

# 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

## Motivation of the work

45,818 Spanish people (versus 38,763 in 2011) had communication disorders [1,2].



Language aids are essential to avoid dependency and isolation.

Natural Language Generation (NLG) may be helpful but keeping in mind that it requires reliable and complete linguistic resources.

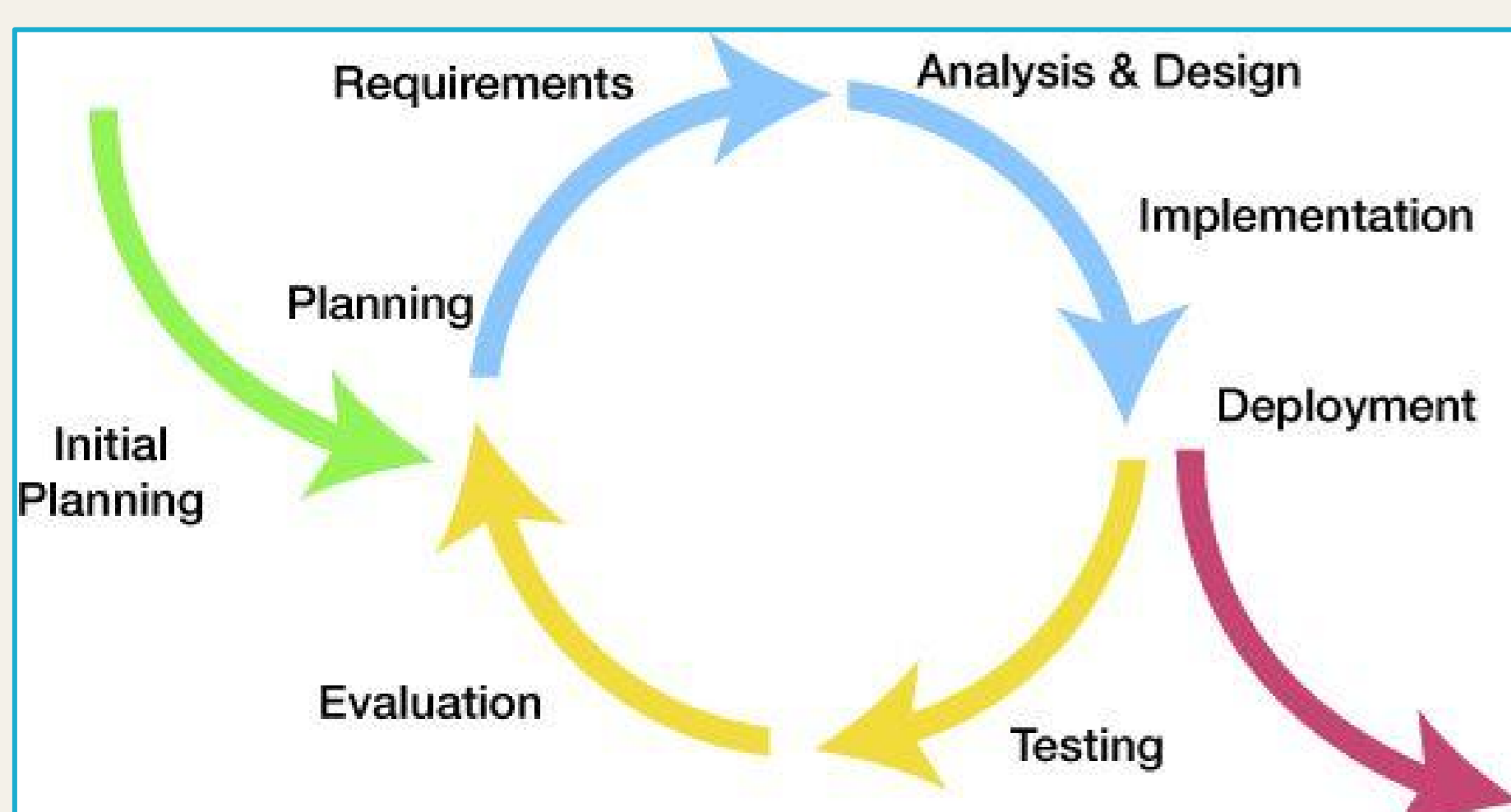
There is no automatic system conducting NLG for Spanish. It will be valuable in many fields of application.

## Thesis objectives

- In-depth analysis of 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) techniques 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.

## Methodology

I applied a Spiral Model of Software Development and Enhancement [3], since it allows me to get prototypes of the desired system during its development.



## Previous research

Spanish SimpleNLG adaptation and extension with automatic performance [4] integrated in *PictoDroid Lite* [5].

Pending of publication in the INLG Conference 2017.

Pictograms selected by the user as system input.

Sentence in natural language as system output.

## Research plan

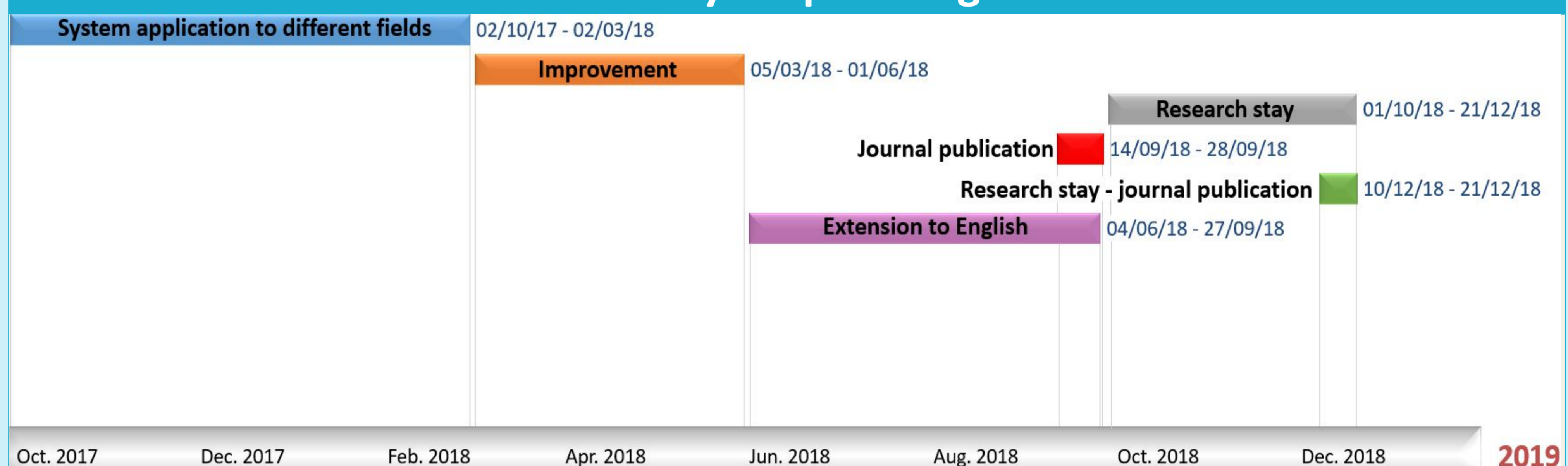
- Familiarisation with the concepts, techniques and algorithms in NLG.
- Design and development of a surface generator.
- System application to different areas.
- Enhancing the system.
- Attendance to national and international conferences.
- Journal publication about the conducted research.
- Research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

## Results & Discussions

- ✓ NLG state of the art.
- ✓ Spanish lexica: Elsa & aLexiS (pending of publication in the INLG Conference 2017).

Resource	Morphology	Syntax	Semantics	Compared to aLexiS (% lemmas)
<i>GilcUB-M</i> [6]	Yes	No	No	-29.12%
<i>Freeling</i> [7]	Yes	Yes	No	-13.09%
<i>TIP Conjugator</i> [8]	Yes (only verbs)	No	No	+9.73%
<i>AnCora-Verb-Es</i> [9]	Yes (only verbs)	Yes	Yes	-83.08%

## Next year planning



## References

- [1] Press release available Oct. 2016 at <http://www.dependencia.imserso.es/InterPresent2/groups/imserso/documents/binario/bdepcd2014.pdf>.
- [2] Survey on disability, personal autonomy and dependency situations. Available Oct. 2016 at <http://www.ine.es/en/prensa/np524.en.pdf>.
- [3] Boehm, B. W. (1988). A spiral model of software development and enhancement. *Computer*, 21(5), 61-72.
- [4] García Méndez, Silvia (2016). Natural language generation library for Spanish. Master Thesis. University of Vigo.
- [5] Available at <http://www.accegal.org/pictodroid-lite/>.
- [6] Molinero, M., Sagot, B., & Nicolas, L. (2009). A morphological and syntactic wide-coverage lexicon for Spanish: The Leffe. In RANLP 2009.
- [7] Lluís Padró, Miquel Collado, Samuel Reese, Marina Lloberes, and Irene Castellón. (2010). Freeling 2.1: Five years of open-source language processing tools. In Proc. of the LREC 2010.
- [8] Francisco Javier Carreras-Riudavets, Z. Hernández-Figueroa, and G. Rodríguez-Rodríguez. (2010). La conjugación de verbos en español y su morfología.
- [9] Juan Aparicio, Mariona Taulé, and M. Antònia Mart. (2008). Ancora-verb: A lexical resource for the semantic annotation of corpora. In Proc. of LREC 2008.