Computational Approaches for Understanding Semantic Constraints on Two-termed Coordination Structures

Julie Kallini(✉) and Christiane Fellbaum

Computer Science Department, Princeton University, Princeton, NJ 08544, USA
jkallini@alumni.princeton.edu, fellbaum@princeton.edu

Abstract. Coordination is a linguistic phenomenon where two or more terms or phrases, called *conjuncts*, are conjoined by a coordinating conjunction, such as *and*, *or*, or *but*. Well-formed coordination structures seem to require that the conjuncts are semantically similar or related. In this paper, we utilize English corpus data to examine the semantic constraints on syntactically *like* coordinations, which link constituents with the same lexical or syntactic categories. We examine the extent to which these semantic constraints depend on the type of conjunction or on the lexical or syntactic category of the conjuncts. We employ two distinct, independent metrics to measure the semantic similarity of conjuncts: WordNet relations and semantic word embeddings. Our results indicate that both measures of similarity have varying distributions depending on the particular conjunction and the conjuncts' lexical or syntactic categories.

Keywords: Coordination · Corpus linguistics · Semantics

1 Introduction

Coordination is the syntactic phenomenon whereby two or more terms or phrases are linked into one larger phrasal structure. We examine two-termed coordination phrases, where two elements (the *conjuncts*) are linked by the coordinating conjunctions *and*, *or*, or *but*, as in example (1).

(1) a. The president will [$_{VP}$ understand the criticism] and [$_{VP}$ take action].
b. Would you like [$_{NP}$ soup] or [$_{NP}$ salad] with your meal?
c. The new student was [$_{AP}$ intelligent] but [$_{AP}$ lazy].

The widely accepted Law of the Coordination of Likes (LCL) [17], which was proposed to account for the syntactic constraints on coordination, requires that the conjuncts belong to the same lexical or syntactic category.[1] While the LCL

[1] We focus on the coordination of lexical categories like nouns, verbs, and adjectives, as well as syntactic (phrasal) categories such as noun phrases, verb phrases and adjective phrases. For simplicity, we refer to both as *categories* in this paper.

P. Sojka et al. (Eds.): TSD 2022, LNAI 13502, pp. 64–76, 2022.
https://doi.org/10.1007/978-3-031-16270-1_6

accounts for the acceptability of the examples in (1), there are cases where it is not restrictive enough, as it would allow structures such as those in (2) [12]:

(2) a. * John ate with [$_{NP}$ his mother] and [$_{NP}$ good appetite].
b. * John [$_{AdvP}$ probably] and [$_{AdvP}$ unwillingly] went to bed.

The ungrammaticality here results from the semantic nature of these coordinations rather than their syntax. A stronger version of the LCL would require that conjuncts must also be semantically compatible. In this case, the prepositional phrase "with his mother" expresses accompaniment, whereas "with good appetite" expresses manner, so coordinating "his mother" and "good appetite" in (2a) produces a zeugma [12]. In (2b), the semantic difference between two adverbs (manner vs. epistemic) seems to account for the unacceptability.

Previous work explored the syntactic properties of coordination through a corpus-based approach [7], but a similar examination of the semantic constraints on coordination remains an open challenge. This paper explores the semantic properties of coordination structures through a large-scale quantitative corpus analysis. We study syntactically *like* coordinations, where the conjuncts have the same categories, and measure semantic constraints in terms of WordNet relations and word embeddings, which provide two independent measures of semantic similarity. We investigate whether the constraints depend on the type of conjunction (*and*, *or*, *but*) or on the categories of the conjuncts (noun, verb, adjective, adverb). A broader goal is to share data that may inform linguistic hypotheses about coordination.

2 Background and Related Work

Traditional linguistic analyses have given a thorough treatment of various semantic use cases of coordination; the three main types are often referred to as *conjunctive*, *disjunctive*, and *adversative* coordinations [4,5]. Conjunctive coordination links equal elements and is signalled by English *and*; disjunctive coordination usually indicates mutually exclusive options and is signalled by English *or*; and adversative coordination displays semantic contrast and is signalled by English *but*. However, the three conjunctions are not limited to these functions.

Quirk et al. [13] note that *and* is the conjunction with the most general meaning and usage and that it can take on several different connotations in context. For instance, *and* can link semantically contrastive elements and be replaced by *but* to produce a phrase with equivalent meaning, as in "she tried hard *and* failed." Quirk et al. also point out that *or* can be logically equivalent to *and* when following a negative, as demonstrated by the semantic equivalence of (3a) and (3b) [13].

(3) a. He doesn't have long hair *or* wear jeans.
b. He doesn't have long hair, *and* he doesn't wear jeans.

A similar replacement can take place with permissive modals, as in "the play can be performed in public *or* private theaters." Furthermore, *or* is not constrained to disjunctive scenarios. In (4), the conjuncts linked by *or* do not necessarily represent mutually exclusive options.

(4) a. He is good at painting with watercolors *or* with oils.
b. You can boil an egg *or* make a sandwich.

With regard to computational approaches to understanding coordination, previous work has focused on syntax rather than semantics. While the LCL mentioned in the introduction overgenerates with regard to semantics, it is too restrictive with regard to syntax, as it rules out perfectly acceptable coordinations with syntactically unlike conjuncts [12,14]:

(5) a. Pat is [NP a Republican] and [AP proud of it].
b. John is [AP healthy] and [PP in good shape].

In a previous paper, we examined *unlike* category coordinations using constituency-parsed corpora to identify coordination structures and determine the distributions of unlike phrasal category combinations [7]. The findings show, broadly, that noun phrases tend to coordinate with subordinate clauses and that the first conjunct tends to be shorter in length than the second conjunct, supporting an anti-symmetric account for the syntactic structure of coordination.

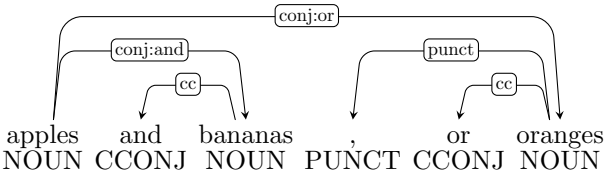
3 Approach

This paper focuses on the *semantic* properties of coordination, which to our knowledge have not yet been explored through a computational approach. We extract coordinate structures from hand-annotated Universal Dependencies corpora, and employ two methods to measure the similarity of conjuncts: WordNet's paradigmatic relations [3,10], and Google's Word2Vec semantic vectors [8,9], which reflect syntagmatic similarity. With these metrics, we investigate whether semantic relatedness correlates with particular conjunctions or categories.

3.1 Universal Dependencies Corpora

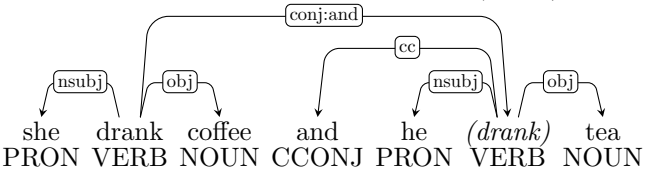
We examine corpora annotated within the Universal Dependencies (UD) project, which aims to provide a consistent dependency treebank annotation across many languages [11]. The CONJ relation links the first conjunct to all subsequent conjuncts, and all coordinating conjunctions are attached to the immediately following conjunct by the CC relation. We utilize the enhanced dependencies of UD v2, which augment the CONJ dependency labels between conjuncts by explicitly including their coordinating conjunction in the label. This feature is useful for disambiguating conjuncts in nested coordination phrases where more than one conjunction is involved, as in (6).

(6) UD v2 enhanced coordination annotation.



The UD v2 dependencies also elegantly annotate ellipsis constructions by using null nodes to represent elided material, such as in example (7). This representation aids the disambiguation of ellipsis constructions from simple coordinations of constituents.[2]

(7) UD v2 enhanced ellipsis annotation, where *(drank)* represents a null node.



We extract coordinations from three UD v2 corpora with enhanced dependencies: the English Web Treebank (EWT) [15], the Georgetown University Multilayer corpus (GUM) [1,18], and the English portion of the Parallel Universal Dependencies (PWT) treebanks [19]. Table 1 provides details about each corpus.

Table 1. Word counts, sentence counts, and example sources for each corpus we use.

Corpus	Words	Sentences	Example media/sources
EWT	254,825	16,621	weblogs, newsgroups, emails, reviews, etc.
GUM	135,886	7,397	interviews, news stories, academic writings, etc.
GUMReddit	16,356	895	Reddit posts
PUD	21,176	1,000	news, wikipedia

3.2 Coordination Extraction

Our coordination extraction script requires input files in the CoNLL-U format, the format in which UD annotations are provided. Sentences are represented using one or more lines, where each line corresponds to a single token or word. Ten fields fully describe each token or word, but for coordination extraction, we are only concerned with a subset of fields: the word `ID`, `FORM`, `LEMMA`, `UPOS`, `HEAD`,

[2] UD v2 also handles shared modifiers, such as the adjective *old* in "old men and women," using a distinct type of annotation.

DEPREL (basic universal dependency relation to the HEAD), and DEPS (enhanced dependency graph). We use a CoNLL-U parser to process corpus files into nested Python dictionaries [16].

Table 2. The usage and applicable lexical categories for each WordNet relation in our semantic analysis.

Relation	Usage	Categories
Synonymy	Are the conjuncts within the same synset?	N, V, Adj, Adv
Co-hyponymy	Are the conjuncts co-hyponyms?	N, V
Antonymy	Are the conjuncts antonyms?	Adj, Adv
Hypernymy	Is the first conjunct a hypernym of the second, or is the second conjunct a hypernym of the first?	N, V

Due to the nature of coordination annotation in UD, in which subsequent conjuncts are dependents of the first conjunct, we maintain coordination phrases as a dictionary mapping first conjunct IDs to sets of subsequent conjunct IDs. For each token in a sentence, the script searches the token's DEPS field for any dependencies of the form CONJ:CC, where CC is the lemma of a coordinating conjunction. If such a dependency is present, the current token is a conjunct of a coordination phrase, and the corresponding head is the first conjunct of that phrase. There can be only one such CONJ:CC dependency; we have checked this in the corpora programmatically, and one can also reason that it is impossible for a word to be a secondary conjunct of more than one coordination phrase. Importantly, the enhanced dependencies also indicate when a conjunct has been elided and thus should be excluded from the semantic analysis.[3]

3.3 Semantic Analysis

On the pragmatic assumption that conjuncts must be related in meanings, we examine and measure their semantic relatedness using two different, independent resources that capture paradigmatic and syntagmatic relatedness, respectively, WordNet [3,10] and Google's Word2Vec word embeddings [8,9]. We include coordination phrases with like conjuncts from the following open-class lexical categories: nouns, verbs, adjectives, and adverbs.

WordNet-based Similarity. WordNet's structure allows us to compare conjuncts in terms of "classical" semantic relations: synonymy, antonymy, and hypernymy/hyponymy. We expect many conjuncts to be co-hyponyms (as in *beer and wine*) or antonyms (as in *right and left*), since *and*, *or*, and *but* generally serve to conjoin elements with similar or contrasting meanings. We expect to find few synonyms (as in *cars and automobiles*), since conjoining words with

[3] Our code is available at https://github.com/jkallini/SemanticCoordinationAnalysis.

near-identical meanings seems redundant and uninformative.[4] We also expect to find relatively few conjoined words that are in a hypernymy/hyponymy relation (as in *roses and flowers*) except in cases where the hypernym in the second conjunct is modified and thus denotes co-hyponyms, as in *roses and other flowers*. We gather frequency data by counting coordination phrases that contain the basic presence or absence of these relations, and so the conjuncts' relative placement in the WordNet hierarchy does not affect our analysis. Table 2 summarizes the WordNet relations that we use for coordination semantic analysis in this project.

A challenge that accompanies the use of WordNet to analyze semantic relationships between conjuncts is *word sense disambiguation* (WSD), or the problem of selecting the correct sense/synset for strings that have multiple meanings. To handle ambiguous strings, we test the WordNet relations on all possible pairs of synsets corresponding to the two conjuncts of a coordination phrase.

Table 3. Summary of frequencies and chi-square tests comparing the presence of synonymy, antonymy, and co-hyponymy across the conjunctions *and*, *or*, and *but*. Statistically significant results are in bold.

Conjunction	Synonymy		Antonymy		Co-hyponymy	
	yes	no	yes	no	yes	no
'and' n (%)	160 (3.0)	5209 (97.0)	38 (6.2)	576 (93.8)	836 (17.6)	3919 (82.4)
'or' n (%)	23 (3.4)	651 (96.6)	**16 (23.2)**	53 (76.8)	104 (17.2)	501 (82.8)
'but' n (%)	19 (3.4)	542 (96.6)	1 (1.9)	53 (98.1)	94 (18.5)	413 (81.5)
Chi-square Test	$\chi^2(2, N = 5867) = 0.600$ $p = .741$		$\chi^2(2, N = 737) = 28.613$ $\mathbf{p < .001}$		$\chi^2(2, N = 5867) = 0.378$ $p = .828$	

Embedding-Based Similarity. We measure semantic relations among conjuncts with Google's pre-trained Word2Vec word embeddings [8,9]. Speakers commonly conjoin words referring to concepts from a given semantic domain (as in *students and teachers*) that are not reflected by a WordNet-style relation. We ask whether the distributional similarity captured by semantic vectors is reflected in coordinate structures.

4 Results

We first present general statistics about coordination from our corpus data. Our corpora include 6,892 like-category, two-termed coordination phrases, and in 6,641 (96.4%) of these coordinations, both conjuncts are present in WordNet. 27 coordinations (0.4%) include one elided conjunct; we exclude these from the semantic analysis. 5,579 coordinations (80.9%) use *and* as the coordinating conjunction; 723 (10.5%) use *or*, and 572 (8.3%) use *but*. The coordinating conjunction *nor* is only present in 18 coordinations (0.3%), so we exclude it from our analysis. For the results of our semantic analysis detailed in the next sections, we consider p-values less than .05 to be statistically significant.

[4] To avoid potential false positives for synonymy, we filter out coordinations in which both conjuncts have the same lemma, as in "he ran *faster and faster*."

4.1 WordNet Analysis

We begin with the bidirectional WordNet relations: synonymy, antonymy, and co-hyponymy. We observe how the presence of these relations is affected by the coordination phrase's conjunction or the conjuncts' lexical categories.

Table 3 summarizes the results of several chi-square tests of independence to examine the association between the different WordNet relations and the three coordinating conjunctions (*and*, *or*, *but*). The relation between conjunctions and the presence of antonymy was found to be significant, with the coordinating conjunction *or* having the largest proportion of coordinations in which the two conjuncts are antonyms.[5]

Similarly, Table 4 summarizes the results of several chi-square tests of independence examining the association between WordNet relations and the closed-class categories (noun, verb, adjective, adverb). For each WordNet relation, the association between the presence of the relation and the category of the conjuncts was found to be significant. Verbal categories had the largest proportion of synonymy and co-hyponymy, and adverbs had the largest proportion of antonymy.

We also performed an analysis of hypernymy. The first conjunct of the phrase was a hypernym of the second conjunct in 334 coordinations, and the second conjunct was a hypernym of the first in 372 coordinations. Using a chi-square test for goodness-of-fit, we did not find a significant difference in the distribution of the two types of hypernymy relations, $\chi^2(1, N = 706) = 2.045$, $p = 0.153$.

4.2 Word Embedding Analysis

We next present the results of our embedding-based analysis. Table 5 presents summary statistics of cosine similarity between conjuncts for each coordinating conjunction. A one-way ANOVA was performed to compare the effect of the coordinating conjunction on cosine similarity of the conjuncts. The one-way ANOVA revealed that there was a statistically significant difference in cosine similarity between groups ($F(2, 6531) = 13.613$, $p < .001$). Tukey's HSD Test for multiple comparisons found that the mean value of cosine similarity was significantly different between *and* coordinations and *or* coordinations ($p < .001$, 95% C.I. $= [.016, .046]$), and between *or* coordinations and *but* coordinations ($p = .004$, 95% C.I. $= [-.064, -.021]$). There was no statistically significant difference between *and* coordinations and *but* coordinations ($p = .223$).

[5] Previous corpus analyses have shown that antonymous word pairs co-occur within the same sentence with frequencies far higher than chance [2,6].

Table 4. Summary of frequencies and chi-square tests comparing the presence of synonymy, antonymy, and co-hyponymy across the four main lexical categories. Statistically significant results are in bold.

Conjunction	Synonymy		Antonymy		Co-hyponymy	
	yes	no	yes	no	yes	no
NOUN *n* (%)	40 (1.6)	2407 (98.4)	-	-	285 (11.6)	2162 (88.4)
VERB *n* (%)	**147 (4.3)**	3289 (95.7)	-	-	**750 (21.8)**	2686 (78.2)
ADJ *n* (%)	14 (2.2)	618 (97.8)	38 (6.0)	594 (94.0)	-	-
ADV *n* (%)	2 (1.9)	104 (98.1)	**17 (16.0)**	89 (84.0)	-	-
Chi-square Test	$\chi^2(2, N = 6621) = 35.893$ **$p < .001$**		$\chi^2(1, N = 738) = 11.814$ **$p = .001$**		$\chi^2(1, N = 5883) = 101.474$ **$p < .001$**	

Table 5. Summary statistics of cosine similarity between conjuncts for different coordinating conjunctions.

	N	Mean	Std. Dev.	Std. Err.	Min	Max
and	5334	.258	.157	.002	-.156	.967
or	670	.289	.180	.007	-.069	.964
but	530	.246	.142	.006	-.073	.734

We also compared cosine similarity between conjuncts of different lexical categories. Table 6 presents summary statistics of cosine similarity between conjuncts for each lexical category. A one-way ANOVA was performed to compare the effect of the conjuncts' category on their cosine similarity. The one-way ANOVA revealed that there was a statistically significant difference in cosine similarity between groups ($F(3, 6547) = 83.590$, $p < .001$). Tukey's HSD Test for multiple comparisons found that the mean value of cosine similarity was significantly different between all pairs of groups, shown in Table 7. Adverbial conjuncts had the highest cosine similarity on average, while verbal categories had the lowest on average.

Table 6. Summary statistics of cosine similarity between conjuncts for different categories.

	N	Mean	Std. Dev.	Std. Err.	Min	Max
NOUN	2474	.266	.178	.003	-.156	.965
VERB	3331	.242	.132	.002	-.104	.764
ADJ	646	.309	.166	.007	-.075	.893
ADV	100	.443	.224	.022	-.035	.967

Table 7. Results of Tukey's HSD test for cosine similarity between conjuncts of different lexical categories. The table displays the difference between group means, confidence intervals, and p-values for each comparison.

	Difference	95% C.I	p
NOUN vs. VERB	−.024	[−.035, −.016]	< .001
NOUN vs. ADJ	.043	[.025, .061]	< .001
NOUN vs. ADV	.177	[.137, .218]	< .001
VERB vs. ADJ	.067	[.050, .084]	< .001
VERB vs. ADV	.202	[.161, .242]	< .001
ADJ vs ADV	.135	[.092, .178]	< .001

5 Discussion

This section provides an in-depth discussion of the results presented in Sect. 4. We begin with the analysis of WordNet relations, followed by a discussion of embedding-based similarity.

5.1 WordNet Analysis

Table 8 contains several examples of synonymy, antonymy, and co-hyponymy pulled from the corpora. The relation between the presence of synonymy and the type of conjunction within a coordination phrase was not found to be statistically significant; the same was true for co-hyponymy. Overall, few examples of synonymy were found, which supports our hypothesis that conjoining words with very similar meanings is unnecessary and uninformative. With regard to co-hyponymy, while one might expect *and*-coordinations to have the highest percentage of co-hyponymy since *and* usually conjoins equal elements, *and*-coordinations had a lower percentage than *or*- and *but*-coordinations. This result supports the claims by Quirk et al. discussed in Sect. 2; if *and* is often used as a general-purpose conjunction independent of the meaning of the conjuncts, there would be weaker semantic constraints on *and*-coordinations and thus no strong correlation between *and* and the presence of a particular WordNet relation.

This result also suggests that the conjunction *or* is not restricted to its role as a disjunctive coordinator. As mentioned in Sect. 2, *or* may be used to conjoin options that are not mutually exclusive, as in "do you have any *brothers or sisters*." The data highlights other special use-cases for *or*; for instance, *or* is commonly used in appositive phrases, where one noun is used to define or modify another noun. One example from the corpora is shown in (8), where the two conjuncts are synonyms, and the second conjunct defines the first.

(8) Corn, *or* maize, [...] formed the basis of their diet.

While such cases show the versatility of *or*, most of its usages in our corpora overwhelmingly reflect its role as a disjunctive coordinator; the relation between

Table 8. Example coordinations for each bidirectional WordNet relation, conjunct category, and conjunction. Cells for which the given WordNet relation does not apply are filled with 'N/A'. Empty cells indicate that no samples were found.

Coordination Type		Synonymy	Antonymy	Co-hyponymy
NOUN	*and*	It [...] is still valuable for its many [N **examples**] **and** [N **exercises**].	N/A	Many [N **books**] **and** [N **articles**] in moral philosophy start with the observation [...]
	or	[N **Corn**], **or** [N **maize**], domesticated by 5000 BCE, formed the basis of their diet.	N/A	If you study [N **physics**] **or** [N **chemistry**] then you should describe the real world.
	but	-	N/A	Winter is definitely low [N **season**], [...], **but** also an ideal [N **time**] to save money [...]
VERB	*and*	Steven jiggled the handle [...], [V **turning**] **and** [V **twisting**] it most professionally and murmuring encouragements.	N/A	[...] he stayed up all night [V **writing**] letters to his Republican friends **and** [V **composing**] what would become his mathematical testament [...]
	or	These can be [V **bought**] at garden centers **or** [V **purchased**] online.	N/A	A healthy ecosystem [...] will [V **reduce**] the chance of these events happening, **or** will at least [V **mitigate**] adverse impacts.
	but	I also [V **think**] the National Endowment for the Arts is a waste, **but** [V **guess**] I would rather see my money go to the NEA [...]	N/A	[...] she [V **began**] appearing in films [...] **but** [V **continued**] to be primarily active in the theatre [...]
ADJ	*and*	I know that a transaction of this magnitude would make anyone [Adj **apprehensive**] **and** [Adj **worried**].	Hundreds of vendors offered products [Adj **new**] **and** [Adj **old**], joined by celebrity guests [...]	N/A
	or	-	Just as concepts can be [Adj **abstract**] **or** [Adj **concrete**], we can make a distinction [...]	N/A
	but	-	Skin will be [Adj **darker**] **but** [Adj **lighter**] than the Silkies.	N/A
ADV	*and*	[Adv **First**] **and** [Adv **foremost**] was the provision of open space for the benefit of townspeople [...]	It's an entirely [Adv **up**] **and** [Adv **down**] experience, however.	N/A
	or	-	Wilson did not work "[Adv **directly**] **or** [Adv **indirectly**]" for the CIA since retiring.	N/A
	but	-	-	N/A

conjunctions and antonymy was found to be significant, with *or* having the largest proportion of coordinations in which the two conjuncts are antonyms.

Coordinations containing *but* did not have a large percentage of any WordNet relation. This might have to do with its tendency to demonstrate contrast between entire clauses rather than individual words, a semantic phenomenon that is not captured using word-level measures of similarity.

Now we discuss the associations between the bidirectional WordNet relations and the lexical categories of the conjuncts within a coordination phrase. For each WordNet relation, the association between the presence of the relation and the category of the conjuncts was found to be significant, with verbal categories having the largest proportion of synonymy and co-hyponymy, and adverbs having the largest proportion of antonymy. Although it is not entirely clear why verbal

conjuncts tend to be more semantically similar, it is expected that the analyses of synonymy and co-hyponymy should be complementary. One hypothesis is that repeating similar verbs in conjoined predicates can express emphasis, as in example (9) taken from the corpora.

(9) Trust me, and most especially, [$_V$ trust] and [$_V$ believe] yourself.

The high frequency of antonymous conjuncts in coordinations of adverbs seems to stem from common phrases involving contrasting adverbs, such as "back and forth," "up and down," "here or there," and "more or less."

Finally, we discuss the hypernymy relation. We hypothesized that hypernymy would be attested in certain contexts, and that the second conjunct would more often be a hypernym of the first, as in "I bought *strawberries and other fruit*." The determiner *other* contextualizes that, in addition to strawberries, the rest of the items purchased are also kinds of fruit and therefore hyponyms of fruit, i.e., lexically unspecified co-hyponyms of the first conjunct. While the second conjunct was more often a hypernym of the first in our data, the difference in the distribution of the direction of the hypernymy relation was not statistically significant, suggesting that semantic asymmetry between the two conjunct positions is not very prominent.

5.2 Word Embedding Analysis

Our analysis of conjunct similarity using Word2Vec word embeddings revealed that *or*-coordinations had a significantly higher average cosine similarity between the conjuncts compared to *and*- and *but*-coordinations. This result complements the previous result regarding the large proportion of antonymous coordinations that use *or* as the conjunction. Word embeddings are created such that words that appear in similar contexts will have similar word vectors [8]. Direct antonyms often appear in similar contexts; for instance, *large* and *small* can interchangeably describe the size of an object, and as a result, they have similar word vectors. The vectors for *large* and *small* have a cosine similarity of .733, which is higher than the cosine similarity of near-synonyms like *large* and *big* (.556). Since *or* typically conjoins antonymous conjuncts, *or*-coordinations will have a high average cosine similarity between the conjuncts.

The nature of how word embeddings capture word similarity also accounts for the high average cosine similarity for adverbial conjuncts. Since a large percentage of adverbial coordination phrases contained antonymy, it follows that these coordination phrases would have a high cosine similarity between the conjuncts. Verbal coordinations had a significantly lower cosine similarity between conjuncts, despite having the largest proportion of synonymous coordinations; this can again be accounted for by the semantic tools we use. Near-synonymous verbs in WordNet like *grow* and *develop* or *print* and *publish* have similar word vectors, but the similarity is not quite as large as antonymous word vectors.

6 Conclusion

This paper analyzed the semantics of two-termed coordination phrases through a computational corpus analysis. We explore the differences in meaning between the two conjunct positions and the possible relationships they share by utilizing two representations of words: WordNet and word embeddings. The results show that *and* is a general-purpose coordinator that can conjoin conjuncts in various semantic relationships. The conjunction *or* is primarily used as a disjunctive coordinator, although it is not limited to this function. The relationships and similarities of the conjuncts also depend on their lexical categories. We see this analysis as a step towards a fuller understanding of speakers' real-world usage of coordination phrases.

References

1. Behzad, S., Zeldes, A.: A cross-genre ensemble approach to robust Reddit part of speech tagging. In: Proceedings of the 12th Web as Corpus Workshop (WAC-XII), pp. 50–56 (2020)
2. Fellbaum, C.: Co-occurrence and antonymy. Int. J. Lexicogr. **8**(4), 281–303 (1995)
3. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press, Cambridge (1998)
4. Haspelmath, M.: Coordinating constructions: an overview. In: Coordinating Constructions. Typological Studies in Language, vol. 58, pp. 3–39. John Benjamins Publishing Company (2004). https://doi.org/10.1075/tsl.58.03has
5. Haspelmath, M.: Coordination. In: Shopen, T. (ed.) Language Typology and Syntactic Description, vol. 2, 2 edn., chap. 1. Cambridge University Press (2007). https://doi.org/10.1017/CBO9780511619434
6. Justeson, J.S., Katz, S.M.: Co-occurrences of antonymous adjectives and their contexts. Comput. Linguist. **17**(1), 1–19 (1991)
7. Kallini, J., Fellbaum, C.: A corpus-based syntactic analysis of two-termed unlike coordination. In: Findings of the Association for Computational Linguistics: EMNLP 2021, pp. 3998–4008. Association for Computational Linguistics, Punta Cana, Dominican Republic, November 2021. https://doi.org/10.18653/v1/2021.findings-emnlp.335, https://aclanthology.org/2021.findings-emnlp.335
8. Mikolov, T., Chen, K., Corrado, G.S., Dean, J.: Efficient estimation of word representations in vector space (2013). https://arxiv.org/abs/1301.3781
9. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Neural Information Processing Systems (NeurIPS) (2013). https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf
10. Miller, G.A.: Wordnet: a lexical database for English. Commun. ACM **38**(11), 39–41 (1995)
11. Nivre, J., et al.: Universal dependencies v2: an evergrowing multilingual treebank collection. In: Proceedings of the 12th Language Resources and Evaluation Conference, pp. 4034–4043. European Language Resources Association, Marseille, France, May 2020. https://aclanthology.org/2020.lrec-1.497
12. Prażmowska, A.: Is unlike coordination against the law (of the coordination of likes)? (2015)

13. Quirk, R., Greenbaum, S., Leech, G., Svartvik, J.: A Comprehensive Grammar of the English Language. Longman, London (1985)
14. Sag, I.A., Gazdar, G., Wasow, T., Weisler, S.: Coordination and how to distinguish categories. Nat. Lang. Linguistic Theory **3**(2), 117–171 (1985). https://www.jstor.org/stable/4047644
15. Silveira, N., et al.: A gold standard dependency corpus for English. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014), pp. 2897–2904. European Language Resources Association (ELRA), Reykjavik, Iceland, May 2014. https://www.lrec-conf.org/proceedings/lrec2014/pdf/1089_Paper.pdf
16. Stenström, E.: CoNLL-U parser. https://github.com/EmilStenstrom/conllu/ (2021)
17. Williams, E.S.: Transformationless grammar. In: Linguistic Inquiry, vol. 12, pp. 645–653 (1981)
18. Zeldes, A.: The GUM corpus: creating multilayer resources in the classroom. Lang. Resour. Eval. **51**(3), 581–612 (2016). https://doi.org/10.1007/s10579-016-9343-x
19. Zeman, D., et al.: CoNLL 2017 shared task: multilingual parsing from raw text to universal dependencies. In: Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, pp. 1–19. Association for Computational Linguistics, Vancouver, Canada, August 2017. https://doi.org/10.18653/v1/K17-3001, https://aclanthology.org/K17-3001