What to Make of *make*? Sense Distinctions for Light Verbs

Julie Kallini and **Christiane Fellbaum**
Department of Computer Science
Princeton University
jkallini@alumni.princeton.edu,
fellbaum@princeton.edu

Abstract

Verbs like *make*, *have* and *get* present challenges for applications requiring automatic word sense discrimination. These verbs are both highly frequent and polysemous, with semantically "full" readings, as in *make dinner*, and "light" readings, as in *make a request*. Lexical resources like WordNet encode dozens of senses, making discrimination difficult and inviting proposals for reducing the number of entries or grouping them into coarser-grained supersenses. We propose a data-driven, linguistically-based approach to establishing a motivated sense inventory, focusing on *make* to establish a proof of concept.

From several large, syntactically annotated corpora, we extract nouns that are complements of the verb *make*, and group them into clusters based on their Word2Vec semantic vectors. We manually inspect, for each cluster, the words with vectors closest to the centroid as well as a random sample of words within the cluster. The results show that the clusters reflect an intuitively plausible sense discrimination of *make*. As an evaluation, we test whether words within a given cluster co-occur in coordination phrases, such as *apples and oranges*, as prior work has shown that such conjoined nouns are semantically related. Conversely, noun complements from different clusters are less likely to be conjoined. Thus, coordination provides a similarity metric independent of the contextual embeddings used for clustering. Our results pave the way for a WordNet sense inventory that, while not inconsistent with the present one, would reduce it significantly and hold promise for improved automatic word sense discrimination.

1 Background and Related Work

Jespersen coined the term *light verb* to denote verbs like *have*, *take* and *make* that carry little (but not zero) semantic information and that select for a noun, verb, or adjective complement to form a complex predicate. In their light verb use, these verbs are semantically bleached versions of main verbs as in (1a) and (1b), respectively:

(1) a. She made an attempt to prove the theorem.
    b. She made a birthday party for her best friend.

English light verbs usually have a corresponding simple full verb (e.g., *attempt*), but there are a number of subtle semantic distinctions between the light verb construction and the full verb (for a discussion see Kearns (2002)).

Automatic word sense disambiguation often relies on look-up in lexical resources like WordNet, where one confronts the challenge of dozens of different senses. WordNet includes 49 senses, an inventory that is often criticized by its users, but that is on fact lower than the number of sense distinction found in other lexical resources. Thus, Merriam Websters lists 25 main senses of the transitive verb, most of them with multiple subsenses. Even more vexing is the fact that light and full verb uses of *make* are not distinguished. Different proposals for grouping senses into semantically underspecified clusters have been made (Hughes and Prakash, 2006; Wei et al., 2015), but different automatic or manual efforts have resulted in multiple sense inventories that overlap only partially.

We propose a novel evaluation plan that is motivated by a previous study of coordination structures. In such structures, two constituents are conjoined by a coordinating conjunction, such as *and* or *or*. Prior work has shown that conjoined nouns are semantically related as measured via various WordNet relations like synonymy, antonymy, and co-hyponymy (Kallini and Fellbaum, 2022). This makes anomalous utterances, such as *apples and/or texting gloves*, or instances of zeugma, as in *she* *made a salad and a mess in the kitchen*, unlikely or humorous. To our knowledge, previous attempts at sense distinctions via argument selection have considered only single noun complements of a verb, a difficult task given that light verbs combine with a large number of nouns. Our focus in this paper is on *make*, but we expect our analysis to extend straightforwardly to other light verbs.

2 Approach

We distinguish different senses of *make* by examining its nominal complements, or nouns that it selects as a direct object. We reason that these noun complements must be sufficiently semantically similar for the verb phrases headed by *make* to be well formed, and that grouping these nouns can reveal distinct uses of *make* that point to different senses. To achieve this aim, we extract complements from dependency corpora and find groupings by clustering their word embeddings.

2.1 Universal Dependencies Corpora

We extract complements of *make* from corpora annotated within the Universal Dependencies (UD) project, which aims to provide a consistent dependency treebank annotation across many languages (Nivre et al., 2020). We use several English UD corpora to identify complements, and these corpora which are listed and detailed in Table 1.

UD annotates direct objects of verbs with the OBJ dependency relation. An example sentence showing the dependency relation between a form of *make* and its direct object is shown in Figure 1. Our complement extraction script requires input files in the CoNLL-U format, the typical format in which UD corpora are provided. In the CoNLL-U format, sentences are represented using one or more lines, where each line corresponds to a single token or word. Several fields are used to describe each token or word, but we mainly use the `HEAD` field, which is a pointer to the word token's head in the sentence, and the `DEPREL` field, which represents the basic universal dependency relation to the head. If the `HEAD` of a word token is a form of the verb *make*, and its `DEPREL` relation is OBJ, then it is a direct object and thus a complement of *make*. We use a CoNLL-U parser to process corpus files into nested Python dictionaries (Stenström, 2021) and perform this check for each token in the corpora to extract complements.

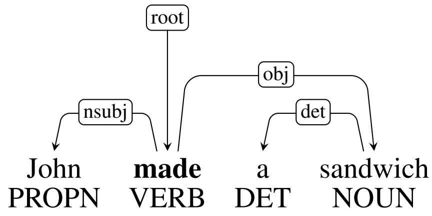

Figure 1: A sentence that uses the OBJ relation in UD to link *make* to its direct object.

2.2 Complement Clustering

To find groupings of complements, we perform k-means clustering on the complements' word embeddings. We use Google's Word2Vec word embeddings, which are 300-dimensional vectors pretrained on the Google News dataset (Mikolov et al., 2013a,b). We present two clustering analyses in this paper. As a first simple method, we run k-means clustering with $k = 30$ clusters on the unaltered 300-dimensional word vectors corresponding to the complements of *make*. In the second method, we also use principal component analysis (PCA) to reduce the embedding dimensionality for the complements' vectors and extract features that are relevant to the cluster structure, and we measure inertia to find an optimal value of k for clustering. PCA constructs a set of uncorrelated directions, or "components," that are ordered by their variance. Previous work has shown that removing features with low variance using PCA provides a filter that results in a more robust clustering, i.e. clusters with clearer structure that are less sensitive to noise (Ben-Hur and Guyon, 2003).

Based on cumulative explained variance, we determined that there is important information to be gained from the first 150 principal components, so we use the first 150 PCA features for the second clustering analysis. Along with PCA, we additionally performed an analysis of inertia, which measures how well the data is captured by clustering for different values of k. After trying values of $k \in [1, 30]$ we chose $k = 15$ clusters. For both clustering approaches, we used quantile outlier detection to filter out clusters that had too many or too few members. This removed clusters corresponding to senses that were either too generic or very specific.

2.3 Evaluation Using Coordination

Our evaluation is motivated by our previous work showing that pairs of nouns conjoined in coordina-

Corpus	Words	Sentences	Complements	Example media/sources
EWT	254,825	16,621	197	weblogs, newsgroups, emails, reviews, etc.
GUM	135,886	7,397	145	interviews, news stories, academic writings, etc.
GUMReddit	16,356	895	25	Reddit posts
LinES	94,217	5,243	109	fiction, nonfiction, spoken media
Atis	61,879	5,432	39	airline travel information
ParTUT	49,633	2,090	53	legal documents, news stories, webpages, etc.
PUD	21,176	1,000	21	news, wikipedia

Table 1: Word counts, sentence counts, *make* complement counts, and example sources for each corpus we use (Silveira et al., 2014; Zeldes, 2017; Behzad and Zeldes, 2020; Zeman et al., 2017)

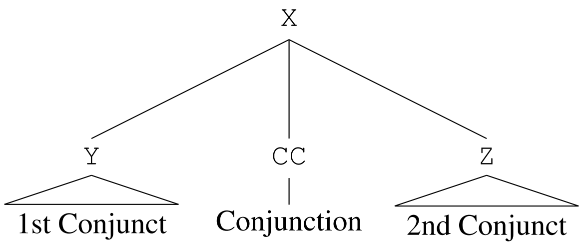

Figure 2: Simple ternary-branching coordination with Penn Treebank-style constituency annotations.

tion phrases are semantically similar; if the complements within a single cluster are sufficiently semantically similar in their functions as well as their contextual embedding representations, then we expect these complements to co-occur in coordination structures. To derive coordination data, we analyzed both manually and automatically parsed constituency corpora with Penn Treebank-style annotations, provided by Kallini and Fellbaum. Figure 2 shows an example of a simple instance of coordination in a constituency tree.

We performed this coordination analysis for two lexically-rich clusters, and we indeed found it to be the case that complements from the same cluster would more often co-occur in coordination structures. This result is detailed in the next section.

For less lexically-rich clusters, we devise an additional evaluation plan inspired by coordination. We use an independent similarity metric to compare complements within clusters as well as complements between clusters. First, we generate complement pairs. For instance, take A and B to be distinct clusters. We can measure the similarity of the complements within these two clusters by generating a list of complement pairs, $A \times B$. The average similarity of complement pairs in $A \times B$ should be less than the average similarity of complement pairs in $A \times A$ or $B \times B$. We use Wu-Palmer similarity as the metric for comparison and the Lesk algorithm for word sense disambiguation.

3 Results and Discussion

In total, we found 493 noun complements of *make* in the corpora after removing stopwords and tokens that are not present in the Word2Vec dictionary. The clusters found using simple k-means clustering with $k = 30$ clusters are summarized in Table 2. Outlier clusters have been removed from this table, so we present a reduced set of 26 clusters. The clusters found using k-means clustering with $k = 15$ clusters using PCA are summarized in Table 3. Figure 3 presents a visualization of complement clusters from this second analysis using the first two PCA components.

The second clustering analysis motivates a significantly reduced sense inventory while aligning with senses of *make* currently present in WordNet. For instance, there is a clear cluster for cases where *make* corresponds to cooking or preparing food (cluster #7 in Table 3). The cluster including complements like *impact, donation, and contribution* roughly correspond to its "give" meaning. The cluster with noun complements related to "mistakes" relates to the sense of "causing" or giving rise to an event.

However, our first analysis with a larger number of clusters captures some meaningful distinctions that are lost with a smaller value of k. For instance, this analysis provides a cluster of complements like *statue* and *sculpture* that correspond to the sense of "building" or "creating." The cluster containing *money* presents the sense of "gaining," and the cluster with complements such as *progress* and *improvement* presents the sense of "reaching for a goal."

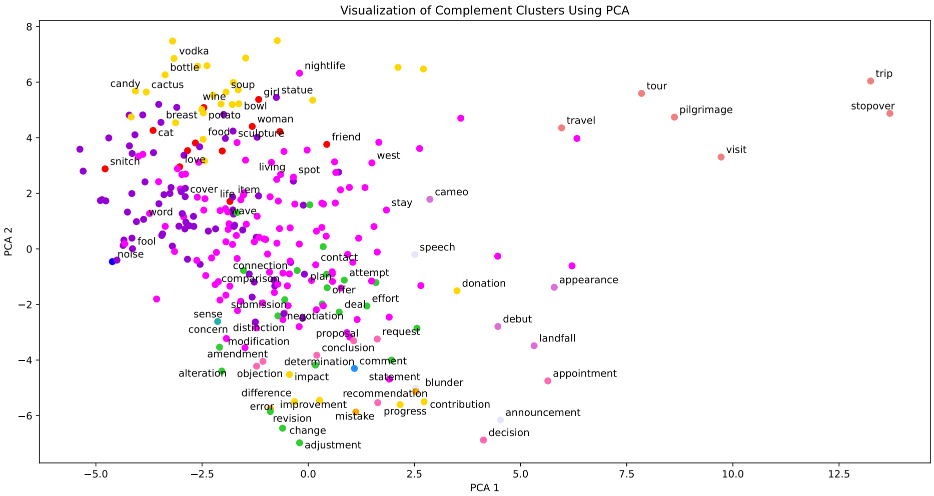

Figure 3: Visualization of complement clusters using the first two PCA components.

Cluster #	Size	Centroid Words	Sample Words
0	3	coup, coup_d'_etat, coup$_d$'_état	coup
1	5	entry, metastasis, breast	metastasis, breast, entry
2	16	word, phrase, language	word, reference, lyric
3	12	noise, ambient_noise, noises	noise, sound
4	25	sense, impression, feel	sense, assumption, representation
5	25	change, adjustment, alter	alteration, revision, change
6	20	decision, recommendation, announcement	conclusion, agreement, request
7	3	statue, bronze$_s$tatue, sculpture	statue, sculpture
8	13	friend, mother, daughter	child, love, mother
9	4	comment, leave	comment
10	3	cat, pet, bird	pet, cat, bird
11	19	effort, attempt, endeavor	project, plan, amendment
12	7	vodka, bottle, brandy	wine, bottle, vodka
13	7	contribution, donation, contributions	contribution, donation
14	4	reservation, reservations	reservation
15	11	money, funds, dollars	money, profit, buck
16	11	mistake, blunder, error	blunder, mistake, error
17	19	dessert, sandwich, soup	lunch, cheeseburger, food
18	10	debut, appearance, debuts	cameo, debut, appearance
19	14	joke, laugh, chuckle	chatter, mischief, joke
20	9	difference, disparity, discrepancy	distinction, gap, impact
21	5	appointment, appointments	appointment
22	7	progress, strides, improvement	recovery, improvement, progress
23	11	statement, remarks, press_release	statement, speech, filling
24	6	adaptation, adaption, film	adaptation, film
25	33	deal, agreement, offer	sale, package, transfer

Table 2: Size, word vectors close to the centroid, and a sample of cluster member words for 26 clusters created from basic k-means clustering.

Cluster #	Size	Sample Words
0	22	friend, life, love, girl, cat
1*	144	spot, stay, wave, west, nightlife
2	24	modification, alteration, change, adjustment, revision
3	114	comparison, sculpture, statue, cover, distinction
4*	4	comment
5	25	tour, travel, visit, pilgrimage, trip
6	9	noise
7	30	vodka, soup, wine, potato, food
8	21	objection, conclusion, proposal, submission, decision
9	11	blunder, error, mistake
10	33	deal, negotiation, effort, offer, attempt
11	12	debut, landfall, appearance, cameo
12	13	statement, announcement, speech
13	11	sense
14	20	impact, donation, difference, contribution, improvement

* Cluster identified as an outlier based on size.

Table 3: Size and sample words for each of the 15 clusters created from k-means clustering with PCA.

3.1 Evaluation

For the evaluation using coordination structures, we picked two clusters and tested whether complements within those clusters tended to co-occur in coordination phrases pulled from separate, independent corpus data. We chose clusters 3 and 7 since these were lexically-rich compared to some others that were large but contained repeated entries. The results show, generally, that complements from within the same cluster tend to coordinate more often than complements paired from different clusters. We found 26 instances of coordination where both conjuncts were members of cluster #3, such as "meaning and reference" and "writing and language." We found even more for cluster #7, since this cluster contains many types of food; there were 93 instances of coordination where both conjuncts were from cluster #7, such as "lunch or dinner," and "wine or cocktails." There were fewer (21) coordinations where conjuncts came from different clusters, such as "money and food."

We extended this initial analysis to cover the other clusters by generating complement pairs and measuring their Wu-Palmer similarity. Figure 4 shows that complement pairs where both complements are within the same cluster have a higher average Wu-Palmer similarity than pairs where the complements are members of different clusters, as shown by the brightness of the diagonal in the heatmap. These two evaluation steps generally show that the clusters represent nouns that are not only semantically similar based on contextual

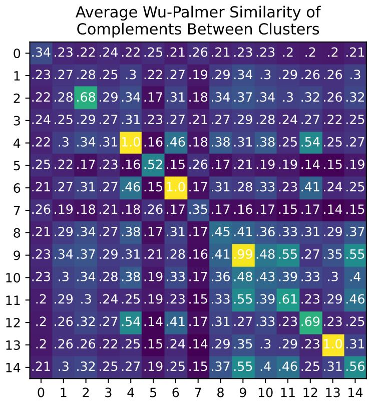

Figure 4: Average Wu-Palmer similarity for complement pairs between clusters from k-means clustering with PCA.

embeddings but also on their functional similarity.

4 Conclusion

We showed that a clustering analysis of complements of the light verb *make* can pave the way towards reducing WordNet's large sense inventory for this verb. Furthermore, we provided a novel evaluation method using coordination structures to test the robustness of the complement clustering. Future directions may apply this approach straightforwardly to other light verbs whose large sense inventories in WordNet have stymied word sense disambiguation efforts.

References

Shabnam Behzad and Amir Zeldes. 2020. A cross-genre ensemble approach to robust Reddit part of speech tagging. In *Proceedings of the 12th Web as Corpus Workshop (WAC-XII)*, pages 50–56.

Asa Ben-Hur and Isabelle Guyon. 2003. Detecting stable clusters using principal component analysis. *Methods in molecular biology*, 224:159–82.

Christopher Thad Hughes and Sushant Prakash. 2006. Clustering wordnet senses utilizing modified and novel similarity metrics.

Otto Jespersen. 1965. *A Modern English Grammar on Historical Principles: Syntax/Completed by Niels Hailslund*. Allen & Unwin.

Julie Kallini and Christiane Fellbaum. 2021. A corpus-based syntactic analysis of two-termed unlike coordination. In *Findings of the Association for Computational Linguistics: EMNLP 2021,*, pages 3998–4008, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Julie Kallini and Christiane Fellbaum. 2022. Computational approaches for understanding semantic constraints on two-termed coordination structures. In *Proceedings of the 25th International Conference on Text, Speech and Dialogue,*, pages 64–76, Cham. Springer International Publishing.

Kate Kearns. 2002. Light verbs in english.

Tomas Mikolov, Kai Chen, Greg S. Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Neural Information Processing Systems (NeurIPS)*.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4034–4043, Marseille, France. European Language Resources Association.

Natalia Silveira, Timothy Dozat, Marie-Catherine de Marneffe, Samuel Bowman, Miriam Connor, John Bauer, and Chris Manning. 2014. A gold standard dependency corpus for English. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2897–2904, Reykjavik, Iceland. European Language Resources Association (ELRA).

Emil Stenström. 2021. CoNLL-U parser. `https://github.com/EmilStenstrom/conllu/`.

Tingting Wei, Yonghe Lu, Huiyou Chang, Qiang Zhou, and Xianyu Bao. 2015. A semantic approach for text clustering using wordnet and lexical chains. *Expert Systems with Applications*, 42(4):2264–2275.

Amir Zeldes. 2017. The GUM corpus: Creating multilayer resources in the classroom. *Language Resources and Evaluation*, 51(3):581–612.

Daniel Zeman et al. 2017. 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*, pages 1–19, Vancouver, Canada. Association for Computational Linguistics.