Collective Elaboration of a Coreference Annotated Corpus for Portuguese Texts

Evandro Fonseca[1], Vinicius Sesti[1], Sandra Collovini[1] , Renata Vieira[1], Ana Luísa Leal[2] and Paulo Quaresma[3]

evandro.fonseca@acad.pucrs.br, vinicius.sesti@acad.pucrs.br, sandra.abreu@acad.pucrs.br, renata.vieira@pucrs.br, analeal@umac.mo, pq@di.uevora.pt

[1]Pontifical Catholic University of Rio Grande do Sul (PUCRS)
[2]University of Macau
[3]University of Évora

Abstract. This paper describes the collaborative creation of a corpus with coreference annotation for Portuguese. The annotation was performed using the coreference annotation CORP, and the editing tool CorrefVisual. The texts were automatically annotated and manually revised by Portuguese speakers. As a result a new corpus for coreference studies was produced for Portuguese.

1 Introduction

In this paper we describe the creation of a collaborative annotated corpus. Seven teams participated in the task. The texts were chosen by the teams themselves. As a result of this task, we created 'Corref-PT', a coreference corpus for Portuguese. The texts submitted by the teams were first annotated with CORP [10], a nominal coreference resolution tool for Portuguese. Then, the editing tool CorrefVisual [28] was used for the manual revision of the previously annotated texts. Agreement was measured with Kappa, considering the concordance among team members and across teams.

The paper is organized as follows. Section 2 presents an overview of the problem of coreference resolution. Related work is presented in Section 3. Section 4 presents an overview of the corpus submission and information about participating teams. Section 5 describes the corpus annotation, including the distribution of texts among annotators, annotation tools and annotation agreement. Section 6 describes the results of this Ibereval task: the Corref-PT corpus, its metrics and a brief discussion regarding the annotation process and its problems. Finally, in Section 7, the conclusions and future work are presented.

2 Coreference Resolution

Coreference resolution basically consists of finding different references to a same entity in a text, as in the example: "A França resiste como único país da União

Européia a não permitir o patenteamento de genes". The noun phrases [único país da União Européia a não permitir o patenteamento de genes] and [A França] are considered coreferent. In other words, they belong to the same coreference chain.

Coreference resolution may provide important input for other NLP tasks. One example is the area of entity relation extraction, since coreference links may be useful for extracting implicit relations [12]. Consider the following sentence:

"[Barack Obama], said today that the climate changes are a great threat for the planet". [The United States president] ...". When identifying and creating a coreference relation between [Barack Obama] and [the United States president], it is possible to infer a relation between the entities [Barack Obama] and [United States] (in which Barack Obama is the president of the United States). Also, when we link Barack Obama with the president, it is possible to classify him as a person, as well as to say that he has a relation with the United States.

3 Related Work

Coreference resolution is very important in understanding texts; thus, it is a crucial step in many high-level natural processing tasks, ranging from information extraction to text summarization or machine translation [30]. In general, the evaluation of systems devoted to this task depends on reference corpora (golden standards). There are, for example, English coreference annotated corpora that have been used in coreference resolution tracks such as SemEval, ACE and CoNLL [3,29,24,8,23,22]. SemEval (Evaluation Exercises on Semantic Evaluation) includes, among others tasks, the Coreference Resolution task [24], considering multiple languages (Catalan, Dutch, English, German, Italian and Spanish). This task involved automatically detecting full coreference chains, composed of named entities, pronouns, and full noun phrases. The datasets used in SemEval task were extracted from five corpora: 1) the AnCora corpora [25] for Catalan and Spanish; 2) the KNACK-2002 corpus [16] for Dutch; 3) the OntoNotes Release 2.0 corpus for English [23]; 4) the TurBa-D/Z corpus [15] for German; and 5) the LiveMemories corpus [26] for Italian.

CoNLL-2011 Coreference Task included a closed (limited to using the distributed resources) and an open track (unrestricted use of external resources). The task was to automatically identify mentions of entities and events in texts and to link the coreferring mentions together to form coreference chains. For this, the participants could use information from other structural layers including parsing, semantic roles, word sense and named entities. It was based on OntoNotes 4.0 [22].

The OntoNotes is a large-scale corpus of general anaphoric coreference not restricted to noun phrases or to a specified set of entity types [23,22]. In addition to coreference, the corpus provides other layers of annotation: syntactic trees; propositions structures of verbs; partial verb and noun word senses; and 18 named entity types. OntoNotes is a multi-lingual resource with annotations available in three languages: English, Chinese and Arabic.

OntoNotes corpus is of crucial important for data modeling of linguistically easier cases of coreference. Complex cases are being investigated more recently, one of the main reasons for this is the lack of appropriated datasets [29]. The ARRAU dataset is a multi-domain corpus with large-scale annotations of various linguistic phenomena related to anaphora. A second release of the ARRAU is presented in [29], and the authors not only focused on increasing the number of documents, but also invested a considerable effort into improving the data quality. The data is manually labeled for tasks such as coreference resolution, bridging, mention detection, referentiality an genericity. The documents were annotated for anaphoric information, using the MMAX (Multi-Modal Annotation in XML) tool, which is specific for corpus annotation, with main focus in the annotation of coreference [21]. The annotation followed the ARRAU guidelines, which focused on a more detailed representation of linguistic phenomena related to anaphoric and coreference. The authors present the main differences between ARRAU and two coreference corpus: ACE and OntoNotes. The difference between these corpora stands out, ARRAU considers different types of noun phrases, including markables that do not participate in coreference chains (singletons and non-referentials). Also, this corpus combines coreference with bridging, and for the third release of ARRAU, the authors plan to focus on bridging.

One of the difficulties for the creation of annotated corpora is the availability of specialists for this task. An alternative is crowd-sourcing approach, which uses a non-expert crowd to annotate text, driven by cost, speed and scalability [17]. In [3] Phrase Detectives, an interactive online game for creating annotated anaphoric coreference corpora using GWAP (game-with-a-purpose) approach is presented. The Phrase Detectives Corpus 1.0 contains 45 documents from Wikipedia articles and narrative text, with 6,452 markables.

HAREM is a joint evaluation effort for Portuguese (Avaliação de Sistemas de Reconhecimento de Entidades Mencionadas) [27]. This contest had the purpose of studying expressions regarding proper names (mentioned entities). The Second HAREM took took place in 2008 and it included the task of identifying the semantic relations between mentioned entities, called ReRelEM track (Reconhecimento de Relações entre Entidades Mencionadas). This was concerned with the automatic detection of relations between named entities in a document [11]. ReRelEM, although maintaining the restriction to named entities, is also a source of coreference annotation, since the authors proposed the detection of relations between named entities, including coreference, represented by the relation of *Identity* (entities with the same referent, defined to all the categories and whose instances must had the same category).

Another related Portuguese corpus is the Summ-it corpus [4,1]. It is a corpus gathering annotations of various linguistic levels, including coreference, but also morphological, syntactic and rhetorical relations. Summ-it has a total of 560 coreference chains with an average of 3 noun phrases for chain, where the largest chain has 16 members (noun phrases). Recently, a new version of Summ-it corpus was enriched with two layers: named entities and the relations that

occur between these entities [6,5], this version is called Summ-it++[1] and is described in [1]. The coreference information is the same from the original Summ-it corpus. However, other layers of linguistic (morpho-syntactic) information were generated by other tools and converted to a new format based on SemEval [24]. Garcias's corpus also contains coreference annotation, but only for Person entities [13]. It is also given in the SemEval format. It is a multilingual corpus[2] including Portuguese, Galician and Spanish. One of the motivations for this collaborative task of creating an annotated corpus is, therefore, to increase the number of annotated coreference data for Portuguese. Instead of creating such annotated corpus from scratch, we adopted a different methodology, we proposed the edition of coreference chains produced by coreference resolution tool.

4 Corpus submission and participating teams

The general objective of the proposed task was a collective elaboration of a Portuguese annotated corpus for nominal coreference. For that, each participant team submitted a corpus of their own interest. Seven teams submitted their corpus. The resulting corpus is composed by journalistic texts [20]; by miscellaneous texts (books, magazines, journalistic, among others) [7]; and Wikipedia dump articles, selected randomly. The corpus is further described in Section 6.1.

The first phase of the task consisted of corpus submission by participant teams. Each participant team submitted around 30 texts written in Portuguese, considering domains of their own interest. The proposed average size for these texts was 1200 tokens each. Plus, each team justified the reason(s) of corpus choice, including the related studies. Seven groups submitted texts for annotation. Three main text sets were submitted, as described below and detailed in Table 1.

- CSTNews[20] is a corpus developed for multi-document summarization and used for several studies in Portuguese, mainly for researches on discourse phenomena. This was divided in five parts, one for each group from USP.
- A sample of the larger corpus PAROLE [7], compiled in the scope of the European project LE-PAROLE. For each language involved in the project, a 20 million word corpus was built with harmonized design, composition and codification, including a 250.000 word subcorpus, tagged with POS information and revised manually.
- Wikipedia articles written in Portuguese language. This corpus is an extract composed by 30 entire articles, each with more than 1100 and less than 1400 words, randomly selected from the Wikipedia dump from 26/03/2017.

There was a training phase, when participants got used to the editing tool[28]. We provided one text annotated by CORP[10]. Each team's members revised the coreference chains and could ask questions about the task.

[1] http://www.inf.pucrs.br/linatural/summit_plus_plus.html
[2] http://gramatica.usc.es/ marcos/coling14.tar.bz

Team	Corpus		Texts
USP 1	CST-News	1/5	28
USP 2		2/5	28
USP 3		3/5	28
USP 4		4/5	28
USP 5		5/5	25
UFBA	Wikipedia		30
EVORA	Le-Parole		12

Table 1. Submitted texts

Finally, there was the annotation phase. First, all texts were annotated with CORP, then, each team received its own corpus plus a few extra texts included for measuring team level agreement (according to Table 2). The corpus annotation phase is described in detail in the next section.

5 Corpus Annotation

5.1 Text Distribution among Annotators

The corpora were received and distributed among team members in a way to allow agreement measures. For that we used first a set of three texts chosen by the organizers for calculating inter team level agreement; secondly a subgroup of four texts from each submitted corpus should be annotated by all members in its respective team. In Table 2, we exemplify how we organized the distribution of texts. This example considers a scenario of a team with three annotators and a corpus of sixteen texts. Each member annotated one of our chosen texts for inter team agreement (TK1, TK2 and TK3), whereas four texts of the submitted corpus were replicated to all annotators of that team (TG1, TG2, TG3 and TG4).

Participant 1	**Participant 2**	**Participant 3**
TK1	**TK2**	**TK3**
TG1	**TG1**	**TG1**
TG2	**TG2**	**TG2**
TG3	**TG3**	**TG3**
TG4	**TG4**	**TG4**
TG5	TG6	TG7
TG8	TG9	TG10
TG11	TG12	TG13
TG14	TG15	TG16

Table 2. Distribution scheme

The texts were then annotated with coreference and distributed among each team. The annotation consisted in editing the generated chains. Next, we de-

scribe CORP, the coreference resolution tool [10], and CorrefVisual [28], the editing tool used in this task.

5.2 Annotation tools

The annotation task is based on previous annotation generated by a coreference resolution tool and the edition of the generated chains with the help of an editing tool, as described below.

CORP is a coreference resolution tool for Portuguese [9] which was built on the basis of deterministic rules, in the line with previous tools proposed for English [19,18]. An important difference from these previous works for English is, however, the inclusion of semantic knowledge, which is provided by Onto.PT [14]. The tool produces 2 outputs: the first in XML, containing the original text, the list of sentences, tokens, part-of-speech, coreference chains (Figure 1) and single mentions. This format allows the interoperability with other applications. The second output format is given in HTML for the visualization of generated coreference chains, which can be seen through the tool's web interface[3]. A desktop version is also available for download[4].

CorrefVisual is a tool developed in order to allow the edition of coreference chains annotated with CORP. It provides a user-friendly graphical interface for visualizing and replacing NPs in other coreference chains. It also allows the editing of noun phrases, creation and deletion of chains and persistency of changes.

The interface displays information in three different main panels: the first displays the text and selected noun phrases; the second displays coreference chains, each in a particular subpanel; and the third displays single (non-coreferent) noun-phrases (unique mentions). Each chain is associated with one color in order to show the different chains.

Upon selection of noun phrases, they are highlighted in the text according to their chain's color. In figure 2, one chain is highlighted. CorrefVisual is available for download[5].

[3] http://ontolp.inf.pucrs.br/corref/
[4] http://www.inf.pucrs.br/linatural/wordpress/index.php/recursos-e-ferramentas/corp-coreference-resolution-for-portuguese/
[5] http://www.inf.pucrs.br/linatural/wordpress/index.php/recursos-e-ferramentas/correfvisual/

```
<Cadeia_5>
  <sn id="4" tokens="15...16" nucleo="França" sintagma="a França" Categoria="ORGANIZAÇÃO|LOCAL" sentenca="1">
    <word_15 token="a" lemma="o" pos="art" features="F=S"/>
    <word_16 token="França" lemma="França" pos="prop" features="F=S"/>
  </sn>
  <sn id="5" tokens="19...22" nucleo="país" sintagma="único país de a União Européia a" Categoria="ORGANIZAÇÃO|LOCAL"
  sentenca="1">
    <word_19 token="único" lemma="único" pos="adj" features="M=S"/>
    <word_20 token="país" lemma="país" pos="n" features="M=S"/>
    <word_21 token="de" lemma="de" pos="prp" features="-"/>
    <word_22 token="a" lemma="o" pos="art" features="F=S"/>
    <word_23 token="União_Européia" lemma="" pos="prop" features="F=S"/>
    <word_24 token="a" lemma="a" pos="prp" features="-"/>
  </sn>
  <sn id="27" tokens="107...108" nucleo="França" sintagma="A França" Categoria="ORGANIZAÇÃO|LOCAL" sentenca="6">
    <word_107 token="A" lemma="o" pos="art" features="F=S"/>
    <word_108 token="França" lemma="França" pos="prop" features="F=S"/>
  </sn>
  <sn id="28" tokens="110...112" nucleo="país" sintagma="o único país" Categoria="ORGANIZAÇÃO|LOCAL" sentenca="6">
    <word_110 token="o" lemma="o" pos="art" features="M=S"/>
    <word_111 token="único" lemma="único" pos="adj" features="M=S"/>
    <word_112 token="país" lemma="país" pos="n" features="M=S"/>
  </sn>
  <sn id="34" tokens="128...129" nucleo="país" sintagma="o país" Categoria="ORGANIZAÇÃO|LOCAL" sentenca="7">
    <word_128 token="o" lemma="o" pos="art" features="M=S"/>
    <word_129 token="país" lemma="país" pos="n" features="M=S"/>
  </sn>
</Cadeia_5>
```

Fig. 1. CORP - XML coreference chains

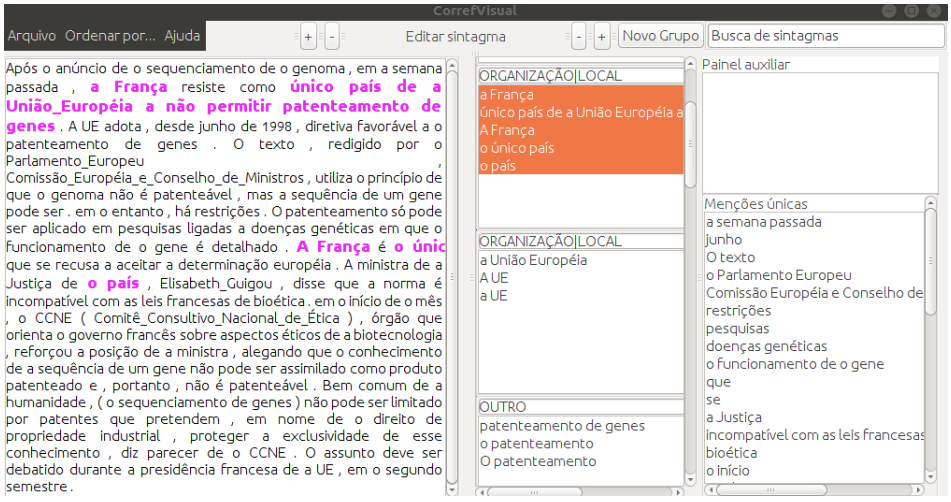

Fig. 2. CorrefVisual - mentions in a chain are highlighted in the text panel.

5.3 Annotation agreement

We measured annotation agreement on the basis of Kappa statistics. Kappa is usually used to measure concordance among canonical elements. For the coreference task, we need to calculate the agreement of complex elements: coreference

chains. Basically, a coreference chain may have two or more noun phrases. Thus, for the correct calculation of agreement, we need to transform these chains into items that may be analysed as a category.

One way of perform that is to transform each chain into coreference pairs. That is, for the chain C={a,b,c}, wherein 'a', 'b' and 'c' represent noun phrases, we represent it as follows: P={(a,b),(a,c),(b,c)}.

To perform the calculation, we need to consider the set of documents (D) and the set of annotators (A). For example, for a set of documents D={$d1$, $d2$, $d3$} and set of annotators A={$a1$, $a2$, $a3$}, we create, for each document d_x belonging to the set of documents D, the set $\cup_{dx}$, where $\cup_{dx}$ is the union of all coreference chains annotated for that document; such that U_x={$dx_{a1} \cup dx_{a2} \cup dx_{a3}$}.

Assuming that annotator $a1$ has created two coreference chains: $c1_{a1}$={a, b, c}, $c2_{a1}$={d, f}, and annotators $a2$ and $a3$ have considered only one, $c1_{a2}$={a,b,c}, $c1_{a3}$={a,b,c}, while d and f are annotated as non-coreferent by both, the resulting union set is U_{d1} = {a, b, c, d, f }.

Then we transform the union set into pairs and determine which pairs are considered coreferent or not by each annotator. The set of pairs is $P_{U\,d1}$= {(a,b), (a,c), (a,d), (a,f), (b,c), (b,d), (b,f), (c,d), (c,f), (d,f)}.

In Table 3, we can see the Kappa calculation of this example. Each pair represents an item to be classified as Coreferent or Non-Coreferent. The pairs (a, b), (a, c) and (b, c) appear in the same coreference chain for three annotators, indicating that they considered them coreferent. The pairs (a, f), (b, d), (b, f), (c, d) and (c, f) were considered non-coreferent by all anotators. For the pair (d, f), there was a disagreement between the annotators. Thus, the Coreferent class receives '1' and the Non-Coreferent class receives '2'. This process made for document $d1$ is repeated for other documents. We calculate Kappa [2] from the values represented in Table.

Pair	**Coreferent**	**Non-Coreferent**	**S**
a,b	3	0	1
a,c	3	0	1
a,d	0	3	1
a,f	0	3	1
b,c	3	0	1
b,d	0	3	1
b,f	0	3	1
c,d	0	3	1
c,f	0	3	1
d,f	1	2	0.333
N=10	C1=10	C2=20	Z=9.333

Table 3. Dataset for 'd_1', 'a_1', 'a_2' and 'a_3'

5.4 Kappa Results

Table 4 shows the resulting Kappa for each team and across teams. The lowest concordance was 0.41 and the highest was 0.64 for teams (intragroup). Kappa was 0.51 when calculated among different teams (intergroup). For intergroup agreement, only six teams were considered due to a few missing annotation texts.

According to the interpretation given in Table 5 [31], the resulting Kappa indicates mainly moderate agreement, which is in line with what was expected for such a challenging task.

Team	Members	Overlay Texts	Kappa
USP 1	3	4	0.51
USP 2	4	4	0.48
USP 3	3	4	0.55
USP 4	3	4	0.64
USP 5	3	4	0.57
UFBA	3	4	0.43
EVORA	2	2	0.41
INTERGROUP	6	3	0.51

Table 4. Concordance intra and intergroup

Kappa	Agreement
<0	Less than chance agreement
0.01 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 0.99	Almost perfect agreement

Table 5. Interpretation of Kappa

6 Corref-PT

As a result of this IBEREVAL task, we obtained a coreference corpus for Portuguese: Corref-PT. The corpus was annotated as an effort made by seven teams, with a total of twenty-one Portuguese native speakers annotators, varying among students and professors in the area of computational linguists. The corpus is available in CORP's XML (Figure 1) and SemEval format [24] used by other well known coreference corpora, such as Ontonotes [22], Summ-it++ [1] and Garcia's corpus [13]. Corref-PT is available for download[6].

[6] http://www.inf.pucrs.br/linatural/wordpress/index.php/recursos-e-ferramentas/corref-pt/

In Table 6, we show the SemEval format. It is available in a single file, containing all texts. Each text document is contained within a "#begin document ID" line and another line containing only "#end document". Each sentence's information is organized vertically, with one token per line, and a blank line after the last token of each sentence. The information associated with each token is available in columns (separated by a tab character - "\t"). The annotation columns contain, respectively: Token's ID in sentence; the word or multiword itself; lemma; each word's Part-of-speech tagging; features (gender and number); Head, denoting if the word is a head word in the NP (if so, this field receives '0') and coreference information, where each coreferent noun phrase starts with "(", followed by the chain's ID. Note that the ") " just occurs in the last NP token. Basically, coreferent NPs receives the same chain ID.

ID	Token	Lemma	POS	Feat	Head	Coref
1	Segundo	segundo	prp	_	_	_
2	informações	informar	n	F=P	0	_
3	de	de	prp	_	_	_
4	a	o	art	F=S	_	_
5	assessoria	assessoria	n	F=S	0	_
6	de	de	prp	_	_	_
7	o	o	art	M=S	_	(2
8	apresentador	apresentador	n	M=S	0	2)
9	,		,	_	_	_
10	ele	ele	pron-pers	M=3S=NOM	0	(2)
11	não	não	adv	_	_	_
12	poderia	poder	v-fin	COND=3S	_	_
13	comparecer	comparecer	v-inf	_	_	_
14	a	a	prp	_	_	_
15	o	o	art	M=S	_	_
16	Deic		prop	M=S	0	_
17	em	em	prp	_	_	_
18	a	o	art	F=S	_	(5
19	quarta-feira	quarta-feira	n	F=S	0	5)
20	...					

Table 6. Corref-PT - Semeval format

6.1 Corpus Metrics

Corref-PT is composed by texts from the CSTNews corpus [20]; from the Parole corpus (miscellaneous texts from books, magazines, journalistic, among others) [7]; Wikipedia articles, selected randomly; and also a few scientific texts from

Fapesp Magazine[7]. Metrics about number of texts, tokens, mentions, coreferential mentions, coreference chains and chains sizes are shown in Table 7.

Corpus	Texts	Tokens	Mentions	Coreferent Mentions	Coreference Chains	Largest Chain	Avg. Chain Size
CST-News	137	54445	14680	6797	1906	25	3.6
Le-Parole	12	21607	5773	2202	573	38	3.8
Wikipedia	30	44153	12049	4973	1308	53	3.8
Fapesp Magazine	3	3535	1012	496	111	33	4.5
Total	182	123740	33514	14468	3898	53	3.7

Table 7. Corref-PT - Corpus Metrics

6.2 Annotation task evaluation

The annotators evaluated the task regarding a few issues inquired through a survey on Google Forms. Fifteen of the 21 participants sent their answers. They were asked about their confidence level in the annotation, whether the previous automatic annotation was helpful for the task and about the necessity of noun phrase edition for the task (considering that noun phrase identification was made automatically by a parser). We can see in Figure 3 that few annotators had high confidence in their annotation. Most participants were not sure about this issue.

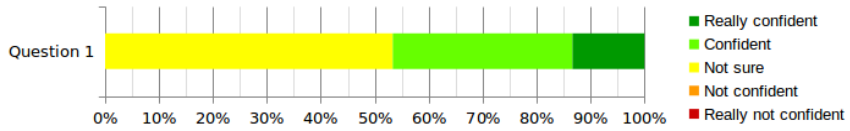

Fig. 3. Question 1 - confidence level

Regarding previous annotation (Figure 4), most participants were ambivalent whether this helps or not the process, but a greater number thought it was helpful.

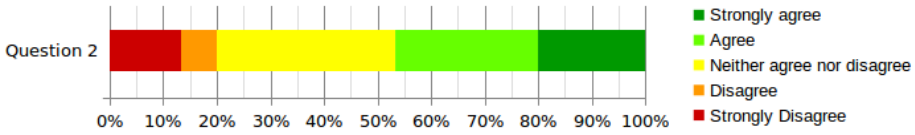

Fig. 4. Question 2 - usefulness of previous annotation

[7] http://revistapesquisa.fapesp.br/

Regarding noun phrase edition (Figure 5), 60% of participants strongly agreed that is indispensable for the annotation task. That is indeed a crucial pre-processing requirement for build the chains, that the references are correctly identified. The main problem here was that the task was in fact mostly fixed regarding mention detection, and it was based on the parser's NP chunks. Suggestions given by the annotators were most related to CorrefVisual's usability - one major problem was related to noun phrase edition. That was very difficult to handle by the annotators, since the mention detection is required for identifying coreference chains correctly, but the tool was not primarily meant for that.

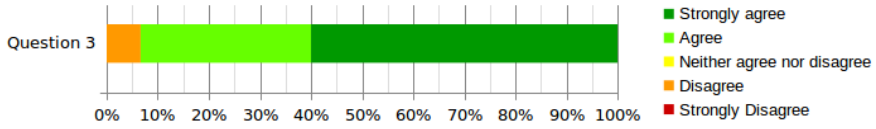

Fig. 5. Question 3 - noun phrase edition required

7 Conclusion

In this paper, we presented a collaborative coreference annotation task which resulted in a coreference corpus for Portuguese with nearly 4000 chains. Considering Summ-it++, a previous available resource of the kind, with around 500 chains, we now have a coreference annotated corpus with 8 times as many chains. The resource is available both in the SemEval format and in CORP's XML[8]. The annotated corpus can be visualized in the CorrefVisual tool[9]. For the next steps, we have to improve questions regarding automatic mention detection, which seems to be a major pre-processing issue for this task, and similarly we have also to improve the ways for their manual editing, if we consider further annotation tasks. Regarding the annotation agreement, we can see that there is mainly moderate agreement. However, as future work, a revision of this annotation should be done in order to improve the quality of annotation.

References

1. A. Antonitsch, A. Figueira, D. Amaral, E. Fonseca, R. Vieira, and S. Collovini. Summ-it++: an enriched version of the summ-it corpus. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*,

[8] http://www.inf.pucrs.br/linatural/wordpress/index.php/recursos-e-ferramentas/corref-pt/

[9] http://www.inf.pucrs.br/linatural/wordpress/index.php/recursos-e-ferramentas/correfvisual/

pages 2047–2051, Paris, France, 2016. European Language Resources Association (ELRA).
2. J. Carletta. Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics*, 22(2):249–254, 1996.
3. J. Chamberlain, M. Poesio, and U. Kruschwitz. Phrase detectives corpus 1.0 crowdsourced anaphoric coreference. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 2039–2046, Paris, France, 2016. European Language Resources Association (ELRA).
4. S. Collovini, T. I. Carbonel, J. T. Fuchs, J. C. Coelho, L. Rino, and R. Vieira. Summ-it: Um corpus anotado com informações discursivas visando a sumarização automática. In *Proceedings of V Workshop em Tecnologia da Informação e da Linguagem Humana , Rio de Janeiro, RJ, Brasil*, pages 1605–1614, 2007.
5. S. Collovini de Abreu and R. Vieira. Relp: Portuguese open relation extraction. *Knowledge Organization*, 44(3):163–177, 2017.
6. D. O. F. do Amaral and R. Vieira. NERP-CRF: uma ferramenta para o reconhecimento de entidades nomeadas por meio de conditional random fields. 6(1):41–49, 2014.
7. M. F. B. do Nascimento, A. Mendes, and L. Pereira. Providing on-line access to portuguese language resources: Corpora and lexicons. In *Proceedings of the International Conference on Language Resources and Evaluation , Portugal*, 2004.
8. G. Doddington, A. Mitchell, M. Przybocki, L. Ramshaw, S. Strassel, and R. Weischedel. The automatic content extraction (ace) program: Tasks, data, and evaluation. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, and R. Silva, editors, *Proceedings of the 4th International Conference on Language Resources and Evaluation – LREC 2004*, pages 837–840, Lisboa, 2004.
9. E. B. Fonseca, V. Sesti, A. Antonitsch, A. A. Vanin, and R. Vieira. Corp - uma abordagem baseada em regras e conhecimento semântico para a resolução de correferências. *Linguamatica*, 9(1):3–18, 2017.
10. E. B. Fonseca, R. Vieira, and A. Vanin. Corp: Coreference resolution for portuguese. In *12th International Conference on the Computational Processing of Portuguese, Demo Session (PROPOR)*, 2016.
11. C. Freitas, C. Mota, D. Santos, H. G. Oliveira, and P. Carvalho. Second HAREM: advancing the state of the art of named entity recognition in portuguese. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC, Valletta, Malta*, 2010.
12. R. Gabbard, M. Freedman, and R. Weischedel. Coreference for learning to extract relations: yes, virginia, coreference matters. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 288–293. Association for Computational Linguistics, 2011.
13. M. Garcia and P. Gamallo. Multilingual corpora with coreferential annotation of person entities. In *Proceedings of the 9th edition of the Language Resources and Evaluation Conference - LREC*, pages 3229–3233, 2014.
14. H. Gonçalo Oliveira. *Onto. PT: Towards the Automatic Construction of a Lexical Ontology for Portuguese*. PhD thesis, Ph. D. thesis, Univ. of Coimbra/FST, 2012.
15. E. W. Hinrichs, S. Kübler, and K. Naumann. A unified representation for morphological, syntactic, semantic, and referential annotations. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, CorpusAnno '05,

pages 13–20, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.
16. V. Hoste and G. De Pauw. Knack-2002: a richly annotated corpus of dutch written text. In *Proceedings of The Fifth international conference on Language Resources and Evaluation*, pages 1432–1437, Genoa, Italy, 2006. European Language Resources Association, European Language Resources Association.
17. J. Howe. *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. Crown Publishing Group, New York, NY, USA, 1 edition, 2008.
18. H. Lee, A. Chang, Y. Peirsman, N. Chambers, M. Surdeanu, and D. Jurafsky. Deterministic coreference resolution based on entity-centric, precision-ranked rules. volume 39, pages 885–916. Computational Linguistics - MIT Press, 2013.
19. H. Lee, Y. Peirsman, A. Chang, N. Chambers, M. Surdeanu, and D. Jurafsky. Stanford's multi-pass sieve coreference resolution system at the conll-2011 shared task. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*. Association for Computational Linguistics, 2011.
20. E. G. Maziero, M. L. del Rosario Castro Jorge, and T. A. S. Pardo. Identifying multidocument relations. In *Natural Language Processing and Cognitive Science, Proceedings of the 7th International Workshop on Natural Language Processing and Cognitive Science, NLPCS 2010, In conjunction with ICEIS 2010, Funchal, Madeira, Portugal, June 2010*, pages 60–69, 2010.
21. C. Müller and M. Strube. Mmax: A tool for the annotation of multi-modal corpora. In *Proceedings of the 2nd IJCAI Workshop on Adaptive Text Extraction and Mining - IJCAI 2001*, Seattle, Washington, 2001.
22. S. Pradhan, L. Ramshaw, M. Marcus, M. Palmer, R. Weischedel, and N. Xue. Conll-2011 shared task: Modeling unrestricted coreference in ontonotes. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–27. Association for Computational Linguistics, 2011.
23. S. S. Pradhan, E. Hovy, M. Marcus, M. Palmer, L. Ramshaw, and R. Weischedel. Ontonotes: A unified relational semantic representation. In *Proceedings of the International Conference on Semantic Computing*, ICSC '07, pages 517–526, Washington, DC, USA, 2007. IEEE Computer Society.
24. M. Recasens, L. Màrquez, E. Sapena, M. A. Martí, M. Taulé, V. Hoste, M. Poesio, and Y. Versley. Semeval-2010 task 1: Coreference resolution in multiple languages. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 1–8. Association for Computational Linguistics, 2010.
25. M. Recasens and M. A. Martí. Ancora-co: Coreferentially annotated corpora for spanish and catalan. *Language Resources and Evaluation*, 44(4):315–345, 2010.
26. K. J. Rodríguez, F. Delogu, Y. Versley, E. Stemle, and M. Poesio. Anaphoric annotation of wikipedia and blogs in the live memories corpus. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, editors, *Proceedings of the International Conference on Language Resources and Evaluation - LREC*. European Language Resources Association, 2010.
27. D. Santos, N. Cardoso, N. Seco, and R. Vilela. Breve introduçao ao harem. *HAREM, a primeira avaliaçao conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentaçao e actas do encontro, Linguateca*, 2007.
28. M. d. O. Tubino and M. M. S. Silva. Visualização, manipulação e refinamento de correferência em língua portuguesa. Trabalho de conclusão de curso, Pontifícia Universidade Católica do Rio Grande do Sul, 2015.

29. O. Uryupina, R. Artstein, A. Bristot, F. Cavicchio, K. Rodriguez, and M. Poesio. ARRAU: Linguistically-Motivated Annotation of Anaphoric Descriptions. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 2058–2062, Portorož, Slovenia, 2016. European Language Resources Association (ELRA).
30. K. van Deemter and R. Kibble. What is coreference, and what should coreference annotation be? In *Proceedings of the Workshop on Coreference and Its Applications*, CorefApp '99, pages 90–96, Stroudsburg, PA, USA, 1999. Association for Computational Linguistics.
31. A. J. Viera, J. M. Garrett, et al. Understanding interobserver agreement: the kappa statistic. *Fam Med*, 37(5):360–363, 2005.