

Automated Paraphrasing of Portuguese Informal into Formal Language^{*}

Anabela Barreiro¹, Ida Rebelo-Arnold²,
Jorge Baptista^{1,3}, Cristina Mota¹, and Isabel Garcez⁴

¹ INESC-ID Lisboa, L2F-Spoken Language Lab, R. Alves Redol 9, 1000-029 Lisboa, Portugal

² Universidad de Valladolid UVa, Plaza del Campus S/N 47011 Valladolid, España

³ Universidade do Algarve, Campus de Gambelas, P-8005-139 Faro, Portugal

⁴ Universidade de Lisboa, Alameda da Universidade, 1600-214 Lisboa, Portugal

anabela.barreiro@inesc-id.pt

imdamotoar@funge.uva.es

jbaptis@ualg.pt

cmota@ist.utl.pt

isabelgarcez@campus.ul.pt

Abstract. This paper presents the automation process of paraphrasing and converting Portuguese constructions typical of informal or spoken language into a formal written language. We illustrate this automation process with examples extracted from the e-PACT corpus that involve the placement of clitic pronouns in verbal compound contexts. Our task consists in paraphrasing and normalizing, among others, constructions such as *vou-lhe/posso-lhe fazer uma surpresa* into *vou/posso fazer-lhe uma surpresa* ‘lit: I will/can_to him/her make a surprise / I will/can make_to him/her a surprise; I will/can make him/her a surprise’, where the clitic pronoun *lhe* migrates from an enclitic position after the first verb of the verbal compound to an enclitic position after the main verb, which is the verb responsible for the selection of that pronominal argument. The first verb is either an auxiliary verb or a volitive verb, e.g. *querer* ‘want’. This is a standard revision procedure in EP. Cases like this represent linguistic phenomena where in general language students and language users get confused or stumble. The paper focuses on general language where the phenomena being observed occur, describes examples of interest found in the corpus, and presents an automatic solution for the normalization of informal syntactic inadequacies found in the researched structures into standard formal writing structures through the application of very generic transformational grammars.

1 Introduction

Automation of content revision is one of the most desired functions for a reviewer or an editor, especially for those tedious informal writing "imperfections" that are time-consuming and prevent the efficient and faster revision of author texts. Here, the term

^{*} This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013, exploratory project eSPERTo EXPL/MHC-LIN/2260/2013, and post-doctoral grant SFRH/BPD/91446 /2012.

"imperfection" does not necessarily mean a grammatical error, but the use of informal constructions that are typical of the speech modality, such as the manifestation of orality in written production of professional writers. A paraphraser with automatic normalization and/or revising functions could be used as a learning application for language students, among other applications. In this paper we present the process of converting informal or "less polished" ways of expression into formal expressions employed in written texts, as we wish to create a standard form like those contained in authoring and style guides, e.g. technical guides, used to obtain a high quality publication.

We illustrate this automated process with Portuguese compound verbal predicate constructions (henceforth, *verbal compounds*) involving sequences of two (sometimes more) verbs and a clitic pronoun, where the clitic is an argument of the second verb. The clitic can be placed either attached to the verb it depends on, e.g. *queria ver-te* 'I wanted to see you' and this is the structure that style books and grammars usually recommended for use as "correct", i.e. in formal discourse; or be moved next to the first verb, e.g. *queria-te ver* (EP), *te queria ver* (BP), which is often deemed as "less formal" or even a somewhat "lax" use. While the second verb of the verbal compound is a "full" verb a.k.a. *distributional* verb (that is, a lexical item selecting arguments and with a word sense definable intensionally), the first verb can be an *auxiliary* verb, in the sense defined by [12, pp.393-396], which are often designated as *verbal periphrasis* (in Portuguese *locuções verbais*⁵, e.g. *estou a ver-te* vs. *estou-te a ver* (EP), *te estou a ver* (BP); or a full verb, including the so-called *volitive* verbs like *querer* 'want', *desejar* 'wish', and other verbal constructions. In all these cases, the normalization requires the clitic pronoun to migrate to an enclitic position and attached to the second verb of the verbal compound, e.g., *eu quero-o ver* > *eu quero vê-lo*. In the "correct" example, the infinitive verb undergoes a change of *ver* > *vê* and the clitic pronoun undergoes a change of *o* > *lo*, an orthographic rule motivated by phonetic reasons.

In Natural Language Processing (NLP), most parsers process Portuguese auxiliary verbs in as much the same way as any other verb, i.e., as an ordinary, full verb; see, for example, the parse trees produced by the PALAVRAS [9]⁶ and the LxParser [27]⁷. A different proposal is presented by [4], who process verbal auxiliary constructions in a different way, distinguishing the auxiliary from the main verb, and take into account the different placement options of the clitic pronouns. In fact, auxiliary verbs require an adequate proposal of systematization that takes into account not only the lexical, but also the semantico-syntactic properties of these verbs. Within the scope of the description of the Lexicon-Grammar of (European) Portuguese Verbs [1–3] provide a list of 100+ auxiliary verbal constructions (among 330+ auxiliary-like verbal constructions). In this way, it will be possible to create lists of occurrences and build local grammars that can be used by both human and machine users.

It is important to highlight that all the verbs illustrated and analyzed in this paper form a locution with another verb (the main verb). In many co-occurrences, the meaning

⁵ For a comprehensive overview of the topic, relevant work can be found in [13, 22, 21]. It is also worth mentioning the proposals of [16] for the system of auxiliary verbs in French.

⁶ <http://visl.sdu.dk/visl/pt/parsing/automatic/dependency.php>

⁷ <http://lxcenter.di.fc.ul.pt/services/pt/LXParserPT.html>

of the main verb often gets an aspectual value. There are also verbs whose meanings are constructed with the co-occurrence of a preposition followed by another verb.

Because the topic of our research is so broad in scope and our corpus includes such a wide variety of cases of difficult categorization and computational treatment, we decided to focus only on verbal compounds involving clitics. The examples illustrated in the paper have been extracted from the e-PACT corpus [5], which is made of two novels authored by David Lodge, and the paraphrastic alignments were performed by using the CLUE-Aligner tool [7], which we have been used in paraphrase alignment research works.⁸ The corpus contains both simple and non-standard examples, including sentences typical of dialogue or informal pieces of informal communication, which characterize the type of literary texts that constitute the corpus. We analyzed a certain amount of occurrences in the corpus and created a typology of categories of verbal compounds. Then, we used those categories to create generic local grammars that served as the basis for automated processing of paraphrases. The non-standard/standard pairs of paraphrastic contrasts resulting from this study will be geared towards their integration into the eSPERTo paraphrasing tool, which, among other applications, aims to enable the adaptation and revision of texts. Currently, eSPERTo is integrated in an on-line application that provides paraphrastic suggestions to help Portuguese language learners. As this tool evolves, it is envisaged that its resources will be used in text production and revision.⁹ Another experimental application involves the construction of a dataset of paraphrastic contrasts among the distinct varieties of the Portuguese language, an indispensable resource for variety conversion/adaptation [6, 23]. These efforts are in line with the proposal to create an international standard of Portuguese [24, 25]. Finally, as an initial approach, we have started addressing the topic of teaching learners formal/informal language through the use of chatbots as teacher-agents.

It is relevant to mention that, even though the e-PACT corpus is not optimal, it is the best publicly available resource that suits our purposes, because it contains aligned parallel sentences that are translations of the same literary texts, and these sentences often contain informal language. The lack of parallel corpora of paraphrases in general, but especially for Portuguese, is a necessity that has not been addressed in sufficient supply. Another instrumental factor is that the parallel sentences in e-PACT correspond to two different varieties of the Portuguese language, the European (EP) and the Brazilian (BP) varieties, which we have been contrasting in recent work [6]. These key characteristics are pivotal to both variety adaptation and revision. In this paper, we focus on text revision, but the paper serves the two purposes, conversion of informal EP/BP into formal EP/BP, and adaptation of the BP into the EP variety and vice-versa. We argue that we have made a small but positive contribution to the improvement of editing and revision standards as well as the automation of specific transformations of informal into formal discourse.

⁸ For the purpose of saving space in this paper, we present all examples in a conventional manner, i.e., marked in bold in regular enumerated examples.

⁹ The utility of eSPERTo's paraphrasing capabilities has been explored in two other applications described in [17]: (i) in a question-answering system to increase the linguistic knowledge of an intelligent conversational virtual agent, and (ii) in a summarization tool to assist the paraphrasing task.

2 Related Work

The verbal compounds that are the object of our study and research have the particularity of including a clitic pronoun either in the EP or in the BP sentences or having that clitic implied in a paraphrase of the verbal compound constructions in one or the other Portuguese variety (cf. example (2)). In Portuguese, a clitic pronoun plays a syntactic role at the sentence level, and follows different placement ordering rules depending on the language variety (EP or BP), the number and the semantics of the predicates, co-occurrence with a preposition, among other factors.

There are several studies that focused on the acquisition of clitics in EP, of which [11] and [26] are just two examples among others that have been referenced in recent related work [23]. Those studies are mainly related to difficulties in the performance of using clitics in the early stages of language acquisition. Clitic acquisition difficulties are materialized, in particular, by choices outside the norm for its placement in sentences. When we look at our data, it seems that the hesitations and difficulties extend into adulthood, and there are patterns of variation in the selection and position of the clitics in any corpus of oral modality or simply of written transcription of this modality, where informality is recurrent in all paths of modern writing, including social media, but also more "serious" channels of communication, such as newspapers, opinion articles, and literary writing, among others.

In BP, in turn, several studies focus on the observation of the spontaneous constructions of more or less schooled speakers involving the use of clitics, cf. [10, 18–20], among others. This observation reveals a distance between the two varieties with respect to application of the rules for selection and placement of clitics in Portuguese. All this has an impact on both the work of reviewers and translators and on language learning, either for Portuguese as a native/mother language (PML) or Portuguese as a foreign language (PFL). eSPERTo can be used in a language learning environment, where both PML and PFL students can learn how to produce and apply precise paraphrases (i.e., semantically equivalent sentences). Therefore, the resources herein created can help assisting writers in text production, revision, or adaptation, but they can also be valuable in a classroom environment.

In this paper, we continue a previous line of research [6], where a first general introduction to a broader task of gathering EP–BP paraphrastic variants has been put forward, followed by a narrowed approach to the question of EP–BP paraphrases involving the third person clitic with dative value, *lhe* [23]. In this study, we focus on the alignment of verbal compound constructions, when these constructions deal with clitic pronouns. Our small experiment shows that the methodology and approach is feasible in a larger stand-alone project as long as an enough amount of adequate corpora is available to provide the wide-coverage necessary for an efficient normalization process, as that required for a large-scale paraphrasing system. These data will also constitute the building blocks for the creation of grammars that can be applied to multiple cases.

3 Placement of Clitics in Verbal Compounds

Clitics in Portuguese can move to the left or right either of the auxiliary or the main verb. Some of the nuances to clitic placement in verbal compounds will be illustrated

in this paper with examples from the e-PACT corpus. Part of the difficulties in establishing paraphrase categories are related to the approximate value of seemingly "equivalent" constructions. The examples illustrate that in each EP–BP paraphrastic pair, one sentence contained a verbal compound with a clitic and the other sentence contained a paraphrase of it. Sometimes the paraphrase presents a quite different verbal compound structure, which may not even include the clitic pronoun appearing in the semantically-equivalent sentence.

3.1 PROCLDAT or ACC VAUX-ter VPARTPASS

Examples (1)–(3) represent important contrasts with the evidently productive rule of EP enclisis. They occur in the presence of the auxiliary *ter* ‘have’ (VAUX-ter) and are probably the model that generates the incorrectness in the construction *lhes voltava a telefonar* ‘called them back’. This is a case of a false analogy because, in fact, the enclitic placement rule should have been applied here, e.g., *voltava a telefonar-lhes* ‘he would call them again’. In the BP paraphrase the clitic pronoun is completely avoided by using a more loose transformation.

- (1) *EN* - *It was rumoured that he collected the phone numbers of likely-sounding girls and **called them back** after the programme to make dates.*
EP - *Dizia-se que colecionava os números das raparigas que mais lhe agradavam e **lhes voltava a telefonar** depois, a marcar encontros.*
BP - *Diziam até que ele colecionava números de telefone de garotas com voz macia **para ligar mais tarde** e marcar encontros.*

In example (2), the paraphrase in BP, [N VAUXter NP[*boa viagem*]] ‘N have NP[a good trip/flight]’ (simplified ‘[Y have good X]’) presents an inversion of the topic to avoid the use of the 3rd person clitic requested for the verb *agradar* ‘to please’ as a paraphrase of the EP [NP[a *viagem*] VMAIN *agradou* PREPa N] ‘[NP[the trip/flight] VMAIN pleased N]’ (simplified ‘[X please Y]’). In BP, the different lexical selection explains the absence of ENCLITDAT. In the EP sentence, the presence of the clitic pronoun *lhes* is suppressed in BP by the inversion of the topic. The verb *agradar* in Portuguese requires the use of the preposition *a* (PREPa), not required by the verb *please* in English. The EP paraphrase is more formal while the one in BP is more neutral. The pronoun *lhe* can never be attached to a past participle in [VAUXter + VPP] auxiliary constructions.

- (2) *EN* - *and he hopes they have enjoyed the flight*
EP - *e diz esperar que a viagem **lhes tenha agradado**.*
BP - *e ele desejava que **tivessem tido uma boa** viagem*

In example (3), EP also presents a more formal paraphrase (more close to the original English wording) than BP. The variation from one paraphrase into another presumes a translator’s choice. In detail, the paraphrase in BP selects the same lexical item as in EP, *mudar* ‘change’ which occurs with the reflexive pronoun *se*, but with a compound personal infinitive and PROCLIT from clitic to the main verb (VMAIN). However, the verb *mudar-se* (*de X para Y*) ‘move from X to Y’ is ambiguous, i.e., the reflexive (*-se*)

is optional (the sentence would still be correct if the reflexive pronoun was omitted as in *tivessem mudado para...*). This occurrence (less formal in BP) is attested, however, in the BP grammar that rejects the use of clitics prior to VAUX. The variety determines the order of the clitic. In an EP subordinate clause *se* appears before VAUX.

- (3) *EN* - though they **moved** in due course **to** better insulated accommodation
EP - embora mais tarde **se tivessem mudado para** uma habitação bem isolada
BP - mesmo depois de **terem se mudado para** acomodações mais isoladas

3.2 VAUX PREP VINF+ENCLITDAT-lhe versus VAUX2 lhe VGER NP

In example (4), the EP verbal compound is standard, but its paraphrase in BP is much closer in structure to the original English source, which makes it sound odd. There is no evidence as whether this was an intentional faithfulness to the original, or an unsuccessful attempt to use controlled language. The paraphrase in EP consists of the periphrastic construction [*continuar a* + VINF ENCLDAT] ‘continue to’. In BP, the somewhat complex paraphrase involves the modal auxiliary *dever* ‘should’ followed by the adverb *ainda* ‘still’ followed by the construction [VAUX-*estar* PROCL-lhe VGER *causando* NP]. The entire sequence of elements in BP has as a semantic axis the aspectual notion of action in progress, identical to that of the EP paraphrase that is expressed in a much simpler and more concise construction. This example illustrates the need, already mentioned in this paper, to construct grammars for the specific purpose of generating paraphrases that are suitable for reviewers, editors and students of PFL. We cannot state categorically whether the BP version is due to a recurring use of the construction in this variety or whether it is simply a result of poor interpretation by the translator. In addition, it may include not only DAT *lhe*, but also all ACC when the main verb is in the infinitive form, VINF. This rule applies even in the presence of a negation *não* ‘not’ that precedes the verb in the position VAUX in the verbal compound. The verb *continuar* ‘continue’ is a VAUX (*ter*, *ser*, etc.) typical of verbal periphrasis, therefore it assigns aspectual meaning to the main verb *doer* ‘ache’, occupying the position of a non-typical auxiliary, such as in *não conseguiram dominá-la* ‘they were unable to dominate her’.

- (4) *EN* - There’s no bally reason why [] **should be giving you any more pain**.
EP - Não há a mínima razão para [] **continuar a doer-lhe**
BP - Não há um pingo de razão por que [] **deva ainda estar lhe causando essa dor**

3.3 PREP-a VINF+REFLPRO-se > PROCLITse VGER

In example (5), EP determines enclisis while BP determines proclisis. It is interesting to note that both varieties maintain the aspectual notion of progression. This notion is doubly represented, either by the selection of PREP-*a* VINF in EP and a gerundive VGER in BP with the ellipse of the auxiliary *estar* in both constructions, and by the lexical selection, by which, both reflexive verbs *formar-se* ‘form’ and *preparar-se* ‘prepare’ express the notion of an ongoing action. These do not correspond to paraphrases in the transformational sense defined by Maurice Gross [14, 15], however, the alignment task provides candidates to be added as paraphrastic pairs to a paraphrasing system. This is

an important and necessary formalization that proposes to systematize the paraphrases between EP and BP, although their establishing seems, at first glance, rather complex. The importance of this example resides in the fact that the opposition PREP-*a* VINF > VGER is a distinctive mark between the two varieties of Portuguese. Thus, it becomes necessary to offer exhaustive lists of paraphrastic possibilities always with the utmost care so that the meaning between the paraphrases remains of high quality, independently on whether our goal is to establish a controlled version of Portuguese, to assist in the revision task, to support text editing or to teach PFL.

- (5) *EN* - *I sense a storm of depression flickering on the horizon, and a tidal wave of despair gathering itself to swamp me.*
EP - *Sinto uma tempestade de depressão avolumar-se no horizonte e uma maré de desespero a formar-se para me engolir.*
BP - *Pressinto a chegada de uma tempestade de depressão se formando no horizonte e uma onda de desespero se preparando para me engolir.*

4 Normalization of Informal into Formal Language

Based on the main characteristics pointed out in Section 3 regarding the placement of clitics in verbal compounds in several contexts: (i) co-occurring with modals (VMOD) in relative sub-clauses; (ii) several cases of the use of proclitics or enclitics in formal and informal contexts (3.1); (iii) co-occurring with aspectual verbs (VASP) in periphrastic constructions (3.2); or (iv) co-occurring with aspectual verbs with the progressive meaning (3.3), we propose here a local grammar that allows the normalization of an informal verbal compound construction, where the enclitic pronoun appears after a verb (V). This verb can be an auxiliary (VAUX) or any other verb form (VASP, VMOD, etc.). This informal verbal construction is normalized into an equivalent formal construction, where the clitic is moved to a position after the main verb in the infinitive form (VINF) illustrated in Figure 1.

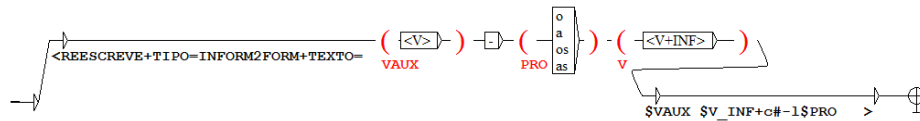


Fig. 1: Grammar to normalize informal into formal language with clitics

Based on the proposed grammar, a few hundreds of normalization/paraphrasing procedures take place. These standard paraphrases can integrate the eSPERTo paraphrasing system after validation by a linguist. Figure 2 illustrates the revision capability within eSPERTo, where a sentence written in an informal or less careful language, can be revised with suggestions that are more polished, or correspond to a more standardized norm of the written language. For example, for the sentence *A menina generosa queria-o surpreender todos os dias* ‘The generous girl wanted to surprise him everyday’, eSPERTo presents as a conversion option for the informal verbal compound with

clitic *queria-o surpreender*, its formal equivalent *queria surpreendê-lo* ‘wanted to surprise him’. The paraphrasing system offers this paraphrasing suggestion to the user, where the clitic migrates from an enclitic position attached to the verb *querer* ‘want’ to an enclitic position attached to the main verb. This transformation requires the infinitive form of the main verb *surpreender* ‘surprise’ to change to *surpreendê-* before the enclitic ACC pronouns *-lo*, *-la*, *-los*, *-las*, an orthographic rule motivated by phonetic reasons, as in previous examples (cf. Section 1).

eSPERTo - System for Paraphrasing in Editing and Revision of Text

The screenshot displays the eSPERTo web interface, which is divided into two main panels: 'Parameters' on the left and 'Input file or text (click to show/hide)' on the right.

Parameters Panel:

- Demo mode:** ☐
- Interface idiom:** English (dropdown)
- Resources idiom:** Portuguese (dropdown)
- Dictionary:** PT-Dict_NEW (dropdown)
- Sample text:** SAN (dropdown)
- Paraphrasing:** Check all (checked), Uncheck all (button)
- ☐ Active > Passive
- ☐ Passive > Active
- ☐ Simple adverb > Compound
- ☐ Compound adverb > Simple
- ☐ Nominal/adjectival predicate > Verb
- ☐ Nominal/adjectival predicate >
- Nominal/adjectival predicate:**
 - ☐ Verb > Nominal/adjectival predicate
 - ☐ Relative construct > Adjective
 - ☐ Possessive > Relative construct
 - ☐ Synonyms
 - ☐ Human intransitive adjective
 - ☐ Predicate nouns with Vsuf fazer
 - ☐ Predicate nouns with Vsuf ser de
 - ☐ PE > PB
 - ☐ PB > PE
 - ☐ SAL
 - ☒ Informal > Formal
- Debug:** ☐
- Process results** (button)

Input file or text (click to show/hide) Panel:

- Choose file:** Browse file (button)
- Insert text in the text box**
- Text input box: A menina generosa queria-o surpreender todos os dias.
- Process results** (button)

Results (click to show/hide) Panel:

- Text input box: A menina generosa [**queria-o** **surpreender**] todos os dias .
- Save paraphrased text** (button)
- Dropdown menu: **queria surpreendê-lo**
- Text input box: Suggest your own paraphrase >

Fig. 2: Conversion of an informal verbal compound with a clitic pronoun into a formal equivalent where the clitic appears after the main verb

5 Conclusions and Future Work

Language style revision is an important feature of the eSPERTo project, whose main focus is the development of an innovative paraphrasing system with capacity to produce semantically equivalent sentences and ways of expression, always aiming the improvement of the quality of each text. In this paper, we tried to established some categories defined on the basis of the syntactic structure for verbal compound constructions involving clitics. We have done this study based on pairs of paraphrastic constructions extracted from sentences of two David Lodge’s novels translated into EP and BP.

It is important to notice that, especially in literary texts, translators often use a free translation, which (optimally) preserves the meaning of the original text, but involves the restructuring of syntax, sometimes with a flexible use of the lexicon or expressions to deliver a natural articulation of words in the target language. As a result, the translated text may appear "soft" or more or less idiomatic with regards to the original text. In

this process, even professional human translators may introduce errors turning a particular part of a translation unfaithful to the original. All in all, translation can be seen as a paraphrasing process using words in another language, where the introduction of different words and structures can create some distance between source and target languages. In this sense, in our study, paraphrases assuming full semantic equivalence compete with paraphrases that retain an approximate conceptual equivalence [8]. The first ones are indispensable for precision, but we cannot dispense the second ones because they also play an important role in paraphrasing tasks, namely in revision or style change.

Data extracted from corpora, although useful and containing statistical significance, require linguistic analysis and categorization of patterns/structures and semantic equivalence. We hope that our attempt to define a typology and use linguistic knowledge to normalize informal constructions will continue because it reveals a crucial task in developing a revision or language improvement tool. This paper sheds light on the necessity to include a formal/informal feature into several applications for text editing and revision, including in a language learning environment, in which learners need to grasp both formal and informal ways of communication. In the near future, we will discuss the topic of including chatbots as agents for teaching learners the differences between formal and informal language based on the learner's written input. For very formal language texts, the chatbots can suggest more informal phrases, or the other way around. This topic will be explored in the scope of enetCollect COST Action, where chatbots will play a role in a language learning application.

References

1. Baptista, J.: ViPER: A Lexicon-Grammar of European Portuguese Verbs. In: Radimsky, J. (ed.) Proceedings of the 31st International Conference on Lexis and Grammar. pp. 10–16. Università degli Studi di Salerno (Italy)/University of South Bohemia in Nové Hradý (Czech Republic), Nové Hradý (Czech Republic) (2012)
2. Baptista, J.: ViPER: uma base de dados de construções léxico-sintáticas de verbos do Português Europeu. In: Silva, F., Falé, I., Pereira, I. (eds.) Actas do XXVIII Encontro da APL - Textos Seleccionados. pp. 111–129. APL/Colibri, Lisboa (2013)
3. Baptista, J., Mamede, N.: Dicionário gramatical de verbos do português europeu (October 2016), (in print)
4. Baptista, J., Mamede, N., Gomes, F.: Auxiliary verbs and verbal chains in European Portuguese. In: Pardo, T., Branco, A., Klautau, A., Vieira, R., de Lima, V.L.S. (eds.) Computational Processing of the Portuguese Language. pp. 110–119. No. 6001 in Lecture Notes in Computer Science / Lecture Notes in Artificial Intelligence, PROPOR 2010, Springer, Berlin (2010)
5. Barreiro, A., Mota, C.: e-PACT: eSPERTo Paraphrase Aligned Corpus of EN-EP/BP Translations. *Tradução em Revista* 1(22), 87–102 (2017)
6. Barreiro, A., Mota, C.: Paraphrastic Variance between European and Brazilian Portuguese. In: Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects, VarDial@COLING 2018, Santa Fe, New Mexico, USA. Association for Computational Linguistics (2018)
7. Barreiro, A., Raposo, F., Luís, T.: CLUE-Aligner: An Alignment Tool to Annotate Pairs of Paraphrastic and Translation Units. In: Proceedings of the 10th Edition of the Language Resources and Evaluation Conference. pp. 7–13. LREC 2016, European Language Resources Association (2016)

8. Barzilay, R., McKeown, K.R.: Extracting paraphrases from a parallel corpus. In: Proceedings of the 39th Annual Meeting on Association for Computational Linguistics. pp. 50–57. Association for Computational Linguistics (2001)
9. Bick, E.: The Parsing System “PALAVRAS”. Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Aarhus University Press (2000)
10. Castilho, A.: O português do Brasil. In: Ilari, R. (ed.) *Linguística Românica*. pp. 237–269. Fundamentos 83, Ática (2011)
11. Costa, J., Grolla, E.: Pronomes, clíticos e objetos nulos: dados de produção e compreensão. In: *Aquisição de língua materna e não materna: questões gerais e dados do português*. pp. 177–199. Language Science Press, Berlin, Germany (2017)
12. Cunha, C., Lindley-Cintra, L.: *Nova Gramática do Português Contemporâneo*. João Sá da Costa, 1ª ed edn. (1986)
13. Gonçalves, A.: Predicados complexos verbais em contexto de infinitivo não-preposicionado do português europeu. Ph.D. thesis, Universidade de Lisboa (2000)
14. Gross, M.: *Méthodes en syntaxe: régime des constructions complétives*. Actualités scientifiques et industrielles, Hermann (1975)
15. Gross, M.: Les bases empiriques de la notion de prédicat sémantique. *Langages* **15**(63), 7–52 (1981)
16. Gross, M.: La fonction sémantique des verbes supports. *Travaux de Linguistique: Revue Internationale de Linguistique Française* **37**(1), 25–46 (1998)
17. Mota, C., Barreiro, A., Raposo, F., Ribeiro, R., Curto, S., Coheur, L.: eSPERTo’s Paraphrastic Knowledge Applied to Question-Answering and Summarization. In: Barone, L., Monteleone, M., Silberstein, M. (eds.) *Automatic Processing of Natural-Language Electronic Texts with NooJ: 10th International Conference, NooJ 2016, České Budějovice, Czech Republic, International Publishing, Cham* (2016)
18. Naro, A.J., Scherre, M.M.P.: *Origens do Português Brasileiro*. Parábola (2007)
19. Neves, M.H.M.: *Gramática do português falado*. Unicamp (1999)
20. Neves, M.H.M.: *Gramática de usos do português*. Unesp (2000)
21. Paiva Raposo, E.: Verbos auxiliares. In: Paiva Raposo, E., Nascimento, M., Mota, M., Segura, L., Mendes, A. (eds.) *Gramática do Português*. vol. 2, pp. 1221–1281. Fundação Calouste Gulbenkian, Lisboa (2013)
22. Pontes, E.: *Verbos Auxiliares em Português, Perspectivas Linguísticas*, vol. 10. Vozes, Petrópolis, RJ (1973)
23. Rebelo, I., Quaresma, A.B.P.: EP–BP Paraphrastic Alignments of Verbal Constructions Involving the Clitic Pronoun *lhe*. In: *International Conference on Computational Processing of Portuguese. PROPOR 2018*, Springer (2018 (forthcoming))
24. Santos, D.: Como estudar variantes do português e, ao mesmo tempo, construir um português internacional?, presentation at Contact, Variation and Change: corpora development and analysis of Iberoromance language varieties workshop
25. Santos, D.: Portuguese language identity in the world: adventures and misadventures of an international language. In: Khachaturyan, E. (ed.) *Language - Nation - Identity: The questione della lingua in an Italian and non-Italian context*, pp. 31–54. Cambridge Scholars Publishing (2015)
26. Silva, C.G.A.G.: *Assimetrias na Aquisição de Clíticos Diferenciados em Português Europeu*. Master’s thesis, Universidade Nova de Lisboa, Lisboa, Portugal (2008)
27. Silva, J., Branco, A., Castro, S., Reis, R.: Out-of-the-box robust parsing of portuguese. In: Pardo, T., Branco, A., Aldebaro, K., Vieira, R., Strube de Lima, V. (eds.) *Computational Processing of Portuguese. Proceedings of the 9th International Conference on the Computational Processing of Portuguese (PROPOR’2010)*. Lecture Notes in Computer Science / Lecture Notes in Artificial Intelligence, vol. 6001, pp. 75–85. PROPOR 2010, Springer, Porto Alegre RS, Brazil (April 2010 2010)