

Grammar specification for the recognition of temporal expressions

Estela Saquete[†] and Patricio Martínez-Barco[‡]

[†]Laboratorio Multimedia

[‡]Departamento de Lenguajes y Sistemas Informáticos

Universidad de Alicante

Estela.Saquete@ua.es

patricio@dlsi.ua.es

Keywords: temporal expressions, anaphora resolution, coreference, grammars

Abstract

This paper tells about the recognition of temporal expressions and the resolution of their temporal reference. A proposal of the units we have used to face up this tasks over a restricted domain is shown. We work with newspapers' articles in Spanish, that is why every reference we use is in Spanish. For the identification and recognition of temporal expressions we base on a temporal expression grammar and for the resolution on a dictionary, where we have the information necessary to do the date operation based on the recognized expressions. In the evaluation of our proposal we have obtained successful results for the examples studied.

1. Introduction

Nowadays, one of the most important tasks of the Natural Language Processing (LNP) that must be solved in applications like Information Extraction (IE) [Gaizauskas(1995), Muñoz(1998)], Information Retrieval (IR) or Summarising [Spark-Jones(1993)], is the resolution of the linguistic coreference.

In outline, the resolution of the linguistic coreference can be classified in:

- a) Pronoun Resolution and
- b) Definite Description Resolution (DD)

To sum up, in pronoun resolution we can find papers about personal, possessive, demonstrative and omitted pronoun resolution. We can stand up the papers by Baldwin (1997), Kennedy and Bougarev (1996), Lappin y Leass (1994), Mitkov (1998) and in Spanish the ones by Ferrández *et al.* (1999).

In Definite Descriptions we can distinguish between *identity type*, *part of* and *synonymy* definite descriptions, and resolution of coreference of the temporal and locative expressions. The *identity type*, *part of* and *synonymy* DD have been treated in Spanish in the papers by Muñoz (2000). However, there is no outstanding paper about the treatment and resolution of temporal expressions in Spanish.

In this paper a proposal of a grammar for the recognition of temporal expressions in restricted texts in Spanish is presented, as well as an approximation to the resolution of the coreference introduced by them. The domain is restricted to articles published in newspapers. In a text of that kind of domain there are dates with typical representations like, for example: "23/01/2000" o "23 de enero del 2000" (23rd of January of 2000), but we can find references to dates named previously too, for example: "Dos días antes" (Two days before), "La semana anterior" (The previous week), etc. This kind of coreference should be solved and transformed to dates with a determined format for a more efficient analysis of the text. For that, we use a grammar for the recognition of temporal expressions, that is shown in the section 2 of this paper, and a dictionary where we keep the numeric equivalence of the expressions and

that permits to transform these expressions in a determined format that we have chosen: dd/mm/aaaa, as it is shown in the section 3 of the paper.

2. Proposal of a system for the recognition of temporal expressions and its temporal reference

In Figure 1 the graphic representation of the system proposed for the recognition of temporal expressions and for the resolution of its references is shown, according to the temporal model propose. The texts are tagged with lexical and morphological information and this information is the input to the parser proposed. This parser is implemented using an ascending technique (chart parser) and it is based on the grammar we show below. Once the parser recognizes the temporal expressions in the text, these are introduced into the resolution unit, which will update the value of the reference according to the date it is referring. We can find anaphoric temporal expressions and not anaphoric temporal expressions. The way to differentiate between them is using the grammar below.

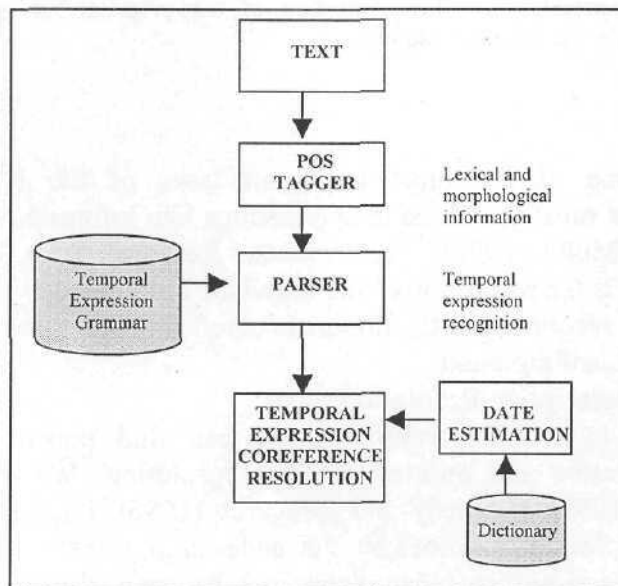


Figure 1. Graphic representation of the system proposed.

2.1. Grammar for the recognition of temporal expressions

The parser used a grammar based on two different kinds of rules. On one hand there are rules for the date recognition (complete temporal expressions) and on the other hand there are rules for the temporal reference recognition (incomplete temporal expressions that need the location of another temporal referent to be understood). So that, the grammar proposed is divided in two parts that we briefly describe below:

2.1.1 Date Recognition

There is a large amount of date formats. In Table 1 it can be observed a sample of some of the rules that have been defined in our system for the date recognition. The nomenclature in which the rules are above-mentioned is the one used in the data dictionaries of the information systems. The words that

appear between brackets means that they are optional. The words between square brackets, [], means that only one of them appear in the rule. The words between keys, { }, means that they will be repeated one or more times in the rule.

Fecha → dd + "/" + mm + "/" + (aa)aa (12/06/1975) (06/12/1975)
Fecha → dd + "-" + mes + "-" + (aa)aa (12-junio-1975) (12-Jun.-1975)
Fecha → dd + "de" + mm + "de" + (aa)aa (12 de junio de 1975) (12 th of June of 1975)
Fecha → ("E") + diasemana+ dd + "de"+ mes + "de" + (aa)aa (El domingo 12 de junio de 1975) (Sunday, 12 th of June of 1975)
Dd → ["01" "02" "03" "..." 31"]
Dia → ["uno" "dos" ... "treinta y uno"] [one two ... thirty one]
Mm → ["01" "02" "03" "..." 12"]
Mes → ["enero" "febrero" "marzo" "abril" "mayo" "junio" "julio" "agosto" "septiembre" "octubre" "noviembre" "diciembre"]
(January February March April May June July August September October November December)
A → ["1" "2" "3" "..." 9" "0"]
Diasemana → ["lunes" "martes" "miércoles" "jueves" "viernes" "sábado" "domingo"]
(Monday Tuesday Wednesday Thursday Friday Saturday Sunday)

Table 1. Sample of rules for the dates recognition

2.1.2 Reference Recognition

There are two types of temporal references that should be treated: the time adverbs and the nominal phrases that are referring to temporal relationships. In Table 2 we show some of the rules used for the detection of every kind of reference.

Time Adverbs	referencia → "ayer" (yesterday) referencia → "mañana" (tomorrow) referencia → "anteayer" (the day before yesterday) referencia → "anoche" (last night)
Temporal Nominal Phrases	referencia → "el" + "próximo" + ["día" "mes" "año"] (the next day month year) referencia → "un" + ["día" "mes" "año"] + "después" (a day month year later) referencia → num + ["días" "meses" "años"] + "después" (num days months years later) referencia → "un" + ["día" "mes" "año"] + "antes" (a day month year before) referencia → num + ["días" "meses" "años"] + "antes" (num days months years before) referencia → "dentro" + "de" + "un" + ["día" "mes" "año"] (within a day month year) referencia → "dentro" + "de" + num + ["días" "meses" "años"] (within num days months years) referencia → "el" + "pasado" + ["día" "mes" "año"] (the last day month year) referencia → "el" + ["día" "mes" "año"] + "siguiente" (the next day month year) referencia → "los" + num + ["días" "meses" "años"] + "siguientes" (the num next days months years) referencia → "el" + ["día" "mes" "año"] + "pasado" (the last day month year) referencia → "los" + num + ["días" "meses" "años"] + "pasados" (the last num days months years) num → ["dos" "tres" "cuatro" "cinco" ...] (two three four five ...)

Table 2. Sample of rules for the reference recognition

3. Coreference resolution based on a temporal model

For the coreference resolution we use a dictionary that contains the interpretation for every reference named before. In some cases the references are

estimated using the newspaper's date (FechaP). Others refer to a date named before in the text that is being analysed. For these cases, a temporal model that allows to know over what date the dictionary operations are going to be done, is defined. This model is based on the two rules below and it is only applicable to these dates that are not FechaP, since for FechaP there is nothing to resolve:

1. By default, the newspaper's date is used as a base referent (temporal expression) if it exists, if not, the system date is used.
2. In case of finding a new date, compare the reference with this date and update the temporal expression with the new value.

Moreover, some functions will be used, like Dia(date), that provides the day of the date that is passed as parameter, Mes(date), that provides the month of the date introduced and Año(date), that provides the year. *Dial* y *DiaF* represent the first and the last day of the month.

The operation "+1" used in some entries of the dictionary is able to interpret the dates to give a correct date. That is, if the function Mes(date) returns 12 and the operation "+1" is applied over that value, the return value will be 01 and one year will be increased.

In Table 3 some of the entries of this dictionary are shown:

REFERENCE	DICIONARY ENTRY
"ayer" (yesterday)	Dia(FechaP) -1 / Mes(FechaP) / Año(FechaP)
"mañana" (tomorrow)	Dia(FechaP) +1 / Mes(FechaP) / Año(FechaP)
"anteayer" (the day before yesterday)	Dia(FechaP) -2 / Mes(FechaP) / Año(FechaP)
"anoche" (last night)	Dia(FechaP) -1 / Mes(FechaP) / Año(FechaP)
"el" + "próximo" + "día" (the next day)	Dia(FechaP)+1 / Mes(FechaP) / Año(FechaP)
"un" + "mes" + "después" (a month later)	[Dial/Mes(fechaAnterior)+1/Año(fechaAnterior) -- DiaF/Mes(fechaAnterior) +1/Año(fechaAnterior)]
num + "años" + "después" (num years later)	[01/01/ Año(fechaAnterior) + num -- 31/12/ Año(fechaAnterior) + num]
"un" + "día" + "antes" (a day before)	Dia(fechaAnterior)-1/Mes(fechaAnterior)/Año(fechaAnterior)
num + "meses" + "antes" (num months before)	[Dial/Mes(fechaAnterior) -num / Año(fechaAnterior) -- DiaF/ Mes(fechaAnterior) -num / Año(fechaAnterior)]
"dentro" + "de" + "un" + "año" (within a year)	[01/01/ Año(fechaAnterior) +1 -- 31/12/ Año(fechaAnterior) +1]
"dentro" + "de" + num + "días" (within num days)	Dia(fechaAnterior)+num / Mes(fechaAnterior) / Año(fechaAnterior)
"el" + "pasado" + "día" (the last day)	Dia(fechaAnterior)-1/Mes(fechaAnterior) / Año(fechaAnterior)
"el" + "mes" + "siguiente" (the next month)	[Dial / Mes(fechaAnterior) +1 / Año(fechaAnterior) -- DiaF / Mes(fechaAnterior) +1 / Año(fechaAnterior)]
"los" + num + "años" + "siguientes" (the num years later)	[01/01/Año(fechaAnterior) -- 31/12 / Año(fechaAnterior) +num]
"el" + "día" + "pasado" (the last day)	Dia(fechaAnterior)-1/Mes(fechaAnterior) / Año(fechaAnterior)
"los" + num + "meses" + "pasados" (the num last months)	[Dial/Mes(fechaAnterior) - num / Año(fechaAnterior) -- DiaF/Mes(fechaAnterior) - 1 / Año(fechaAnterior)]

Table 3. Sample of some of the entries of the dictionary

Moreover, the dictionary has the relationship between numeric and string expressions of days and months, that is, one have the value 1 and July is 07.

The unit that makes the estimation of the dates will accede to the right entry in the dictionary in each case and it will apply the function specified obtaining a date in the format dd/mm/aaaa or a range of dates. So, at that point the coreference will have been resolved.

Two examples showing the behaviour of the system are proposed below:

- (a) *La oficina de Congresos de la Universidad ha propuesto 5 congresos para [este año], sin embargo, el crecimiento para [el próximo año] será superior a los 15. Por otro lado, el Director de la oficina ofrece [mañana] una conferencia.* (The University Conference Office has proposed 5 conferences for [this year], however, the increase for [the next year] will be over 15. On the other hand, the Office Manager offers [tomorrow] a lecture)

We assume that the newspaper's date is 25/04/2000. The system, for the reference [el próximo año] (*the next year*), will return [01/01/2001-31/12/2001]. For the reference [mañana] (*tomorrow*) it will return 26/04/2000.

- (b) *El 25 de enero de 1999, la Oficina de Congresos de la Universidad de Alicante propuso dos congresos que fueron cancelados [dos días después]* (The 25th of January of 1999, the University Conference Office of Alicante proposed two conferences that were cancelled two days later)

In this case there is a date in the text before the reference. This date will be translated as 25/01/1999, so that the reference [dos días después] (*two days later*) will be translated as 27/01/1999.

4. System evaluation

In the system above, two different units have been implemented. On one hand a parser has been implemented in LPA Prolog, because this language based on logic programming provides simple tools for the grammar construction making the implementation of parsers easier. However, Visual Basic was the programming language used for the implementation of the temporal coreference resolution unit because it includes a large amount of tools for handling dates and defining date operations. The definition of an interface has allowed the interconnection between the two units.

Finally, the evaluation of the system has been done with a sample of approximately 200 references, distributed among 16 articles. The articles used belong to digital edition on the Internet of two Spanish newspapers that are *El País Digital* y *ABC*. The articles describe international news that happened the 22nd and 23rd of May of 2000 and the number of references that every article have is in the result table of the evaluation.

To evaluate the system we used two measures, precision and recall. The first one is defined as the number of successes of the system divided by the number of references that have been treated. Recall is the number of successes divided by the total of real references existing in the text.

Precision= N Successes / N Ref. Treated

Recall= N Successes /N Real Ref.

The results obtained for the articles are shown in Table 4.

ARTICLE	PRECISION	RECALL	REAL REF.
Article 1	96,15%	80,64%	31
Article 2	77,7%	77,7%	9
Article 3	100%	84,2 %	19
Article 4	100%	100%	5
Article 5a	75%	100%	3
Article 5b	100%	100%	1
Article 6	100%	18,75%	16
Article 7	100%	63,63%	11
Article 8	96,77%	100%	30
Article 9	94,11%	88,88%	18
Article 10a	100%	75%	12
Article 10b	100%	90%	10
Article 10c	100%	75%	8
Article 11	100%	72,72%	11
Article 12	90%	90%	10
Article 13	85,71%	92,30%	13
Article 14	100%	100%	10
Article 15	100%	100%	6
Article 16	100%	92,85%	14
TOTAL	95,59 %	82,28 %	237

Table 4. Evaluation

The total has been calculated according to the number of successes being 195, the number of treated references is 204 and the number of total references is 237.

Although the obtained results are highly successful, we have detected some fails that have been deeply analysed. As it can be observed by the results, our system could be improved in some aspects. Below, a study of the problems detected and their possible improvements is shown:

- The unit that resolve the temporal references is not able to resolve undetermined temporal references like: "hace unos cuantos días" (some days before). A human could give an approximate interpretation of its reference, but never a specific date o range of dates.
- In the newspaper's articles, sometimes there are expressions like : " el sábado hubo cinco accidentes en la N-III" (*Saturday there were five accidents in the N-III road*). To resolve this kind of references we should need context information of the sentence where the reference is. That information could be the time of the sentence's verb. If the verb is a past verb, that indicates us that we have to solve a reference like "el sábado pasado" (*last Saturday*), whereas if it is a future verb it refers to "el sábado próximo" (*the next Saturday*). Because of our system does not use semantic or context information we assume this kind of reference refers to the last day, not the next, because the news usually tell us facts occurred previously.

5. Conclusions

In this paper a system for temporal expressions recognition and their reference resolution has been presented, based on a temporal model proposed. The system has two different units: the parser based on a temporal expression grammar, which allows to identify these kind of expressions and a coreference resolution unit which is based

in a dictionary and make a transformation of the expressions to dates, resolving their reference in this way. The evaluation of the system shows successful results of precision and recall for our proposal.

For future works, it is pretended to extend the system with the temporal references that are not treated in this paper. Moreover, the study of the verbal forms in the sentences where the references are found will improve the efficiency of the system solving some kind of expressions.

6. References

Baldwin, B. (1997) CogNIAC: high precision coreference with limited knowledge and linguistic resources. In *Proceedings of ACL/EACL workshop on Operational factors in practical, robust anaphora resolution (ACL97)*, pp 38-45.

Ferrández, A., Palomar, M., & Moreno, L. (1999). An empirical approach to Spanish anaphora resolution. *Machine Translation*, 14: 2-3.

Gaizauskas, R., Wakao, T., Humphreys, K., Cunningham, H., & Wilks, Y. (1995). Description of the LaSIE system as used for MUC-6. In *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, Morgan Kaufmann, pp 207-220

Kennedy, C. & Boguraev, B. (1996). Anaphora for Everyone: Pronominal Anaphora resolution without a Parser. In *Proceedings of 16th International Conference on Computational Linguistic (COLING '96)*, pp 113-118.

Lappin, S. y Leass, H.J. (1994). An Algorithm for pronominal anaphora resolution. *Computational Linguistic*, 20 (4): 535-561.

Mitkov, R. (1998). Robust pronoun resolution with limited knowledge. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistic and 17th International Conference on Computational Linguistic (COLING-ACL '98)*, pp 869-875.

Muñoz, R., Palomar, M., & Ferrández, A. (2000). Processing of Spanish Definite Descriptions. In *Proceedings of the Mexican International Conference in Artificial Intelligence*. Lectures Notes in Artificial Intelligence, 1973. Acapulco (Mexico). Springer-Verlag, pp 526-537.

Muñoz, R., Montoyo, A., Llopis, F. & Suárez, A. (1998) Reconocimiento de entidades en el sistema EXIT. *Procesamiento del Lenguaje Natural*, 23: 47-53

Spark-Jones, K. (1993). What may be in a summary? In *Information Retrieval*, 93: Von der Modellierung zur Anwendung, pp 9-26.