety of other manually annotated corpora for a number of languages, including UD corpora (Przepiórkowski et al. 2024b). However, in each case, the sparseness of data[7] resulted in some tendencies being only very weakly statistically significant and/or in the need to aggregate results in a way that might have influenced the final results.

To ameliorate this problem, a large portions of the COCA corpus of American English Davies 2008–2025 were automatically parsed with the

[6] See, e.g., Temperley and Gildea 2018 and references there, as well as Futrell et al. 2020.

[7] The relevant corpora had roughly between 15K (K = thousand) and 2,250K sentences, which translates into between 5K and 90K extracted coordinations.

Stanza dependency parser (Qi et al. 2020) trained on UD and SUD corpora, as well as with the Berkeley Neural Parser (BNP; Kitaev and Klein 2018, Kitaev et al. 2019) with the `benepar_en3` constituency model (Przepiórkowski et al. 2024a,b).[8] As these parsers produced much noise, implementing various filters removing obviously wrong parses was necessary. Nevertheless, the final results confirmed the earlier results based on manually parsed corpora, and – as was expected – all results turned out to be highly statistically significant.

While these relatively recent results have so far been only published in the proceedings of conferences devoted to NLP tools and resources (ACL, LREC-COLING, TLT), they are of vital importance for theoretical analyses of coordination, as they provide a novel argument not only against the most common treatment of coordination in corpora (both UD and SUD) and in a relatively niche theoretical dependency framework (Mel'čuk 1988, 2009), but also – by extension – against the asymmetric approaches almost universally assumed in Chomskyan linguistics.

7 Conclusion

This paper demonstrates that bridging the gap between theoretical and computational linguistics can be fruitful for both, but especially for theoretical linguistics. The awareness of traditional NLP resources (corpora, dictionaries) and tools (especially, parsers) among theoretical linguistics is too little, given how useful they can be for constructing and verifying theoretical arguments.

We conclude by noting that the NLP tools and resources used in the investigation of coordination described in this paper all date from the pre-LLM era. This is a feature, not a bug. It is not clear to us how LLMs could similarly support theoretical linguists in theory development.

[8] The relevant portions had roughly between 20M (M = million) and 70M sentences, which translates into 10–15M extracted binary coordinations.

References

Anne Abeillé and Rui Chaves. 2024. Coordination. In Stefan Müller, Anne Abeillé, Robert D. Borsley, and Jean-Pierre Koenig, editors, *Head-Driven Phrase Structure Grammar: The Handbook*, 2nd edition, pages 775–829. Language Science Press, Berlin.

Vit Baisa and Vit Suchomel. 2012. Large corpora for Turkic languages and unsupervised morphological analysis. In *First Workshop on Language Resources and Technologies for Turkic Languages, LREC 2012*, pages 28–32, Istanbul, Turkey. European Language Resources Association (ELRA).

Joan Bresnan, editor. 1982. *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge, MA.

Benjamin Bruening. 2025. Selectional violations in coordination (a response to Patejuk and Przepiórkowski 2023). *Linguistic Inquiry*, 56(3):439–483.

Benjamin Bruening and Eman Al Khalaf. 2020. Category mismatches in coordination revisited. *Linguistic Inquiry*, 51(1):1–36.

Miriam Butt and Tracy Holloway King, editors. 2012. *The Proceedings of the LFG'12 Conference*. CSLI Publications, Stanford, CA.

Miriam Butt and Tracy Holloway King, editors. 2015. *The Proceedings of the LFG'15 Conference*. CSLI Publications, Stanford, CA.

Noam Chomsky. 1957. *Syntactic Structures*. Mouton, The Hague.

Mary Dalrymple, John J. Lowe, and Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford University Press, Oxford.

Mark Davies. 2008–2025. The Corpus of Contemporary American English (COCA). Available online at https://www.english-corpora.org/coca/.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Jessica Ficler and Yoav Goldberg. 2016. Coordination annotation extension in the Penn Tree Bank. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 834–842, Berlin, Germany.

Richard Futrell, Roger P. Levy, and Edward Gibson. 2020. Dependency locality as an explanatory principle for word order. *Language*, 96(2):371–412.

Kim Gerdes, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. 2018. SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD. In (Marneffe et al. 2018), pages 66–74.

Kim Gerdes, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. 2021. Starting a new treebank? Go SUD! In *Proceedings of the Sixth International Conference on Dependency Linguistics (DepLing, Syntax Fest 2021)*, pages 35–46, Sofia, Bulgaria.

Jan Hajič, Jarmila Panevová, Eva Hajičová, Petr Sgall, Petr Pajas, Jan Štěpánek, Jiří Havelka, Marie Mikulová, Zdeněk Žabokrtský, Magda Ševčíková Razímová, and Zdeňka Urešová. 2006. Prague Dependency Treebank 2.0 (PDT 2.0).

Richard Hudson. 1984. *Word Grammar*. Blackwell, Oxford.

Richard Hudson. 1990. *English Word Grammar*. Blackwell, Oxford.

Miloš Jakubíček, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý, and Vít Suchomel. 2013. The TenTen corpus family. In *7th International Corpus Linguistics Conference CL*, pages 125–127.

Daniel Janus and Adam Przepiórkowski. 2007. Poliqarp: An open source corpus indexer and search engine with syntactic extensions. In *Proceedings of the ACL 2007 Demo and Poster Sessions*, pages 85–88, Prague.

Hiroshi Kanayama, Na-Rae Han, Masayuki Asahara, Jena D. Hwang, Yusuke Miyao, Jinho D. Choi, and Yuji Matsumoto. 2018. Coordinate structures in Universal Dependencies for head-final languages. In (Marneffe et al. 2018), pages 75–84.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extending verbnet with novel verb classes. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation, LREC 2006*, pages 1027–1032, Genoa. ELRA.

Nikita Kitaev, Steven Cao, and Dan Klein. 2019. Multilingual constituency parsing with self-attention and pre-training. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3499–3505, Florence, Italy. Association for Computational Linguistics.

Nikita Kitaev and Dan Klein. 2018. Constituency parsing with a self-attentive encoder. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2676–2686, Melbourne, Australia. Association for Computational Linguistics.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19:313–330.

Marie-Catherine de Marneffe, Teresa Lynn, and Sebastian Schuster, editors. 2018. *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*. Association for Computational Linguistics.

Igor Mel'čuk. 1988. *Dependency Syntax: Theory and Practice*. The SUNY Press, Albany, NY.

Igor Mel'čuk. 2009. Dependency in natural language. In Alain Polguère and Igor Mel'čuk, editors, *Dependency in Linguistic Description*, pages 1–110. John Benjamins, Amsterdam.

Alan Boag Munn. 1993. *Topics in the Syntax and Semantics of Coordinate Structures*. Ph.D. dissertation, University of Maryland.

Ad Neeleman, Joy Philip, Misako Tanaka, and Hans van de Koot. 2023. Subordination and binary branching. *Syntax*, 26(1):41–84.

Agnieszka Patejuk. 2015. *Unlike Coordination in Polish: An LFG Account*. Ph.D. dissertation, Institute of Polish Language, Polish Academy of Sciences, Cracow.

Agnieszka Patejuk. 2023. Coordination. In Mary Dalrymple, editor, *Handbook of Lexical Functional Grammar*, pages 309–374. Language Science Press, Berlin.

Agnieszka Patejuk and Adam Przepiórkowski. 2012a. A comprehensive analysis of constituent coordination for grammar engineering. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, pages 2191–2207, Mumbai, India.

Agnieszka Patejuk and Adam Przepiórkowski. 2012b. Lexico-semantic coordination in Polish. In (Butt and King 2012), pages 461–478.

Agnieszka Patejuk and Adam Przepiórkowski. 2012c. Towards an LFG parser for Polish: An exercise in parasitic grammar development. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012*, pages 3849–3852, Istanbul, Turkey. European Language Resources Association (ELRA).

Agnieszka Patejuk and Adam Przepiórkowski. 2014. Structural case assignment to objects in Polish. In *The Proceedings of the LFG'14 Conference*, pages 429–447, Stanford, CA. CSLI Publications.

Agnieszka Patejuk and Adam Przepiórkowski. 2015. An LFG analysis of the so-called reflexive marker in Polish. In (Butt and King 2015), pages 270–288.

Agnieszka Patejuk and Adam Przepiórkowski. 2017a. Filling the gap. In *The Proceedings of the LFG'17 Conference*, pages 327–347, Stanford, CA. CSLI Publications.

Agnieszka Patejuk and Adam Przepiórkowski. 2017b. POLFIE: współczesna gramatyka formalna języka polskiego. *Język Polski*, XCVII(1):48–64.

Agnieszka Patejuk and Adam Przepiórkowski. 2018. Predicative constructions with infinitival and clausal subjects. In *The Proceedings of the LFG'18 Conference*, pages 304–324, Stanford, CA. CSLI Publications.

Agnieszka Patejuk and Adam Przepiórkowski. 2019. Coordination of unlike grammatical functions. In *Proceedings of the Fifth International Conference on Dependency Linguistics (DepLing, SyntaxFest 2019)*, pages 26–37. Association for Computational Linguistics.

Agnieszka Patejuk and Adam Przepiórkowski. 2023. Category mismatches in coordination vindicated. *Linguistic Inquiry*, 54(2):326–349.

Carl Pollard and Ivan A. Sag. 1987. *Information-Based Syntax and Semantics, Volume 1: Fundamentals*. CSLI Publications, Stanford, CA.

Carl Pollard and Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. Chicago University Press / CSLI Publications, Chicago, IL.

Adam Przepiórkowski. 2022. Coordination of unlike grammatical cases (and unlike categories). *Language*, 98(3):592–634.

Adam Przepiórkowski, Mirosław Bańko, Rafał L. Górski, and Barbara Lewandowska-Tomaszczyk, editors. 2012. *Narodowy Korpus Języka Polskiego*. Wydawnictwo Naukowe PWN, Warsaw.

Adam Przepiórkowski, Mirosław Bańko, Rafał L. Górski, Barbara Lewandowska-Tomaszczyk, Marek Łaziński, and Piotr Pęzik. 2011. National Corpus of Polish. In *Proceedings of the 5th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, pages 259–263, Poznań, Poland.

Adam Przepiórkowski, Magdalena Borysiak, and Adam Głowacki. 2024a. An argument for symmetric coordination from Dependency Length Minimization: A replication study. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 1021–1033, Torino, Italy. ELRA and ICCL.

Adam Przepiórkowski, Magdalena Borysiak, Adam Okrasiński, Bartosz Pobożniak, Wojciech Stempniak, Kamil Tomaszek, and Adam Głowacki. 2024b. Symmetric dependency structure of coordination: Crosslinguistic arguments from dependency length minimization. In *Proceedings of the 22nd Workshop on Treebanks and Linguistic Theories (TLT 2024)*, pages 11–22, Hamburg,Germany. Association for Computational Linguistics.

Adam Przepiórkowski, Elżbieta Hajnicz, Anna Andrzejczuk, Agnieszka Patejuk, and Marcin Woliński. 2017. Walenty: gruntowny składniowo-semantyczny słownik walencyjny języka polskiego. *Język Polski*, XCVII(1):30–47.

Adam Przepiórkowski, Elżbieta Hajnicz, Agnieszka Patejuk, Marcin Woliński, Filip Skwarski, and Marek Świdziński. 2014. Walenty: Towards a comprehensive valence dictionary of Polish. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, pages 2785–2792, Reykjavík, Iceland. European Language Resources Association (ELRA).

Adam Przepiórkowski and Agnieszka Patejuk. 2012. The puzzle of case agreement between numeral phrases and predicative adjectives in Polish. In (Butt and King 2012), pages 490–502.

Adam Przepiórkowski and Agnieszka Patejuk. 2015. Two representations of negation in LFG: Evidence from Polish. In (Butt and King 2015), pages 322–336.

Adam Przepiórkowski and Agnieszka Patejuk. 2018. Arguments and adjuncts in Universal Dependencies. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*, pages 3837–3852, Santa Fe, NM. (Best position paper at COLING 2018).

Adam Przepiórkowski and Agnieszka Patejuk. 2019. Nested coordination in Universal Dependencies. In *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)*, pages 58–69. Association for Computational Linguistics.

Adam Przepiórkowski and Agnieszka Patejuk. 2023. Filling gaps with Glue. In *The Proceedings of the LFG'23 Conference*, pages 223–240. PubliKon.

Adam Przepiórkowski and Agnieszka Patejuk. 2025. Prenominal adverbs, or apparent selectional violations in coordination. *Linguistic Inquiry*, Early Access:1–29.

Adam Przepiórkowski and Michał Woźniak. 2023. Conjunct lengths in English, Dependency Length Minimization, and dependency structure of coordination. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15494–15512, Toronto, Canada. Association for Computational Linguistics.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108.

Ivan A. Sag, Gerald Gazdar, Thomas Wasow, and Steven Weisler. 1985. Coordination and how to distinguish categories. *Natural Language and Linguistic Theory*, 3(2):117–171.

Sebastian Schuster and Christopher D. Manning. 2016. Enhanced English Universal Dependencies: An improved representation for natural language understanding tasks. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation, LREC 2016*, pages 2371–2378, Portorož, Slovenia. European Language Resources Association (ELRA).

Berke Şenşekerci and Adam Przepiórkowski. 2024. Coordination of unlikes in Turkish. In *The Proceedings of the LFG'24 Conference*, pages 207–225. PubliKon.

Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Reidel, Dordrecht.

David Temperley and Daniel Gildea. 2018. Minimizing syntactic dependency lengths: Typological/cognitive universal? *Annual Review of Linguistics*, 4:67–80.

Philipp Weisser. 2020. On the symmetry of case in conjunction. *Syntax*, 23(1):42–77.

Edwin Williams. 1981. Transformationless grammar. *Linguistic Inquiry*, 12(4):645–653.

Marcin Woliński. 2015. Deploying the new valency dictionary Walenty in a DCG parser of Polish. In *Proceedings of the Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT 14)*, pages 221–229, Warsaw. Institute of Computer Science, Polish Academy of Sciences.

Marcin Woliński. 2019. *Automatyczna analiza składnikowa języka polskiego*. Wydawnictwa Uniwersytetu Warszawskiego, Warsaw.

Appendix A. Results of Przepiórkowski and Woźniak 2023

(See next page.)

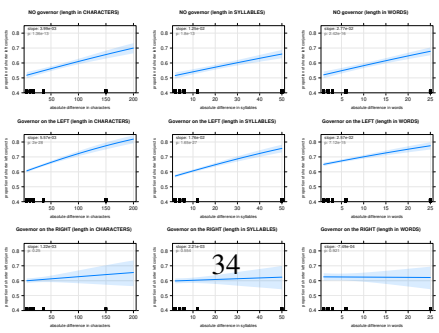


Figure 3: Modelled proportions of coordinations in PTB_{fc} with left conjuncts shorter, depending on the absolute difference of conjunct lengths, with confidence bands (Przepiórkowski and Woźniak 2023).