

A Larger Annotated Corpus of Portuguese Coreference

Evandro Fonseca

Blip

evandro.fonseca@blip.ai

Renata Vieira

University of Évora

renata.vieira@gmail.com

Abstract

Coreference resolution is a crucial task in natural language processing (NLP) that aims to identify and link expressions in a text that refer to the same entity. However, the lack of annotated data for coreference resolution in Portuguese has hindered the development of robust and accurate systems for this language. In this paper, we present an assessment of coreference annotation utilizing large language models (LLMs) for Portuguese: LLM-PREF is proposed to annotate coreference in Portuguese texts. It was evaluated and compared to a system previously proposed in the literature. The results show that although the model’s world knowledge and inference capacity are quite rich - allowing it to recognize complex coreference patterns, including the pronominal anaphora phenomenon - it does not excel the previously developed rule based system.

1 Introduction

The Coreference Resolution task is challenging for Natural Language Processing, considering the required linguistic knowledge and the sophistication of language processing techniques involved. Even though it is a demanding task, a motivating factor in the study of this phenomenon is its usefulness.

Several Natural Language Processing tasks may benefit from their results, such as named entity recognition, relation extraction between named entities, summarization, sentiment analysis, among others. Coreference Resolution is a process that consists of identifying certain terms and expressions that refer to the same entity. For example, in the sentence “Joe Biden drops out of presidential race. The president says that...” we can say that [the president] is a coreference of [Joe Biden]. By grouping these referential terms, we form coreference groups, more commonly known as coreference chains.

The Coreference Resolution task for Portuguese has received some attention in past years (Coreixas,

2010; Silva, 2011; Garcia and Gamallo, 2014; Fonseca et al., 2014, 2015, 2016b; Fonseca, 2018). Most of these models were built using machine learning approaches. The most recent Portuguese coreference system, proposed by Fonseca (2018), was inspired by Lee et al. (2013).

Due to the scarcity of coreference tools for Portuguese and the recent advancements of large language models, this paper proposes the evaluation of LLMs to annotate Portuguese texts, specifically the FAPESP parallel corpus, with coreferences.

This paper is organized as follows: Section 2 presents related work on Portuguese coreference resolution; Section 3 describes the proposed annotation strategy; Section 4 presents the evaluation; Section 5 presents corpus annotation and result analysis; and finally, Section 6 presents our conclusions and future work.

2 Related Work

When dealing with the Coreference Resolution task for Portuguese, it is possible to find many corpora with some kind of coreference annotation, such as: HAREM (Freitas et al., 2010), Garcia’s corpus (Garcia and Gamallo, 2014), Summit (Collovini et al., 2007), Summ-it++ (a new and enriched version of Summ-it (Fonseca et al., 2016a)), and Corref-PT (Fonseca et al., 2017), a semi-automatically annotated corpus.

HAREM contains annotation of named entities and their identity relations; its main purpose was the evaluation of Named Entity Recognition (NER) systems. The corpus contains manually annotated named entities distributed in ten semantic categories. Relations between these named entities have also been annotated manually, in four types: identity, inclusion, placement, and other. Garcia’s corpus contains coreference annotation for person entities.

Summ-it contains noun phrase coreference an-

notation, being thus the corpus with the most complete coreference chains. It was semi-automatically annotated with morphosyntactic information and manually annotated with coreference. Besides coreference, the texts were also manually annotated with rhetorical relations. Also, for each text, there are manual and automatically generated summaries. Summ-it++ is an enriched version of Summ-it, adding two new annotation layers: named entities and relations between named entities. In addition, the format was changed to the SemEval (Recasens et al., 2010) format. Corref-PT (Fonseca et al., 2017) was built through a collaborative task and has 182 texts annotated semi-automatically with coreferences.

Despite the existence of many corpora, it is still difficult to produce rich coreference models using machine learning approaches due to the insufficient amount of available samples. Therefore, in this paper, we present a corpus annotated by an LLM, as an alternative for the construction of a large corpus with coreferences for Portuguese. A large corpus would allow for the production of well-trained coreference models.

Recent research on large language models (LLMs) highlights their potential and limitations in performing coreference tasks. Studies such as Gan et al. (2024) and Liu et al. (2024) emphasize the need for improvements in coreference resolution and comprehension of long contexts. Hicke and Mimno (2024) focuses on improving coreference annotation in English literary texts by exploring techniques to enhance the capabilities of LLMs in addressing the complexities inherent in literary language. Despite these efforts, challenges persist in developing rich coreference models due to limited sample availability, particularly for languages like Portuguese.

In this paper, we test whether a Portuguese coreference annotation generator based on LLMs can achieve good results without specific coreference training, comparing its performance to the latest proposed approach for Portuguese by Fonseca (2018).

3 Annotation Generation

LLM-PREF is the method proposed in this paper, based on GPT-4o (OpenAI, 2023). A detailed prompt was developed, based on a set of guidelines for annotating coreference chains. In other words, instructions to ensure that the annotation is per-

formed while respecting the coherence and cohesion of the text, as described in Fonseca (2018). Additionally, the prompt indicates a structured JSON output, as can be seen in 3.1. The prompt was built in Portuguese, we present its original version.

3.1 Prompt

Você é um assistente que realiza anotação linguística de cadeias de correferência. Uma menção é considerada correferente de outra quando ambas se referem ao mesmo referente, ou seja, a mesma entidade, objeto ou conceito. Essa relação de correferência é importante para garantir a coesão e a coerência de um texto, pois evita repetições desnecessárias e ambiguidades. Você precisa anotar as cadeias seguindo a seguinte estrutura JSON de exemplo:

```

1 "chains" : [ {
2   "chain_0" : [ {
3     "np" : "FAPESP",
4     "start" : 0,
5     "end" : 5
6   } ],
7   {
8     "np" : "Pesquisa FAPESP",
9     "start" : 0,
10    "end" : 14
11  } ],
12  ... ]
13 },
14 "chain_1" : [ {
15   "np" : "Lula",
16   "start" : 0,
17   "end" : 4
18 },
19 {
20   "np" : "Luis inácio lula da silva"
21   ,
22   "start" : 100,
23   "end" : 125
24 },
25 ... ]
26 }
27 ]

```

As regras para que uma menção seja correferente de outra incluem:

- Mesmo referente:** Como mencionado anteriormente, para que duas menções sejam correferentes, é necessário que elas se refiram ao mesmo referente. Isso significa que ambas devem se relacionar com a mesma entidade, objeto ou conceito.
- Uso de pronomes:** Uma das formas mais comuns de estabelecer a correferência entre

duas menções é por meio do uso de pronomes. Por exemplo, se em um texto é mencionado o nome de uma pessoa do gênero masculino (referente) e, em seguida, utiliza-se o pronome “ele” para se referir a essa mesma pessoa, temos uma correferência estabelecida.

3. **Uso de sinônimos:** Outra forma de estabelecer a correferência é por meio do uso de sinônimos. Por exemplo, se em um texto é mencionado o termo “cachorro” e, em seguida, utiliza-se o termo “animal de estimação” para se referir ao mesmo ser, temos uma correferência estabelecida.
4. **Proximidade textual:** A proximidade entre as menções também é um fator importante para estabelecer a correferência. Geralmente, quanto mais próximas as menções estiverem no texto, maior será a probabilidade de serem correferentes.
5. **Coerência e contexto:** A correferência também deve estar em conformidade com a coerência e o contexto do texto. Isso significa que as menções devem fazer sentido em relação ao restante do texto e ao tema abordado.
6. **Uso de conectivos:** Alguns conectivos, como “este”, “aquele”, “o mesmo”, entre outros, também podem ser utilizados para estabelecer a correferência entre duas menções.
7. **Coesão:** Além de estabelecer a correferência, é importante que as menções sejam coesas, ou seja, que haja uma conexão lógica entre elas. Isso garante a fluidez e a clareza do texto.

Em resumo, para que uma menção seja correferente de outra, é necessário que ambas se refiram ao mesmo referente, estejam próximas no texto, façam sentido em relação ao contexto e sejam coesas. O uso de pronomes, sinônimos e conectivos também pode ajudar a estabelecer essa relação de correferência. Analise o texto abaixo e anote todas suas cadeias de correferência lembrando que cada cadeia de correferência só pode ter menções de uma mesma entidade. Caso a entidade seja única, crie uma cadeia com apenas uma menção. É importante anotar todos os sintagmas nominais, nenhuma menção pode ficar para trás.

Table 1: Metrics for LLM-PREF

	MUC	B ³	CEAF _m	CEAF _e
Precision	62,0	50,1	57,3	32,6
Recall	29,3	21,2	32,0	25,4
F-measure	39,8	29,8	41,1	28,6
CoNLL	32,7			

Table 2: BLANC metric for LLM-PREF

Coreferential Links			Non-Coreferential Links		
P	R	F	P	R	F
63.8	38.9	48.4	69.9	64.3	65.3

Table 3: Metrics for Fonseca (2018)

	MUC	B ³	CEAF _m	CEAF _e
Precision	53,7	50,3	49,5	44,2
Recall	52,7	47,7	53,1	57,7
F-measure	53,2	48,9	51,3	50,0
CoNLL	50,7			

Table 4: BLANC metric for Fonseca (2018)

Coreferential Links			Non-Coreferential Links		
P	R	F	P	R	F
55.8	15.1	23.8	80.7	96.7	88.0

4 Evaluation

To perform the evaluation, five texts from the Summ-it corpus (Collovini et al., 2007), a manually annotated corpus with coreference, were considered. We applied five metrics from CoNLL scorer (Pradhan et al., 2014).

- MUC (Vilain et al., 1995) measures how many clusters of mentions are necessary to cover the standard chains.
- B-CUBED (Bagga and Baldwin, 1998) based on mentions, it generates results by considering the present and absent mentions of each entity in a given chain.
- CEAF (Luo, 2005) uses the alignment between entities (CEAF_e) or mentions (CEAF_m) to calculate their results
- BLANC (Recasens and Hovy, 2011) evaluates both coreference links and non-coreference links.

Each metric favors a specific feature. Additionally, we calculate CoNLL, which is the average of MUC, B³, and CEAF_e.

The evaluation process was based on SemEval, a widely recognized corpus format for the coreference resolution task (Pradhan et al., 2011). In Table

5, the SemEval format is presented. Each line is composed of a token followed by its annotation layers (Antonitsch et al., 2016). Regarding coreferences, each noun phrase starts using “(” followed by the chain ID. Note that the “)” just occurs in the last NP token. Basically: coreferent NPs receives the same chain ID.

Table 1 presents the LLM-PREF results.

Coreference is difficult to evaluate, as we can see in the different results given by each metric. Depending on the phenomena observed, a different metric is proposed. LLM-PREF is based on the same underlying principles of Fonseca (2018), since the prompt was built upon their previous work. We can see in general higher precision and lower recall in LLM-PREF. The significantly lower recall observed in LLM-PREF compared to Fonseca (2018) is likely due to the mention detection step. While the previous rule-based system employs a dedicated and elaborate pipeline using the CoGrOO parser to explicitly extract noun phrases, our approach relies entirely on a single prompt to simultaneously identify and resolve mentions, naturally increasing the chances of missing entities. Overall, despite the higher precision, the previous rule-based system still surpasses the LLM output in terms of general metrics.

However, if we look at the BLANC metric (Bilateral Assessment of Noun Phrase Coreference), which considers coreferential and non-coreferential links separately, we see an advantage for LLM-PREF in the coreferential links. Basically, a non-coreference link is formed by two mentions that are not coreferences to each other. The BLANC metric aims to reward correct coreference chains, proportionally to their length.

Tables 3 and 4 present the results for Fonseca (2018)’s model on the same subset.

5 Corpus Annotation and Result Analysis

The Portuguese texts of the FAPESP parallel corpus (Aziz and Specia, 2011), consisting of 3,840 articles were annotated with an LLM. A total of 105 texts had to be excluded due to prompt size limitations. The resulting annotated corpus is composed of 3,735 texts, 104,348 coreference chains, and 226,973 noun phrases. Although the automatic annotation is not perfect, this resource may be useful for future research in coreference. It can be curated to produce higher quality annotations, and it can be used as a source of examples of corefer-

ence chains.

To produce the entire corpus, 18 million tokens were expended. The LLM utilized 10,416,029 tokens for input prompts, and 8,274,124 tokens for completions, resulting in a total of 18,690,153 tokens consumed. Essentially, in each iteration, the input tokens consist of the sum of prompt tokens and the text content, while the completion tokens represent the entire output, specifically the JSON output. This substantial token count underscores the extensive computational effort involved in processing the corpus.

It is important to highlight that the original FAPESP Parallel Corpus is distributed under a Creative Commons Attribution-NonCommercial (CC BY-NC 2.0) license. Therefore, the new coreference layer generated in this work inherits these non-commercial and attribution constraints, ensuring an ethical distribution and use of the corpus by the NLP research community.

Regarding the result analysis, we investigated several chains to identify the main errors in the annotation process. In chain C1, it is clear that the model mistakenly combined two separate entities, likely due to hallucination. Chain C2 presents a similar case where three distinct entities were linked.

- C1={[Peru], [Bolívia]} (see context in Figure 1)
- C2={[a prefeitura do Rio], [o Instituto Municipal de Urbanismo Pereira Passos], [a Secretaria do Meio Ambiente]} (see context in Figure 2)

Unfortunately, understanding the reasons behind this is quite challenging. Additionally, we identified chains that are partially correct, such as C3. In C3, it is evident that Professor Hermógenes could be considered part of the group of 105 researchers, but he does not represent the same entity.

- C3={[105 pesquisadores], [eles],[professor Hermógenes de Freitas Leitão]} (see context in Figure 3)

In C4 and C5, we found a more complex case. In addition to splitting the chains of “José de Souza Martins” and “Martins,” the model invented an

Table 5: SemEval annotation scheme

ID	Token	Lemma	PoS	Feat	Head	NE	Rel	Coref
1	A	o	art	F=S	–	–	–	–
2	opinião	opinião	n	F=S	0	–	–	–
3	é	ser	v-fin	PR=3S=IND	–	–	–	–
4	de	de	prp	–	–	–	–	–
5	o	o	art	M=S	–	–	–	(2)
6	agrônomo	agrônomo	n	M=S	0	–	–	–
7	Miguel_Guerra	–	prop	M=S	0	PES	(9)	–
8	,	–	–	–	–	–	–	–
9	de	de	prp	–	–	–	–	–
10	a	o	art	F=S	–	–	–	–
11	UFSC	–	prop	F=S	0	ORG	(9)	(3)

Figure 1: Context of C1 chain

Houve um programa do governo dedicado ao crescimento econômico, com uma estratégia de incorporação, cujo lema era “Integrar para não entregar”, lembram? Existia uma mentalidade naquele momento na América Latina de que sem uma política de ocupação dessas áreas elas seriam entregues a outros. Não só no Brasil. Os governos do [Peru[21]] e da [Bolívia[21]] tinham programas idênticos de integração nacional....

Figure 2: Context of C2 chain

Pensando nisso acabamos de fazer um trabalho pioneiro na América Latina com [a prefeitura do Rio[10]] com equipes de pesquisadores de geomorfologia, oceanografia, engenharia e saúde, entre outros. O objetivo foi obter um diagnóstico da vulnerabilidade potencial da cidade. O estudo foi patrocinado pelo [Instituto Municipal de Urbanismo Pereira Passos[10]] e chama Rio, próximos 100 anos – O aquecimento global e a cidade. Estudo semelhante foi realizado logo após para todo o estado do Rio por solicitação da [Secretaria do Meio Ambiente[10]]. São trabalhos de avaliação de impactos devidos a mudanças climáticas do nível do mar.

Figure 3: Context of C3 chain

Iniciado em 1994, o Projeto Flora Fanerogâmica do Estado de São Paulo integra [105 pesquisadores[6]] de 23 instituições de todo o estado na identificação das plantas com flores da região. De 1994 a 1997 [eles[6]] realizaram 111 expedições científicas e 58 visitas ... A iniciativa do projeto Flora Fanerogâmica foi do [professor Hermógenes de Freitas Leitão[6]], que morreu em fevereiro de 1996, quando dirigia uma atividade de campo da pesquisa.

additional reference of “José de Souza Martins” and missed some mentions, such as “Talcott Parsons,” resulting in incorrect linkage with the pronoun “ele.” Furthermore, the same chain contains an excessive number of incorrect pronominal links. It seems that the coreference task can be challenging, even for large language models.

- C4={ [José de Souza Martins], [professor José de Souza Martins], [José de Souza Martins], [ele], [ele], ... } (see context in Figure 4)

- C5={ [Martins], [Martins] } (see context in

Figure 4)

Despite some inconsistencies found in the last presented chains, interesting coreference chains were found, see examples C6 and C7. In these chains, we can see that the model’s world knowledge and inference capacity are quite rich, allowing it to recognize complex coreference patterns, including the pronominal anaphora phenomenon.

- C6¹={ [Dieter Muehe], [geógrafo baiano Dieter Carl Ernst Heino

¹this chain can be found in 911_3648.json file

Figure 4: Context of C4 and C5 chains

[José de Souza Martins[1]] A sociologia que examina as margens, os sonhos e a esperança Mariluce Moura e Marcos de Oliveira A escrita do [professor José de Souza Martins[1]] , 69 anos, é de surpreendente ... indo da própria e diferenciada trajetória pessoal desse filho de operários, [ele[1]] próprio um trabalhador muito precoce ... o professor [Martins[8]] não acredita que os sonhos sejam domínio exclusivo da psicanálise e da teoria freudiana. Em meio a tamanha riqueza de reflexões... há pouco lançado numa segunda edição pela editora Contexto. É um dos 27 livros publicados por [Martins[8]], considerado por [ele[1]] mesmo central em sua obra sociológica ... Em tempo: [Martins[8]] foi professor da Cátedra Simón Bolívar na Universidade de Cambridge, na Inglaterra... Como é essa relação entre pesquisa sociológica e linguagem? - Eu aprendi sociologia no grupo... Isso era muito próprio da sociologia dos anos 1950, 60. O Talcott Parsons fez assim e era sucesso, portanto. Até o dia em que Wright Mills, outro sociólogo importante, disse que era preciso traduzir Parsons para o inglês (ora, [ele[1]] tinha escrito em inglês!). Parsons foi derrotado na revolta estudantil de 1968...

Muehe],[ele],[professor titular], [este filho de alemães nascido em Maragogipe, no Recôncavo Baian],[Ganhador do Prêmio Conrado Wessel 2003 na categoria Ciência Aplicada ao Mar, Dieter Muehe],[senhor]}
(see context in Figure 5)

- C7={[[a tuberculose],[essa doença infecto-contagiosa],[a doença]}(see context in Figure 6)

6 Conclusion

This paper presents an alternative for annotating coreference using large language models. The motivation is based on the critical shortage of comprehensive coreference corpora for the Portuguese language, providing means for the development and enhancement of coreference resolution models. A corpus comprising 3,735 texts from the FAPESP corpus (Aziz and Specia, 2011), with 104,348 coreference chains, and 226,973 noun phrases was annotated using GPT-4o (OpenAI, 2023). GPT was used due to its availability in the work environment.

The evaluation, although limited in scope, demonstrated promising results. Despite some errors in the annotation, the overall quality of the coreference chains points to some capability of LLMs for identifying complex coreference patterns. Coreference resolution can potentially improve the performance of various natural language processing tasks such as named entity recognition, relation extraction, summarization, and sentiment analysis. This corpus can contribute to further studies on coreference.

Future work will involve expanding the evaluation to include a larger set of texts and further

refining the prompts for the annotation process to minimize errors. By making Coref/FAPESP publicly available², we hope to foster further research and development in Portuguese coreference resolution and contribute to the broader NLP community.

References

- A. Antonitsch, A. Figueira, D. Amaral, E. Fonseca, R. Vieira, and S. Collovini. 2016. Summ-it++: an enriched version of the summ-it corpus. In *Proceedings of 10th edition of the Language Resources and Evaluation Conference*, Portorož, Slovenia.
- Wilker Aziz and Lucia Specia. 2011. Fully automatic compilation of a Portuguese-English parallel corpus for statistical machine translation. In *STIL 2011*, Cuiabá, MT.
- Amit Bagga and Breck Baldwin. 1998. Algorithms for scoring coreference chains. In *Proceedings of the first International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference*, pages 563–566, Granada, Spain.
- Sandra Collovini, Thiago I Carbonel, Juliana Thiesen Fuchs, Jorge César Coelho, Lúcia Rino, and Renata Vieira. 2007. Summ-it: Um corpus anotado com informações discursivas visando a sumarização automática. In *V Workshop em Tecnologia da Informação e da Linguagem Humana*.
- Tatiane Coreixas. 2010. Resolução de correferência e categorias de entidades nomeadas. Master’s thesis, Pontifícia Universidade Católica Do Rio Grande Do Sul.
- E. B. Fonseca, R. Vieira, and A. Vanin. 2016a. Adapting an entity centric model for portuguese coreference resolution. In *Proceedings of the 10th Annual Conference on Language Resources and Evaluation (LREC 2016)*, In Press.

²The corpus is available at: <https://github.com/evandrofonsecatake/llm-pref>

Figure 5: Context of C6 chain

Geógrafo [Dieter Muehe[1]] diz que é fundamental monitorar o litoral e o oceano para saber o que realmente vai mudar no clima... A resposta do [geógrafo baiano Dieter Carl Ernst Heino Muehe[1]] ao pedido de entrevista de Pesquisa FAPESP soou quase como um blefe. “Venha até meu apartamento, em Niterói, para conversarmos. Estou aposentado e vou pouco ao Fundão”, disse [ele[1]] se referindo ao o campus da Universidade Federal do Rio de Janeiro (UFRJ), onde é [professor titular[1]] e ainda orienta doutorandos. Aos 70 anos, a voz baixa e aparentemente tímida ao telefone dava a impressão de que se tratava de um pesquisador cansado, dedicado, a essa altura da existência, apenas a criar netos. A realidade de [Dieter Muehe[1]] é bem diferente do que sua discrição deixa ver. Durante a entrevista, [este filho de alemães nascido em Maragogipe, no Recôncavo Baiano[1]], colocou sobre a mesa dois livros da maior importância para quem estuda a costa do país... Os dois livros tiveram a ativa participação de [Dieter Muehe[1]] como coordenador dos números... “Só com mais informações é que saberemos, nos próximos anos, o que vai realmente mudar no clima e quais as conseqüências para as populações”, alerta. [Ganhador do Prêmio Conrado Wessel 2003 na categoria Ciência Aplicada ao Mar, Dieter Muehe[1]] tem uma filha e dois netos.... Então teremos definitivamente fixados os nossos limites marítimos. O [senhor[1]] é geógrafo, mas seu trabalho abrange todas as frentes de pesquisa? Sim....

Figure 6: Context of C7 chain

Pesquisadores de Ribeirão Preto desenvolvem a primeira vacina gênica contra [a doença[3]] Desde que o alemão Robert Koch anunciou a descoberta do bacilo d[a tuberculose[3]], em 1882, a prevenção e o tratamento d[a doença[3]] desafiam cientistas em todo o mundo. E pela dimensão que [essa doença infecto-contagiosa[3]] alcança na atualidade, torna-se particularmente importante o desenvolvimento de uma nova vacina contra [a tuberculose[3]] por uma equipe de pesquisadores ... considerada de terceira geração, que poderá ser aplicada no controle d[a tuberculose[3]], caso seja comprovada em humanos a mesma eficiência já atestada em animais. Segundo Célio Silva, o trabalho começou em 1990, quando ele foi para Londres fazer seu ... não conferia proteção satisfatória contra [a tuberculose[3]]. ... o agente causador d[a tuberculose[3]], se esconde dentro das células humanas e não é atingido pela ação dos anticorpos, seria necessário estimular os linfócitos T CD8, capazes de destruir especificamente as células infectadas pelos bacilos. ... agente causador d[a doença[3]] Reforçou essa posição o fato de entre um terço e metade da população mundial já estar infectada com o bacilo d[a tuberculose[3]]. ... Um dos problemas mais sérios relacionados com o controle d[a tuberculose[3]] é o aparecimento de bacilos que apresentam resistência a vários dos medicamentos utilizados no tratamento, como a isoniazida, pirazinamida, estreptomina e rifampicina, entre outros. ... Os casos mais comuns de imunossupressão associados com [a tuberculose[3]] são os indivíduos com Aids, estressados, que tomam drogas imunossupressoras, alcoólatras e desnutridos, entre outros. ...ela pode até erradicar [a tuberculose[3]] em nosso meio”, afirma o pesquisador. ... O tratamento d[a tuberculose[3]] feito com drogas antimicobacterianas é de longa duração - demora pelo menos seis meses.

- E. B. Fonseca, R. Vieira, and A. Vanin. 2016b. Improving coreference resolution with semantic knowledge. In *Proceedings of the 12th International Conference on the Computational Processing of Portuguese (PROPOR 2016)*, In Press.
- Evandro Fonseca, Vinicius Sesti, Sandra Collovini, Renata Vieira, Ana Luísa Leal, and Paulo Quaresma. 2017. Collective elaboration of a coreference annotated corpus for portuguese texts. In *Proceedings of II workshop on Evaluation of Human Language Technologies for Iberian Languages*, volume 1881, pages 68–82, Murcia, Spain.
- Evandro B Fonseca, Renata Vieira, and Aline A Vanin. Resolução de coreferência em língua portuguesa: Pessoa, local e organização.
- Evandro B Fonseca, Renata Vieira, and Aline A Vanin. 2014. Coreference resolution in portuguese: Detecting person, location and organization. In *Journal of the Brazilian Computational Intelligence Society*, volume 12, pages 86–97.
- Evandro Brasil Fonseca. 2018. Resolução de coreferência nominal usando semântica em língua portuguesa.
- Evandro Brasil Fonseca, Renata Vieira, and Aline Vanin. 2015. Dealing with imbalanced datasets for coreference resolution. In *The Twenty-Eighth International Flairs Conference*.
- Cláudia Freitas, Cristina Mota, Diana Santos, Hugo Gonçalo Oliveira, and Paula Carvalho. 2010. Second harem: Advancing the state of the art of named entity recognition in portuguese. In *The seventh international conference on Language Resources and Evaluation*.
- Yujian Gan, Juntao Yu, and Massimo Poesio. 2024. Assessing the capabilities of large language models in coreference: An evaluation. In *Joint 30th International Conference on Computational Linguistics and 14th International Conference on Language Resources and Evaluation, LREC-COLING 2024*, pages 1645–1665. European Language Resources Association (ELRA).

- Marcos Garcia and Pablo Gamallo. 2014. Multilingual corpora with coreferential annotation of person entities. In *Proceedings of the 9th edition of the Language Resources and Evaluation Conference*, pages 3229–3233.
- Rebecca MM Hicke and David Mimno. 2024. [lions: 1] and [tigers: 2] and [bears: 3], oh my! literary coreference annotation with llms. *arXiv preprint arXiv:2401.17922*.
- Heeyoung Lee, Angel Chang, Yves Peirsman, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 2013. Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Computational Linguistics*, 39(4):885–916.
- Yanming Liu, Xinyue Peng, Jiannan Cao, Shi Bo, Yanxin Shen, Tianyu Du, Sheng Cheng, Xun Wang, Jianwei Yin, and Xuhong Zhang. 2024. Bridging context gaps: Leveraging coreference resolution for long contextual understanding. *arXiv preprint arXiv:2410.01671*.
- Xiaoqiang Luo. 2005. On coreference resolution performance metrics. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Vancouver, Canada.
- OpenAI. 2023. [GPT-4 technical report](#). *CoRR*, abs/2303.08774.
- Sameer Pradhan, Xiaoqiang Luo, Marta Recasens, Eduard H. Hovy, Vincent Ng, and Michael Strube. 2014. [Scoring coreference partitions of predicted mentions: A reference implementation](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 30–35, Baltimore, MD, USA.
- Sameer Pradhan, Lance Ramshaw, Mitchell Marcus, Martha Palmer, Ralph Weischedel, and Nianwen Xue. 2011. 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.
- Marta Recasens and Eduard H. Hovy. 2011. [BLANC: implementing the rand index for coreference evaluation](#). *Natural Language Engineering*, 17(4):485–510.
- Marta Recasens, Lluís Màrquez, Emili Sapena, M Antònia Martí, Mariona Taulé, Véronique Hoste, Massimo Poesio, and Yannick Versley. 2010. 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.
- Jefferson Fontinele da Silva. 2011. Resolução de coreferência em múltiplos documentos utilizando aprendizado não supervisionado. Master’s thesis, Universidade de São Paulo.
- Marc Vilain, John Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman. 1995. A model-theoretic coreference scoring scheme. In *Proceedings of the 6th Conference on Message Understanding*, pages 45–52, Columbia, Maryland.