A Preliminary Study of Statistically Predictive Syntactic Complexity Features and Manual Simplifications in Basque

Itziar Gonzalez-Dios, María Jesús Aranzabe, Arantza Díaz de Ilarraza
IXA NLP Group
University of the Basque Country (UPV/EHU)
`itziar.gonzalezd@ehu.eus`

Abstract

In this paper, we present a comparative analysis of statistically predictive syntactic features of complexity and the treatment of these features by humans when simplifying texts. To that end, we have used a list of the most five statistically predictive features obtained automatically and the Corpus of Basque Simplified Texts (CBST) to analyse how the syntactic phenomena in these features have been manually simplified. Our aim is to go beyond the descriptions of operations found in the corpus and relate the multidisciplinary findings to understand text complexity from different points of view. We also present some issues that can be important when analysing linguistic complexity.

1 Introduction and Related Work

Linguistic complexity has gained a lot of attention in the last years from different points of view. In Natural Language Processing (NLP), for example, there are two main research lines that deal with text complexity: readability assessment (RA) and automatic text simplification (ATS). RA seeks to analyse the complexity of the texts to classify them according to a level by means of analysis of linguistic features and machine learning techniques (DuBay, 2004; Benjamin, 2012; Zamanian and Heydari, 2012). On the other hand, ATS deals with the complexity of the texts in order to give a more accessible and simple equivalent version by reducing its lexical and syntactic complexity (Shardlow, 2014; Siddharthan, 2014). In the last years, parallel corpora of original and simplified texts have also been created in order to study the phenomena considered as complex and the changes performed when simplifying (Caseli et al., 2009; Bott and Saggion, 2014; Brunato et al., 2015).

In Basque, linguistic complexity has been studied from both neurolinguistic and computational points of view. From a neurolinguistic point of view, the internal word ordering (Laka and Erdozia, 2010), the relative clauses (Carreiras et al., 2010), and the effect of phrasal length on sentence word order (Ros et al., 2015) have been analysed. From the computational point of view, text complexity has been studied for readability assessment (Gonzalez-Dios et al., 2014) or automatic text simplification (Aranzabe et al., 2012).

In this paper, from the computational point of view, we want to see if the syntactic features that are statistically predictive of readability have been considered by humans when simplifying texts and how they have been treated. To that end, we analyse the phenomena involved in the most five predictive syntactic features (the ratios of conditional, concessive, purpose, temporal and relative clauses) in the Corpus of Basque Simplified Texts (CBST) (Gonzalez-Dios, 2016) in the two simplification approaches contained (structural and intuitive). This way, we want to make a linguistic analysis of the simplified texts beyond the description of the corpus and relate them to computational and neurolinguistic studies when possible. So, we want to start understanding the grounds of linguistic complexity from a multidisciplinary point of view.

This paper is structured as follows: in Section 2 we present the resources and the methodology we have followed in this analysis and in Section 3 we show the treatment of the of the phenomena in the

corpus CBST. In Section 4 we show how the phenomena have been simplified and in Section 5 we expose some issues that are necessary to keep on with the analysis of text complexity. Finally, we conclude and outline the future work in Section 6.

2   Resources and Experimental Set Up

To analyse wheater the syntactic phenomena considered in the statistically predictive features have been treated by humans, we have used two resources: a list of predictive features and a corpus of simplified texts. To get the list of predictive features, we have based on the experiments carried out to assess the readability of Basque texts (Gonzalez-Dios et al., 2014). In these experiments, in addition to create a model to classify texts as simple or complex, a list with most predictive features or linguistic ratios was created using the data mining software *Weka* (Hall et al., 2009) and the rankers Chi Square Ranker and InfoGain Ranker. The features in the list are classified according to their linguistic type: global, lexical, morphological, morpho-syntactic, syntactic and pragmatic.

To train the readability assessment model and get the list of predictive features, two corpora of scientific popularisation domain were used: 1) the *Elhuyar (T-comp)* corpus, which contains 200 texts for adults which, was considered as a complex corpus and 2) the *Zernola (T-simp)* corpus, containing 200 texts for children, was considered as simple. These corpora were analysed with the framework for the automatic analysis of Basque (Aduriz and Díaz de Ilarraza, 2003), the sentence and clause boundary detector (Aranzabe et al., 2013) and the apposition detector (Gonzalez-Dios et al., 2013).

The second resource is the Corpus of Basque Simplified Texts (CBST) (Gonzalez-Dios, 2016). This corpus compiles 227 original sentences[1] of the *Elhuyar (T-comp)* corpus that have been simplified by two language experts following the structural and intuitive approaches. The structural approach was performed by a court translator that followed easy-to-read guidelines inspired by Mitkov and Štajner (2014) such as use simple and short sentences with only a finite verb in each, use high frequency words, use always the same word to refer to same concept, keep logical and chronological ordering, recover elided arguments and verbs, and resolve anaphora. She was also given a set of operations performed in a previous phase of the corpus that included operations such as substitute synonyms, split clauses or reorder clauses. The intuitive approach was performed by a teacher following her experience on teaching Basque as foreign language and her intuition. In both approaches the target audience was open.

So, to see how the phenomena contained in the most predictive syntactic features have been treated in the CBST corpus, we have followed this methodology:

1. We have extracted the syntactic features from the list of predictive features and their significance according to the two rankers (Table 1).

Ratio	Chi square	InfoGain
Conditional clauses / clauses	249.9054	0.7057
Concessive clauses / clauses	189.0847	0.5338
Purpose clauses / clauses	148.062	0.3808
Temporal clauses / clauses	143.678	0.3716
Relative clauses / clauses	77.1949	0.2398
Causal clauses / clauses	61.3733	0.1814
Modal clauses / clauses	60.4716	0.1657
Completive clauses / clauses	54.202	0.1632

Table 1: Most predictive syntactic features.

2. We have extracted automatically the clauses related to the five most predictive features (conditional clauses, concessive clauses, purpose clauses, temporal clauses and relative clauses, henceforth target clauses) from the CBST corpus.

---

[1]In total, there are 785 sentences and 13 303 words.

3. Following the annotation of the corpus, we have performed the analysis of the target clauses by tagging manually if they have been treated or not (binary classification) and how they have been simplified. In the case of the relative clauses, we have added complementary information: we have annotated if the elided element is the object or subject to be able to compare if the treatment of clauses suits to the findings by Carreiras et al. (2010).

The results of the treatment are shown in Section 3

3 Treatment of the Predictive Syntactic Features in the CBST

In the CBST there are 114 clauses (89 sentences) that contain the target clauses. In Table 2 we show the percentages of the target clauses that have been treated in both simplification approaches. In parentheses we show the raw numbers.

Clause type	Quantity	Structural approach	Intuitive approach
Conditional clauses	17	82.35 (14)	35.29 (6)
Concessive clauses	5	100.00 (5)	80.00 (4)
Purpose clauses	22	22.73 (5)	22.73 (5)
Temporal clauses	17	52.94 (9)	58.82 (10)
Relative clauses	53	79.25 (42)	52.83 (28)
Total	114	85.50 (75)	71.82 (63)

Table 2: Percentages of simplified clauses in both structural and intuitive approaches.

Looking at the results of the conditional clauses, we see that there is a big difference depending on the approach. While in the structural approach 82.35 % has been simplified, 35.29 % has been treated in the intuitive. The percentages of both approaches are closer in the case of the temporal clauses (structural: 52.94 %; intuitive: 58.82 %), but they also differ in the relative clauses (structural: 79.25 %; intuitive: 52.83 %). The treatment of the purpose clauses has been slight in both approaches, and we do not have enough data for the analysis of the concessive clauses. According to the simplification approach in general, we see that there have been more simplified target clauses in the structural approach (85.50 %) than in the intuitive (71.82 %).

In Section 4 we present how these target sentences have been manually simplified.

4 Simplification of the Target Clauses the CBST

As we also want to know how the simplification is carried out in the target clauses, we analyse the macro-operations that have been performed. Eight macro-operations[2] are defined in the annotation scheme of the CBST: delete, merge, split, transformation, insert, reordering, no_operation and other. The most found macro-operations in the target clauses are split and transformations. In the split operations sentences are divided, and in the transformation operations different changes are performed in words, phrases or clauses. Transformations can be lexical, morphological, syntactic, discursive, corrections and reformulations. The detailed description of each macro-operation and the operations involved is found in Gonzalez-Dios (2016).

In the following subsections, we present the results according to the macro-operations mainly performed. These macro-operations are split, transformation, split and transformation together in the same clause and only others (delete, insert or reordering). We illustrate the results with examples of original and simplified versions in both approaches.

4.1 Conditional Clauses

In the conditional clauses (Table 3), the most applied macro-operation has been the transformation in both approaches. The syntactic transformations that have been applied are a) the change of the syntactic connective, b) transforming a clause into a phrase, c) transforming a subordinate clause into a main

---

[2]Macro-operations are an abstraction or generalisation of different simplification operations.

clause or d) changing the finite verb with a non finite. There have also been morphological transformations like changing the verb mood from potential or conditional to present or past indicative and lexical transformations.

Approach	Treated	Split	Transformation	Split + Trans	Only others
Structural	14	28.57 (4)	42.86 (6)	14.29 (2)	14.29 (2)
Intuitive	6	0.00 (0)	66.67 (4)	33.33 (2)	0.00 (0)

Table 3: Macro-operations performed in conditional clauses.

In Table 4 we show two examples of conditional clauses. In the first sentence, a split, an insert and a change in the verb mood and tense (from hypothetic to present potential) have been performed in the structural approach while in the intuitive approach only the change in the verb mood and tense (form hypothetic to past indicative) has been carried out. In the second sentence, no operation has been performed in the structural approach but the syntactic connective has been changed (conditional to causal). There has been a reordering too.

Original	Structural	Intuitive
Bidean arrokazko planeta txikiagoren bat aurkitu izan balu, bereganatu egingo zuen (...) If it had probably found a smaller planet made out of stone, it would have appropriated it (...)	*Bidean arrokazko planeta txikiagoren bat aurki dezake; orduan, planetak bereganatu egingo du; (...)* It can find a a smaller planet made out of stone; then, it will appropiate the planet; (...)	*Bidean arrokazko planeta txikiagoren bat aurkitu izan bazuen, bereganatu egingo zuen.* If it had found a smaller planet made out of stone, it would have appropriated it (...)
gene konpontzailea mutatuta badago, ezin du bere funtzioa bete, (...) if the repairing gene is muted it cannot perform its function, (...)	*gene konpontzailea mutatuta badago, ezin du bere funtzioa bete.* if the repairing gene is muted it cannot perform its function, (...)	*ezin dute funtzio hori bete mutatuta daudelako; (...)* it cannot perform its function because they are muted; (...)

Table 4: Example of simplifications of conditional clauses.

4.2 Purpose Clauses

In the purpose clauses (Table 5) the reordering (exactly, the operation named change of the order of clauses) has also been used in the intuitive approach. The transformations that have been applied in both approaches are a) the conversion of subordinate clauses into main or coordinate clauses at syntactic level, b) lexical transformations and c) reformulations.

Approach	Treated	Split	Transformation	Split + Trans	Only others
Structural	5	20.00 (1)	40.00 (2)	40.00 (2)	0.00 (0)
Intuitive	5	0.00 (0)	20.00 (1)	20.00 (1)	60.00 (3)

Table 5: Macro-operations performed in purpose clauses.

In Table 6 we show the simplified versions of two purpose clauses. In the first clause, a split has been performed in the structural approach while no operation was performed in the intuitive (not treated). The second clause was not treated in the structural approach while the ordering of clauses was altered (reordering) in the intuitive.

Original	Structural	Intuitive
Datu horiek eta tesian ateratako ondorio guztiak baliagarriak izango dira jarraipen berezitua egiteko (...) All that data and the results of the PhD thesis will be useful to make a specialised monitoring (...)	*(...) jarraipen berezitua egingo zaie. Horretarako, datu horiek guztiak eta tesiko ondorio guztiak baliatuko dira.* (...) a specialised monitoring will be done to them. To that end, all that data and all the results of of the PhD thesis will be used.	*Datu horiek eta tesian ateratako ondorio guztiak baliagarriak izango dira jarraipen berezitua egiteko (...)* All that data and all the results of the PhD thesis will be useful to make a specialised monitoring (...)
euste-indar bera izateko, eraso-angelua handiagoa izan behar du, (...) to keep the same holding strength, the clearance angle should be bigger (...)	*euste-indar bera izateko, eraso-angelua handiagoa izan behar du, (...)* to keep the same holding strength, the clearance angle should be bigger (...)	*eraso-angeluak handiagoa izan behar du euste-indar bera izateko.* the clearance angle should be bigger to keep the same holding strength (...)

Table 6: Example of simplifications of purpose clauses.

4.3 Temporal Clauses

In the temporal clauses (Table 7), the transformation has also been the most used macro-operation. The syntactic transformations that have been carried out are similar to those found in the conditional sentences: a) the change of the syntactic connective, b) transforming a clause into a phrase, c) transforming a subordinate clause into a main clause or d) changing the finite verb with a non finite). There have also been morphological transformations (e.g. change the persons of the verb) and reformulations.

Approach	Treated	Split	Transformation	Split + Trans	Only others
Structural	9	33.33 (3)	44.44 (4)	11.11 (1)	11.11 (1)
Intuitive	10	10.00 (1)	70.00 (7)	20.00 (2)	0.00 (0)

Table 7: Macro-operations performed in temporal clauses.

In Table 8 we show an example of a temporal clause which has undergone a split and a reordering of phrases and clauses in both approaches. As we see, although the operations performed are the same, the output text is different.

Original	Structural	Intuitive
Printzipio hori gertatzen da hegazkinen hegoetan hegan egiten ari denean, (...) That principle happens in the wings of the plane when it is flying, (...)	*Hegan egiten dugu; orduan, hegazkinen hegoetan gertatzen da printzipio hori.* We fly; then, that principle happens in the wings of the aircrafts.	*Printzipio hori hegazkinen hegoetan gertatzen da; hegazkinak orduan egiten du hegan.* That principle happens in the wings of the aircrafts; the plane flies then.

Table 8: Example of simplifications of temporal clauses.

4.4 Relative Clauses

In the relative clauses (Table 9), although the transformation has been the most used macro-operation, the split has also been important above all in the structural approach. The transformations that have been carried out are, as in the conditional and temporal clauses, a) the change of of the syntactic connective, b)

transforming a clause into a phrase, c) transforming a subordinate clause into a main clause or d) changing the finite verb with a non finite. There have also been lexical transformations and reformulations. The other macro-operations involved are the insertion of non required phrases, the reordering of phrases and clauses and the deletion of the phrases.

Approach	Treated	Split	Transformation	Split + Trans	Only others
Structural	42	33.33 (14)	52.38 (22)	7.14 (3)	7.14 (3)
Intuitive	28	14.29 (4)	71.43 (20)	7.14 (2)	7.14 (2)

Table 9: Macro-operations performed in relative clauses.

In the sentence of Table 10 the same transformation has been performed in both approaches: the finite verb "*eusten dien*" (that keeps them) has transformed into the non-finite verb "*eusteko*" (to keep).

Original	Structural	Intuitive
hegazkin horiei airean eusten dien printzipio fisikoa (...) the physical principle that keeps these planes in the air (...)	*hegazkin horiei airean eusteko printzipio fisikoa (...)* the physical principle to keep these planes in the air (...)	*hegazkin horiei airean eusteko printzipio fisikoa (...)* the physical principle to keep these planes in the air (...)

Table 10: Example of simplifications of relative clauses.

As relative clauses have also been studied from a neurolinguistic point of view by Carreiras et al. (2010), now we compare our findings. They claim that Basque subject relative clauses are harder to process than object relative clauses. So, we want to see if the relative clauses where the subject has been elided have been mainly treated in both approaches. These results are shown in Table 11.

Approach	Subject relative clause	Object relative clause
Structural	80.95 (34/42)	81.82 (9/11)
Intuitive	69.04 (29/42)	9.09 (1/11)

Table 11: Simplification of subject and object relative clauses.

In the case of the structural approach, more than 80 % of the subject and object relative clauses have been simplified. In the intuitive approach, on the other hand, almost 70 % of the subject clauses relative clauses have been simplified while the object relative clauses have hardly been. So, we see that in the case of the intuitive approach the results are consequent to the findings by Carreiras et al. (2010). In the structural approach, however, we think that the guidelines have influenced these results.

5   Discussion

While carrying out these experiments and performing the data analysis, we have found some issues we want to point out. The aim of this analysis is to guide a future study and make a reflection of the data and resources.

The first main issue is the data scarcity problem since in some cases e.g. the concessive clauses we were not able to analyse the results. However, we think that it is important to perform preliminary studies that can raise other questions while bigger corpora are created or compiled.

After having detected contradictory results when comparing the treatment by human editors to the statistically predictive features, we decided to carry out a deeper analysis of other phenomena that are found in the sentences of our target clauses. In this analysis, we want to see e.g. if in the sentences containing purpose clauses there are other phenomena that have been simplified. For example, in Table 12, we present a sentence containing a purpose clause where the relative clause also found has been reformulated.

Original	Structural
Hegan egiteko, pisu hori guztia konpentsatzen duen indar bat behar da.	*Hegan egiteko pisu hori guztia konpentsatu behar da indar baten bidez.*
To fly, it is needed a force that compensates all that weight	To fly all that weight should be compensated by means of a force.

Table 12: A sentence where a purpose clause has not been simplified but other phenomena have been.

These are the cases we show in Table 13: in the percentage of the second column both the target clause and other phenomena have been treated, in the third column only other phenomena have been treated, in the forth column only the target clause and in the fifth column no simplification has performed. We recall that the raw numbers are shown in parentheses.

Sentences containing	Target clause and other phenomena		Other phenomena		Only target clause		No simplification	
	Str.	Int.	Str.	Int.	Str.	Int.	Str.	Int.
Conditional clauses	70.59 (12)	29.41 (5)	17.65 (3)	52.94 (9)	11.76 (2)	5.88 (1)	0.00 (0)	11.76 (2)
Purpose clauses	18.18 (4)	22.73 (5)	77.27 (17)	63.64 (14)	4.55 (1)	0.00 (0)	0.00 (0)	13.64 (3)
Temporal clauses	47.06 (8)	47.06 (8)	41.18 (7)	29.41 (5)	5.88 (7)	11.76 (2)	5.88 (7)	11.76 (2)
Relative clauses	67.92 (36)	50.94 (27)	16.98 (9)	33.96 (18)	13.21 (7)	1.89 (1)	1.89 (1)	13.21 (7)

Table 13: Applied operations in the sentences containing the target clauses.

The combination of the simplification carried out in other phenomena is not, however, homogeneous in all the cases. It is noticeable in the case of purpose clauses, but it is not so noticeable in the rest. To study how different phenomena may affect each other will be our next step. In fact, as we have 114 target clauses in 89 sentences, we find that in 28.09 % of the sentences there is more than one target clause. So, we think that other phenomena in the sentences can be important. To enlarge this analysis, based on our manual analysis, we consider that in the future we should also take into account other parameters such as the length of the target clause or its subordination depth to get to know why a clause has been simplified and not the other.

Another interesting issue is the comparison of approaches, as we have seen that depending on the approach the results vary a lot. The guidelines and the target audience human editors had in mind has influenced their simplification decisions. This has been seen in the case of conditional and relative clauses where the results differ the most. We think that the guideline 'use simple and short sentences with only a finite verb in each'[3] has been important when performing the simplification in the structural approach and according to the intuition this guideline may not be needed. Moreover, in the case of the relative clauses we have seen that it can be sufficient to treat only subject relative clauses as performed in the intuitive approach and also corroborated by the neurolinguistic studies. The ideal situation will be to based on the neurolinguistic works to write simplification guidelines and that is why we think that multidisciplinary work should be encouraged.

As conclusion of the operations performed, we want to point out that both splitting and syntactic transformations such as a) changing the syntactic connective, b) transforming a clause into a phrase, c) transforming a subordinate clause into a main clause or d) changing the finite verb with a non finite seem to be important operations together with e) lexical operations and f) reformulations. These operations should be taken into account in the Basque ATS system although some of them are not feasible or require

---

[3]This guideline or a similar ones are found in many easy-to-read guidelines or recommendations in several languages.

a deeper analysis.

We have also seen in Table 8 that the mere description of the operations is not enough to capture all the changes performed in the sentences. This should lead to a re-thinking of the annotation scheme or annotation methodology.

6   Conclusion and Future Work

In this paper, we have presented the treatment that the phenomena in the statistical predictive syntactic features have undergone in the CBST corpus with the aim of contrasting machine and human perspectives. We have seen that in the structural approach there is a tendency to treat these phenomena, while the data in the intuitive approach is more diverse. In the case of relative clauses, we have related our findings to a neurolinguistic study and we have seen that the results of the intuitive approach agree. Looking at the simplifications macro-operations, we have corroborated the importance of the split and transformation operations in both approaches.

In the future, we would like to carry out an analysis the combination of the phenomena at sentence level to see how they may affect each other and to broaden this study when the CBST corpus be enlarged. In fact, data scarcity has been a problem to draw conclusions about the results, but we want to open a way to analyse the corpora based on different approaches. We also plan to reanalyse and compare our results as new neurolinguistic study are carried out and published, since we consider that multidisciplinary approaches can be helpful to get to understand linguistic complexity.

References

Itziar Aduriz and Arantza Díaz de Ilarraza, 2003. *Inquiries into the Lexicon-syntax Relations in Basque*, chapter Morphosyntactic disambiguation ands shallow parsing in Computational Processing of Basque. University of the Basque Country.

María Jesús Aranzabe, Arantza Díaz de Ilarraza, and Itziar Gonzalez-Dios. 2012. First Approach to Automatic Text Simplification in Basque. In Luz Rello and Horacio Saggion, editors, *Proceedings of the Natural Language Processing for Improving Textual Accessibility (NLP4ITA) workshop (LREC 2012)*, pages 1–8.

María Jesús Aranzabe, Arantza Díaz de Ilarraza, and Itziar Gonzalez-Dios. 2013. Transforming Complex Sentences using Dependency Trees for Automatic Text Simplification in Basque. *Procesamiento de Lenguaje Natural*, 50:61–68.

Rebekah George Benjamin. 2012. Reconstructing Readability: Recent Developments and Recommendations in the Analysis of Text Difficulty. *Educational Psychology Review*, 24(1):63–88.

Stefan Bott and Horacio Saggion. 2014. Text Simplification Resources for Spanish. *Language Resources and Evaluation*, 48(1):93–120.

Dominique Brunato, Felice Dell'Orletta, Giulia Venturi, and Simonetta Montemagni. 2015. Design and Annotation of the First Italian Corpus for Text Simplification. In *The 9th Linguistic Annotation Workshop held in conjuncion with NAACL 2015*, pages 31–41.

Manuel Carreiras, Jon Andoni Duñabeitia, Marta Vergara, Irene de la Cruz-Pavía, and Itziar Laka. 2010. Subject Relative Clauses are not Universally Easier to Process: Evidence from Basque. *Cognition*, 115(1):79–92.

Helena M. Caseli, Tiago F. Pereira, Lucia Specia, Thiago. A. S. Pardo, Caroline. Gasperin, and Sandra Aluísio. 2009. Building a Brazilian Portuguese Parallel Corpus of Original and Simplified Texts. In *the Proceedings of CICLing*, pages 59–70.

William H. DuBay. 2004. The Principles of Readability. *Impact Information*, pages 1–76.

Itziar Gonzalez-Dios, María Jesús Aranzabe, Arantza Díaz de Ilarraza, and Ander Soraluze. 2013. Detecting Apposition for Text Simplification in Basque. In *Computational Linguistics and Intelligent Text Processing*, pages 513–524. Springer.

Itziar Gonzalez-Dios, María Jesús Aranzabe, Arantza Díaz de Ilarraza, and Haritz Salaberri. 2014. Simple or Complex? Assessing the Readability of Basque Texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 334–344, Dublin, Ireland, August. Dublin City University and Association for Computational Linguistics.

Itziar Gonzalez-Dios. 2016. *Euskarazko egitura sintaktiko konplexuen analisirako eta testuen sinplifikazio automatikorako proposamena / Readability Assessment and Automatic Text Simplification. The Analysis of Basque Complex Structures*. Ph.D. thesis, University of the Basque Country (UPV/EHU).

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. 2009. The WEKA Data Mining Software: an Update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18.

Itziar Laka and Kepa Erdozia. 2010. Linearization References Given "Free Word Order"; Subject Preferences Given Ergativity: a Look at Basque. In Torrego E. (ed.)., editor, *Festschrift for Professor Carlos Piera*. Oxford University Press.

Ruslan Mitkov and Sanja Štajner. 2014. The Fewer, the Better? A Contrastive Study about Ways to Simplify. In *Proceedings of the Workshop on Automatic Text Simplification - Methods and Applications in the Multilingual Society (ATS-MA 2014)*, pages 30–40, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.

Idoia Ros, Mikel Santesteban, Kumiko Fukumura, and Itziar Laka. 2015. Aiming at Shorter Dependencies: the Role of Agreement Morphology. *Language, Cognition and Neuroscience*, 30(9):1156–1174.

Matthew Shardlow. 2014. A Survey of Automated Text Simplification. *International Journal of Advanced Computer Science and Applications (IJACSA), Special Issue on Natural Language Processing*, pages 58–70.

Advaith Siddharthan. 2014. A Survey of Research on Text Simplification. *The International Journal of Applied Linguistics*, pages 259–98.

Mostafa Zamanian and Pooneh Heydari. 2012. Readability of texts: State of the art. *Theory and Practice in Language Studies*, 2(1):43–53.