Aligning Sentences from Standard Wikipedia to Simple Wikipedia

William Hwang, Hannaneh Hajishirzi, Mari Ostendorf, and Wei Wu
{wshwang, hannaneh, ostendor, weiwu}@u.washington.edu
University of Washington

Abstract

This work improves monolingual sentence alignment for text simplification, specifically for text in standard and simple Wikipedia. We introduce a method that improves over past efforts by using a greedy (vs. ordered) search over the document and a word-level semantic similarity score based on Wiktionary (vs. WordNet) that also accounts for structural similarity through syntactic dependencies. Experiments show improved performance on a hand-aligned set, with the largest gain coming from structural similarity. Resulting datasets of manually and automatically aligned sentence pairs are made available.

1 Introduction

Text simplification can improve accessibility of texts for both human readers and automatic text processing. Although simplification (Wubben et al., 2012) could benefit from data-driven machine translation, paraphrasing, or grounded language acquisition techniques, e.g. (Callison Burch and Osborne, 2003; Fung and Cheung, 2004; Munteanu and Marcu, 2005; Smith et al., 2010; Ganitkevitch et al., 2013; Hajishirzi et al., 2012; Kedziorski et al., 2014), work has been limited because available parallel corpora are small (Petersen and Ostendorf, 2007) or automatically generated are noisy (Kauchak, 2013).

Wikipedia is potentially a good resource for text simplification (Napoles and Dredze, 2010; Medero and Ostendorf, 2009), since it includes *standard* articles and their corresponding *simple* articles in English. A challenge with automatic alignment is that standard and simple articles can be written independently so they are not strictly parallel, and have very different presentation ordering. A few studies use editor comments attached to Wikipedia edit logs to extract pairs of simple and difficult words (Yatskar et al., 2010; Woodsend and Lapata, 2011). Other methods use text-based similarity techniques (Zhu et al., 2010; Coster and Kauchak, 2011; Kauchak, 2013), but assume sentences in standard and simple articles are ordered relatively.

In this paper, we align sentences in standard and simple Wikipedia using a greedy method that, for every simple sentence, finds the corresponding sentence (or sentence fragment) in standard Wikipedia. Unlike other methods, we do not make any assumptions about the relative order of sentences in standard and simple Wikipedia articles. We also constrain the many-to-one matches to cover sentence fragments. In addition, our method takes advantage of a novel word-level semantic similarity measure that is built on top of Wiktionary (vs. WordNet) which incorporates structural similarity represented in syntactic dependencies. The Wiktionary-based similarity measure has the advantage of greater word coverage than WordNet, while the use of syntactic dependencies provides a simple mechanism for approximating semantic roles.

Here, we report the first manually annotated dataset for evaluating alignments for text simplification, develop and assess a series of alignment methods, and automatically generate a dataset of sentence pairs for standard and simple Wikipedia. Experiments show that our alignment method significantly outperforms previous methods on the hand-aligned

211

Good	Apple sauce or applesauce is a puree made of apples.	Applesauce (or applesauce) is a sauce that is made from stewed or mashed apples.
Good Partial	Commercial versions of applesauce are really available in supermarkets	It is easy to make at home, and *it is also sold already made in supermarkets as a common food.*
Partial	*Applesauce* is a sauce that is *made from* stewed and mashed apples.	*Applesauce is made by* cooking down apples with water or apple cider to the desired level.

Table 1: Annotated examples: the matching regions for partial and good partial are italicized.

set of standard and simple Wikipedia article pairs. The datasets are publicly available to facilitate further research on text simplification.

2 Background

Given comparable articles, sentence alignment is achieved by leveraging the sentence-level similarity score and the sequence-level search strategy.

Sentence-Level Scoring: There are two main approaches for sentence-level scoring. One approach, used in Wikipedia alignment (Kauchak, 2013), computes sentence similarities as the cosine distance between vector representations of tf.idf scores of the words in each sentence. Other approaches rely on word-level $\sigma(w, w')$ semantic similarity scores $s(W, W') = \frac{1}{Z} \sum_{w \in W} \max_{w' \in W'} \sigma(w, w') idf(w)$. Previous work use WordNet-based similarity (Wu and Palmer, 1994; Mohler and Mihalcea, 2009; Hosseini et al., 2014), distributional similarity (Guo and Diab., 2012), or discriminative similarity (Hajishirzi et al., 2010; Rastegari et al., 2015).

In this paper, we leverage pairwise word similarities, and introduce two novel word-level semantic similarity metrics and show that they outperform the previous metrics.

Sequence-Level Search: There are several sequence-level alignment strategies (Shieber and Nelken, 2006). In (Zhu et al., 2010), sentence alignment between simple and standard articles is computed without constraints, so every sentence can be matched to multiple sentences in the other document. Two sentences are aligned if their similarity score is greater than a threshold. An alternative approach is to compute sentence alignment with a sequential constraint, i.e. using dynamic programming (Coster and Kauchak, 2011; Barzilay and Elhadad, 2003). Specifically, the alignment is computed by a recursive function that optimizes alignment of one or two consecutive sentences in one article to sentences

in the other article. This method relies on consistent ordering between two articles, which does not always hold for Wikipedia articles.

3 Simplification Datasets

We develop datasets of aligned sentences in standard and simple Wikipedia. Here, we describe the *manually annotated* dataset and leave the details of the *automatically generated* dataset to Section 5.2.

Manually Annotated: For every sentence in a standard Wikipedia article, we create an HTML survey that lists sentences in the corresponding simple article and allow the annotator to judge each sentence pair as a good, good partial, partial, or bad match (examples in Table 1): *Good:* The semantics of the simple and standard sentence completely match, possibly with small omissions (e.g., pronouns, dates, or numbers). *Good Partial:* A sentence completely covers the other sentence, but contains an additional clause or phrase that has information which is not contained within the other sentence. *Partial:* The sentences discuss unrelated concepts, but share a short related phrase that does not match considerably. *Bad:* The sentences discuss unrelated concepts.

The annotators were native speaker, hourly paid, undergraduate students. We randomly selected 46 article pairs from Wikipedia (downloaded in June 2012) that started with the character 'a'. In total, 67,853 sentence pairs were annotated (277 good, 281 good partial, 117 partial, and 67,178 bad). The kappa value for interannotator agreement is 0.68 (13% of articles were dual annotated). Most disagreements between annotators are confusions between 'partial' and 'good partial' matches. The manually annotated dataset is used as a test set for evaluating alignment methods as well as tuning parameters for generating automatically aligned pairs across standard and simple Wikipedia.

4 Sentence Alignment Method

We use a sentence-level similarity score that builds on a new word-level semantic similarity, described below, together with a greedy search over the article.

4.1 Word-Level Similarity

Word-level similarity functions return a similarity score $\sigma(w_1, w_2)$ between words w_1 and w_2. We introduce two novel similarity metrics: Wiktionary-based similarity and structural semantic similarity.

WikNet Similarity: The Wiktionary-based semantic similarity measure leverages synonym information in Wiktionary as well as word-definition cooccurrence, which is represented in a graph and referred to as WikNet. In our work, each lexical content word (noun, verb, adjective and adverb) in the English Wiktionary is represented by one node in WikNet. If word w_2 appears in any of the sense definitions of word w_1, one edge between w_1 and w_2 is added, as illustrated in Figure 1. We prune the WikNet using the following steps: i) morphological variations are mapped to their baseforms; ii) atypical word senses (e.g. "obsolete," "Jamaican English") are removed; and iii) stopwords (determined based on high definition frequency) are removed. After pruning, there are roughly 177k nodes and 1.15M undirected edges. As expected, our Wiktionary based similarity metric has a higher coverage of 71.8% than WordNet, which has a word coverage of 58.7% in our annotated dataset.

Motivated by the fact that the WikNet graph structure is similar to that of many social networks (Watts and Strogatz, 1998; Wu, 2012), we characterize semantic similarity with a variation on a link-based node similarity algorithm that is commonly applied for person relatedness evaluations in social network studies, the Jaccard coefficient (Salton and McGill, 1983), by quantifying the number of shared neighbors for two words. More specifically, we use the extended Jaccard coefficient, which looks at neighbors within an n-step reach (Fogaras and Racz, 2005) with an added term to indicate whether the words are direct neighbors. In addition, if the words or their neighbors have synonym sets in Wiktionary, then the shared synonyms are used in the extended Jaccard measure. If the two words are in each other's synonym lists, then the similarity is set to
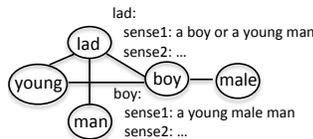

Figure 1: Part of WikNet with words "boy" and "lad".

1 otherwise, $\sigma_{wk}(w_1, w_2) = \sum_{l=0}^{n} J_l^s(w_1, w_2)$, for $J_l^s(w_1, w_2) = \frac{\Gamma_l(w_1) \cap_{syn} \Gamma_l(w_2)}{\Gamma_l(w_1) \cup \Gamma_l(w_2)}$ where $\Gamma_l(w_i)$ is the l-step neighbor set of w_i, and $\cap_{syn}$ accounts for synonyms if any. We precomputed similarities between pairs of words in WikNet to make the alignment algorithm more efficient. The WikNet is available at `http://ssli.ee.washington.edu/tial/projects/simplification/`.

Structural Semantic Similarity: We extend the word-level similarity metric to account for both semantic similarity between words, as well as the dependency structure between the words in a sentence. We create a triplet for each word using Stanford's dependency parser (de Marneffe et al., 2006). Each triplet $t_w = (w, h, r)$ consists of the given word w, its head word h (governor), and the dependency relationship (e.g., modifier, subject, etc) between w and h. The similarity between words w_1 and w_2 combines the similarity between these three features in order to boost the similarity score of words whose head words are similar and appear in the same dependency structure: $\sigma_{ss_{wk}}(w_1, w_2) = \sigma_{wk}(w_1, w_2) + \sigma_{wk}(h_1, h_2)\sigma_r(r_1, r_2)$ where σ_{wk} is the WikNet similarity and $\sigma_r(r_1, r_2)$ represents dependency similarity between relations r_1 and r_2 such that $\sigma_r = 0.5$ if both relations fall into the same category, otherwise $\sigma_r = 0$.

4.2 Greedy Sequence-level Alignment

To avoid aligning multiple sentences to the same content, we require one-to-one matches between sentences in standard and simple Wikipedia articles using a greedy algorithm. We first compute similarities between all sentences S_j in the simple article and A_i in standard article using a sentence-level similarity score. Then, our method iteratively selects the most similar sentence pair $S^*, A^* = \arg\max s(S_j, A_i)$ and removes all other pairs associated with the respective sentences, repeating until all sentences in the shorter document are aligned. The cost of aligning sentences in two articles S, A is $O(mn)$ where m, n are the number of sentences in
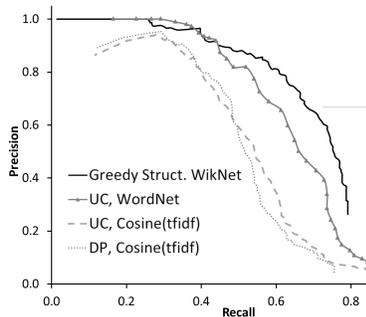
Figure 2: Precision-recall curve for our method vs. baselines.

Good vs. Others	Max F1	AUC
Greedy Struc. WikNet ($sim_G, \sigma_{ss_{wk}}$)	**0.712**	**0.694**
Unconst. WordNet (sim_{UC}, σ_{wd})	0.636	0.651
Ordered Vec. Space (sim_{DP}, s_{cos})	0.564	0.495
Unconst. Vec. Space (sim_{UC}, s_{cos})	0.550	0.509
Good & Good Partial vs. Others		
Greedy Struc. WikNet ($sim_G, \sigma_{ss_{wk}}$)	**0.607**	**0.529**
Unconst. WordNet (sim_{UC}, σ_{wd})	0.515	0.499
Ordered Vec. Space (sim_{DP}, s_{cos})	0.415	0.387
Unconst. Vec. Space (sim_{UC}, s_{cos})	0.431	0.391

Table 2: Max F1, AUC for identifying good matches and identifying good & good partial matches.

articles S and A, respectively. The run time of our method using WikNet is less than a minute for the sentence pairs in our test set.

Many simple sentences only match with a fragment of a standard sentence. Therefore, we extend the greedy algorithm to discover good partial matches as well. The intuition is that two sentences are good partial matches if a simple sentence has higher similarity with a fragment of a standard sentence than the complete sentence. We extract fragments for every sentence from the Stanford syntactic parse tree (Klein and Manning, 2003). The fragments are generated based on the second level of the syntactic parse tree. Specifically, each fragment is a S, SBAR, or SINV node at this level. We then calculate the similarity between every simple sentence S_j with every standard sentence A_i as well as fragments of the standard sentence A_i^k. The same greedy algorithm is then used to align simple sentences with standard sentences or their fragments.

5 Experiments

We test our method on all pairs of standard and simple sentences for each article in the hand-annotated data (no training data is used). For our experiments, we preprocess the data by removing topic names, list markers, and non-English words. In addition, the data was tokenized, lemmatized, and parsed using Stanford CoreNLP (Manning et al., 2014).

5.1 Results

Comparison to Baselines: The baselines are our implementations of previous work: *Unconstrained WordNet* (Mohler and Mihalcea, 2009), which uses an unconstrained search for aligning sentences and WordNet semantic similarity (in particular Wu-Palmer (1994)); *Unconstrained Vector Space* (Zhu et al., 2010), which uses a vector space representation and an unconstrained search for aligning sentences; and *Ordered Vector Space* (Coster and Kauchak, 2011), which uses dynamic programming for sentence alignment and vector space scoring.

We compare our method (*Greedy Structural WikNet*) that combines the novel Wiktionary-based structural semantic similarity score with a greedy search to the baselines. Figure 2 and Table 2 show that our method achieves higher precision-recall, max F1, and AUC compared to the baselines. The precision-recall score is computed for good pairs vs. other pairs (good partial, partial, and bad).

From error analysis, we found that most mistakes are caused by missing good matches (lower recall). As shown by the graph, we obtain high precision (about .9) at recall 0.5. Thus, applying our method on a large dataset yields high quality sentence alignments that would benefit data-driven learning in text simplification.

Table 2 also shows that our method outperforms the baselines in identifying good and good partial matches. Error analysis shows that our fragment generation technique does not generate all possible or meaningful fragments, which suggests a direction for future work. We list a few qualitative examples in Table 3.

Ablation Study: Table 4 shows the results of ablating each component of our method, sequence-level alignments and word-level similarity.

Sequence-level Alignment: We study the contribution of the greedy approach in our method by using word-level structural semantic WikNet similarity $\sigma_{ss(wk)}$ and replacing the sequence-level greedy search strategy with dynamic programming and un-

Good	The castle was later incorporated into the construction of Ashtown Lodge which was to serve as the official residence of the Under Secretary from 1782.	After the building was made bigger and improved, it was used as the house for the Under Secretary of Ireland from 1782.
Good Partial	Mozart's Clarinet Concerto and Clarinet Quintet are both in A major, and generally Mozart was more likely to use clarinets in A major than in any other key besides E-flat major	Mozart used clarinets in A major often.

Table 3: Qualitative examples of the good and good partial matches identified by our method.

Sequence-level	Max F1	AUC
Greedy ($sim_G, \sigma_{ss_{wk}}$)	**0.712$^+$**	**0.694$^+$**
Ordered ($sim_{DP}, \sigma_{ss_{wk}}$)	0.656$^+$	0.610$^+$
Unconstrained ($sim_{UC}, \sigma_{ss_{wk}}$)	0.689	0.689
Word-level	**Max F1**	**AUC**
Structural WikNet ($sim_G, \sigma_{ss_{wk}}$)	**0.712$^+$**	**0.694$^+$**
WordNet (sim_G, σ_{wd})	0.665$^+$	0.663$^+$
Structural WordNet ($sim_G, \sigma_{ss_{wd}}$)	0.685	0.679
WikNet (sim_G, σ_{wk})	0.697	0.669

Table 4: Max F1, AUC for ablation study on word-level and sequence-level similarity scores. Values with the $^+$ superscript are significant with p<0.05.

constrained approaches. As expected, the dynamic programming approach used in previous work does not perform as well as our method, even with the structural semantic WikNet similarity, because the simple Wikipedia articles are not explicit simplifications of standard articles.

Word-level Alignment: Table 4 also shows the contribution of the structural semantic WikNet similarity measure $\sigma_{ss_{wk}}$ vs. other word-level similarities (WordNet similarity σ_{wd}, structural semantic WordNet similarity $\sigma_{ss_{wd}}$, and WikNet similarity σ_{wk}). In all the experiments, we use the sequence-level greedy alignment method. The structural semantic similarity measures improve over the corresponding similarity measures for both WordNet and WikNet. Moreover, WikNet similarity outperforms WordNet, and the structural semantic WikNet similarity measure achieves the best performance.

5.2 Automatically Aligned Data

We develop a parallel corpus of aligned sentence pairs between standard and simple Wikipedia, together with their similarity scores. In particular, we use our best case method to align sentences from 22k standard and simple articles, which were download in April 2014. To speed up our method, we index

the similarity scores of frequent words and distribute computations over multiple CPUs.

We release a dataset of aligned sentence pairs, with a scaled threshold greater than 0.45.[1] Based on the precision-recall data, we choose a scaled threshold of 0.67 (P = 0.798, R = 0.599, F1 = 0.685) for good matches, and 0.53 (P = 0.687, R = 0.495, F1 = 0.575) for good partial matches. The selected thresholds yield around 150k good matches, 130k good partial matches, and 110k uncategorized matches. In addition, around 51.5 million potential matches, with a scaled score below 0.45, are pruned from the dataset.

6 Conclusion and Future Work

This work introduces a sentence alignment method for text simplification using a new word-level similarity measure (using Wiktionary and dependency structure) and a greedy search over sentences and sentence fragments. Experiments on comparable standard and simple Wikipedia articles outperform current baselines. The resulting hand-aligned and automatically aligned datasets are publicly available.

Future work involves developing text simplification techniques using the introduced datasets. In addition, we plan to improve our current alignment technique with better text preprocessing (e.g., coreference resolution (Hajishirzi et al., 2013)), learning similarities, as well as phrase alignment techniques to obtain better partial matches.

[1]`http://ssli.ee.washington.edu/tial/projects/simplification/`

References

[Barzilay and Elhadad2003] Regina Barzilay and Noemie Elhadad. 2003. Sentence alignment for monolingual comparable corpora. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

[Callison Burch and Osborne2003] Chris Callison Burch and Miles Osborne. 2003. Bootstrapping parallel corpora. In *Proceedings of the Human Language Technologies - North American Chapter of the Association for Computational Linguistics Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond - Volume 3 (HLT NAACL PARALLEL)*.

[Coster and Kauchak2011] William Coster and David Kauchak. 2011. Simple english Wikipedia: A new text simplification task. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*.

[de Marneffe et al.2006] Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*.

[Fogaras and Racz2005] Daniel Fogaras and Balazs Racz. 2005. Scaling link-based similarity search. In *Proceedings of the International Conference on World Wide Web (WWW)*, pages 641–650.

[Fung and Cheung2004] Pascale Fung and Percy Cheung. 2004. Mining Very-Non-Parallel Corpora: Parallel Sentence and Lexicon Extraction via Bootstrapping and EM. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

[Ganitkevitch et al.2013] Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. 2013. PPDB: The paraphrase database. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*, pages 758–764, Atlanta, Georgia, June. Association for Computational Linguistics.

[Guo and Diab.2012] Weiwei Guo and Mona Diab. 2012. Modeling semantic textual similarity in the latent space. In *Proceedings of the Conference of the Association for Computational Linguistics (ACL)*.

[Hajishirzi et al.2010] Hannaneh Hajishirzi, Wen-tau Yih, and Aleksander Kolcz. 2010. Adaptive near-duplicate detection via similarity learning. In *Proceedings of the Association for Computing Machinery Special Interest Group in Information Retrieval(ACM SIGIR)*, pages 419–426.

[Hajishirzi et al.2012] Hannaneh Hajishirzi, Mohammad Rastegari, Ali Farhadi, and Jessica Hodgins. 2012. Semantic understanding of professional soccer commentaries. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*.

[Hajishirzi et al.2013] Hannaneh Hajishirzi, Leila Zilles, Daniel S Weld, and Luke S Zettlemoyer. 2013. Joint coreference resolution and named-entity linking with multi-pass sieves. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

[Hosseini et al.2014] Mohammad Javad Hosseini, Hannaneh Hajishirzi, Oren Etzioni, and Nate Kushman. 2014. Learning to solve arithmetic word problems with verb categorization. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

[Kauchak2013] David Kauchak. 2013. Improving text simplification language modeling using unsimplified text data. In *Proceedings of the Conference of the Association for Computational Linguistics (ACL)*.

[Kedziorski et al.2014] Rik Koncel Kedziorski, Hannaneh Hajishirzi, and Ali Farhadi. 2014. Multi-resolution language grounding with weak supervision. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 386–396.

[Klein and Manning2003] Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the Conference of the Association for Computational Linguistics (ACL)*, pages 423–430.

[Manning et al.2014] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of the Conference of the Association for Computational Linguistics: System Demonstrations (ACL)*, pages 55–60.

[Medero and Ostendorf2009] Julie Medero and Mari Ostendorf. 2009. Analysis of vocabulary difficulty using wiktionary. In *Proceedings of the Speech and Language Technology in Education Workshop (SLaTE)*.

[Mohler and Mihalcea2009] Michael Mohler and Rada Mihalcea. 2009. Text-to-text semantic similarity for automatic short answer grading. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL)*.

[Munteanu and Marcu2005] Dragos Stefan Munteanu and Daniel Marcu. 2005. Improving machine translation performance by exploiting non-parallel corpora. *Computational Linguistics*.

[Napoles and Dredze2010] Courtney Napoles and Mark Dredze. 2010. Learning simple wikipedia: a cogitation in ascertaining abecedarian language. In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies Workshop on Computation Linguistics and Writing: Writing Processes and Authoring Aids (NAACL HLT)*.

[Petersen and Ostendorf2007] Sarah Petersen and Mari Ostendorf. 2007. Text simplification for langauge learners: A corpus analysis. In *Proceedings of the Speech and Language Technology in Education Workshop (SLaTE)*.

[Rastegari et al.2015] Mohammad Rastegari, Hannaneh Hajishirzi, and Ali Farhadi. 2015. Discriminative and consistent similarities in instance-level multiple instance learning. In *Proceedings of Computer Vision and Pattern Recognition (CVPR)*.

[Salton and McGill1983] Gerard Salton and Michael McGill. 1983. *Introduction to Modern Information Retrieval*. McGraw-Hill.

[Shieber and Nelken2006] Stuart Shieber and Rani Nelken. 2006. Towards robust context-sensitive sentence alignment for monolingual corpora. In *Proceedings of the Conference of the Association for Computational Linguistics (ACL)*.

[Smith et al.2010] Jason R. Smith, Chris Quirk, and Kristina Toutanova. 2010. Extracting parallel sentences from comparable corpora using document level alignment. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*.

[Watts and Strogatz1998] Duncan J. Watts and Steven H. Strogatz. 1998. Collective dynamics of small-world networks. *Nature*, pages 440–442.

[Woodsend and Lapata2011] Kristian Woodsend and Mirella Lapata. 2011. Wikisimple: Automatic simplification of wikipedia articles. In *Proceedings of the Association for Advancement of Artificial Intelligence Conference on Artificial Intelligence (AAAI)*, pages 927–932, San Francisco, CA.

[Wu and Palmer1994] Zhibiao Wu and Martha Palmer. 1994. Verbs semantics and lexical selection. In *Proceedings of the Conference of the Association for Computational Linguistics (ACL)*.

[Wu2012] Wei Wu. 2012. *Graph-based Algorithms for Lexical Semantics and its Applications*. Ph.D. thesis, University of Washington.

[Wubben et al.2012] Sander Wubben, Antal Van Den Bosch, and Emiel Krahmer. 2012. Sentence simplification by monolingual machine translation. In *Proceedings of the Conference of the Association for Computational Linguistics (ACL)*, pages 1015–1024.

[Yatskar et al.2010] Mark Yatskar, Bo Pang, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2010. For the sake of simplicity: Unsupervised extraction of lexical simplifications from wikipedia. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*.

[Zhu et al.2010] Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych. 2010. A monolingual tree-based translation model for sentence simplification. In *Proceedings of the International Conference on Computational Linguistics (COLING)*.