

Compound Temporal Adverbs in Portuguese and in Spanish*

Jorge Baptista¹, Dolors Català Guitart²

¹ Universidade do Algarve – FCHS, Campus de Gambelas, P – 8000 Faro, Portugal
jbaptis@ualg.pt

<http://w3.ualg.pt/~jbaptis>

² Universitat Autònoma de Barcelona - Departament de Filologia Francesa i
Romànica, edifici B, E - 08193 Belaterra, Barcelona, Spain

dolors.catala@uab.es

<http://seneca.uab.es/filfrirom>

Abstract. This paper reports on an ongoing research on temporal adverbs and deals with the problem of processing a family of Portuguese and Spanish compound temporal adverbs, in a contrastive approach, aiming at building finite state transducers to translate them from one language into the other. Because of the large number of combinations involved and their complexity, it is not easy to list them in full. However, their modularity and relative independence from the surrounding sentence make them especially apt for a formal description using a finite state approach.

1 Introduction

Time-related nouns (*Ntmp*) Pt: *manhã, tarde, noite*/ Sp: *mañana, tarde, noche* ‘morning, afternoon, evening/ night’, here abbreviated as *MTN*, often appear in texts as part of complex temporal adverbs [1], [6], [7]. Those circumstantial adverbs of time are prepositional phrases basically formed by the combination of a preposition, a determinant, a *Ntmp* and eventually by a modifier. A detailed study of the sequences in which they appear shows they enter many combinations with other *Ntmp* and time-related adverbs forming complex temporal adverbs. Most of them are semantically transparent but syntactically frozen.

2 Simple Combinations: *Prep Det Ntmp*

We start with the simplest adverbs, formed by a preposition, a determiner and the *Ntmp* =: *MTN*. Concerning preposition, we have considered those expressing a *basic* temporal localization (Pt: *a, de, em, por*/ Sp: *a, de, en, por*), *duration* (Pt: *durante, ao longo de*/ Sp: *durante, a lo largo de*), *beginning* (Pt: *a partir de, desde*/ Sp: *a partir*

* This research was partly supported by FCT (Project POSI/PLP/34729/99) and EU (Socrates).

E.M. Ranchhod and N.J. Mamede (Eds.): PorTAL 2002, LNAI 2389, pp. 133-136, 2002.

© Springer-Verlag Berlin Heidelberg 2002

de, desde), *end* (Pt: *até a/ Sp: hacia, hasta*) and also those indicating *approximate* indications (Pt: *perto de, por volta de/ Sp: sobre*)¹.

In some cases, there is a complete lexical correspondence between both languages, as can be seen in the following pairs: Pt: *O João fez isso de (manhã + tarde + noite)/ Sp: Juan ha hecho eso de (mañana + tarde + noche)* ‘John did this in the morning, afternoon, evening/ at night’². In both languages these *Ntmp* can combine with the indefinite article, but then the adverbs do not express temporal localization, but rather the duration of the process: Pt: *O João fez isso numa (manhã + tarde + noite)/ Sp: Juan ha hecho eso en una (mañana + tarde + noche)* ‘John did this in one morning, afternoon, evening/ night’. Also, both Portuguese and Spanish do not allow combinations of these three *Ntmp* with *Prep* =: *em/en* with the definite article *la* unless they are followed by a modifier: Pt: *O João fez isso na (manhã + tarde + noite) (*E + de domingo) / Sp: Juan ha hecho eso en la (mañana + tarde + noche) (*E + de domingo)* ‘John did this in the morning/ afternoon/evening/ at night of Sunday.

However, soon many differences appear. With *Prep* =: *a*³, Portuguese admits only two *Ntmp*, while Spanish has none of these forms: Pt: *O João fez isso à (*manhã + tarde + noite)/ Sp: Juan ha hecho eso a la (*mañana + *tarde + *noche)* ‘John did this at the morning/ afternoon/ evening/ at night’.

In both languages, there are three basic demonstratives Pt: *esta, essa, aquela/ Sp: esta, esa, aquella*. There is no adverb with *Prep* =: *a* or *de* with these determiners in either languages. In Spanish the demonstratives can combine with *Prep* =: *en*, and this *Prep* can also be reduced before them. However, in Portuguese several restrictions can be observed: *Prep* =: *em* is not allowed before demonstrative *esta* while *essa* and *aquela* are obligatorily introduced by *em*.

This kind of restriction can vary depending on the preposition introducing the adverb, the *Ntmp* itself and the determiner of the *Ntmp*. Therefore, a very detailed description of such combinations is necessary in order to build finite state transducers able to make adequate translations between the two languages.

3 Complex Combinations

The basic adverbs mentioned above can further accept several modifiers, involving other *Ntmp*, for example:

- (i) the *days of the week*: Pt: *O João fez isso na manhã de sexta-feira/ Sp: Juan hizo eso en la mañana del viernes* ‘John did this on the morning of Friday’;
- (ii) *dates* (day - month - year, either isolated or in combination): Pt: *O João fez isso na manhã de 14 de Abril de 2002/ Sp: Juan hizo eso en la mañana del 14 abril de 2002* ‘John did this on the morning of April 14, 2002’; and
- (iii) a limited set of time adverbs, e.g. Pt: *ontem, hoje* and *amanhã/ Sp: ayer, hoy* and

¹ For clarity of presentation, examples in this paper will be taken mostly from the set of basic time-location adverbs.

² The English literal translation of the examples is only meant to show the syntactic phenomena and its acceptability is irrelevant for the purpose of this paper.

³ In Portuguese some *Prep* are contracted with the articles; in Spanish this morphological contraction usually does not occur, e.g. Pt: *na* (= *em + a*) *manhã de ontem/ Sp: en la mañana de ayer* ‘yesterday morning’.

mañana): Pt: *O João fez isso durante a manhã de ontem*/ Sp: *Juan hizo eso durante la mañana de ayer* ‘John did this during the morning of yesterday’;

These temporal elements can also be combined in (sometimes long) appositive sequences: Pt: *O João fez isso na sexta-feira, dia 14 de Abril de 2002 pela manhã*/ Sp: *Juan hizo eso el viernes, día 14 abril de 2002 por la mañana* ‘John did this on Friday, April 14, 2002 by the morning’.

The *Ntmps* =: *MTN* can also appear as modifiers of time adverbs built around the *Ntmp* =: *hora* ‘hour’: Pt: *O João fez isso às duas (E + horas) da (manhã + tarde + noite)*/ Sp: *Juan hizo eso a las dos (E + horas) de la (mañana + tarde + noche)* ‘John did this at two o’clock of the morning/ afternoon/ night’. Cultural differences between the two languages give rise to different sets of values that can be associated with each *de MTN* modifier. For instance, Spanish speakers would accept: Sp: *Juan hizo eso a las ocho (E + horas) de la (tarde + noche)* ‘John did this at eight o’clock of the afternoon/ night’, but Portuguese speakers would rather use: Pt: *O João fez isso às oito (E + horas) da (*tarde + noite)* ‘John did this at eight o’clock of the night’.

4 Finite State Transducers

Temporal adverbs such as those briefly described above ⁴ constitute a set of linguistic expressions particularly well suited for representation by means of finite state transducers (noted FST) ⁵, both for their modularity and for their relative independence from the sentence in which they appear. The method of representing them in FST has proved both efficient and adequate [1], [6], [7]. Our purpose here, however, is not only to build a set of lexical FST to be used in recognition of adverbs in a text, but also to use them as a starting point to build FST that could provide appropriate translation for each adverb, having Portuguese and Spanish both as source and target languages ⁶. To do this, two separate sets of FST were built, one for each source and target language-pair (i.e. language-pairs *pt* > *sp* and *sp* > *pt*), using *INTEX* ⁷ linguistic development environment both to create the FST and to apply them to corpora ⁸.

⁴ There are even more complex combinations designating time intervals, composed of adverbs expressing beginning and end, e.g.: Pt: *O João trabalhou das 8 (E + horas) da noite de ontem até às 6 (E + horas) da manhã de hoje*/ Sp: *Juan ha trabajado de las ocho (E + horas) de la noche de ayer hasta las 6 (E + horas) de la mañana de hoy* ‘John worked from 8 o’clock of the evening of yesterday until 6 o’clock of the morning of today’, but these were not taken in consideration in this paper.

⁵ See [4], [5], [10], [11], [12], [13] for an overview of the use of FST in linguistic description.

⁶ In our FST, both the *input* and the *output* are inflected forms of the source language and target language, respectively. Therefore, already available linguistic resources for Portuguese [3], [8], [9] and Spanish [2] were not used at this stage. For certain adverbs that include free elements, such as the adjective *fria* (cold) in the otherwise compound adverb *numa (E + fria) manhã de Inverno* ‘in a cold winter morning’ it will be necessary to use them.

⁷ See [12], [13] for a detailed description of *INTEX*.

⁸ The Portuguese corpus consists of the first fragment of the *CETEMPúblico*, taken from the daily newspaper *Público* (<http://cgi.portugues.mct.pt/cetempublico>, 2002-04-02). It forms a text file of 58,790 KB, with 9,632,423 words. The Spanish corpus is somehow similar in content, but a little smaller. It is composed of texts from the CD edition of the newspaper *El Mundo* of 1995 and consists of a text file of 40,497 KB, with 6,540,493 words.

Up to now, a large variety of combinations have been described for both language-pairs. It is difficult to report precise figures because local grammars for expressions involving numerical values, i.e. *hour* and *dates*, generate an overwhelming number of combinations. If we disregard those grammars, the Portuguese FST of compound adverbs represents over 8,500 different combinations, and the Spanish FST over 9,300.

5 Final Remarks

Considering the family of time-adverbs here described for the purpose of translation, the differences between the two languages are mainly syntactic. The forms of the source language have to be mapped onto the target language with extreme detail. For the most part, an equivalent, if somehow slightly different, form exists. Several elements that compose these complex adverbs present some modularity, and are often used in different combinations, which make them particularly apt for a finite state approach.

References

1. Baptista, J: Manhã, tarde, noite. Analysis of temporal adverbs using local grammars. *Seminários de Linguística* 3 (1999) 5–31
2. Blanco, X.: Les dictionnaires électroniques de l'espagnol (DELASs et DELACs). 17^{ème} Colloque Lexiques et Grammaires Comparés (Guernesey, October 4-7, 1998)
3. Eleutério, S., Ranchhod, E., Freire, H. Baptista, J.: A System of Electronic Dictionaries of Portuguese. *Linguisticae Investigationes* 19-1 (1995) 57–82
4. Gross, M.: The Construction of Local Grammars. In Schabes, Y., Roche, E. (eds.): *Finite State Language Processing*. MIT Press/Bradford. Cambridge/ London (1997) 329–354
5. Gross, M.: Construção de gramáticas locais e autómatos finitos. In Ranchhod, (org.) 2001: *Tratamento das Línguas por Computador. Uma Introdução à Linguística Computacional e suas Aplicações*. Caminho, Lisboa (2001) 91–131
6. Maurel, D.: Adverbes de date: étude préliminaire à leur traitement automatique. *Linguisticae Investigationes* 14–1 (1990) 31–63
7. Maurel, D.: Reconnaissance automatique d'un groupe nominal prépositionnel. Exemple des adverbes de date. *Lexique* 11 (1992) 147–161
8. Ranchhod, E.: O uso de dicionários e de autómatos finitos na representação lexical das línguas naturais. In Ranchhod, E. (org.): *Tratamento das Línguas por Computador. Uma Introdução à Linguística Computacional e suas Aplicações*. Caminho, Lisboa (2001) 13–48.
9. Ranchhod, E., Mota, C., Baptista, J.: A Computational Lexicon of Portuguese for Automatic Text Parsing. SIGLEX'99: Standardizing Lexical Ressources. 37th Annual Meeting of the ACL. College Park, Mariland, USA. (1997) 74–81
10. Ranchhod, E. (org.): *Tratamento das Línguas por Computador. Uma introdução à Linguística Computacional e suas Aplicações*. Caminho, Lisboa (2001)
11. Schabes, Y., Roche, E. (eds.): *Finite State Language Processing*. MIT Press/Bradford. Cambridge/ London (1997)
12. Silberztein, M.: *Dictionnaires électroniques et analyse automatique de texts. Le système INTEX*. Masson, Paris (1997)
13. Silberztein, M.: *INTEX Manual*. ASSTRIL, Paris (2000). <http://www.bestweb.net/~intex/downloads/Manual.pdf> (2002-04-02)