Knowing the Author by the Company His Words Keep

Armin Hoenen[1], Niko Schenk[2]

[1]CEDIFOR, [2]Applied Computational Linguistics Lab (ACoLi)

Goethe University, Frankfurt am Main, Germany

`{hoenen,nschenk}@em.uni-frankfurt.de`

Abstract

In this paper, we analyze relationships between word pairs and evaluate their idiosyncratic properties in the applied context of authorship attribution. Specifically, on three literary corpora we optimize word pair features for information gain which reflect word similarity as measured by word embeddings. We analyze the quality of the most informative features in terms of word type relation (a comparison of different constellations of function and content words), similarity, and relatedness. Results point to the extraordinary role of function words within the authorship attribution task being extended to their pairwise relational patterns. Similarity of content words is likewise among the most informative features. From a cognitive perspective, we conclude that both relationship types reflect short distance connections in the human brain, which is highly indicative of an individual writing style.

Keywords: Word pairs, word relations, authorship attribution

1. Introduction

In traditional linguistics, there exists a famous saying that one should know a word by the company it keeps (Firth, 1957), which informally describes the meaning of a particular word by the context in which it occurs. In this paper, we investigate sets of frequently appearing similar words and their relations. To this end, we make use of word embeddings as a substitute and ground our work on related applications of distributed word representations and their applications (Mikolov et al., 2013a; Mikolov et al., 2013b; Mikolov et al., 2013c), which allows us to investigate relationships and even associations between words. Word embeddings have become popular recently and have been successfully applied to a variety of NLP tasks. Generally, word embeddings are believed to capture distributional similarity with an implication to semantic similarity. Kiela et al. (2015), for example, investigate relatedness and similarity and find embeddings to be specializable for both phenomena.

In our work, we focus our word pair-based methodology on the particular domain of authorship attribution. From an interpretation perspective, we conjecture that our proposed method allows to inspect not only single words in isolation and ask *what kind of word choices may characterize certain authors* (Marsden et al., 2013), but also to focus on properties of latent syntactic and semantic relations *between* words.

2. Word Relations & Human Cognition

In this study a special focus is laid on word pairs and their relations. Those can be understood as edges in a wider network and thus represent a first step towards modeling true semantic connections: Although finding a direct neurological correlate for words is a rather complex and controversial endeavour, many methods and theories even in psychology implicitly or explicitly operate with such networks. In computational linguistics, for instance, word nets have become a popular way of rationalizing relationships between concepts manifested through words. Especially theories on the mental lexicon belong to this sphere, cf. Elman (2004); Libben and Jarema (2002).

Accepting some neural implication of words, a most fundamental distinction is between function words and content words. The former carry few to no meaning, have been characterized as belonging to closed classes not easily incorporating neologisms and finally they function as markers of grammatical, syntactic or discourse related functions. Content words on the other hand are the meaningful referential actants or arguments around which all relations are built. Cognitively, both fundamental word classes imply different activity localized at different regions in the brain, see for instance Gordon and Caramazza (1982); Bradley and Garrett (1983); Friederici (1985); Diaz and McCarthy (2009). Shore (1995) referring to Bates et al. (1988) mention a left vs. right hemisphere contrast on the one hand and an anterior vs. posterior opposition on the other.

One possible hypothesis is thus that *function word - content word* (FC) connections display other, distinct properties from *content word - content word* (CC) and *function word - function word* (FF) connections, when inspecting word pair relations.

Neurologically, the latter two (CC & FF) could happen to be local connections (which could point to dendritic connectivity), whilst the former could span longer distances (pointing to axonic connectivity). In neurocomputation it is well perceived and taken as a basis for models that neurons have two main types of connectors passing on electrical signals: dendrites and axons (Hameroff, 2010). Dendrites transmit activation via synaptic gaps to neighboring cells. Activation spreads through neighbor's neighbors and so forth but decreases in strength with distance. New dendrites are formed even in adult life (Tavosanis, 2012) and thus their connectivity patterns may be more subject to differences over time and between individuals. Axons (usually one per neuron) are threshold dependent (i.e. they only "fire" once the activation of the cell has surpassed a critical activation level). They transmit and reinforce the signal information over longer distances (Purves et al., 2006, pp.1050). Consequently, if function words and content words are located in different brain regions, their connectivity type should not be characterized exclusively by dendritic

connections. Since neuroimaging techniques may be hardly able to visualize differentiatingly on-the-fly axonic activation pathways in contrast to dendritic ones the claims made here are most likely better testable through other than direct neurological imaging techniques.[1] If to this end distributional similarity would entail or be correlated with the cognitive connection between distributionally similar words, the properties of such connections could be reflected by the textual data. If FF and CC connections roughly represented neural short distance conncetions whereas FC connections represented long distance connections, if furthermore short distance connections are characterized more by dendritic connectivity, then a plausible hypothesis investigated in this paper could be that FF and CC connections are more plastic (they are more subject to interindividual differences) than FC connections which involve less flexible axonic connectivity. In turn this would entail that in authorship attribution those features (CC and FF) would be more informative, since they are more individual.

Authorship attribution is the task of identifying the unknown author of a textual document. An overview of current authorship attribution methods is given by Stamatatos (2009), consider also Rudman (1997); Baayen et al. (2002); Burrows (2002); Koppel and Schler (2003); Argamon (2008); Luyckx and Daelemans (2008); Spracklin et al. (2008); Luyckx (2011); Pennebaker (2011); Rybicki and Eder (2011); Smith and Aldridge (2011); Marsden et al. (2013); Evert et al. (2015); Eder et al. (2016); Markov et al. (2016). While in authorship attribution studies have focussed on various different aspects and linguistic levels, a look onto word pairs with an implication to word type has—to the best of our knowledge—not yet been taken. We fill this gap and look at the connection types when using the author as a class label in a machine learning setup by optimizing the feature vectors. This allows us to distinguish between more and less informative features.

3. Related Work

Relationships between two and more words have been researched for instance in psycho- and computational linguistics. Naturally, binary word relations are a subset of n-ary word relations. The identification of pairwise semantic word relations such as synonymy, hypernymy, meronymy, but also noun-compound relations, relations between named entities or semantically typed relations are among the areas of research where word pairs have been focussed (Hearst, 1992; Pantel et al., 2004; Roark and Charniak, 1998; Berland and Charniak, 1999; Costello et al., 2006; Strube and Ponzetto, 2006).

The specific type of relations, we will focus on here is function words and content words with an implication to authorship. Typically function words are high frequency words and the most frequent words of texts strongly tend to be function words, cf. Islam and Hoenen (2013). In determining word pair relations, a study using high frequency words vs. low frequency words is reported by Davidov and Rappoport (2006). Their focus is not authorship attribution but

the identification of word categories. A more fine grained analysis using word class instead of the rough binary difference function-content word was for instance pursued by Hasegawa et al. (2004), who found the relations between named entities to be informative or Widdows and Dorow (2002), who extract informative noun noun relations.

In the psycholinguistic and linguistic literature, some studies focus on the distinction between function and content words albeit with no explicit focus on authorship and seldom at word pairs. For instance, Corver and van Riemsdijk (2001) inspect some syntactic, distributional and lexical patterns and structures for function and content words. Bell et al. (2009) consider the predictability of English content and function words in discourse. On a related note, word pair features have received special attention in the recognition of implicit discourse relations (Biran and McKeown, 2013). Lexical access latencies, which are correlated with frequency are investigated for instance by Segalowitz and Lane (2000). While thus in the computational linguistics literature, word pair relations have been extensively researched, a focus or look to the categories function vs. content word is rather rare. On the other hand (psycho-) linguistic literature often applied these categories, but a focus on word pairs is rather rare.

Looking to authorship attribution, traditionally function words have been used much, moreover the application of both function and content words is not new, see for instance Koppel et al. (2009). Garcia and Martin (2006) look at the ratio of function to content words of a text. As a third (sub)field, authorship attribution thus uses both the distinction between function and content words and focusses on authorship, but to our best knowledge, the use of pairwise word patterns is rather rare and such a use under the distinction between content and function words novel.

4. Experiment

4.1. Corpora

We use two corpora provided by the computational stylistics group[2] since they cover two languages for which large word nets do exist: an English prose corpus, *A Short Collection of British Fiction* and a German one, *German Prose*. A third corpus comes from the Japanese Institute for Japanese Language and Linguistics,[3] the Meiroku corpus, containing short Japanese newspaper articles from the 19th century.[4] The three corpora vary in a large number of parameters, such as number of texts (German 66, English & Japanese 26), the number of authors (German 21, English 10, Japanese 13), the sizes of the respective texts, the time span of text creation, the genre and most of all in language and writing system, for more details see Table 1. The variety is of such a dimension, that any similar result obtained on all corpora has a very low likeliness to be caused by inherent corpus or sampling similarity.

---

[1] Though slightly outdated, Koch and Zador (1993) describe how dendrites are barely visualizable on microscopes at the time of publication.

[2] https://sites.google.com/site/computationalstylistics/

[3] https://www.ninjal.ac.jp/english/database/type/corpora/

[4] http://pj.ninjal.ac.jp/corpus_center/cmj/meiroku/

language	mean (T)	sd (T)	mean (S)	range (T)
EN	$245,292$	$182,066$	$16,501$	$41,129 - 973,341$
DE	$84,187$	$101,221$	$5,070$	$13,993 - 607,144$
JP	$1,949$	816	91	$533 - 4,863$

Table 1: Authorship corpora employed in this study and their properties. T = tokens, S = sentences.

4.2. Collecting Word Pairs

As a preprocessing step, we lowercase all texts and delete all punctuation symbols as we are interested in word relations only. For each text, we construct word embeddings using *word2vec*[5] with default settings. For any word in a specific text T_i being part of a corpus C, one can obtain the m most similar words according to the embeddings of this text alone, $0 < m < |T_i| - 1$. The word embedding vectors v_j for all words $w_j \in T_i$ of a specific text have a fixed number of dimensions (here 100). We choose the n most frequent words of the union of all texts in the corpus and denote this set MFW, $0 < n < \sum_{i=1}^{|C|} |T_i|, T_i \in C$. For each most frequent word $mfw_i \in MFW$, we collect the m most similar neighbors[6] for each text T_i. If the most frequent word is not in the text, the set contains m times ϵ. Otherwise, similarity for the current most frequent word with each other word $\in T_i$ is defined through some established vector similarity (here: cosine similarity) between the word pairs' vectors, $\delta(w_j, w_k) = cos(v_j, v_k)$. Thus, for each text T_i, for each word $w_i \in MFW$, we obtain a set with m most similar neighbors in T_i. The superset containing all such sets from one text is called W_t and the set of all W_t is called W_c.[7]

In the next step, we collect the set of all unique word pairs (w_i, n_i), where w_i is a most frequent word and n_i is any (most similar) neighbor of w_i in any text, $w_i \in MFW$, $n_i \in W_c$, $w_i \neq n_i$. The set of all unique such pairs U is used to construct a $|U|$-dimensional vector. For each word pair $(u_1, u_2) \in U$ with u_1 the most frequent word and u_2 a similar neighbor, a vector for each text is constructed and assigned 1 if the respective text's W_t in the set corresponding to the actual most frequent word's (u_1) neighbors contains u_2, otherwise the value is 0. For optimization in a machine learning setting, we assign the author as class label to the so constructed vector.

For an illustration see Table 2. Columns represent the feature vectors per text, which are used for optimization. 1 means the text at hand contains this word pair as one which according to the embeddings of the text has a non trivial similarity relation (the right word is among the n most similar 'neighbors' of the left, most frequent word given the embeddings of the actual text). 0 means such a similarity relation is absent. The last row contains the author of

---

[5] https://code.google.com/archive/p/word2vec/

[6] We call these words neighbors, which is a more or less arbitrary choice made in order to emphasize on conceptual and functional proximity.

[7] A further level of abstraction allowing also cross language comparisons based on word embeddings would be the use of (normalized) ranks or rank distances in a neighbor vector instead of tokens.

Feature	T_1	T_2	$\ldots$
the-then	1	0	
basket-grandma	1	0	
because-ten	0	1	
green-red	1	0	
car-automobile	1	1	
car-engine	0	0	
...	...	...	
Class	author1	author2	

Table 2: Author/text-wise representation of feature vectors.

the text as class label and the first column makes features (word pairs) explicit. We choose the most frequent $1,000$ words and the most similar 100 neighbors, since they have been shown to yield good performances on an authorship attribution experiment (Hoenen, 2017). We then conduct an optimization for Information Gain (IG) on the above described binary vectors (1 per text), using the WEKA machine learning environment (Hall et al., 2009). IG reduces the set of dimensions of the original vectors to the most informative ones by iteratively analysing additive performative gains (and ranks features for their informativity). We analyze the most informative features (word pairs) looking at their connection type (FC,CC,FF). In order to gain more insight, we also analyze the semantic similarity of CC pairs. For English, we additionally look at synonymy and relatedness.

For identifying function words, we bootstrap a function word lexicon, applying a heuristic: we extract all words not tagged as one of the tags starting with or equalling (NN,NE,ADV,ADJA,ADJD,TRUNC,V,FM,XY) for German from the Tiger corpus (Brants et al., 2004); for English from the Brown corpus (Francis and Kucera, 1979) all words not tagged as or having a tag starting with (v,to+vb,rb,n,jj,fw, punctuation-tags); for Japanese, *meishi, doushi, keyoushi, keijoushi, eitango, romajimon, kanbun* and *fukushi* were excluded. We give the proportions of incidences of each connection type before and after optimization.

For semantic similarity of CC pairs, we look at all features, where both words are present in WordNet (Miller, 1995; Fellbaum, 1998) or GermaNet (Hamp and Feldweg, 1997; Henrich and Hinrichs, 2010). We then compute the average similarity of all synsets of both, according to Jiang and Conrath (1997), which has been shown by Gurevych and Niederlich (2005) to have a high correlation with human judgements. We compute the proportion of wordpairs similar above a threshold (here 0.5) among all WordNet or GermaNet pairs.

For computational modeling, we employ the German standard interfaces provided from GermaNet, for English we

use WS4J[8] and for Japanese the JAWJAW wrapper of the Japanese WordNet.[9] We extracted the first translation of each term, then we proceeded with the English terms.[10] For English, we additionally look at synonymy and relatedness, using the same datasets as Kiela et al. (2015). These are: a) lists from Nelson et al. (2004) reflecting experiments on free association pointing to how related any two terms are and b) a resource on synonymy, the MyThes thesaurus developed by the OpenOffice.org project.[11] Synonymy, compare Freitag et al. (2005) and sources therein, refers to words which appear in extremely similar contexts and can be used to express the same meaning. We look at all word pairs, where both words were related according to those resources and give numbers for how many of those were carried over into the optimization.

5.   Results

The investigated numbers of features ($|U|$) were $251,348$ initial features for German, $668,242$ for English and $2,696$ for Japanese. An optimization through IG reduces the set of features in such a way, that the classification result is optimized (better with the new set, than with the old full set) if possible. The algorithm assigns each feature an estimated informativity and retains only those in the set of optimized features which surpass a certain threshold of informativity. WEKA records these IG assignments per feature and allows to inspect them in closer detail. Here, exemplarily we look at which features survive if we set the IG threshold at 0.1 including quite uninformative but no contradictory features, and 0.5 at the level of which already less than 10% of the features are present. The retained features are thus the most informative features (best characterize) for the author, cf. Tables 3–6 for the results.[12]

In Tables 3-6, the column *retained* gives the percentage of features surpassing a certain informativity level (IG thresh). The reduced feature set at IG thresh 0.5 contained overall $39,345$ features for English, where the entire feature set had contained $668,242$, thus the proportion in column retained is $\frac{39,345}{668,242} = 0.06$.

Column FF in Table 3 gives the proportion of all FF features, which at the current level of informativity have been retained within the more effective reduced feature set, given (denominator) all FF features in the entire set. FF features in entire feature set: $17,590$ (EN), $23,334$ (DE), $1,270$ (JP).

Analogously, in Table 4, column FC gives the proportion of all retained FC features. For English, the entire feature set contained $173,394$ FC features. Within the reduced feature set $12,608$ features were FC features, thus $\frac{12,608}{173,394} = 0.07$ is the percentage of FC features retained given all FC features measured. Initially there were $173,394$ FC features

Corpus	IG thresh	retained	FF	Diff
English	0.1	0.97	0.99	+0.02
German	0.1	0.72	0.93	+0.23
Japanese	0.1	0.9	0.92	+0.02
English	0.5	0.06	0.14	+0.08
German	0.5	0.007	0.018	+0.11
Japanese	0.5	0.06	0.1	+0.04

Table 3: Statistics on function word pair features.

Corpus	IG thresh	retained	FC	Diff
English	0.1	0.97	0.98	+0.01
German	0.1	0.72	0.71	−0.01
Japanese	0.1	0.9	0.9	+0
English	0.5	0.06	0.07	+0.01
German	0.5	0.007	0.007	+0
Japanese	0.5	0.06	0.03	−0.03

Table 4: Statistics on function word–content word pair features.

for English, $124,121$ (DE), 1310 (JP).

Table 5 is analogous to the above tables, CC gives the percentage of retained CC features in contrast to overall retained features. $477,258$ (EN) CC features initially, $103,893$ (DE), 116 (JP).

For English, using the above mentioned data sets, synonymy and relatedness could be tested, results shown in Table 6. As in the above tables, percentages of retained synonymic/related features given all synonymic/related features in the initial set are contrasted to the overall rate of feature retention.

Finally, Table 7 shows the percentage of features which are captured by a word net (both words of feature present) and have a semantic word net based similarity of more than 0.5. Column Sim gives their proportion on the entire initial feature set, while Sim retained gives the percentage in the reduced set of retained features (optimization). For German $12,510$ word pairs of the initial set were such that both words were covered by GermaNet. $5,600$ of those were similar above 0.5 semantic similarity (not IG thresh). This ratio is given in column Sim. In the reduced feature set, a subset of $10,310$ word pair features were covered by GermaNet with $4,716$ being similar above the similarity threshold. Thus through optimization the proportion of similar features given testable features increased, the amount of increase is displayed in column Diff. A column named Diff is used analogously also in the other tables and allows to see which features increase proportions in the optimized feature sets.

6.   Discussion

The results (especially in columns labeled Diff) were similar in all three languages and corpora. This is despite vast differences in the corpora and in the features extracted (for the time being in a binary representation). It entails that some independent explanation should hold for the observed phenomenon. The proportion of informative FF connections was larger than for FC connections, which in turn

---

[8]https://code.google.com/archive/p/ws4j/

[9]http://nlpwww.nict.go.jp/wn-ja/index.en.html

[10]JAWJAW provides similarity sets, but we found none of the actual pairs of the data.

[11]https://www.openoffice.org/linguacomponent/thesaurus.html

[12]For illustration, percentages are rounded to the second decimal digit or where necessary to the third.

Corpus	IG thresh	retained	CC	Diff
English	0.1	0.97	0.96	−0.01
German	0.1	0.72	0.68	−0.04
Japanese	0.1	0.9	0.69	−0.21
English	0.5	0.06	0.05	−0.01
German	0.5	0.007	0.005	−0.02
Japanese	0.5	0.06	0.009	−0.051

Table 5: Statistics on content word pair features.

Corpus	IG thresh	retained	syn/rel	Diff
English/syn	0.1	0.97	0.99	+0.02
English/syn	0.5	0.06	0.14	+0.08
English/rel	0.1	0.97	0.99	+0.02
English/rel	0.5	0.06	0.16	+0.1

Table 6: Statistics on informative synonyms (MyThes thesaurus) and informative related terms Nelson et al. (2004).

was larger than for CC connections. This would contradict the initial hypothesis that FF and CC connections show more individual plasticity and are thus author informative.

The current results could for instance be correlated with frequency rather than connection type. However, such a conclusion does not necessarily reflect the overall picture. Whereas there are very few and mostly very frequent function words implying a denser network and shorter pairwise paths (more effective dendritic coactivation), content words form much larger networks. This entails that only a restricted number of them might be susceptible to their neighbors spreading (dendritic) activation. These could be semantically similar terms which could align for instance with graded priming effects. The results in all three languages have shown that despite being overwhelmingly CC connections, semantically similar word pairs are author informative just as FF connections. Furthermore, for English this could be shown to hold true also for related and synonymic terms which further strengthens the assumption, that some kind of semantic relationship entails proximity. Finally, similar CCs were author informative despite their lower frequencies.

Additionally, the more informative the features became the larger was the proportion of retained similar word pair features (this trend was robust and is not just a consequence of the displayed IG limits). This may entail that what authors convey to be similar concepts is within the limits of simi-

larity of the overarching language largely individual. Such a conclusion would nicely align with hypothesis of larger general semantic transparency and adaptability, see for instance related hypotheses Eger et al. (2016). Since cross linguistically, the structures of synonymy (and ultimately word nets) seem to differ to certain extents, it seems plausible that interindividual differences to similar extents can exist within speakers of the same language. However, function words and functional morphemes such as regular endings are rigid and shared between speakers as to the mapping of form and function.

Furthermore as suspected above, if similarity is rather realized as dendritic short distance connection neurologically, a logical consequence would be a large plasticity since dendrons can adapt even in adult life, probably quicker so than axons the architecture and components of which are more complex. FF connections seem to be highly individual (again they could refect short distance dendritic connectivity). This time another plausible factor is their high frequency through which different pairs become more easily distinguishable. The important role of function words in authorship attribution, see for instance Burrows (2002); Pennebaker (2011) would be extended stating that not only frequencies and choice of function words are largely individual, but also patterns of their combination.

Yarowsky (1995); Yarowsky (1993) show that the trend for words to exhibit only one sense in a collocation is much less pronounced in collocations with function words. Thus, in his context it was also a crucial feature of word pair relations whether they were function or content words. Techniques such as pattern extraction for flexible patterns as presented in Schwartz et al. (2013) use word groups which can be larger than word pairs. These contain at least two high frequency words (often probably function words) and they have been shown to perform well and add value (as opposed to using only character n-gram or word n-gram features) on the identification of authors of very short Twitter messages, a disproportionally difficult authorship attribution task. This finding would align well with the here hypothesized findings.

7. Conclusion and Outlook

We presented a method, which uses word embeddings to identify pairwise word relations based on distributional criteria. We used these gained from single texts to optimize for recognition of an author and analysed the relations of author informative word pairs. Looking at these in terms of relation type showed similar patterns across three heterogeneous languages and corpora. The results point to similarity of content words (independent of word choice and lexicon) being subject to individual (authorial) fluctuation within the general characteristics of a language system although the extent of variability is not clear. Furthermore, the important role of function words for authorship attribution could be extended to relational patterns between them. A possible interpretation of FF connections and CC connections for similar words which deserves further research is that they could reflect short distance connections in the brain and by that token be highly adaptable and individual.

Corpus	IG thresh	Sim	Sim retained	Diff
English	0.1	0.001	0.001	±0
German	0.1	0.45	0.46	+0.01
Japanese	0.1	0.95	0.97	+0.02
English	0.5	0.001	0.002	+0.001
German	0.5	0.45	0.56	+0.11
Japanese	0.5	0.95	1	+0.05

Table 7: Statistics on word pair features retained, where both words according to the wordNets of their languages are similar above a threshold (here:0.5).

8. Bibliographical References

Argamon, S. (2008). Interpreting Burrows's delta: geometric and probabilistic foundations. *Literary and Linguistic Computing*, 23(2):131–147.

Baayen, H., van Halteren, H., Neijt, A., and Tweedie, F. (2002). An experiment in authorship attribution. In *6th JADT*, pages 29–37.

Bates, E., Bretherton, I., and Snyder, L. (1988). *From first words to grammar*. Cambridge University Press.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational english. *Journal of Memory and Language*, 60(1):92–111.

Berland, M. and Charniak, E. (1999). Finding parts in very large corpora. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 57–64. Association for Computational Linguistics.

Biran, O. and McKeown, K. (2013). Aggregated Word Pair Features for Implicit Discourse Relation Disambiguation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Volume 2: Short Papers*, pages 69–73, Sofia, Bulgaria.

Bradley, D. C. and Garrett, M. F. (1983). Hemisphere differences in the recognition of closed and open class words. *Neuropsychologia*, 21(2):155–159.

Brants, S., Dipper, S., Eisenberg, P., Hansen-Schirra, S., König, E., Lezius, W., Rohrer, C., Smith, G., and Uszkoreit, H. (2004). Tiger: Linguistic interpretation of a german corpus. *Research on Language and Computation*, 2(4):597–620.

Burrows, J. (2002). 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing*, 17(3):267–287.

Corver, N. and van Riemsdijk, H. (2001). *Semi-lexical categories: the function of content words and the content of function words*, volume 59. Walter de Gruyter.

Costello, F. J., Veale, T., and Dunne, S. (2006). Using wordnet to automatically deduce relations between words in noun-noun compounds. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 160–167. Association for Computational Linguistics.

Davidov, D. and Rappoport, A. (2006). Efficient unsupervised discovery of word categories using symmetric patterns and high frequency words. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 297–304. Association for Computational Linguistics.

Diaz, M. T. and McCarthy, G. (2009). A comparison of brain activity evoked by single content and function words: An fmri investigation of implicit word processing. *Brain Research*, 1282:38—-49.

Eder, M., Kestemont, M., and Rybicki, J. (2016). Stylometry with R: A package for computational text analysis. *R Journal*, 16(1).

Eger, S., Hoenen, A., and Mehler, A. (2016). Language classification from bilingual word embedding graphs. In *Proceedings of COLING 2016*. ACL.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in cognitive sciences*, 8(7):301–306.

Evert, S., Proisl, T., Vitt, T., Schöch, C., Jannidis, F., and Pielström, S. (2015). Towards a better understanding of Burrows's Delta in literary authorship attribution. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, pages 79–88, Denver, Colorado, USA. Association for Computational Linguistics.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Bradford Books.

Firth, J. R. (1957). *Papers in linguistics, 1934-1951*. Oxford University Press.

Francis, W. N. and Kucera, H. (1979). Brown corpus manual. Technical report, Department of Linguistics, Brown University, Providence, Rhode Island, US.

Freitag, D., Blume, M., Byrnes, J., Chow, E., Kapadia, S., Rohwer, R., and Wang, Z. (2005). New experiments in distributional representations of synonymy. In *Proceedings of the Ninth Conference on Computational Natural Language Learning*, pages 25–32. Association for Computational Linguistics.

Friederici, A. D. (1985). Levels of processing and vocabulary types: Evidence from on-line comprehension in normals and agrammatics. *Cognition*, 19(2):133–166.

Garcia, A. M. and Martin, J. C. (2006). Function words in authorship attribution studies. *Literary and Linguistic Computing*, 22(1):49–66.

Gordon, B. and Caramazza, A. (1982). Lexical decision for open-and closed-class words: Failure to replicate differential frequency sensitivity. *Brain and Language*, 15(1):143–160.

Gurevych, I. and Niederlich, H. (2005). Computing semantic relatedness of GermaNet concepts. In *Bernhard Fisseni, Hans-Christian Schmitz, Bernhard Schröder & Petra Wagner (Eds.), Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen:Proceedings of Workshop "Applications of GermaNet II" at GLDV'2005*, pages 462–474. Peter Lang.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD Explorations*, 11(1):10–18.

Hameroff, S. (2010). The "conscious pilot"—dendritic synchrony moves through the brain to mediate consciousness. *Journal of biological physics*, 36(1):71–93.

Hamp, B. and Feldweg, H. (1997). Germanet - a lexical-semantic net for german. In *In Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 9–15.

Hasegawa, T., Sekine, S., and Grishman, R. (2004). Discovering relations among named entities from large corpora. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 415. Association for Computational Linguistics.

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th con-

ference on Computational linguistics-Volume 2, pages 539–545. Association for Computational Linguistics.

Henrich, V. and Hinrichs, E. (2010). Gernedit - the germanet editing tool. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).

Hoenen, A. (2017). Using word embeddings for computing distances between texts and for authorship attribution. In *Proceedings of the 22nd International Conference on Applications of Natural Language to Information Systems (NLDB)*.

Islam, Z. and Hoenen, A. (2013). Source and translation classification using most frequent words. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 1299–1305.

Jiang, J. J. and Conrath, D. W. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. In *Proc of 10th International Conference on Research in Computational Linguistics, ROCLING'97*.

Kiela, D., Hill, F., and Clark, S. (2015). Specializing Word Embeddings for Similarity or Relatedness. In Lluís Màrquez, et al., editors, *EMNLP*, pages 2044–2048. The Association for Computational Linguistics.

Koch, C. and Zador, A. (1993). The function of dendritic spines: devices subserving biochemical rather than electrical computation. *Journal of Neuroscience*, 13:413–413.

Koppel, M. and Schler, J. (2003). Exploiting stylistic idiosyncrasies for authorship attribution. In *Proceedings of IJCAI'03 Workshop on Computational Approaches to Style Analysis and Synthesis*, volume 69, page 72.

Koppel, M., Schler, J., and Argamon, S. (2009). Computational methods in authorship attribution. *Journal of the Association for Information Science and Technology*, 60(1):9–26.

Libben, G. and Jarema, G. (2002). Mental lexicon research in the new millennium. *Brain and language*, 81(1):2–11.

Luyckx, K. and Daelemans, W. (2008). Authorship attribution and verification with many authors and limited data. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 513–520. Association for Computational Linguistics.

Luyckx, K. (2011). *Scalability Issues in Authorship Attribution*. Proefschriften UA-LW: taalkunde. VUB University Press.

Markov, I., Gómez-Adorno, H., Sidorov, G., and Gelbukh, A. (2016). Adapting cross-genre author profiling to language and corpus. In *Proceedings of the CLEF*, pages 947–955.

Marsden, J., Budden, D., Craig, H., and Moscato, P. (2013). Language Individuation and Marker Words: Shakespeare and His Maxwell's Demon. *PLoS ONE*, 8(6):63–88.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient Estimation of Word Representations in Vector Space. *Proceedings of Workshop at ICLR*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed Representations of Words and Phrases and their Compositionality. In C. J. C. Burges, et al., editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.

Mikolov, T., tau Yih, W., and Zweig, G. (2013c). Linguistic Regularities in Continuous Space Word Representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT-2013)*. Association for Computational Linguistics.

Miller, G. A. (1995). Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41.

Nelson, D. L., Mcevoy, C. L., and Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, pages 402–407.

Pantel, P., Ravichandran, D., and Hovy, E. (2004). Towards terascale knowledge acquisition. In *Proceedings of the 20th international conference on Computational Linguistics*, page 771. Association for Computational Linguistics.

Pennebaker, J. W. (2011). *The secret life of pronouns*. Bloomsbury Press.

Purves, W., Sadava, D., Hillis, D., Orians, G., and Heller, H. (2006). *Biologie*. 7th Edition, German version edited by J. Makl. Spektrum.

Roark, B. and Charniak, E. (1998). Noun-phrase co-occurrence statistics for semiautomatic semantic lexicon construction. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 2*, pages 1110–1116. Association for Computational Linguistics.

Rudman, J. (1997). The state of authorship attribution studies: Some problems and solutions. *Computers and the Humanities*, 31(4):351–365.

Rybicki, J. and Eder, M. (2011). Deeper Delta across genres and languages: do we really need the most frequent words? *Literary and Linguistic Computing*, 26(3):315–321.

Schwartz, R., Tsur, O., Rappoport, A., and Koppel, M. (2013). Authorship attribution of micro-messages. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1880–1891.

Segalowitz, S. J. and Lane, K. C. (2000). Lexical access of function versus content words. *Brain and language*, 75(3):376–389.

Shore, C. M. (1995). *Individual differences in language development*, volume 7. Sage Publications.

Smith, P. W. H. and Aldridge, W. (2011). Improving Authorship Attribution: Optimizing Burrows' Delta Method. *Journal of Quantitative Linguistics*, 18(1):63–88.

Spracklin, L., Inkpen, D., and Nayak, A. (2008). Using the complexity of the distribution of lexical elements as a feature in authorship attribution. In *LREC*.

Stamatatos, E. (2009). A Survey of Modern Authorship Attribution Methods. *J. Am. Soc. Inf. Sci. Technol.*, 60(3):538–556.

Strube, M. and Ponzetto, S. P. (2006). Wikirelate! computing semantic relatedness using wikipedia. In *AAAI*, volume 6, pages 1419–1424.

Tavosanis, G. (2012). Dendritic structural plasticity. *Developmental neurobiology*, 72(1):73–86.

Widdows, D. and Dorow, B. (2002). A graph model for unsupervised lexical acquisition. In *Proceedings of the 19th international conference on Computational linguistics-Volume 1*, pages 1–7. Association for Computational Linguistics.

Yarowsky, D. (1993). One sense per collocation. In *Proceedings of the ARPA Human Language Technology Workshop, Princeton*, pages 266–271.

Yarowsky, D. (1995). Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, pages 189–196. Association for Computational Linguistics.