

CONTRIBUTION TO NATURAL LANGUAGE GENERATION FOR SPANISH

Author: Silvia García Méndez

Thesis Advisors: Enrique Costa Montenegro, Milagros Fernández Gavilanes

Affiliation: GTI -atlanTtic, University of Vigo

1. Motivation of the work

Natural Language Generation (NLG) applications:

- For helping people with expression problems.
- **For ensuring the human-machine communication.**
- Generation of reports and summaries.

There is no automatic system conducting NLG for Spanish.

2. Thesis objectives

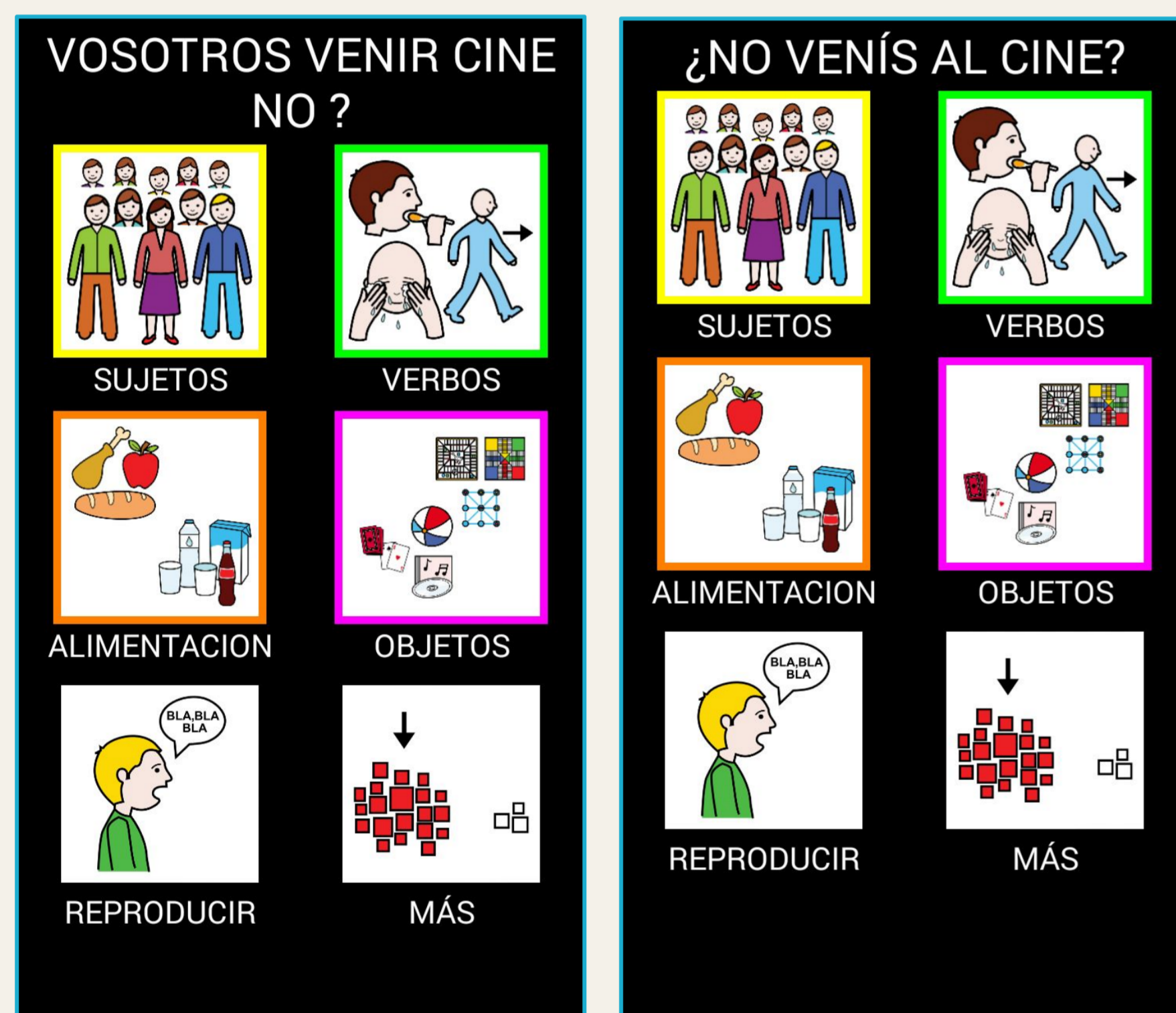
- ✓ In-depth analysis of the NLG state of the art.
- ✓ Build a lexicon for Spanish with large coverage and high precision including a wide range of linguistic data.
- ✓ Development of the first version of an automatic NLG system based on linguistic knowledge and statistics for Spanish.
- ✓ Application of Natural Language Processing (NLP) to improve our system.
- ✓ Test the automatic performance and flexibility in different fields of application.
 - Extending our system to English.
 - Test the Spanish and English versions within a communication application.
 - Evaluate the possibility of adapting the system to other languages.

3. Research plan

- ✓ Familiarisation with the concepts, techniques and algorithms in NLG.
- ✓ Design and development of a surface generator.
- ✓ System application to different areas.
- ✓ Enhancing the system.
- ✓ Attendance to national and international conferences.
 - Journal publication about the conducted research.
 - Research stay in an NLG European research group and journal publication.
 - System adaptation to English.

4. Previous research

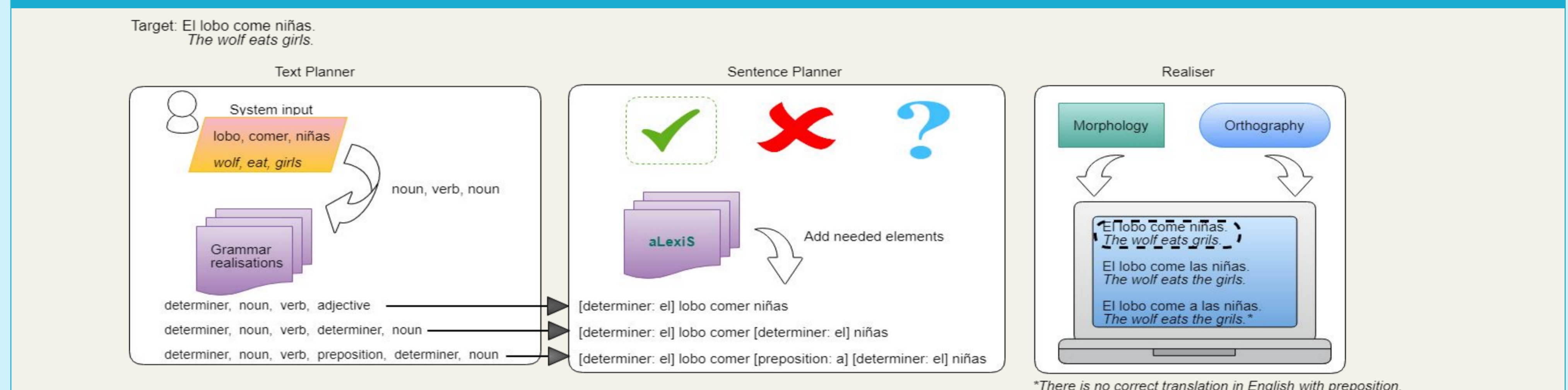
Spanish SimpleNLG version with automatic performance [1] integrated in *PictoDroid Lite* [2].



Pictograms selected by the user as system input. Sentence in natural language as system output.

Winner of the **XI Edition Vodafone Connecting for Good to Innovation in Telecommunications Award (2017)**.

5. System overview



6. Results & Discussions

- NLG hybrid system independent from third-party libraries.
- Dataset specially designed for automatic NLG.
- **Promising results for correctness precision and inter-agreement measures among annotators.**

Accuracy	An. 1	An. 2	An. 3	An. 4	An. 5
Annotator 1	-	0.812	0.616	0.668	0.734
Annotator 2		-	0.664	0.664	0.690
Annotator 3			-	0.707	0.672
Annotator 4				-	0.664
Annotator 5					-

Accuracy measures between pairs of annotators.

Automatic generation success	Enhanced SimpleNLG			
	Correct	Incorrect	Total	
Our NLG Library	Correct	346	390	736
	Incorrect	17	195	212
	Total	363	585	948

Our NLG system vs Enhanced SimpleNLG - automatic generation success.

More detailed description in [3].

Knowledge transfer to the business world

- Automatic data-to-text generation.
- Task focused on business reports.
- Aiming to track market changes.

7. Publications & Conferences

- **Journal publication under review [3].**
- Several conferences attended to present the undergoing work [4,5,6].
- Co-author of several other publications on the NLP topic [7, 8,9,10].
- Invited speaker at the V TADEGa Conference (2017).
- Scientific committee of NEGES (2017) and WGML (2018).
- Presence at the Mobile World Congress (2018) making a demo of the system.

8. Next Year Planning



9. References

- [1] García-Méndez, S. (2016). Natural language generation library for Spanish. Master Thesis. University of Vigo.
- [2] Available at <http://www.accegal.org/pictodroid-lite/>.
- [3] García-Méndez, S.; et al. (2018). Natural Language Generation library for automatic Spanish text creation. Expert Systems with Applications. UNDER REVIEW.
- [4] García-Méndez, S.; et al. (2017). Lexicon for Natural Language Generation in Spanish Adapted to Alternative and Augmentative Communication. Proceedings of the LiRA@INLG.
- [5] García-Méndez, S.; et al. (2017). Machine Learning para la Generación de Contenido en Lenguaje Natural. Proceedings of the WGML.
- [6] García-Méndez, S.; et al. (2018). Automatic Natural Language Generation Applied to AAC for Online Video Content Services using SimpleNLG for Spanish. Proceedings of the Web4All@The Web Conference.
- [7] Álvarez-López, T.; Fernández-Gavilanes, M.; García-Méndez, S.; et al. (2016). GTI at TASS 2016: Supervised Approach for Aspect Based Sentiment Analysis in Twitter. Proceedings of TASS@SEPLN.
- [8] Fernández-Gavilanes, M.; Juncal-Martínez, J.; Álvarez-López, T.; García-Méndez, S.; et al. (2016). Sistemas NLP para el análisis de sentimiento y detección de aspectos basados en machine learning. Proceedings of the WGML.
- [9] Álvarez-López, T.; Fernández-Gavilanes, M.; Costa-Montenegro, E.; Juncal-Martínez, J.; García-Méndez, S.; Bellot, P. (2017). A Book Reviews Dataset for Aspect Based Sentiment Analysis. Proceedings of the LTC.
- [10] Fernández-Gavilanes, M.; Juncal-Martínez, J.; García-Méndez, S.; et al. (2018). Creating emoji lexica from unsupervised sentiment analysis of their descriptions. Expert Systems with Applications.