# Learning to Simplify Sentences Using Wikipedia

**William Coster**
Computer Science Department
Pomona College
`wpc02009@pomona.edu`

**David Kauchak**
Computer Science Department
Pomona College
`dkauchak@cs.pomona.edu`

## Abstract

In this paper we examine the sentence simplification problem as an English-to-English translation problem, utilizing a corpus of 137K aligned sentence pairs extracted by aligning English Wikipedia and Simple English Wikipedia. This data set contains the full range of transformation operations including rewording, reordering, insertion and deletion. We introduce a new translation model for text simplification that extends a phrase-based machine translation approach to include phrasal deletion. Evaluated based on three metrics that compare against a human reference (BLEU, word-F1 and SSA) our new approach performs significantly better than two text compression techniques (including T3) and the phrase-based translation system without deletion.

## 1 Introduction

In this paper we examine the sentence simplification problem: given an English sentence we aim to produce a simplified version of that sentence with simpler vocabulary and sentence structure while preserving the main ideas in the original sentence (Feng, 2008). The definition what a "simple" sentence is can vary and represents a spectrum of complexity and readability. For concreteness, we use Simple English Wikipedia[1] as our archetype of simplified English. Simple English Wikipedia articles represent a simplified version of traditional English Wikipedia articles. The main Simple English

Wikipedia page outlines general guidelines for creating simple articles:

- *Use Basic English vocabulary and shorter sentences. This allows people to understand normally complex terms or phrases.*

- *Simple does not mean short. Writing in Simple English means that simple words are used. It does not mean readers want basic information. Articles do not have to be short to be simple; expand articles, add details, but use basic vocabulary.*

The data set we examine contains aligned sentence pairs of English Wikipedia[2] with Simple English Wikipedia (Coster and Kauchak, 2011; Zhu et al., 2010). We view the simplification problem as an English-to-English translation problem: given aligned sentence pairs consisting of a normal, unsimplified sentence and a simplified version of that sentence, the goal is to learn a sentence simplification system to "translate" from normal English to simplified English. This setup has been successfully employed in a number of text-to-text applications including machine translation (Och and Ney, 2003), paraphrasing (Wubben et al., 2010) and text compression (Knight and Marcu, 2002; Cohn and Lapata, 2009).

Table 1 shows example sentence pairs from the aligned data set. One of the challenges of text simplification is that, unlike text compression where the emphasis is often on word deletion, text simplifica-

---

[1]http://simple.wikipedia.org

[2]http://en.wikipedia.org/

| | | | |
|---|---|---|---|
| a. | Normal: | Greene agreed that she could earn more by *breaking away from* 20th Century Fox. | |
| | Simple: | Greene agreed that she could earn more by *leaving* 20th Century Fox. | |
| b. | Normal: | The crust and *underlying relatively rigid* mantle make up the lithosphere. | |
| | Simple: | The crust and mantle make up the lithosphere. | |
| c. | Normal: | They *established themselves here and* called that port Menestheus's port. | |
| | Simple: | They called the port Menestheus's port. | |
| d. | Normal: | Heat engines are often confused with the cycles they attempt to mimic. | |
| | Simple: | *Real* heat engines are often confused with the *ideal engines or* cycles they attempt to mimic. | |
| e. | Normal: | *In 1962 ,* Steinbeck *received* the Nobel Prize *for* Literature. | |
| | Simple: | Steinbeck *won* the Nobel Prize *in* Literature *in 1962*. | |

Table 1: Example aligned sentences from English Wikipedia and Simple English Wikipedia. *Normal* refers an English Wikipedia sentence and *Simple* to a corresponding Simple English Wikipedia sentence.

tion involves the full range of transformation operations:

**deletion:** "*underlying relatively rigid*" in **b.**, "*established themselves here and*" in **c.** and the comma in **d.**

**rewording:** "*breaking away from*" → "*leaving*" in **a.** and "*received*" → "*won*" in **e.**

**reordering:** in **e.** "*in 1962*" moves from the beginning of the sentence to the end.

**insertion:** "*ideal engines or*" in **d.**

Motivated by the need to model all of these different transformations, we chose to extend a statistical phrase-based translation system (Koehn et al., 2007). In particular, we added phrasal deletion to the probabilistic translation model. This addition broadens the deletion capabilities of the system since the base model only allows for deletion *within* a phrase. As Kauchak and Coster (2011) point out, deletion is a frequently occurring phenomena in the simplification data.

There are a number of benefits of text simplification research. Much of the current text data available including Wikipedia, news articles and most web pages are written with an average adult reader as the target audience. Text simplification can make this data available to a broader range of audiences including children, language learners, the elderly, the hearing impaired and people with aphasia or cognitive disabilities (Feng, 2008; Carroll et al., 1998). Text simplification has also been shown to improve

the performance of other natural language processing applications including semantic role labeling (Vickrey and Koller, 2008) and relation extraction (Miwa et al., 2010).

## 2 Previous Work

Most previous work in the area of sentence simplification has not been from a data-driven perspective. Feng (2008) gives a good historical overview of prior text simplification systems including early rule-based approaches (Chandrasekar and Srinivas, 1997; Carroll et al., 1998; Canning et al., 2000) and a number of commercial approaches. Vickrey and Koller (2008) and Miwa et al. (2010) employ text simplification as a preprocessing step, though both use manually generated rules.

Our work extends recent work by Zhu et al. (2010) that also examines Wikipedia/Simple English Wikipedia as a data-driven, sentence simplification task. They propose a probabilistic, syntax-based approach to the problem and compare against a baseline of no simplification and a phrase-based translation approach. They show improvements with their approach on target-side only metrics including Flesch readability and n-gram language model perplexity, but fail to show improvements for their approach on evaluation metrics that compare against a human reference simplification. In contrast, our approach achieves statistically significant improvements for three different metrics that compare against human references.

Sentence simplification is closely related to the

problem of sentence compression, another English-to-English translation task. Knight and Marcu (2002) were one of the first to formalize text compression as a data-driven problem and proposed a probabilistic, noisy-channel model and decision tree-based model for compression. Galley and McKeown (2007) show improvements to the noisy-channel approach based on rule lexicalization and rule Markovization. Recently, a number of approaches to text compression have been proposed that score transformation rules discriminatively based on support vector machines (McDonald, 2006; Cohn and Lapata, 2009) and conditional random fields (Nomoto, 2007; Nomoto, 2008) instead of using maximum likelihood estimation. With the exception of Cohn and Lapata (2009), all of these text compression approaches make the simplifying assumption that the compression process happens only via word deletion. We provide comparisons with some of these systems, however, for text simplification where lexical changes and reordering are frequent, most of these techniques are not appropriate.

Our proposed approach builds upon approaches employed in machine translation (MT). We introduce a variant of a phrase-based machine translation system (Och and Ney, 2003; Koehn et al., 2007) for text simplification. Although MT systems that employ syntactic or hierarchical information have recently shown improvements over phrase-based approaches (Chiang, 2010), our initial investigation with syntactically driven approaches showed poorer performance on the text simplification task and were less robust to noise in the training data.

Both English Wikipedia and Simple English Wikipedia have received recent analysis as a possible corpus by for both sentence compression and simplification. Yamangil and Nelken (2008) examine the history logs of English Wikipedia to learn sentence compression rules. Yatskar et al. (2010) learn a set of candidate phrase simplification rules based on edit changes identified in both Wikipedias revision histories, though they only provide a list of the top phrasal rules and do not utilize them in an end-to-end simplification system. Napoles and Dredze (2010) provide an analysis of the differences between documents in English Wikipedia and Simple English Wikipedia, though they do not view the

data set as a parallel corpus.

## 3 Text Simplification Corpus

Few data sets exist for text simplification and data sets for the related task of sentence compression are small, containing no more than a few thousand aligned sentence pairs (Knight and Marcu, 2002; Cohn and Lapata, 2009; Nomoto, 2009). For this paper, we utilized a sentence-aligned corpus generated by aligning English Wikipedia with Simple English Wikipedia resulting in 137K aligned sentence pairs. This data set is larger than any previously examined for sentence simplification and orders of magnitude larger than those previously examined for sentence compression.

We give a brief overview of the corpus generation process here. For more details and an analysis of the data set, see (Coster and Kauchak, 2011). Throughout this article we will refer to English Wikipedia articles/sentences as **normal** and Simple English Wikipedia articles as **simple**.

We aligned the normal and simple articles at the document level based on exact match of the title and then removed all article pairs that were stubs, disambiguation pages, meta-pages or only contained a single line. Following a similar approach to previous monolingual alignment techniques (Barzilay and Elhadad, 2003; Nelken and Shieber, 2006), we then aligned each simple paragraph to any normal paragraph that had a normalized TF-IDF cosine similarity above a set threshold. These aligned paragraphs were then aligned at the sentence level using a dynamic programming approach, picking the best sentence-level alignment from a combination of the following sentence-level alignments:

- normal sentence inserted
- normal sentence deleted
- one normal sentence to one simple sentence
- two normal sentences to one simple sentence
- one normal sentence to two simple sentence

Following Nelken and Shieber (2006), we used TF-IDF cosine similarity to measure the similarity between aligned sentences and only kept aligned sentence pairs with a similarity threshold above 0.5. We

found this thresholding approach to be more intuitive than trying to adjust a skip (insertion or deletion) penalty, which has also been proposed (Barzilay and Elhadad, 2003).

## 4 Simplification Model

Given training data consisting of aligned normal-simple sentence pairs, we aim to produce a translation system that takes as input a normal English sentence and produces a simplified version of that sentence. Motivated by the large number and importance of lexical changes in the data set, we chose to use a statistical phrase-based translation system. We utilized a modified version of Moses, which was originally developed for machine translation (Koehn et al., 2007).

Moses employs a log-linear model, which can be viewed as an extension of the noisy channel model and combines a phrase-based translation model, an n-gram language model, as well as a number of other models/feature functions to identify the best translation/simplification. The key component of Moses is the phrase-based translation model which decomposes the probability calculation of a normal sentence simplifying to a simple sentence as the product of individual phrase translations:

$$p(\text{simple}|\text{normal}) = \prod_{i=1}^{m} p(\bar{s}_i|\bar{n}_i)$$

where each $\bar{s}_i$ is a phrase (one or more contiguous words) in the simple sentence and $\bar{s}_1, \bar{s}_2, ..., \bar{s}_m$ exactly cover the simple sentence. $\bar{n}_i$ are similarly defined over the normal sentence. $p(\bar{s}_i|\bar{n}_i)$ denotes the probability of a normal phrase being translated/simplified to the corresponding simplified phrase. These phrasal probabilities are extracted from the sentence pairs based on an EM-learned word alignment using GIZA++ (Och and Ney, 2000).

Phrase-based models in machine translation often require that both phrases in the phrasal probabilities contain one or more words, since phrasal deletion/insertion is rare and can complicate the decoding process. For text simplification, however, phrasal deletion commonly occurs: 47% of the sentence pairs contain deletions (Coster and Kauchak,

2011). To model this deletion, we relax the restriction that the simple phrase must be non-empty and include in the translation model probabilistic phrasal deletion rules of the form $p(NULL|\bar{n}_i)$ allowing for phrases to be deleted during simplification.

To learn these phrasal deletions within Moses, we modify the original word alignment output from GIZA++ before learning the phrase table entries in two ways:

1. If one or more contiguous normal words are unaligned in the original alignment, we align them to NULL appropriately inserted on the simple side

2. If a set of normal words $N$ all align to a single simple word $s$ and there exists an $n \in N$ where $n = s$ then for all $n' \in N : n' \neq n$ we align them to NULL.

This second modification has two main benefits. Frequently, if a word occurs in both the normal and simple sentence and it is aligned to itself, no other words should be aligned to that word. As others have noted, this type of spurious alignment is particularly prevalent with function words, which tend to occur in many different contexts (Chen et al., 2009). Second, even in situations where it may be appropriate for multiple words to align to a single word (for example, in compound nouns, such as *President Obama → Obama*), removing the alignment of the extra words, allows us to delete those words in other contexts. We lose some specificity with this adaptation because some deletions can now occur independent of context, however, empirically this modification provides more benefit than hindrance for the model. We conjecture that the language model helps avoid these problematic cases.

Table 2 shows excerpts from an example sentence pair before the alignment alteration and after. In the original alignment ", *aka Rodi*" is unaligned. After the alignment processing, the unaligned phrase is mapped to NULL allowing for the possibility of learning a phrasal deletion entry in the phrase table. We also modified the decoder to appropriately handle NULL mappings during the translation process.

Table 3 shows a sample of the phrasal deletion rules learned. These rules and probabilities were learned by the original phrase-table generation code

| | |
|---|---|
| Normal: | Sergio Rodriguez Garcia , *aka Rodri ,* is a spanish   footballer ... |
| Simple: | Sergio Rodriguez Garcia                    is a spanish football player ... |
| Modified Simple: | Sergio Rodriguez Garcia     NULL     is a spanish football player ... |

Table 2: Example output from the alignment modification step to capture phrasal deletion. Words that are vertically aligned are aligned in the word alignment.

| Phrase-table entry | | | prob |
|---|---|---|---|
| , | $\rightarrow$ | NULL | 0.057 |
| the | $\rightarrow$ | NULL | 0.033 |
| of the | $\rightarrow$ | NULL | 0.0015 |
| or | $\rightarrow$ | NULL | 0.0014 |
| however , | $\rightarrow$ | NULL | 0.00095 |
| the city of | $\rightarrow$ | NULL | 0.00034 |
| generally | $\rightarrow$ | NULL | 0.00033 |
| approximately | $\rightarrow$ | NULL | 0.00025 |
| , however , | $\rightarrow$ | NULL | 0.00022 |
| , etc | $\rightarrow$ | NULL | 0.00013 |

Table 3: Example phrase-table entries learned from the data and their associated probability.

of Moses after the word alignment was modified. The highest probability rules tend to delete punctuation and function words, however, other phrases also appeared. 0.5% of the rules learned during training are deletion rules.

## 5   Experiments

We compared five different approaches on the text simplification task:

**none:**   Does no simplification. Outputs the normal, unsimplified sentence.

**K & M:**   Noisy-channel sentence compression system described in Knight and Marcu (2002).

**T3:**   Synchronous tree substitution grammar, trained discriminatively (Cohn and Lapata, 2009).

**Moses:**   Phrase-based, machine translation approach (Koehn et al., 2007).

**Moses+Del:**   Our approach described in Section 4 which is a phrase-based approach with the addition of phrasal deletion.

From the aligned data set of 137K sentence pairs, we used 124K for training and 1,373 for testing

with the remaining 12K sentences used during development. We trained the n-gram language model used by the last four systems on the simple side of the training data.[3]  T3 requires parsed data which we generated using the Stanford parser (Klein and Manning, 2003). Both Moses and Moses+Del were trained using the default Moses parameters and we used the last 515 sentence pairs from the training set to optimize the hyper-parameters of the log-linear model for both Moses variants. T3 was run with the default parameters.

Due to runtime and memory issues, we were unable to run T3 on the full data set.[4]  We therefore present results for T3 trained on the largest training set that completed successfully, the first 30K sentence pairs. This still represents a significantly larger training set than T3 has been run on previously. For comparison, we also provide results below for Moses+Del trained on the same 30K sentences.

### 5.1   Evaluation

Since there is no standard way of evaluating text simplification, we provide results for three different automatic methods, all of which compare the system's output to a reference simplification. We used BLEU (Papineni et al., 2002), which is the weighted mean of n-gram precisions with a penalty for brevity. It has been used extensively in machine translation and has been shown to correlate well with human performance judgements.

We also adopt two automatic measures that have been used to evaluate text compression that compare the system's output to a reference translation

---

[3]See (Turner and Charniak, 2005) for a discussion of problems that can occur for text compression when using a language model trained on data from the uncompressed side.

[4]On 30K sentences T3 took 4 days to train. On the full data set, we ran T3 for a week and at that point the discriminative training was using over 100GB of memory and we terminated the run.

| System | BLEU | word-F1 | SSA |
|---|---|---|---|
| none | 0.5937 | 0.5967 | 0.6179 |
| K & M | 0.4352 | 0.4352 | 0.4871 |
| T3* | 0.2437 | 0.2190 | 0.3651 |
| Moses | 0.5987 | 0.6076 | 0.6224 |
| Moses+Del | 0.6046 | 0.6149 | 0.6259 |

Table 4: Performance of the five approaches on the test data. All differences in performance are statistically significant. * - T3 was only trained on 30K sentence pairs for performance reasons.

(Clarke and Lapata, 2006): simple string accuracy measure (a normalized version of edit distance, abbreviated SSA) and F1 score calculated over words. We calculated F1 over words instead of grammatical relations (subject, direct/indirect object, etc.) since finding the relation correspondence between the system output and the reference is a non-trivial task for simplification data where reordering, insertions and lexical changes can occur. Clarke and Lapata (2006) showed a moderate correlation with human judgement for SSA and a strong correlation for the F1 measure.

To measure whether the difference between system performance is statistically significant, we use bootstrap resampling with 100 samples with the t-test (Koehn, 2004).

### 5.2 Results

Table 4 shows the results on the test set for the different evaluation measures. All three of the evaluation metrics rank the five systems in the same order with Moses+Del performing best. All differences between the systems are statistically significant for all metrics at the $p = 0.01$ level. One of the challenges for the sentence simplification problem is that, like sentence compression, not making any changes to the system produces reasonable results (contrast this with machine translation). In the test set, 30% of the simple sentences were the same as the corresponding normal sentence. Because of this, we see that not making any changes (*none*) performs fairly well. It is, however, important to leave these sentences in the test set, since not all sentences need simplification and systems should be able to handle these sentences appropriately.

Both of the text compression systems perform poorly on the text simplification task with results that are significantly worse than doing nothing. Both of these systems tended to bias towards modifying the sentences (T3 modified 77% of the sentences and K & M 96%). For K & M, the poor results are not surprising since the model only allows for deletion operations and is more tailored to the compression task. Although T3 does allow for the full range of simplification operations, it was often overly aggressive about deletion, for example T3 simplified:

> *There was also a proposal for an extension from Victoria to Fulham Broadway station on the district line , but this was not included in the bill .*

to "*it included .*" Overall, the output of T3 averaged 13 words per sentence, which is significantly lower than the gold standard's 21 words per sentence. T3 also suffered to a lesser extent from inappropriately inserting words/phrases, which other researchers have also noted (Nomoto, 2009). Some of these issues were a results of T3's inability to cope with noise in the test data, both in the text or the parses.

Both Moses and Moses+Del perform better than the text compression systems as well as the baseline system, *none*. If we remove those sentences in the test set where the simple sentence is the same as the normal sentence and only examine those sentences where a simplification should occur, the difference between the phrase-based approaches and *none* is even more significant with BLEU scores of 0.4560, 0.4723 and 0.4752, for *none*, Moses and Moses+Del respectively.

If we compare Moses and Moses+Del, the addition of phrasal deletion results in a statistically significant improvement. The phrasal deletion was a common operation in the simplifications made by Moses+Del; in 8.5% of the test sentences, Moses+Del deleted at least one phrase. To better understand this performance difference, Table 5 shows the BLEU scores for sentences where each respective system made a change (i.e. the output simplification is different than the input). In both cases, when the systems make simplifications on sentences that should be simplified, we see large gains in the output over doing nothing. While Moses improves over the baseline of doing nothing by 0.047 BLEU,

| | | BLEU | |
|---|---|---|---|
| System | Case | *none* | output |
| Moses | correct change | 0.4431 | 0.4901 |
| | incorrect change | 1 | 0.8625 |
| Moses+Del | correct change | 0.4087 | 0.4788 |
| | incorrect change | 1 | 0.8706 |

Table 5: BLEU scores for Moses and Moses+Del on sentences where the system made a change. "correct change" shows the score where a change was made by the system as well as in the reference and "incorrect change" where a change was made by the system, but not the reference.

we see an even larger gain by Moses+Del with a difference of 0.07 BLEU.

For completeness, we also trained Moses+Del on the same 30K sentences used to train the T3 system.[5] Using this training data, Moses+Del achieved a BLEU score of 0.5952. This is less than the score achieved when using the full training data, but is significantly better than T3 and still represents a small improvement over *none*.

Table 6 shows example simplifications made by Moses+Del. In many of the examples we see phrasal deletion during the simplification process. The output also contains a number of reasonable lexical changes, for example in *a*, *d* and *e*. Example *b* contains reordering and *e* shows an example of a split being performed where the normal sentence is turned into two simplified sentences. This is not uncommon in the data, but can be challenging to model for current syntactic approaches. The examples also highlight some of the common issues with the approach. Examples *a* and *f* are not grammatically correct and the simplification in *f* does not preserve the original meaning of the text. As an aside, the normal sentence of example *d* also contains an omission error following "as" due to preprocessing of the data, resulting from ill-formed xml in the articles.

### 5.3 Oracle

In the previous section, we looked at the performance of the systems based on the best translations suggested by the systems. For many approaches, we can also generate an n-best list of possible translations. We examined the simplifications in this n-

---

| | BLEU | |
|---|---|---|
| System | original | oracle |
| Moses | 0.5987 | 0.6317 |
| Moses+Del | 0.6046 | 0.6421 |

Table 7: BLEU score for the original system versus the best possible "oracle" translations generated by greedily selecting the best translation from an n-best list based on the reference simplification.

best list to measure the potential benefit of reranking techniques, which have proved successful in many NLP applications (Och et al., 2004; Ge and Mooney, 2006), and to understand how well the underlying model captures the phenomena exhibited in the data. For both of the phrase-based approaches, we generated an n-best list of size 1000 for each sentence in the test set. Using these n-best lists, we generated an "oracle" simplification of the test set by greedily selecting for each test sentence the simplification in the n-best list with the best sentence-level BLEU score.

Table 7 shows the BLEU scores for the original system output and the system's oracle output. In all cases, there is a large difference between the system's current output and the oracle output, suggesting that utilizing some reranking technique could be useful. Also, we again see the benefit of the phrasal deletion rules. The addition of the phrasal deletion rule gives the system an additional dimension of flexibility, resulting in a more varied n-best list and an overall higher oracle BLEU score.

## 6 Conclusions and Future Work

In this paper, we have explored a variety of approaches for learning to simplify sentences from Wikipedia. In contrast to prior work in the related field of sentence compression where deletion plays the dominant role, the simplification task we examined has the full range of text-to-text operations including lexical changes, reordering, insertions and deletions.

We implemented a modified phrase-based simplification approach that incorporates phrasal deletion. Our approach performs significantly better than two different text compression approaches, including T3, and better than previous approaches on a similar data set (Zhu et al., 2010). We also showed

| | | |
|---|---|---|
| a. | normal: | *Critical reception* for The Wild has been negative. |
| | simplified: | *Reviews* for The Wild has been negative. |
| b. | normal: | Bauska is a town in Bauska county , in the *Zemgale* region of *southern Latvia* . |
| | simplified: | Bauska is a town in Bauska county , in the region of *Zemgale* . |
| c. | normal: | LaBalme is a commune *in the Ain department in eastern France* . |
| | simplified: | LaBalme is a commune . |
| d. | normal: | Shadow of the Colossus *, released in Japan as ,* is a Japanese-developed action-adventure video game *developed and published* by Sony computer entertainment for the Playstation 2. |
| | simplified: | Shadow of the Colossus is a Japanese-developed action-adventure video game *made* by Sony computer entertainment for the Playstation 2. |
| e. | normal: | Nicolas Anelka is a French *footballer who currently* plays as a striker for Chelsea *in the English premier league* . |
| | simplified: | Nicolas Anelka is a French *football player* . *He* plays for Chelsea . |
| f. | normal: | *Each* edge of *a tesseract is* of the same length. |
| | simplified: | *Same* edge of the same length. |

Table 6: Example simplifications. "normal" is the the unsimplified input sentence and "simplified" the simplification made by Moses+Del.

that the incorporation of phrasal deletion into the simplification process results in statistically significant improvements over a traditional phrase-based approach.

While we obtained positive results using a phrase-based approach, we still believe that incorporating some additional hierarchical structure will help the simplification process, particularly since one of the goals of simplification is to reduce the grammatical complexity of the sentence. Also, as seen in some of the examples above, the phrase-based model can produce output that is not grammatically correct. Though T3 did not perform well, many other syntax-based models exists that have been successful in machine translation.

There are a number of research questions motivated by this work in related areas including the scalability of discriminative trained rule sets, the impact of the language model training source (simple vs. normal English), document-level simplification and applications of text simplification. Our hope is that this new simplification task will spur a variety of related research inquiries.

## Acknowledgments

## References

Regina Barzilay and Noemie Elhadad. 2003. Sentence alignment for monolingual comparable corpora. In *Proceedings of EMNLP*.

Yvonne Canning, John Tait, Jackie Archibald, and Ros Crawley. 2000. Cohesive generation of syntactically simplified newspaper text. In *Proceedings of TSD*.

John Carroll, Gido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. 1998. Practical simplification of English newspaper text to assist aphasic readers. In *Proceedings of AAAI Workshop on Integrating AI and Assistive Technology*.

Raman Chandrasekar and Bangalore Srinivas. 1997. Automatic induction of rules for text simplification. In *Knowledge Based Systems*.

Yu Chen, Martin Kay, and Andreas Eisele. 2009. Intersecting multilingual data for faster and better statistical translations. In *Proceedings of HLT/NAACL*.

David Chiang. 2010. Learning to translate with source and target syntax. In *Proceedings of ACL*.

James Clarke and Mirella Lapata. 2006. Models for sentence compression: A comparison across domains, training requirements and evaluation measures. In *Proceedings of ACL*.

Trevor Cohn and Mirella Lapata. 2009. Sentence compression as tree transduction. *Journal of Artificial Intelligence Research*.

Will Coster and David Kauchak. 2011. Simple English Wikipedia: A new simplification task. In *Proceedings of ACL (Short Paper)*.

Lijun Feng. 2008. Text simplification: A survey. CUNY Technical Report.

Michel Galley and Kathleen McKeown. 2007. Lexicalized Markov grammars for sentence compression. In *Proceedings of HLT/NAACL*.

Ruifang Ge and Raymond Mooney. 2006. Discriminative reranking for semantic parsing. In *Proceedings of COLING*.

Dan Klein and Christopher Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of ACL*.

Kevin Knight and Daniel Marcu. 2002. Summarization beyond sentence extraction: A probabilistic approach to sentence compression. *Artificial Intelligence*.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of ACL*.

Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proceedings of EMNLP*.

Ryan McDonald. 2006. Discriminative sentence compression with soft syntactic evidence. In *Proceedings of EACL*.

Makoto Miwa, Rune Saetre, Yusuke Miyao, and Jun'ichi Tsujii. 2010. Entity-focused sentence simplification for relation extraction. In *Proceedings of COLING*.

Courtney Napoles and Mark Dredze. 2010. Learning simple Wikipedia: A cogitation in ascertaining abecedarian language. In *Proceedings of HLT/NAACL Workshop on Computation Linguistics and Writing*.

Rani Nelken and Stuart Shieber. 2006. Towards robust context-sensitive sentence alignment for monolingual corpora. In *Proceedings of AMTA*.

Tadashi Nomoto. 2007. Discriminative sentence compression with conditional random fields. In *Information Processing and Management*.

Tadashi Nomoto. 2008. A generic sentence trimmer with CRFs. In *Proceedings of HLT/NAACL*.

Tadashi Nomoto. 2009. A comparison of model free versus model intensive approaches to sentence compression. In *Proceedings of EMNLP*.

F.J. Och and H. Ney. 2000. Improved statistical alignment models. In *Proceedings of ACL*.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.

Franz Josef Och, Kenji Yamada, Stanford U, Alex Fraser, Daniel Gildea, and Viren Jain. 2004. A smorgasbord of features for statistical machine translation. In *Proceedings of HLT/NAACL*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of ACL*.

Jenine Turner and Eugene Charniak. 2005. Supervised and unsupervised learning for sentence compression. In *Proceedings of ACL*.

David Vickrey and Daphne Koller. 2008. Sentence simplification for semantic role labeling. In *Proceedings of ACL*.

S. Wubben, A. van den Bosch, and E. Krahmer. 2010. Paraphrase generation as monolingual translation: Data and evaluation. In *Proceedings of the International Workshop on Natural Language Generation*.

Elif Yamangil and Rani Nelken. 2008. Mining Wikipedia revision histories for improving sentence compression. In *ACL*.

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 HLT/NAACL (Short Paper)*.

Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych. 2010. A monolingual tree-based translation model for sentence simplification. In *Proceedings of COLING*.