



Grammatical modeling of nominal ellipsis in Spanish

Hazel Barahona

Instituto de Literatura y Ciencias del Lenguaje
Pontificia Universidad Católica de Valparaíso
Projecto FONDECyT 1171033
Viña del Mar, Chile
hhgg09@yahoo.es

Walter Koza

Instituto de Literatura y Ciencias del Lenguaje
Pontificia Universidad Católica de Valparaíso
Projecto FONDECyT 1171033
Viña del Mar, Chile
walter.koza@pucv.cl



Organization

1. Objective
2. Ellipsis description and theoretical framework
3. Methodology
4. Results
5. Final remarks

Objective

- To develop a computer model for a generative syntax of the nominal ellipsis for Spanish.
 - The resulting algorithm allows us to:
 - create elided structures according to the Phonetic Form, and
 - replace elements elided in natural language texts, establishing an interpretative structure (EI) according to the Logical Form.

Basic concepts of ellipsis

- On descriptive grammars (Bosque & Demonte, 1999; Brucart, 1999; RAE, 2009), the ellipsis is defined as a mechanism to limit lexical redundancy.
 - (1) Antonio compró un auto y yo compré una moto. / Antonio bought a car and I bought a motorcycle.
 - (2) Antonio vendrá, pero yo no sé cuándo Antonio vendrá. / Antonio will come, but I don't know when Antonio will come.
- The nominal ellipsis is the omission of the complement head of the determinant phrase.
 - (3) El hijo de Juan y el hijo de Pedro./ Juan's son and Pedro's son
 - (4) La tasa de natalidad del 2019 y la Ø del 2020. / 2019 birth rate and 2020's birth rate

Theoretical framework

- Distributed morphology (Halle & Marantz, 1993)
- Partial identity theory of ellipsis (Saab, 2008)
- Syntax of silence (Merchant, 2001)

Methodology

- Formal description of nominal ellipsis
- Develop deletion and replacement rules
- Computer implementation

Formal description of ellipsis

(5) El hijo de Juan y el de [E] Pedro fueron al cine. /The son of Juan and the ~~son~~ of Pedro went to the movies.

[[PRENOM]₁ N [POSTNOM]₁]SDI Coord [[PRENOM]₂ N [POSTNOM]₂]SDII

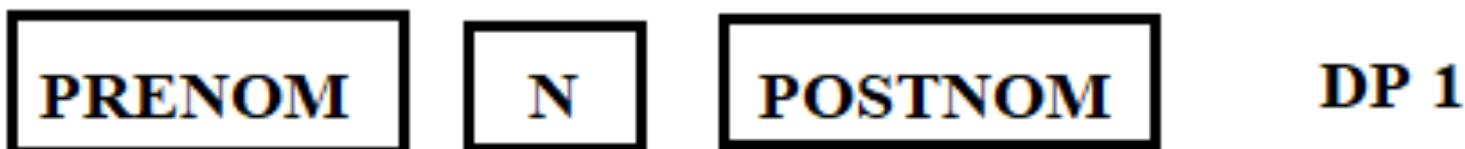
(6) El hijo de Juan fue al cine, pero el [E] de Pedro eligió el circo. /The son of Juan went to the movies, but the ~~son~~ of Pedro chose to go to the circus.

[[PRENOM]₁ N [POSTNOM]₁]SDI [V + ARG (ADJ)]₁ Coord [[PRENOM]₂ N [POSTNOM]₂]SDII [V + ARG (ADJ)]₂

(7) El presidente de Francia homenajeó al [E] de Perú. / The president of France paid tribute to the ~~president~~ of Perú.

[[[PRENOM]₁ N [POSTNOM]₁]SDI]_arg0 PRED [[[PRENOM]₁ N [POSTNOM]₁]SDI]_arg1

Ellipsis structure



SOMETHING



Deletion rules

- Given a coordination A^B , where A and B are DPs and B has a structure equal to A, with a NP complement whose root head N is identical to the root head N of A complement, remove NP head of B.
- Given a Predicative Argument Structure (PAS) A, with arguments $\alpha, \beta, \gamma\dots$, where α has a structure equal to the other A argument to its right and this one has a complement NP whose root head N is equal to the NP head in α , remove the head of $\beta, \gamma\dots$ NP.

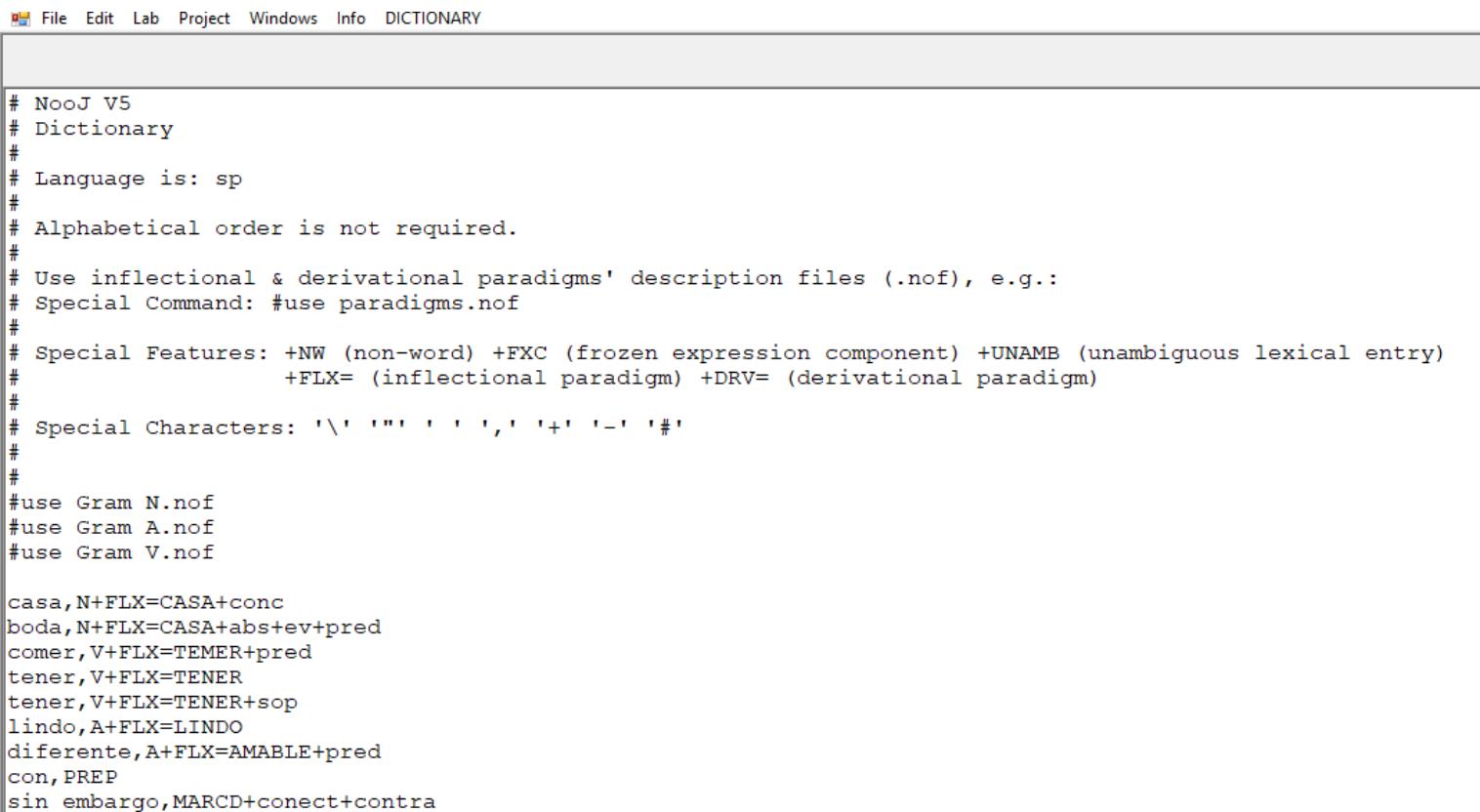
Replacement rules

- Given a coordination A^B , where B is a DP equal to A, and the NP head N of B complement is elided:
 - Copy root and gender of NP head N of A complement
 - Copy Determinant number of B
- Given a PAS A: $P(\alpha, \beta \gamma \gamma)$, with an object to the right of α that presents the elision of its NP head:
 - Copy root and gender of NP head complement of α
 - Copy Determinant number of $\beta, \gamma \dots$ (according to object elided)

Computer implementation

- Dictionary (RAE,2012)
- Computer grammars

Dictionary



NooJ V5
Dictionary

Language is: sp

Alphabetical order is not required.

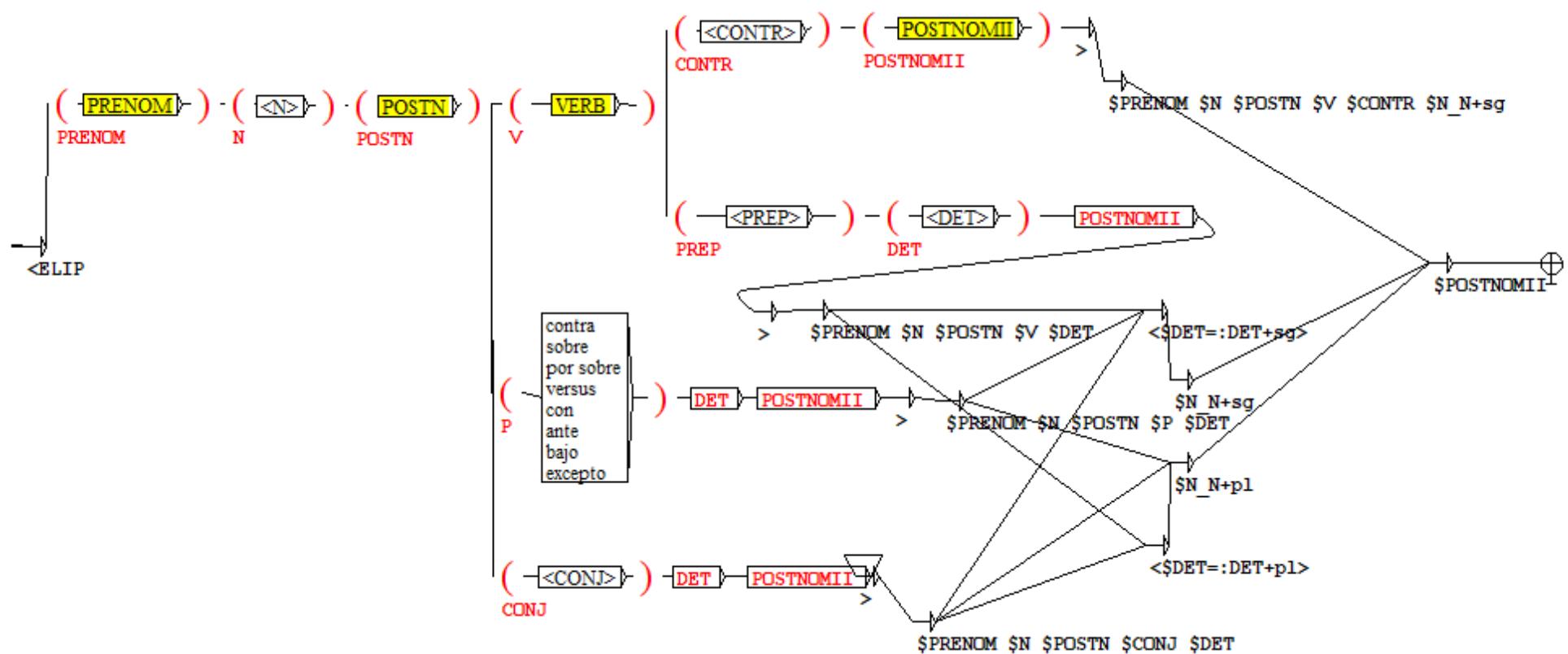
Use inflectional & derivational paradigms' description files (.nof), e.g.:
Special Command: #use paradigms.nof

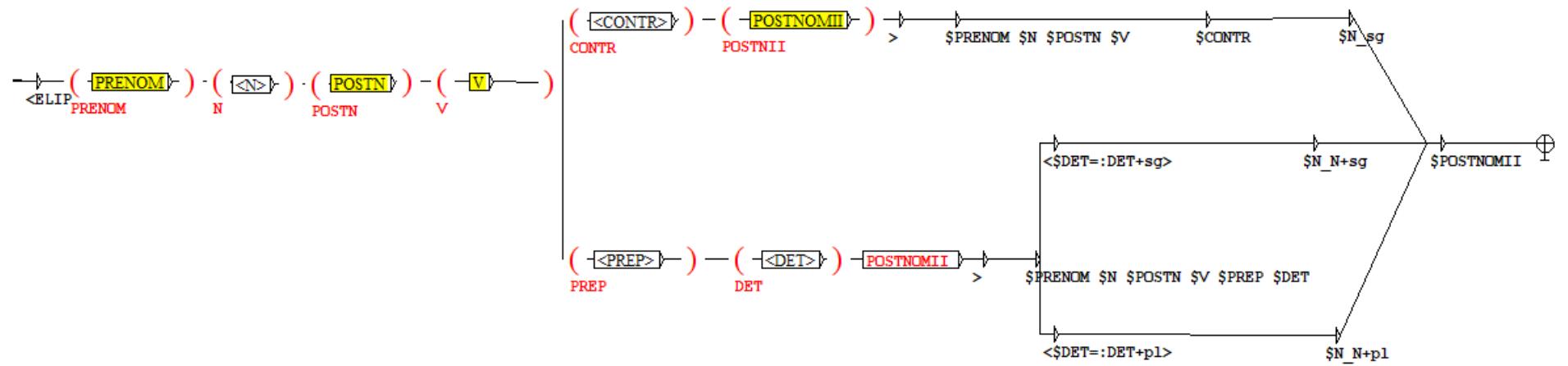
Special Features: +NW (non-word) +FXC (frozen expression component) +UNAMB (unambiguous lexical entry)
+FLX= (inflectional paradigm) +DRV= (derivational paradigm)

Special Characters: '\t\"!\"!\"!\"!\"+'-'#\n#

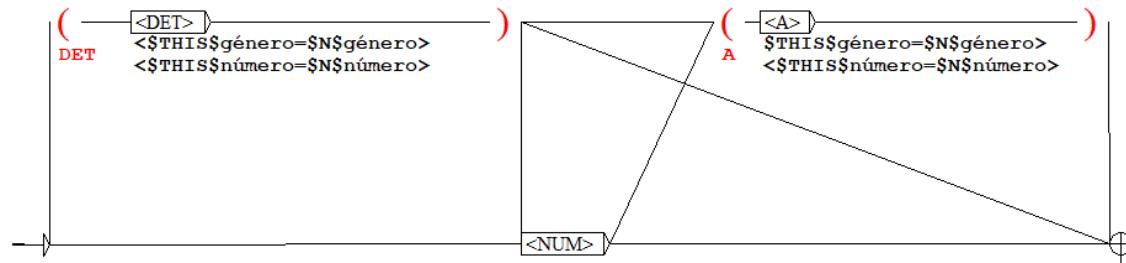
#use Gram N.nof
#use Gram A.nof
#use Gram V.nof

casa,N+FLX=CASA+conc
boda,N+FLX=CASA+abs+ev+pred
comer,V+FLX=TEMER+pred
tener,V+FLX=TENER
tener,V+FLX=TENER+sop
lindo,A+FLX=LINDO
diferente,A+FLX=AMABLE+pred
con,PREP
sin embargo,MARCD+conect+contra

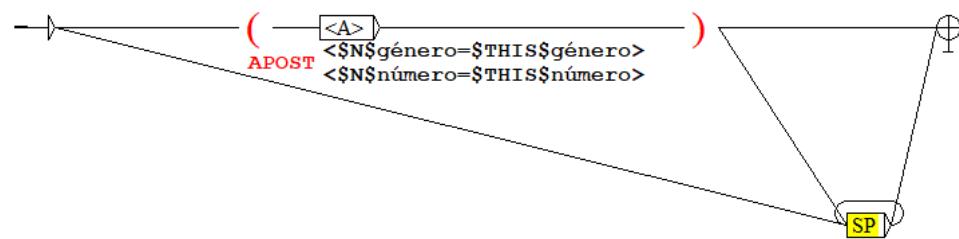




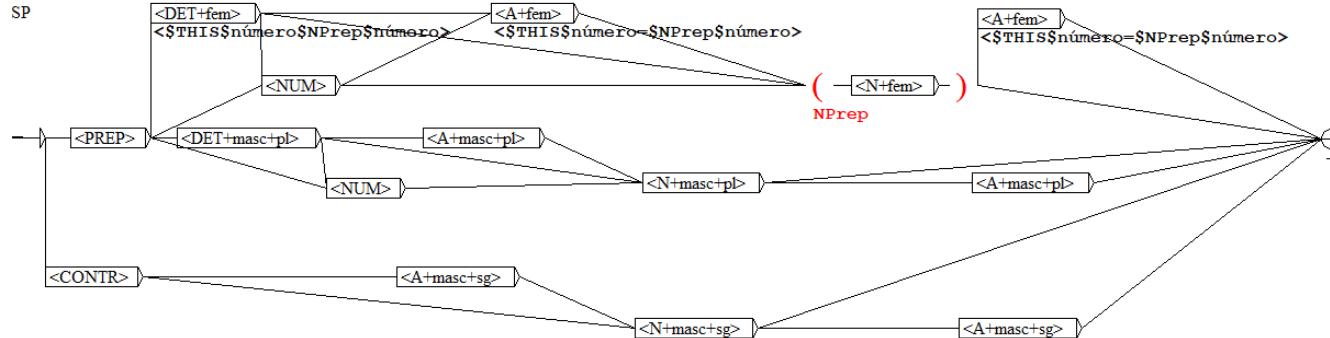
PRENOM

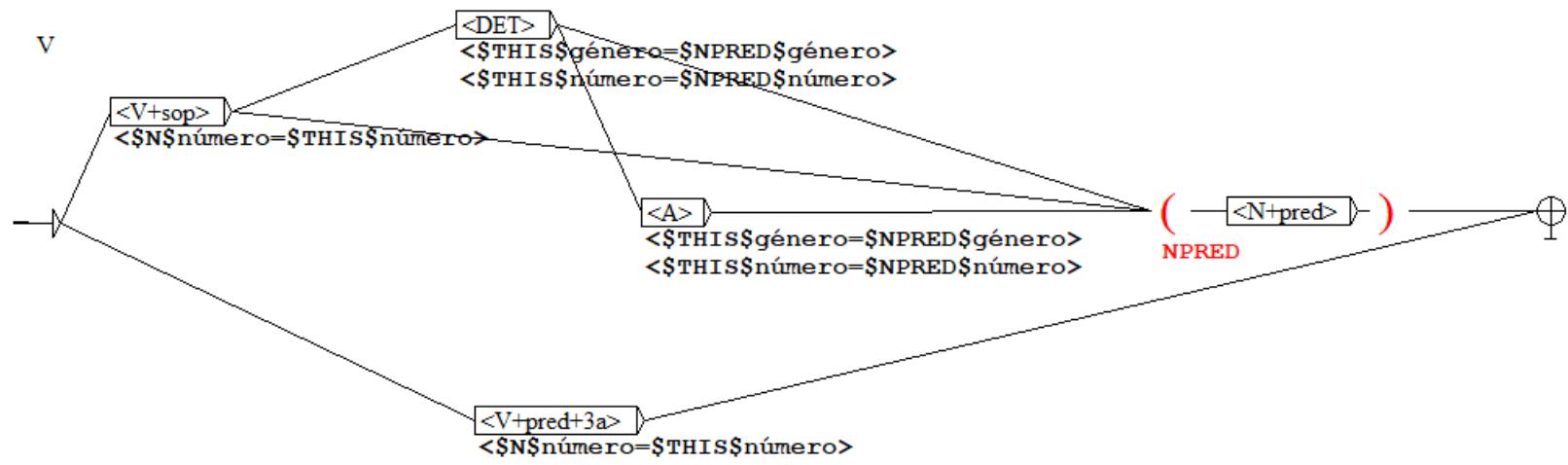


POSTN



SP





Testing

- The methodology was tested on a corpus of 100 sentences with nominal elisions from natural language texts (from Internet).
- The results were evaluated with precision, coverage and F measurement.

| Community Edition - [[Modified] Untitled]

Edit Lab Project Windows Info TEXT

1 + / 1 TUs Characters ^ Language is "Spanish(sp)".
Tokens Text Delimiter is: \n (NEWLINE)
Digrams 8 tokens including:
8 word forms

v Text Annotation Structure < >

cuidar mi salud y la de Juan

| | | | | | |
|---------------------|-----------------|---|-----------------------|--|--------------------|
| 5 | 12 | 15 | 21 | 23 | 26 |
| cuidar,V+tiempo=inf | ELIP | | | | |
| | mi,DET+tipo=pos | salud,N+dominio=gral+género=fem+número=sg | y,CONJ+tipo=coord+cop | la,DET+tipo=def+género=fem+número=sg | de,PREP+tipo=pos+1 |
| | | salud,N+dominio=gral+género=masc | | la,PRON+tipo=cl+caso=ac+género=fem+número=sg | |

| | | |
|--|--|--|
| <p>y la de tu familia y las de transformación social</p> | <p>La lucha independentista y la de transformación social/<ELIP>La lucha independentista y lucha de transformación social</p> | <p>Las luchas independentistas y las</p> |
| <p>Chile homenajeó al de Perú</p> | <p>El presidente de Chile homenajeó al de Perú/<ELIP>El presidente de Chile homenajeó al presidente de Perú</p> | <p>El hijo de Juan y</p> |
| <p>Chile homenajeó al de Perú</p> | <p>El hijo de Juan y el de María/<ELIP>El hijo de Juan y hijo de María</p> | <p>La vía sanguínea y la</p> |
| <p>Juan y el de María y la de transmisión general</p> | <p>El hijo de Juan y el de María/<ELIP>El hijo de Juan y hija de María</p> | <p>La vía sanguínea y la</p> |
| <p>y el de menor frecuencia y el de edad paterna vida y los de muerte y las de raza blanca y las con números pares</p> | <p>La vía sanguínea y la de transmisión general/<ELIP>La vía sanguínea y vía de transmisión general</p> | <p>Los costos de los paquetes</p> |
| <p>y el de edad femenina el de años de escolaridad vida y el de muerte vida y el de muerte la de un daño menor</p> | <p>Los costos de los paquetes estándares y el de los paquetes/<ELIP>Los costos de los paquetes estándares y costo de los paquetes</p> | <p>Los costos de los paquetes</p> |
| | <p>El promedio de edad masculina y la de edad femenina/<ELIP>El promedio de edad masculina y promedio de edad femenina</p> | <p>Los costos de los paquetes</p> |
| | <p>La tasa de complicaciones mayores y la de complicaciones/<ELIP>La tasa de complicaciones mayores y tasa de complicaciones</p> | <p>Los costos de los paquetes</p> |
| | <p>El promedio de edad materna y el de edad paterna/<ELIP>El promedio de edad materna y promedio de edad paterna</p> | <p>Los costos de los paquetes</p> |
| | <p>La medida de los embarazos y la de las gestaciones/<ELIP>La medida de los embarazos y medida de las gestaciones</p> | <p>Los costos de los paquetes</p> |
| | <p>La tasa de complicaciones postoperatorias y la de readmisiones/<ELIP>La tasa de complicaciones postoperatorias y tasa de readmisiones</p> | <p>Los costos de los paquetes</p> |
| | <p>La proporción de ovocitos maduros y la de fecundación/<ELIP>La proporción de ovocitos maduros y proporción de fecundación</p> | <p>Los costos de los paquetes</p> |
| | <p>El promedio de edad masculina y el de edad femenina/<ELIP>El promedio de edad masculina y promedio de edad femenina</p> | <p>Los costos de los paquetes</p> |
| | <p>El promedio de edad de las mujeres y el de años de escolaridad/<ELIP>El promedio de edad de las mujeres y promedio de años de escolaridad</p> | <p>Los costos de los paquetes</p> |
| | <p>El riesgo de vida y el de muerte/<ELIP>El riesgo de vida y riesgo de muerte</p> | <p>Los costos de los paquetes</p> |
| | <p>Las posibilidades de una lesión y la de un daño menor/<ELIP>Las posibilidades de una lesión y posibilidades de un daño menor</p> | <p>Los costos de los paquetes</p> |
| | <p>Las posibilidades de una lesión y la de un daño menor/<ELIP>Las posibilidades de una lesión y posibilidad de un daño menor</p> | <p>Los costos de los paquetes</p> |
| | <p>El tiempo medio de no recurrencia y el de no recidiva/<ELIP>El tiempo medio de no recurrencia y medio de no recidiva</p> | <p>Los costos de los paquetes</p> |
| | <p>El tiempo medio de no recurrencia y el de no recidiva/<ELIP>El tiempo medio de no recurrencia y media de no recidiva</p> | <p>Los costos de los paquetes</p> |

Results

- Unidentified ellipsis: 18
- Wrong recognition: 0

| | |
|-----------|-------|
| Coverage | 82% |
| Precision | 100% |
| F measure | 90.10 |

Final considerations

- The results prove that the rules created are useful for the automatic treatment of nominal ellipsis in Spanish.
- Work is being done to identify omissions due to some units not being listed on the NooJ dictionary, as well as the complexity of syntactic rules.

(8) Puerto de Montevideo "arrebata" un 30% de carga al Ø de Buenos Aires./ Port of Montevideo "snatches" 30% of cargo from the one in Buenos Aires

(9) La brutal diferencia del César Pinares de Universidad Católica al Ø de Colo Colo. / The brutal difference between César Pinares of Universidad Católica and of Colo Colo

Future work

- Optimize grammars
- Analyze cases of nominal ellipsis on peripheral positions:

(10) Según los jugadores de Boca, los Ø de Ríver festejaron desmedidamente. / According to Boca players, Ríver ones celebrated wildly”
- Formalize the automatic identification of gender variation in common noun:

(11) El estudiante de química fue peor que los Ø de biología.
/The student of chemistry was worse than those Ø of biology

References

- Brucart, J. (1999). La elipsis. En Ignacio Bosque & Violeta Demonte (eds.), *Gramática Descriptiva de la lengua español*. vol. 1, cap. 43: 2787-2863
- Lobeck, A. (1995). *Ellipsis: Functional Heads, Licensing and Identification*. New York: Oxford University Press.
- Merchant, J. (1999). The Syntax of Silence: Sluicing, Islands and Identity in Ellipsis. Tesis de Doctorado, Universidad de Santa Cruz.
- Merchant, J. (2001). *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford University Press: Oxford.
- Merchant, J. (2009). Variable Island Repair under Ellipsis. En Kyle Johnson (ed.) *Topics in Ellipsis*. Cambridge: Cambridge University Press.
- Merchant, J. (2016). Ellipsis: A survey of analytical approaches. Ms., University of Chicago. For Jeroen van Craenenbroeck and Tanja Temmerman (eds.), *Handbook of ellipsis*, Oxford University Press: Oxford.
- Saab, A. (2008). Hacía una teoría de la identidad en la elipsis. *Tesis de Doctorado*. Universidad de Buenos Aires.
- Silberztein, M. (2016). *Formalizing Natural Languages: The Nooj Approach*. Wiley-Iste
- Silberztein, M. (2005). NooJ: A Linguistic Annotation System for Corpus Processing. Demo. In *Proceedings of the HLT/EMNLP2005 Conference*. Vancouver.

Thank you

- Any questions?