

Simplifying metaphorical language for young readers: A corpus study on news text

Magdalena Wolska

Eberhard Karls Universität Tübingen
LEAD Graduate School
magdalena.wolska@uni-tuebingen.de

Yulia Clausen

Ruprecht-Karls-Universität Heidelberg
Computational Linguistics
yulia.clausen@gmail.com

Abstract

The paper presents first results of an ongoing project on text simplification focusing on linguistic metaphors. Based on an analysis of a parallel corpus of news text professionally simplified for different grade levels, we identify six types of simplification choices falling into two broad categories: preserving metaphors or dropping them. An annotation study on almost 300 source sentences with metaphors (grade level 12) and their simplified counterparts (grade 4) is conducted. The results show that most metaphors are preserved and when they are dropped, the semantic content tends to be preserved rather than dropped, however, it is reworded without metaphorical language. In general, some of the expected tendencies in complexity reduction, measured with psycholinguistic variables linked to metaphor comprehension, are observed, suggesting good prospect for machine learning-based metaphor simplification.

1 Motivation and problem statement

Text simplification is the process of meaning preserving reduction of discourse complexity whose purpose is to adapt text for specific populations of readers, for instance, children or language learners. The idea has been around since “My Weekly Reader” in the 1920s and Palmer’s work (1932) and over the past 20 years has attracted attention of the computational linguistics community. While broadly interpreted “lexical simplification” – in general understood as substitution of “difficult” words with “simpler” ones – is a common component of automated simplification systems (see, for instance, (Siddharthan, 2014)), studies of text sim-

plification dedicated to specific lexis-related semantic phenomena are lacking. One class of such understudied phenomena are those related to figurative language; a surprising gap in the simplification research considering that metaphors have been shown to cause difficulties in text comprehension and that developing metaphor interpretation competence is a complex developmental process (for an overview, see, for instance, (Winner, 1997)). Since automated systems are trained on corpora of simplified text, understanding patterns of metaphor simplification based on corpus data could help improve simplification models.

In this paper we present a study that is our first step in this direction. We analyze linguistic metaphors in a corpus of news texts professionally simplified for different grade levels. While editors’ guidelines instructed to avoid vivid metaphors, such as “paint into a corner”, our goal was to find out whether, and if so, how, linguistic metaphors in general are simplified by professional editors. Since ultimately we want to build automated metaphor simplification models, the purpose of this study is to investigate whether metaphors in a corpus of professionally simplified text, that is, potential training data, are simplified in systematic ways. Specifically, we were interested in two questions: 1) What types of discourse modifications do editors perform when simplifying metaphorical language? (in other words, whether a well-defined set of classes for the metaphor simplification task can be specified). 2) Do professional editors simplify metaphor phenomena in systematic ways? (if not, training simplification models using machine learning based on corpus data may not be promising).

The paper’s structure follows the data-driven methodology adopted for this study: We first define the criteria used to identify the phenomenon in question: linguistic metaphor. Next, we present

the setup of an annotation study and a typology of simplification choices derived based on an analysis of a corpus of simplified news text. Finally, we present results of an exploratory analysis of the annotated data.

2 Data

2.1 The source corpus

Our data comes from Newsela,¹ a company producing professionally simplified news articles in English and Spanish intended for classroom use. Each Newsela article is available at 5 reading levels spanning grades 2 through 12 of the US school system (elementary school (grades K-4), middle school (grades 5-8), and high school (grades 9-12)). Two levels were used for this first study: the source articles (we will refer to this version as V0) and the most simplified version (V4), since between these versions we expect to see most differences.²

Documents were sampled from a subset of Newsela compiled by Xu et al. This is a parallel corpus of 1130 documents from the English portion of Newsela where each article has been automatically aligned sentence-wise with the four simplified versions using Jaccard similarity; for details on the aligned corpus see (Xu et al., 2015).

2.2 Sample selection

The sample of V0 (source) and V4 (simplified) sentences was drawn from the Xu et al.’s corpus as follows: As shown in Figure 1, different Newsela versions span multiple unevenly distributed grade levels. In order to avoid effects due to differences between grade levels within versions, from V0 only articles at grade level 12 were used and from V4 only articles at grade level 4 (the largest subsets). One sentence from each V0 document was selected with its corresponding V4 sentence(s); only sentences that were not identical between V0 and V4 were included in the sample. Sampling was randomized across all documents to avoid effects due to specific editors’ decisions. This resulted in 582 V0 sentences. Automatic sentence alignments between the versions were manually checked and corrected where necessary; for instance, unaligned V4 sentences were linked appropriately, as in the following example (“i” marks

¹<https://newsela.com>

²Analysis of metaphor simplification across other levels is planned as further work.

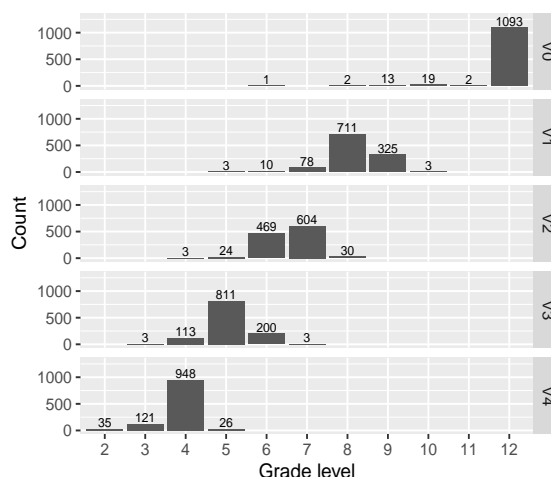


Figure 1: Distribution of grade levels by version in the (Xu et al., 2015) corpus³

the inserted segment):

- V0 Parts of the nation experienced severe but not unprecedented drought during the study, the researchers noted, which might have reduced the amount of rain sustaining their wetlands and ponds
- V4-i Parts of the nation had very little snow or rain while the study was going on.
- V4 That might have meant that there was less water in the wetlands and ponds where amphibians live.

The resulting corpus comprises 582 V0 sentences and their V4 counterparts correctly aligned; 267 alignments have been manually corrected.

2.3 Metaphor identification

We identify linguistic metaphors using Steen et al.’s (2010) refined Metaphor Identification Procedure known as MIPVU.⁴ MIPVU provides guidelines for annotation of potentially metaphorical words, where “words” are linguistic units which receive a separate part-of-speech tag. Phrasal verbs, compounds, and proper names (multiword expressions) can be treated as lexical units as exceptions. For the simplification study we focus on the most common classes of content words: nouns and verbs.

In MIPVU, a lexical unit is considered to be metaphorically used when its meaning in a given context can be contrasted as well as understood in comparison with a more basic meaning that it can have in other contexts. MIPVU strives not to determine the most basic meaning of a word,

³Both plots were created using R’s (R Core Team, 2017) ggplot2 package (Wickham, 2009)

⁴“VU” stands for Vrije Universiteit Amsterdam where the authors of the procedure are based.

	Simplification	Source sentence	Simplified sentence
Preserved	same metaphor	...like the magnetized nails, unable to resist a powerful magnetic force in the galactic bulge ...	Like the magnetized nails, they would have been unable to resist a powerful magnetic force in the galactic bulge ...
	other metaphor	Obama also has grappled publicly with reconciling King's teachings on nonviolence ...	Obama has wrestled publicly with living up to King's teachings on nonviolence ...
	phrase with metaphor(s)	But now she's struggling to obtain documents required by the new law.	But now she's having a hard time getting the papers that the new law requires.
Dropped	content	Our goal is to provide Internet service to people in areas that can't afford to throw down fiber lines ...	Our goal is to provide Internet service to people in areas that can't afford \emptyset usual Internet lines ...
	dropped		
	changed to non-metaphor	In exchange for a 4 percent piece of their companies, entrepreneurs in the program will gain access people in the program will give up a 4 percent share of their companies. In exchange they will get ...
	phrase without metaphor(s)	Utah officials say that since 2008, highway crashes have dropped annually on stretches of rural Interstate ...	They say there have been fewer accidents where the speed limit was raised.

Table 1: Metaphor simplification types.

but rather a meaning that is more basic than the one in the given context. A more basic sense is defined as a “more concrete, specific, and human-oriented sense in the contemporary language use” (Steen et al., 2010, p. 35). A corpus-based dictionary, here: the Macmillan English Dictionary for Advanced Learners,⁵ is consulted for the basic and the contextual senses of lexical units. Two senses of one lexical unit are considered significantly distinct if they are listed under separate numbers in the dictionary. MIPVU defines three metaphor types: indirect (example (1)), direct (2) and implicit (3):

Indirect metaphors occur when contrast as well as comparison exists between the contextual and a more basic meaning:

- (1) *Political cartoons engage and enrage more than articles do because they are visual and transcend language barriers.*

Direct metaphors display no contrast between the contextual and a more basic meaning. In this case contextual meaning is the basic meaning and comparison is expressed explicitly, for instance, by the so-called metaphor flags (words such as *like*, *as*, *so-called*, *-shaped*):

- (2) *Like the magnetized nails, they would have been unable to resist a powerful magnetic force in the galactic bulge ...*

Implicit metaphors represent words pointing back to recoverable metaphorical material:

- (3) *...unable to resist a powerful magnetic force in the galactic bulge around when it was forming stars around 8-13 billion years ago.*

⁵<http://www.macmillandictionary.com>

Measure	Count
No. of sentences containing metaphors	272
No. of metaphor occurrences	416
Verbs	267
Nouns	149
Mean No. of metaphors per sentence	1.53
No. of unique lexemes	314

Table 2: Quantitative information on the annotated metaphor set

In the present study we focus on indirect metaphors (the prevalent type; see (Steen et al., 2010)) and identify metaphorical uses of all nouns and verbs in the sampled original sentences (V0).

Metaphor annotation proceeded as follows: Identification of candidate metaphor occurrences was carried out by one of the authors. All unclear cases were marked and discussed by both authors until agreement was reached. If agreement could not be reached, the case was excluded from further analysis. The final set of metaphorical word uses comprises only clear cases as per MIPVU.⁶

Quantitative information on the annotated metaphors is summarized in Table 2.

⁶As one of the reviewers pointed out, MIPVU is not an easy protocol to apply. It is precisely for this reason that, since the focus of the present study was not on the metaphor identification task, but on simplification types, we opted to select only clear cases of metaphors in V0, as per agreement on metaphor status by both authors. Because this agreement was reached though a discussion in all cases, inter-annotator agreement on metaphor annotation was not calculated. We are planning to conduct a separate metaphor identification study as part of further work.

3 Simplification types

Identification and annotation of simplification types proceeded as follows: Both authors initially analyzed and discussed smaller subsets of the metaphor-annotated corpus (20-30 instances). Once the set of types stabilized to the final set (below) one author annotated all the remaining instances and the other author 99 instances in total (1 erroneous instance had to be excluded). Both annotators are non-native, but fluent, speakers of English. Inter-annotator agreement on the common subset of 99 instances was 0.93 proportion agreement ($\kappa=0.87$) and was deemed reliable. The 7 disagreement cases in the common subset were discussed and resolved for the analysis.⁷

A typology of editors' simplification choices was derived in a data-driven fashion starting off of two basic options: a metaphor can be **preserved** in the simplified version or **dropped**. Corpus analysis revealed three subtypes of metaphor-related discourse modifications within each of these high-level categories: A metaphorically used word can be preserved unchanged (**same metaphor**), replaced with another single word used metaphorically (**other metaphor**), or reworded using multi-word phrasing containing metaphor(s) (**phrase with metaphor(s)**). It can be dropped by replacing it with a single different word in a more basic sense (**changed to non-metaphor**), with multi-word phrasing not containing metaphors (**phrase without metaphor(s)**), or the meaning portion expressed by the metaphor can be omitted altogether (**content dropped**). Table 1 provides a summary of the simplification types with examples.

4 Corpus analysis

Analysis of the annotated simplification types is split into two parts: We start with a high-level overview of the distribution of the simplification types. Then, we perform an exploratory analysis to investigate how four psycholinguistic variables – age of acquisition (AoA), familiarity, concreteness, and imageability –, previously linked to metaphor comprehension (see, for instance, (Paivio et al., 1968; Paivio and Walsh, 1993; Gibbs, 2006; Ureña and Faber, 2010)) and also used in simplification models (e.g. (Cross-

⁷Further annotation will be conducted and inter-annotator agreement recalculated for categories other than the prevalent Preserved.same in order to reduce class imbalance and to add instances of the smaller classes.

Simplification type	Count
Preserved	288
same metaphor	240
other metaphor	34
phrase with metaphor(s)	14
Dropped	128
changed to non-metaphor	81
content dropped	37
phrase without metaphor	10

Table 3: Distribution of simplification types

ley et al., 2007; Jauhar and Specia, 2012; Crossley et al., 2012; Vajjala and Meurers, 2014)), behave across simplification categories. The scores have been extracted from the MRC Psycholinguistic Database (Wilson, 1988) and the Bristol Norms (Kuperman et al., 2012).

Distribution of simplification types is shown in Table 3. Most metaphors, 69%, are preserved in V4, the majority with the same wording. Where metaphorical words are omitted, they tend to be replaced with their literal counterparts. Rewording consisting of longer phrases is dispreferred.

Distributions of psycholinguistic variables are shown in Figure 2. Since for the automated classification task the class imbalance (see Table 3) will need to be countered, we reduce the class imbalance already for distributions visualization by randomly downsampling the largest class (Preserved.same) to 80 instances such that the two basic classes, Preserved and Dropped, are of the same size; Preserved.same mean was estimated by randomly resampling the 80 instances 20 times. Dependent variables are ordered by complexity of intervention into the source semantics that the manipulation they denote involves; for the Preserved class: preserving same meaning at one end vs. paraphrasing by adding lexical material at the other and for the Dropped class: merely replacing the metaphorical lexeme with a non-metaphorical one vs. omitting content altogether.

Within the Preserved type, low-AoA metaphors tend to be preserved and high-AoA are rephrased. In the Dropped class, low-AoA metaphors are rephrased and high are dropped; also on average, as expected, lower AoA metaphors are preserved and higher dropped. Explicable pattern of Imageability can be observed: within both basic types, the higher the score, the more radical

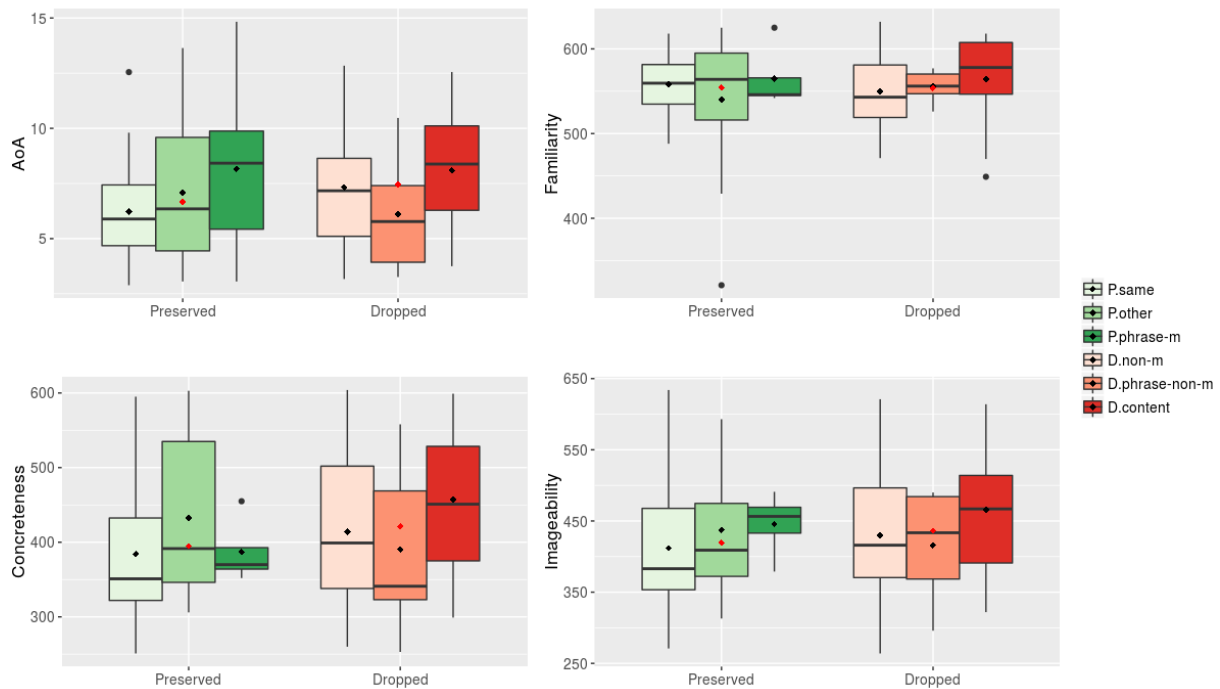


Figure 2: Distribution of psycholinguistic variables by simplification type (legend labels shortened for space reasons; the red dots indicate **P**reserved and **D**ropped group means (20 resamplings of P.same))

the modification (rephrasing and dropping content, respectively); aggregated means display the same pattern. This is consistent with the guideline on avoiding vivid metaphors. The pattern of Concreteness measure is unclear. Familiarity scores are the least discriminating.

5 Discussion and further work

Overall, some of the psycholinguistic variables do exhibit patterns confirming systematicity of professional simplification and good prospect for training machine learning models based on professionally simplified data; Xu et al. (2015) argue likewise. AoA and Imageability exhibit a consistent explicable pattern within and between the two basic types suggesting they can be used as predictors. This is not the case with the Familiarity measure. Interestingly, in the Preserved class, lexical elaboration (Preserved.other) is performed within narrow ranges of 3 of the variables, which could be exploited. The high prevalence of the Preserved class is surprising. On the one hand, it provides a safe default for a basic automated system, on the other hand, sets a high majority-based baseline.

Future work will involve investigating further linguistically and cognitively-motivated variables for metaphor simplification. Likewise, interactions between psycholinguistic variables and their

relation to syntactic complexity variables require further study. We also plan to annotate further data, also at other grade levels, and train models (2-way classification, Preserve vs. Drop, in the first instance). Finally, the categorical approach to metaphor simplification might be entirely reconsidered in view of recent studies showing evidence that the literal-metaphorical distinction is a graded (scalar) phenomenon ((Cameron et al., 2009; Müller and Tag, 2010; Dunn, 2014), among others). Simplification might be thus seen as continuous “reduction of metaphority”⁸.

Acknowledgments

We would like to thank Newsela for making their data available to the community and Wei Xu, Chris Callison-Burch, and Courtney Napoles for sharing their aligned corpus via Newsela; we are planning to make our final annotations available in the same way. We would also like to thank the reviewers for the interesting comments. Magdalena Wolska was supported by the Institutional Strategy of the University of Tübingen (Deutsche Forschungsgemeinschaft, ZUK 63). Part of this work was conducted while she was on a visiting appointment at the Ruprecht-Karls-Universität Heidelberg.

⁸This interesting suggestion for follow-up work and pointer to Dunn’s papers is due to one of the reviewers.

References

- L. Cameron, R. Maslen, Z. Todd, J. Maule, P. Stratton, and N. Stanley. 2009. The discourse dynamics approach to metaphor and metaphor-led discourse analysis. *Metaphor and Symbol* doi:10.1080/10926480902830821.
- S.A. Crossley, D. Allen, and D.S. McNamara. 2012. Text simplification and comprehensible input: A case for an intuitive approach. *Language Teaching Research* doi:10.1177/1362168811423456.
- S.A. Crossley, M.M. Louwse, P.M. McCarthy, and D.S. McNamara. 2007. A linguistic analysis of simplified and authentic texts. *The Modern Language Journal* doi:10.1111/j.1540-4781.2007.00507.x.
- J. Dunn. 2014. Measuring metaphoricity. In *Proceedings of the 52nd Annual Meeting of the ACL*. <http://www.aclweb.org/anthology/P14-2121>.
- R.W. Gibbs. 2006. Metaphor interpretation as embodied simulation. *Mind & Language* doi:10.1111/j.1468-0017.2006.00285.x.
- S.K. Jauhar and L. Specia. 2012. PUOW-SHEF: SimpLex—lexical simplicity ranking based on contextual and psycholinguistic features. In *Proceedings of the 1st Joint Conference on Lexical and Computational Semantics (*SEM-12)*. <http://www.aclweb.org/anthology/S12-1066>.
- V. Kuperman, H. Stadthagen-Gonzalez, and M. Brysbaert. 2012. Age-of-acquisition (AoA) norms for over 50 thousand English words. In *Behavior Research Methods*. <http://crr.ugent.be/archives/806>.
- C. Müller and S. Tag. 2010. The dynamics of metaphor: Foregrounding and activating metaphoricity in conversational interaction. *Cognitive Semiotics* doi:10.1515/cogsem.2010.6.spring2010.85.
- A. Paivio and M. Walsh. 1993. Psychological processes in metaphor comprehension and memory. In *Metaphor and thought*. doi:10.1017/CBO9781139173865.016.
- A. Paivio, J.C. Yuille, and S.A. Madigan. 1968. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology* doi:10.1037/h0025327.
- Harold E. Palmer. 1932. *The Grading and Simplifying of Literary Material*. Institute for Research in English Teaching.
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
- A. Siddharthan. 2014. A survey of research on text simplification. *International Journal of Applied Linguistics* doi:10.1075/itl.165.2.06sid.
- G.J. Steen, A.G. Dorst, J.B. Herrmann, A.A. Kaal, T. Krennmayr, and T. Pasma. 2010. *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. John Benjamins.
- J.M. Ureña and P. Faber. 2010. Reviewing imagery in resemblance and non-resemblance metaphors. *Cognitive Linguistics* doi:10.1515/cogl.2010.004.
- S. Vajjala and D. Meurers. 2014. Readability assessment for text simplification: From analysing documents to identifying sentential simplifications. *ITL—International Journal of Applied Linguistics* doi:10.1075/itl.165.2.04vaj.
- H. Wickham. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer. <http://ggplot2.org>.
- M. Wilson. 1988. MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. In *Behavior Research Methods, Instruments, & Computers*. <http://ota.ox.ac.uk/headers/1054.xml>.
- E. Winner. 1997. *The Point of Words*. Harvard University Press.
- W. Xu, C. Callison-Burch, and C. Napoles. 2015. Problems in Current Text Simplification Research: New Data Can Help. In *Transactions of the Association for Computational Linguistics*. <https://www.transacl.org/ojs/index.php/tacl/article/view/549>.