Towards Temporal Reasoning in Portuguese

Livy Real, Alexandre Rademaker, Fabricio Chalub, Valeria de Paiva
USP, IBM Research, Nuance Communications
livyreal@gmail.com, alexrad@br.ibm.com, fchalub@br.ibm.com, valeria.depaiva@nuance.com

Abstract

This paper describes our ongoing work to create a temporally annotated open Portuguese corpus. We discuss how this task helped to improve and evaluate linked open lexical resources in Portuguese, namely OpenWordNet-PT and TempoWordNet. We use the Linguateca's Bosque corpus, which we annotated with Universal Dependencies (UD2.0) and the system HeidelTime, the state of the art open source time tagging, to build Bosque-T, our proposed temporal corpus.

Keywords: Portuguese, temporal expressions, HeidelTime, WordNet, corpus

1. Introduction

Although time and temporal reasoning pose many problems in language and logic (Steedman, 2005), much improvement has been achieved on temporal information tagging and retrieval in the last decade. At least since the first TempEval, 2007, there has been a concerted effort towards temporal tagging. Systems are performing close to inter-annotator reliability for entity recognition (UzZaman et al., 2014), different domains are being explored (Bethard et al., 2015) and more complex tasks are being addressed, such as temporal relation typing (Derczynski, 2016). While much progress can be found for English temporal processing, the situation for languages other than English is not so positive. However recently, HeidelTime (Strötgen and Gertz, 2015) was made available for 13 languages, including Portuguese, with an automatically built expansion that promises to deal with more than 200 languages.

Here we will concentrate on verifying how much of the traditional wisdom in dealing with time in English and in multilingual projects can be re-purposed, wholesale, for dealing with time in Portuguese. We focus on the HeidelTime system and linguistic linked open resources, namely, OpenWordNet-PT (de Paiva et al., 2012) and TempoWordNet (Dias et al., 2014), linked through the OpenMultilinguaWordNet project (Bond and Foster, 2013). Using non-language-specific tools for bootstrapping the creation of preliminary systems and linguistic resources to less resourced languages is useful in many ways. It creates baselines to compare further work to and it serves to start investigating applications that depend on the kind of data desired. Our intended applications depend on temporal data, so a preliminary investigation of tools and data for dealing with it is a requirement for our project.

We start by investigating what is the state-of-the-art for recognizing time expressions in Portuguese and progress to verify how good our lexical resources are for this first level of investigation. We aim at a fully fledged description of a temporal logic system, similar to the one in (Crouch and de Paiva, 2014), but we need to make sure that the basics (lemmas, word senses, relationships for temporal expressions) are in place for Portuguese.

Steedman (Steedman, 2005) and Crouch (Crouch, 1998) start by discussing what a very naive approach to modelling temporal effects in natural language could be, simply using logical operators for the past and the future. In the simplest possible case this would give us a modal logic with two tense operators, P (for past) and F (for future), applying to propositions ϕ that are evaluated in a model M. When using logic to represent the meanings of natural language sentences, it is assumed that the temporal index of evaluation for the whole proposition is set to the time s at which the utterance is made — the speech time. Thus, for example:

1. "John was in London" is true uttered at s iff $[P(in(john, london))]$ holds,
2. "John is in London" is true uttered at s iff $[in(john, london)]$ holds,
3. "John will be in London" is true uttered at s iff $[F(in(john, london))]$ holds.

in a given M model. The past tense formula evaluated at the speech time s shifts the temporal index to an earlier time — call this the event time — and evaluates the embedded (present tense) proposition relative to the event time. The absence of any operator, as in the present tense formula, means that the speech and event times are identical.

Although there are a number of shortcomings to this particular approach as a linguistic representation, we still want to have for Portuguese the ability to discuss these paradigmatic simple examples of sentences, in the most direct form possible. Thus the direct translations of the sentences above

1. "João esteve em Londres" is true uttered at s iff $[P(in(joao, londres))]$ holds,
2. "João está em Londres" is true uttered at s iff $[in(joao, londres)]$ holds,
3. "João vai estar em Londres" is true uttered at s iff $[F(in(joao, londres))]$ holds.

need to define a completely trivial temporal system in Portuguese, the same way that they do in English. While it seems clear that the *tense systems* are very different in English and Portuguese and that hence temporal markings might need to be modified and adapted, we survey the commonalities between the problems and solutions first. We aim, just like (Costa and Branco, 2012a), to import open good tools we may find to help with the task at hand.

Here we describe first steps towards temporal tagging in Portuguese, that are needed for an eventual temporal reasoning. We start by checking how well HeidelTime works for Portuguese and how much of the needed temporal information is present in OpenWordNet-PT (OWN-PT)(de Paiva et al., 2012), the open wordnet that we have been working on since 2012. Connecting our lexical resources, we use open linked resources (LLOD) (Chiarcos et al., 2012) for the usual reasons: interoperability of existing language resources, e.g. easy retrieval and integration with other resources, easy and local updates, possibilities for crowd sourcing information needs, etc. In particular OWN-PT is linked to OpenMultilinguaWordNet (OMW)(Bond and Foster, 2013), which links several other WordNet projects, including TempoWordNet (TempoWN)(Dias et al., 2014). We expected that the temporal information present in TempoWN would be valuable to improve OWN-PT and to help make sure that the basics are in place to allow temporal extraction in Portuguese, but it is not clear that it is. The contributions of this preliminary investigation are: 1) BosqueT, a Portuguese corpus tagged by HeidelTime and a manual assessment of the data produced; 2) the improvement of OpenWordNet-PT's synsets related to temporal information; 3) an assessment of the quality found in TempoWordNet and of the usefulness of using its linked knowledge for Portuguese processing.

1.1. Related Work

Different approaches to temporal information retrieval arose in the last few years. Many of them are libraries or specific modules of Natural Language Processing pipelines that normalize temporal expressions. A reasonable number of lexical resources have also been constructed for this task. Here we briefly describe some libraries and resources available for Portuguese processing. As usual, most of the work has been done only for English, but we can also find several recent works using a multilingual strategy. Few works are specifically concerned with Portuguese processing and most of those are not open source, unfortunately.

There are not so many open source systems for NLP in Portuguese, but HAREM, the shared evaluation task in Portuguese, did discuss temporal expressions. HAREM (Mota and Santos, 2008) is a series of shared tasks organized by Linguateca [1] for Named Entity Recognition, whose last edition was held in 2008. HAREM's discussions and guidelines for time expressions in Portuguese uses a specific tagset that was built for the state of the art of Portuguese processing at that time. Its aim was to be useful to the Lusophone NLP community. However, the exact tagset used in HAREM is not shared with a large community, which makes the task of comparing HAREM results with any other tools or data quite difficult, as discussed in (Real and Rademaker, 2015).

Other work on Portuguese time expressions includes the LX-TimeAnalyzer (Costa and Branco, 2012b), the STRING system (Mamede et al., 2012) and specifically their temporal analyzer (Hagège et al., 2010). Mostly this work is based on proprietary systems and hence re-using it is not easy. The LX-TimeAnalyzer, for example, is made available for the community in a browsable version,[2] but its code is not open.

Turning to open tools, there is the work on Freeling (Padró and Stanilovsky, 2012) and on the HeidelTime (Strötgen and Gertz, 2015) framework. Freeling offers a date recognition module and two modules for Named Entities recognition, but we have not seen data about their accuracy or precision, either in English or Portuguese. Since HeidelTime offers dates normalization, but also offers other kinds of temporal expressions recognition and uses the same annotation as the TempEval evaluations, we opted to start our investigation with HeidelTime.

2. Resources

Many systems for temporal tagging do not rely on using information present in lexical resources. We believe, as do (Costa and Branco, 2012b), that combining the knowledge of wordnets with the knowledge of temporal oriented systems can improve the quality and coverage of both kinds of systems. This needs to be a two-way road: one can improve the coverage of the lexical resource considering the output of the temporal system and conversely one can improve the temporal tags, if we have more lexical knowledge. For instance, one needs to recognize adverbial expressions – such as *yesterday, today, tomorrow*, respectively *ontem, hoje, amanhã* – and these temporal expressions are not always recognized as such. More difficult is to correctly detect highly ambiguous words, such as *último/last* and *anterior/previous*, similarly ambiguous in Portuguese and English, whether they are used in temporal contexts or not. For this kind of sub-problem, lexical resources can be very helpful. We discuss below the two resources we use in this work, as well the Bosque corpus and the HeidelTime system.

2.1. OpenWordNet-PT

OWN-PT[3] is an open access wordnet for Portuguese, originally developed as a syntactic projection of Universal WordNet (De Melo, 2009). OWN-PT is linked to Bond's collection of open wordnets Open Multilingual Wordnet (OMW) [4] see (Bond and Foster, 2013). These wordnets are of varying size and quality, but the Portuguese version, at approximately half the number of synsets of the English WordNet, is reasonably comprehensive. It is hoped that the ability to connect the different wordnets helps to complete each one individually. There is some evidence for that and this work corroborates it, as it uses temporal information in English to annotate Portuguese synsets. Due to the construction process of this Portuguese wordnet, all the original English synsets are present in OWN-PT, but not all of them have Portuguese words and many glosses and examples are still missing. Automatic translations of glosses are available, and they are being manually checked, but the process is ongoing. We are engaged in completing the translation of the empty OWN-PT synsets, but since this

[1] https://www.linguateca.pt/HAREM/

[2] http://nlxserv.di.fc.ul.pt/lxtimeanalyzer.

[3] http://wnpt.brlcloud.com/wn/

[4] http://compling.hss.ntu.edu.sg/omw/

consists of a long term work, we focus on subsets of synsets related to specific tasks. Considering the synsets related to time expressions seems an interesting and productive idea, which is also related to our work on Portuguese processing of historical data (Paiva et al., 2014).

Princeton WordNet (PWN) classifies as temporal nouns in 1028 synsets, the `noun.time` lexicographer file. Of these, more than 200 synsets still have no Portuguese translations at the moment[5].

2.2. TempoWordNet

TempoWN[6] (Dias et al., 2014) is a lexical knowledge base for temporal analysis where each synset of PWN is assigned an intrinsic temporal value. TempoWN is already linked to OMW, so using its data for improving OWN-PT is easily achieved. Each synset of TempoWN is semi-automatically time-tagged with four labels: atemporal, past, present and future and a confidence level. Temporal classifiers were learned from a set of time-sensitive synsets (manually curated) and then applied to the whole resource to give rise to TempoWN. So, each synset is augmented with its calculated qualitative temporal value. Perhaps the main difference between TempoWN and other resources and tools for temporal expressions recognition is the fact that TempoWN always tags a synset with a temporal value, even if most of the synsets have the 'atemporal' time value assigned.

Using the standard WordNet domain classification for nouns, we know which ones of the 82,115 noun synsets are related to time, the 1028 `noun.time` synsets. However there are no easy ways of determining how many adjectives, verbs and adverbs are time-related. These other parts of speech can also be related to temporal features, but this classification does not exist in Princeton WordNet itself. Thus the use for us of TempoWN and its link to OMW would be to check how many temporal adjectives, adverbs and verbs should be in OWN-PT. We aim to detect, amongst the many adjectives, verbs and adverbs that exist in English and that are empty in Portuguese, the ones that are temporally cogent.

2.3. HeidelTime

HeidelTime[7](Strötgen et al., 2013) is a multilingual, cross-domain temporal tagger that extracts temporal expressions from documents and normalizes them according to the TIMEX3 annotation standard. This standard uses the markup language TimeML (Pustejovsky et al., 2003). HeidelTime uses different normalization strategies depending on the domain of the documents that are to be processed, be them news, narratives (e.g., Wikipedia articles), colloquial (e.g., SMS, tweets), or scientific (e.g., biomedical studies). The tool is a rule-based system and its source code and the resources (patterns, normalization information, and rules) are strictly separated. Since 13 languages are supported with manually developed resources and Portuguese is one of these, we chose to investigate it for our work.

[5]March, 2018.

[6]`https://tempowordnet.greyc.fr/`

[7]`https://github.com/HeidelTime/heideltime`.

2.4. The Bosque corpus

The Bosque corpus is a subset of 'Floresta Virgem', a collection of Portuguese treebanks distributed by Linguateca[8]. According to the creators in their website, the corpus Bosque is "fully revised and corrected in the scope of the project, with a current size of 162,484 lexical units". The Bosque corpus has 9,368 sentences, corresponding to 1,962 different extracts from newspaper text. But many of these 9,368 sentences are no grammatical sentences. Since the corpus was extracted from newswire, there are many headlines that are simply noun phrases like *PT no governo* (The Workers Party (PT) in Power). There are also dialogues, recognizable through the use of the names of the interlocutors, and answers to questions, which tend not to be full grammatical sentences. Still, Bosque is probably the most used corpus in the Lusophone community, it has both Brazilian and European Portuguese variants and has been annotated using several different linguistic theories. Most recently it has been converted to Universal Dependencies version 2.0 (Rademaker et al., 2017). The statistics derived from the Universal Dependencies annotation of the corpus are useful for the work of temporal extraction and the syntactic dependency trees themselves might prove even more useful.

3. Bosque-T

We ran the stand alone version of HeidelTime in our Bosque corpus, creating a temporally annotated corpus in Portuguese. We call this temporally annotated version of the corpus Bosque-T[9]. The main purpose of Bosque-T is to be used as a baseline for future work on temporal extraction. This is similar to the work on TimeBank-PT (Costa and Branco, 2012c), but uses an open source temporal tagging system that is officially the state-of-the-art and that is available to all.

TimeBank-PT is according to its creators 'the result of translating the English corpus used in the first TempEval challenge to the Portuguese language'. While TimeBank-PT is TimeML annotated, it is a translation of an English corpus, not originally Portuguese texts. By contrast, the HAREM data collection is 'truly' Portuguese, but it does not use TimeML guidelines, which have become the 'de facto' standard in temporal annotations. Therefore, as far as we know, our work is the first open corpus that uses the TIMEX3 tagset, from the TimeML temporal markup language, in an original Portuguese corpus.

Out of the 1962 extracts, HeidelTime says 741 have no time annotations at all. Many of the sentences on these extracts do have temporal expressions, but these were not found by the tool. For instance, in the extract[10]

> Em relação ao mesmo mês do ano passado, quando os negócios atingiram 139,8 toneladas de

[8]`http://www.linguateca.pt/floresta/info_floresta_English.html`

[9]Available at `https://github.com/own-pt/portuguese-time`.

[10]In comparison to the same month last year, when business achieved 139,8 tons of gold, the reduction was of 61,37%. The daily average in that month was 6,6 ton, according to data from the Bolsa de Mercadorias e Futuros.

> ouro, a redução é de 61,37%. A média diária naquele mês foi de 6,6 toneladas, segundo dados da Bolsa de Mercadorias e Futuros.

no timex was found. But we should have seen *mesmo mês do ano passado/same month last year* and *média diária/daily average*, which are clearly temporal expressions.
Given that HeidelTime is rule-based, we expected that it would be able to detect all expressions composed by digits or expressions that tend to be always related to time, as the names of the months. But this does not always happen. For example, no timex was found in either of the sentences below.

> A cotação para maio ficou em 20.000 pontos[11]

> Empresa funciona das 9h às 19h, diariamente.[12]

In total HeidelTime identified 2464 tags, 644 unique ones, of different types. Most of the ones identified were dates. Almost 300 timex occurrences were the word *ontem* (yesterday). Several temporal expressions were correctly marked, from full dates such as *dia 23 de maio de 1972* (day 23 of May of 1972) to some complex phrases such as *há cerca de 20 anos* (around 20 years ago).
Nevertheless amongst the expressions found, we also find (interesting) mistakes. In the excerpt[13]

> Manifestações espontâneas em protesto contra o facto de Daniel Cohn-Bendit, líder do Maio de 68, ter sido proibido de residir em França.

the expression *Maio de 68* (May of 68), a well-known French political movement, which is in Wikipedia-PT[14], was tagged as DATE, instead of being considered a named entity.
To see the kinds of issues that are problematic with the tagging, we choose some random 20 extracts from Bosque-T to verify HeidelTime choices on these. Many temporal expressions are missed or half-marked. For example, in the sentence[15]

> A mudança do local de jogo que deve acontecer também na partida contra o Corinthians, no <TIMEX3>próximo</TIMEX3> dia 17 foi determinada pela CBF, que não viu garantias de segurança no estádio santista.

the term *próximo* (next) is correctly tagged, but the actual date *dia 17* (day 17) was not.

Simply looking at the expressions produced by HeidelTime, we can see that a traditional way of referring to the past in Portuguese is missing altogether from the terms produced. For example the sentence[16]

> Monique, 37, disse que descobriu a marquinha, que não é pedra no rim quando se separou do marido, em junho passado.

should have *junho passado* (last June) marked. Not a single *passado* (last, just passed) appears in our HeidelTime terms.
It is also clear that more subtle ways of referring to time are much harder to tag. For example in the sentence [17]

> Eles se dizem oposição, mas ainda não informaram o que vão combater.

the word *ainda* (yet) can be a temporal marker, indicating that a event has not happened so far. These harder, more subtle ways of referring to time, we expected to be missing from the off-the-shelf running of HeidelTime. Also while a full date, such as *dia 23 de maio de 1972* is easy to recognize and tag, a partial date, such as the year *1995* in the sentence [18]

> A seca que atingiu as áreas produtoras de grãos não deve causar grandes estragos na safra <TIMEX3>1994</TIMEX3>/95.

does not get recognized as a date.
Several of the holidays that we have been trying to complete in OWN-PT are not marked by HeidelTime as temporal events, yet. For example the sentence[19]

> Pizzaria oferece cardápio especial para Páscoa.

needed to mark *Páscoa* (Easter) as a temporal noun, as it is marked in English. We recognize that what the HeidelTime developers call "temponyms" (Kuzey et al., 2016) are not fully developed, yet for other languages. They only exist for English, hence given the sentence[20]

> Muito mais do que nos tempos da ditadura, a solidez do PT está, agora, ameaçada.

we did not expect the expression *tempos da ditadura* (dictatorship times) to be marked. However we did expect the word *tempos* (times) to be recognized as a temporal marker and tagged.
We are now in the process of checking the markings we have and verifying their accuracy. We plan to 'triangulate' information provided by OWN-PT for the sentences, with the HeidelTime tags in the near future.

[11] The price for May stood at 20,000 points.

[12] Company operates from 9 am to 7 pm, daily.

[13] Spontaneous demonstrations protesting against the fact that Daniel Cohn-Bendit, leader of May 1968, was banned from residing in France.

[14] https://pt.wikipedia.org/wiki/Maio_de_1968

[15] The change of place for the match, which should happen also in the match against the Corinthians on the next 17th, was determined by the CBF, which did not see guarantees of security measures in the Santos stadium.

[16] Monique, 37, said that she discovered the little mark, not a kidney stone, when she got divorced from her husband last June.

[17] They say they're the opposition, but have not informed us, yet, what they will oppose.

[18] The drought that hit the grain growing areas should not cause a big disaster in the harvest year 1994/95.

[19] Pizzaria offers special menu for Easter

[20] More than in the times of the dictatorship, the existence of the PT is now threatened.

4. Linked Open Data for Temporal Tagging

In this section we discuss how to improve the annotated corpus making use of the linked resources we have at hand. We also mention how OWN-PT can benefit from this work.

Since TempoWN scores all PWN synsets with a temporal value, for this preliminary work, we considered only the synsets whose probability of being `PAST` or `FUTURE` according to TempoWordNet is above 90 percent. This includes more than 3K synsets. Since TempoWN is not manually curated, as PWN and OWN-PT are, we started to manually check the quality of these probability assignments and unfortunately we found many labels that we do not agree with and that do not seem very useful for the present task.

For example, the synset that has the higher probability, 0.998, of being `PAST` is `00012689-a: ideal | constituting or existing only in the form of an idea or mental image or conception`. While one can try to force the interpretation that this abstract image needs to be formed in the past to exist, there is nothing that really connects it to the usual notion of `PAST`.

At first glance, TempoWN has a large coverage that seems to be useful for temporal tagging, but its information is too noisy to be useful. Checking simply the most frequent timex expressions in Bosque-T in TempoWN and OWN-PT, we could complete some missing synsets in Portuguese, but we should not use the extra time score offered by TempoWN. While the synset for *ontem*(yesterday) has more than 0.99 probability of being `PAST` and *agora* (now) also scores 0.99+ possibility of being `PRESENT`, some other probability assignments seem dubious. The synset for *hoje* `00207366-r | today | on this day as distinct from yesterday or tomorrow`, appears in TempoWN with 0.99+ probability of being `FUTURE` and *próximo* `00054212-r | next | at the time or occasion immediately following` has 0.99+ probability of being `PAST`.

We reap the benefits of linked linguistic open data through the connection established between TempoWN, OMW and OWN-PT. However, it is harder to decide if the TempoWN information is useful for the task at hand or not. The markings of adjectives and adverbs should be useful for reasoning with texts in Portuguese, if the probability assignments are reasonable. Many of them seem good, but how to improve TempoWN scores is future work.

Many of the timex expressions found in Bosque-T were missing in OWN-PT at the beginning of this work, for instance the synset `00065748-r | last | most recently`. While in English, this is clearly an adverb, in Portuguese, we need an adverbial phrase to convey the same kind of meaning *por último* ("by last").

For this preliminary work more than 300 temporal synsets were completed in OWN-PT. Many language or culture specific ones are still missing. Some of these empty Portuguese synsets are typical holidays in the United States, such as the synset `15189982-n` for *Father's Day*. There is a holiday called Father's Day (*Dia dos Pais*) in Portuguese. But it happens at different times in Brazil (August) and Portugal (March), while it happens in June in the US and England. Thus, in PWN, this synset holds a relationship with *June*, which only makes sense for the English wordnet. This hints at the issues at the intersection of multilingual and multicultural aspects of lexical and world knowledge. Looking at these translations also helps to notice smaller differences between the languages. A typical and principled difference between the wordnets is that we do not use a prefix like "mid" in the synset `15211711-n` for *mid-May*; we say instead *meados de maio*, which although can be seen as a multi-word expression, is compositional in Portuguese and therefore it may not necessarily be included in a Portuguese lexical base if multilingual alignment was not a previous goal.

5. Conclusions

We presented our ongoing work towards temporal tagging, as a pre-requisite for temporal reasoning in Portuguese. Since not much is available for Portuguese natural language processing, we started by providing an open corpus temporally tagged by the HeidelTime tool, which we call Bosque-T. In the process of analyzing the annotations of Bosque-T, we improved somewhat the coverage of OpenWord-Net temporal synsets and discussed how its link to a temporally-annotated wordnet, TempoWordNet, could be useful for this task.

Due to the different building processes of OpenWordNet-PT and TempoWordNet, the quality of those resources is radically different. While OpenWordNet-PT has less, but reliable information, TempoWordNet offers temporal scoring for every synset of Princeton WordNet, but most of the scores are controversial. We briefly discussed the issues found in Bosque-T, which show that much work still needs to be done to address temporal tagging in Portuguese – at least as far as using open-source tools and resources is considered. We aim to use Bosque-T as a baseline for this future work.

For future work we would like to improve the Portuguese HeidelTime system, using the insights gained from analyzing the issues found in Bosque-T. We also want to manually annotate a small part of the Bosque corpus with the TIMEX3 tagset to make it available as a small golden corpus. Checking how well HeidelTime deals with TimeBank-PT and the HAREM corpora are also possible next steps. Finally maybe one should try a deep analysis of the proposed adaptation of the TimeML guidelines to Portuguese, as proposed by (Hagège et al., 2010).

We are interested in temporal reasoning, not only in temporal Information Retrieval. As a long term goal, we aim to merge temporal information with other linguistic levels. We plan to do so using Bosque-UD, the human revised version of the Bosque corpus annotated with Universal Dependencies. Despite the issues with the quality of TempoWordNet annotations, the mappings provided by the use of linked open data were useful in helping us improve our own annotations. We plan to use the data in the Portuguese DBPedia [21] to help with some of the culturally specific problems, such as named holidays.

[21] `http://pt.dbpedia.org`

6. Bibliographical References

Bethard, S., Derczynski, L., Savova, G., Pustejovsky, J., and Verhagen, M. (2015). Semeval-2015 task 6: Clinical tempeval. *SemEval 2015*.

Bond, F. and Foster, R. (2013). Linking and extending an Open Multilingual Wordnet. In *ACL, 2013*.

Chiarcos, C., Hellmann, S., and Nordhoff, S. (2012). Linking linguistic resources: Examples from the open linguistics working group. *Linked Data in Linguistics. Representing Language Data and Metadata*, Springer:p. 201–216.

Costa, F. and Branco, A. (2012a). Extracting temporal information from Portuguese texts. In *PROPOR 2012*, volume 7243 of *Lecture Notes in Artificial Intelligence*, pages 99–105, Berlin, Germany. Springer.

Costa, F. and Branco, A. (2012b). LX-TimeAnalyzer: A temporal information processing system for Portuguese. Technical Report DI-FCUL-TR-2012-01, Universidade de Lisboa, Faculdade de Ciências, Departamento de Informática.

Costa, F. and Branco, A. (2012c). TimeBankPT: A TimeML annotated corpus of Portuguese. In *LREC'12*, pages 3727–3734, Istanbul, Turkey, May. European Language Resources Association (ELRA).

Crouch, D. and de Paiva, V. (2014). If, not when. *Electronic Notes in Theoretical Computer Science*, 300:3 – 20. IMLA 2013, {UNILOG} 2013.

Crouch, D. (1998). Temporality in natural language. In *ESSLLI class notes*, Saarbruecken, Germany.

De Melo, G.and Weikum, G. (2009). Towards a universal wordnet by learning from combined evidence. In *18th ACM conference on Information and knowledge management*. ACM.

de Paiva, V., Rademaker, A., and de Melo, G. (2012). OpenWordNet-PT: An Open Brazilian Wordnet for Reasoning. In *COLING 2012*.

Derczynski, L. (2016). *Automatically Ordering Events and Times in Text*. Studies in Computational Intelligence. Springer International Publishing.

Dias, G., Hasanuzzaman, M., Ferrari, S., and Mathet, Y. (2014). Tempowordnet for sentence time tagging. In *23rd International Conference on World Wide Web*, pages 833–838. ACM.

Hagège, C., Baptista, J., and Mamede, N. J. (2010). Caracterização e processamento de expressões temporais em português. *Linguamática*, 2:63–76.

Kuzey, E., Strötgen, J., Setty, V., and Weikum, G. (2016). Temponym Tagging: Temporal Scopes for Textual Phrases. In *TempWeb '16*, pages 841–842. ACM.

Mamede, N., Baptista, J., Diniz, C., and Cabarrão, V. (2012). String: An hybrid statistical and rule-based natural language processing chain for portuguese.

Cristina Mota et al., editors. (2008). *Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM*. Linguateca.

Padró, L. and Stanilovsky, E. (2012). Freeling 3.0: Towards wider multilinguality. In *LREC 2012*, Istanbul, Turkey, May. ELRA.

Paiva, V. D., Oliveira, D., Higuchi, S., Rademaker, A., and Melo, G. D. (2014). Exploratory information extraction from a historical dictionary. In *IEEE 10th e-Scienc)*, volume 2, pages 11–18. IEEE, oct.

Pustejovsky, J., Castaño, J., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, A., and Katz, G. (2003). TimeML: robust specification of event and temporal expressions in text. In *(IWCS-5*.

Rademaker, A., Chalub, F., Real, L., Freitas, C., Bick, E., and de Paiva, V. (2017). Universal dependencies for portuguese. In *Depling 2017*, pages 197–206.

Real, L. and Rademaker, A. (2015). Harem and Klue: how to compare two tagsets for named entities annotation. In *NEWS 2015)*, Beijing, China, July.

Steedman, M. (2005). The productions of time: Temporality and causality in linguistic semantics, (available from his webpage).

Strötgen, J. and Gertz, M. (2015). A Baseline Temporal Tagger for all Languages. In *EMNLP 2015*, September.

Strötgen, J., Zell, J., and Gertz, M. (2013). Heideltime: Tuning english and developing spanish resources for tempeval-3. In *SemEval 2013*, pages 15–19, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

UzZaman, N., Llorens, H., Allen, J., Derczynski, L., Verhagen, M., and Pustejovsky, J. (2014). Tempeval-3: Evaluating events, time expressions, and temporal relations. *arXiv:1206.5333v2*.