

Spanish Text Simplification Using Term Familiarity: Applying Principles from English Text Simplification

Gondy Leroy, PhD¹, Brianda Armenta Navarrete¹, Sonia Colina, PhD¹, David Kauchak, PhD²
¹University of Arizona, Tucson, Arizona; ²Pomona College, Pomona, California

Introduction

Health literacy is important for patients; text simplification is one tool to help facilitate the comprehension of information provided to them. Previously, we have developed and validated several features for simplifying English text. We report here on one such feature Term Familiarity¹, and how it applies to Spanish text simplification.

Simplifying Spanish Text Using the Term Familiarity Algorithm

We manually simplified four Spanish texts based on Term Familiarity (Table 1). We identified difficult words based on low term familiarity (low frequency in the LexEsp corpus) and either 1) replaced them with a higher frequency synonym or, when none was available, 2) added an explanation of the word. We integrated short explanations in the text, but added longer explanations between parentheses. We adjusted gender agreement as necessary.

Table 1. Text Statistics

Topic	Avg. Term Familiarity (Freq. in LexEsp)		Terms Simplified in Easy Text	
	Easy (total words)	Difficult (total words)	With Synonyms	With Explanation
Diabetes	1006 (306)	1008 (298)	6	1
Hipertension	1011 (363)	1062 (311)	7	5
Asma	1175 (299)	1162 (296)	11	2
Tabaquismo	632 (468)	717 (404)	9	8
Average	956	987	8	4

Pilot Study and Results

Study Design: We recruited six native Spanish speakers without a medical background. Each read two easy and two difficult texts (Table 1) that were randomly assigned and in random order. On average, twelve (8+4) words were simplified in the texts. For example “En este *artículo* se *revisan* ...” was replaced with “En este *texto*

se *revisan*”. We measured perceived difficulty with a Likert scale (lower is easier), actual difficulty (accuracy) with 10 multiple-choice questions per text (after reading the text), and total reading time.

Table 2. Pilot Study Results

Topic	Perceived Difficulty (1 = Very Easy, 5 = Very Difficult)		Actual Difficulty (Accuracy)		Reading Time (Minutes)	
	Easy	Difficult	Easy	Difficult	Easy	Difficult
Diabetes	3.0	3.0	76	50	42	17
Hipertension	3.3	4.0	85	65	50	24
Asma	3.7	3.5	64	80	15	36
Tabaquismo	2.0	3.0	82	80	34	30
Average	3.0	3.4	77	69	35	27

Results (Table 2): We removed data from 1 subject who spent on average 14 minutes reading, which was 10+ minutes less than the other subjects and infeasible given the text and question length. On average, actual and perceived difficulty were improved with simplification. Reading times were longer for easy texts, probably due to added explanations.

Conclusion and Lessons Learned: Similar to English, term familiarity can be used as the guiding principle to reduce difficulty of Spanish text. While encouraging, there were important lessons learned. In Spanish, increasing overall term familiarity is difficult, i.e. changing one term requires often adjusting other terms for gender and, as a result, the average term familiarity sometimes decreased. Furthermore, the lexicon has enormous impact. In prior studies, we used a smaller lexicon and were unable to simplify text. Finally, adding explanations may weaken improvements in term familiarity. Our approach requires more fine-tuning and the best version will be included in our future online text simplification tool.

Acknowledgments: Research reported in this publication was supported by the National Library Of Medicine of the National Institutes of Health under Award Number R01LM011975. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. G. Leroy, J. E. Endicott, D. Kauchak, O. Mouradi, and M. Just, "User Evaluation of the Effects of a Text Simplification Algorithm using Term Familiarity on Perception, Understanding, Learning and Information Retention," *Journal of Medical Internet Research (JMIR)*, vol. 15, p. e144, 2013.