EP–BP Paraphrastic Alignments

Verbal Predicate Constructions with the Clitic Pronoun *lhe*⋆

Ida Rebelo-Arnold[1], Anabela Barreiro[2], and Paulo Quaresma[3]

[1] Universidad de Valladolid UVa, Plaza del Campus, S/N, 47011 Valladolid, España
[2] INESC-ID Lisboa, L2F-Spoken Language Lab, R. Alves Redol 9, 1000-029 Lisboa, Portugal
[3] Universidade de Évora, R. Romão Ramalho, 59, 7000 Évora, Portugal
imdamotar@funge.uva.es
anabela.barreiro@inesc-id.pt
pq@uevora.pt

Abstract. This paper presents the alignment of verbal predicate constructions with the clitic pronoun *lhe* in the European (EP) and Brazilian (BP) varieties of Portuguese, such as in the sentences *Já* ***lhe*** *arrumaram a bagagem* | ***Sua*** *bagagem está seguramente guardada* "His baggage is safely stowed away", where the EP dative proclisis *lhe* contrasts with the BP possessive pronoun *sua*. We have selected several different paraphrastic contrasts, such as proclisis and enclisis, clitic pronouns co-occurring with relative pronouns and negation-type adverbs, among other constructions to illustrate the linguistic phenomenon. Some differences correspond to real contrasts between the two Portuguese varieties, while others purely represent stylistic choices. The contrasting variants were manually aligned in order to constitute a gold standard dataset, and a typology has been established to be further enlarged and made publicly available. The paraphrastic alignments were performed in the e-PACT corpus using the CLUE-Aligner tool. The research work was developed in the framework of the eSPERTo project.

1 Introduction

In this paper we propose an overview of the use of the clitic pronoun *lhe* in European (EP) and Brazilian (BP) Portuguese. Our methodology consists of applying linguistic knowledge in the alignment of pairs of paraphrastic contrasts between the two varieties, and our main focus is to discuss the different semantico-syntactic behaviour of *lhe* in the constructions in which it occurs, defining a typology for the different "uses".

We have analyzed pairs of paraphrastic units aligned and collected from a subcorpus of e-PACT[4], a written parallel corpus of aligned paraphrases [2]. The illustrative examples presented throughout the paper represent EN–EP and EN–BP translations of the same fiction novels by David Lodge. These novels include some dialogue and informal pieces of oral communication, with a mixture of simple and complex sentences,

⋆ 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.

[4] e-PACT is an acronym for **e**SPERTo **P**araphrase **A**ligned **C**orpus of EN-EP/BP **T**ranslations.

well formed clauses, or sentences containing a standardized use of the EP and BP clitics, including *lhe*, and it also comprises non-standard examples. After having analyzed a certain amount of *lhe* occurrences in the corpus, we proceeded to establish a typology that covers the clitic's most frequent uses. In addition, we have also explored a computational-style annotation where the paraphrastic alignments can be used to create generic local grammars, and will serve as the basis for automated processing of paraphrases. The alignments were performed in the CLUE-Aligner tool[5] [4].

The pairs of paraphrastic contrasts resulting from this study will be geared towards their integration into a paraphrasing tool. Those contrasting pairs would enable the conversion of *lhe* constructions from one variety into another, such as in the sentence ***A Philip*** *só ocorria um nome* | *Apenas um nome* ***lhe*** *veio à cabeça* "Only one name would come into Philip's head", where the non-eliptic complement *A Philip* in EP represents a contrast with the dative proclisis *lhe* in BP. It is important to point out that almost all occurrences found in texts and highlighted here have been mentioned by authors who recognized an ongoing variation pattern between both the EP and the BP uses [7, 8, 11, 13, 14, 16], but none of the cases has been described or categorized in the way it is done in this paper, i.e., from a computational perspective to use in a paraphrase generating system, by employing an alignment tool, using corpora from which the pairs of paraphrastic units were extracted, and analyzing the collected data in order to define a typology of EP–BP variety contrasts.

The research work presented here was developed within the eSPERTo project[6][7] that aims to build an innovative smart, context-sensitive and linguistically enhanced automated paraphrasing system with capacity to produce semantically equivalent sentences and ways of expression to assist writers and language learners in text production, revision, or adaptation. Future developments of eSPERTo aim to enable the adaptation of a text within the different varieties of the Portuguese language, such as EP and BP [3, 5].

2 Related Work

Clitic pronouns are linguistic elements used to express direct and indirect objects, which decline in several cases, nominative, acusative, dative, etc. In Portuguese, a clitic pronoun plays a syntactic role at the phrase level, and follows different placement order rules depending on each variety. In this regard, important contrasts lie in the syntactic preferences of the EP and BP Portuguese varieties. Empirical evidence shows that word order rules are not always clear and understandable to speakers of both Portuguese varieties, so the main relevance in providing paraphrases between these varieties is that they may cause misunderstandings sometimes and paraphrases may help resolve those

[5] www.esperto.l2f.inesc-id.pt/esperto/aligner/index.pl?

[6] https://esperto.l2f.inesc-id.pt/esperto/esperto/demo.pl

[7] Experiments have used eSPERTo to enrich the paraphrastic capabilities in a dialogue system, e.g., to increase the linguistic knowledge of an intelligent virtual agent, and to produce "smart" text reductions in a summarization tool. Recent experiments aim to provide new paraphrastic resources in an e-learning environment, and generate precise paraphrases to be used in machine translation and in professional translation, editing, and proofreading.

misunderstandings. Data reveals that each variety tends to put in evidence its own preferences of clitic usage. Clitics may appear after the verb (enclisis), in the middle of the verb, i.e., between the radical and the ending (mesoclisis), or before the verb (proclisis). Table 1 presents the frequency of clitic pronouns in the two complete novels of David Lodge from which the 40% of sentences that constitute the e-PACT corpus were extracted. In order to obtain these values the Freeling parser [6] was used and the clitic pronouns were identified. Then, a program was developed to count the proclitic and enclitic situations, taking into account the parsing structure of each sentence.

David Lodge		**me**	**te**	**se**	**lhe**	**nos**	**vos**	**lhes**	**a**	**o**	**as**	**os**
Book 1	PT	407	16	331	109	48	0	13	52	74	9	14
	PT-Enclitic	221	9	168	64	23	0	6	38	48	6	11
	PT-Proclitic	186	7	163	45	25	0	7	14	26	3	3
	BR	281	2	285	26	28	0	0	50	53	4	22
	BR-Enclitic	69	1	75	6	7	0	0	35	36	3	18
	BR-Proclitic	212	1	210	20	21	0	0	15	17	1	4
Book 3	PT	29	7	296	127	7	0	10	20	80	3	20
	PT-Enclitic	18	4	146	67	6	0	4	17	52	2	14
	PT-Proclitic	11	3	150	60	1	0	6	3	28	1	6
	BR	22	0	291	41	1	0	0	20	56	5	18
	BR-Enclitic	7	0	98	12	1	0	0	17	31	4	15
	BR-Proclitic	15	0	193	29	0	0	0	3	25	1	3

Table 1: Clitic pronouns in David Lodge novels

In general, in EP most frequently the clitic is joined to the verb on which it depends by means of a hyphen if postposed (enclitic), while in BP it is written as separate word before the verb (proclitic) in declarative sentences. But, there are many nuances to clitic placement, some of which we will illustrate in this paper with examples from corpora. It should be stated that the use of mesoclitics is quite common in EP, and several cases of mesoclitics can be found in our corpus, but none of them represent the clitic *lhe*. Therefore, this phenomenon should be addressed in forthcoming research, but not here. Most contrasts found have been pointed out and, sporadically, contemplated in analyzes by several authors in a more or less detailed way [1, 8–10, 13, 15, 18]. Comparing our perspective with previous works, either of a theoretical or practical nature, syntactic properties are insufficient to deal with clitics in an effective way. Moreover, the use of clitics in Portuguese is too broad a field of research, which we narrowed to the question to EP–BP paraphrases involving the third person clitic with dative value, *lhe*.

From a linguistic perspective, the first grammar-compendium explicits the shift away from traditional grammar rules towards different representations in actual variety uses [11]. We summarize most of the use particularities in EP and in BP contemplated in the follow-on literature mentioned above. On the one hand, EP (i) displays preferably enclisis and only sometimes allows proclisis, accepting even, mesoclisis which is a kind of archaic clitic position; (ii) admits clusterization with dative and accusative function in one lexical item; (iii) avoids nominative personal pronouns with accusative value;

and (iv) displays generalized dative clitics as possessives. On the other hand, BP (i) displays preferably proclisis and only sometimes allows enclisis; (ii) avoids mesoclisis in standard written and spoken language, even if it may be found in a literary corpus; (iii) admits no clusterization of the dative and accusative function; (iv) accepts nominative personal pronouns with accusative value.

From a computational perspective, a set of paraphrastic variants between EP and BP resulting from a previous alignment research task have been described to be used in a paraphrasing system [3]. A high quality paraphrasing system, which has the ambition to include a variety adaptation module to deal with cultural, linguistic and stylistic differences between varieties, requires a large dataset of paraphrastic contrasts among the distinct varieties of the Portuguese language. Such large amount of paraphrastic resources simply does not exist for the Portuguese language. Our broader task consists of gathering paraphrastic variants, including multiwords and other phrasal units, such as the compounds *toda a gente* versus *todo o mundo* "everybody" or the gerundive constructions [*estar a* + V-Inf] versus [*ficar* + V-Ger] (e.g., *estive a observar* | *fiquei observando* "I was observing"), among others. In this paper, we will continue previous line of research [3], but will focus on the alignment of verbal predicate constructions when these constructions co-occur with the clitic pronoun *lhe*. To capture these contrasts in corpora is very important, because the occurrence of linguistic phenomena in texts is indispensable for wide-coverage and for an efficient variety adaptation process. A (semi-)automatic conversion of texts from one variety into another represents a very important function in paraphrasing systems. Furthermore, the resources resulting from the alignment task can be useful and of value for other applications, including language learning, summarization, question-answering, dialogue, plagiarism detection, text authoring and revision, machine translation, among others.

3 The Use of Clitics in Portuguese

Semi-automatic alignments allow us to evaluate the degree of acceptability of the selected paraphrases, because they are made by linguists who are native speakers of EP or BP. Therefore, some comments were made in cases where there is a reasonable distance between varieties, which results in **approximate** paraphrases, semantically less faithful or with different degrees of precision. These characteristics will be, in our view, relevant especially if we consider the application envisaged for them, and the paraphrastic choices will vary according to whether they are intended for teaching Portuguese as a Foreign Language (PFL), as a tool to support the revision and editing of texts or for a search engine with alternatives between varieties. Our paper only points to these uses, concentrating on the description of occurrences and on the possibility of creating lexicon-grammars that systematize them; we leave for a future work the distinction between possible paraphrases in both varieties and paraphrases which are predominantly stylistic, being suitable for either variety.

3.1 Clitics after Adverbials and the Relative Pronoun *que*

EP follows the general rule that a relative *que* will attract the clitic and maintain the proclitic position. In fact, a secondary conclusion could be that the antecedent rule is, at

least in EP, stronger than the tendency to perform enclitic constructions in this variety. Example (1) expresses this feature, by means of what EP selects, a Dative proclitic *lhe* after the relative pronoun *que* while in the BP paraphrase the clitic is elided. In BP, there seems to be some ambiguity with regards to the subject that a larger context clarifies.

(1) $_{EN}$ - *listened to what he took, at the time, to be a very funny parody*
$_{EP}$ - *ouvira o que* ***lhe*** *pareceu ser uma paródia muito divertida* - [V–PRO_DAT(PROCL) ANTEC-QUE]
$_{BP}$ - *ouviu o que parecia ser [ ] uma paródia muito engraçada* - [V-PRO_θ]

3.2 Dative versus Nominative Pronoun

The use of the Dative clitic versus the Nominative clitic is also a common EP–BP contrast. Example (2) illustrates the contrast between the use of a Dative clitic pronoun *lhe* in an enclitic position afer the verb *vendo* "sell" and the use of a prepositional phrase constituted by the preposition *para* with a pronoun in the Nominative case (NOM), *ele*, i.e., *para ele* "to him". This phenomenon may also happen with other prepositions, such as *em* or *com* (e.g., *nunca aconteceu* ***com ele*** "it never happened to/with him").

(2) $_{EN}$ - *I'll sell* ***him*** *my [plane] ticket*
$_{EP}$ - *vendo-****lhe*** *o bilhete* – [V–PRO_DAT(ENCL)]
$_{BP}$ - *vou vender a passagem* ***para ele*** – [V–PREP PRO$_{NOM}$]

On the other hand, among the most interesting occurrences in our study is the occurrence of the dative clitic in EP, which can be paraphrased in BP by the presence of a possessive. This is widely referenced and analyzed in [17], whose reading can broaden understanding of this phenomenon already observed long ago by [11].[8]

4 Typology of *lhe* Constructions in the Corpus

Table 2 summarizes the typology of *lhe* constructions in the e-PACT subcorpus selected by us, contrasting the EP and BP varieties of the Portuguese language.

The **first example** in the Table illustrates, in EP, the verb *sair* followed by two complements each preceded of the preposition *a* "to". The first complement is [+HUM] and the second one is "[+VALUE/currency]". The human complement is expressed by the form of the Dative clitic *lhe* and the monetary value complement is expressed by the preposition followed by the named entity *a 300 dólares*. The second complement of its paraphrase in BP contains no preposition due to the nature of the verb *custar* "cost", which does not select a preposition. There is complete Semantic Identity (Sem), but partial Syntactic Identity (Syn) because of the non-correspondence of all the terms that organize the sequence. The non-match, however, seems to be only in the surface

[8] The translation into null objects in BP seems more frequent than in EP, but this claim requires support from quantitative data. This support could be provided by the analysis of more paraphrases in our corpus or by the use of more data from COMPARA or other English-Portuguese parallel corpora.

Paraphrastic Alignments	Clitics			Paraphrasis Identity		
	VComp	PL	Ant	Lex	Sem	Syn
it would cost him 300 dollars						
$_{EP}$ ia sair-lhe a 300 dólares	Dat	Encl	-	-	+	Parcial
$_{BP}$ ia custar-lhe 300 dólares	Dat	Encl	-		+	
looking out of the window still gives him vertigo						
$_{EP}$ olhar pela janela continua a dar-lhe vertigens	Dat	Encl	-	Parcial	+	T. shift
$_{BP}$ sentía vertigens só de olhar pela janelinha	Ø				+	
with which she prepared his breakfasts						
$_{EP}$ com que lhe preparava os pequenos-almoços	Dat	Procl	Rel	-	+	Parcial
$_{BP}$ com que preparava o seu café da manhã	Poss				+	
it had never happened to him						
$_{EP}$ nunca tal lhe acontecera	Dat	Procl	Neg	+	+	-
$_{BP}$ isso nunca tinha acontecido com ele	Prep+Nom	Encl	Neg	+	+	-
bestowing upon them the title						
$_{EP}$ lhes conferia o título	Dat	Procl	-	-	+	-
$_{BP}$ agraciava-os com o título	Acc	Encl	-	-	+	-

Table 2: Typology of *lhe* Constructions in the e-PACT Subcorpus in EP and BP

structure, since the sequence [V + PREP + N+HUM + N+VALUE/currency] proves adequate to express both paraphrases. Even if, to fully realize the paraphrase, they do not have the same constraints regarding preposition selection in the second argument, the Semantic Identity is maintained integrally.

The **second example** reveals somewhat more complex paraphrases, at first glance, due to the alternation of the topicalized element in the support verb constructions, *dar-lhe vertigens* in EP and *sentia vertigens* in BP, "gives [] vertigo". This alternation is a consequence of the support verb selected in each variant, *dar* "give" in EP or *sentir* "feel" in BP. If we translate both paraphrases in a schematic way, we would have: [this gives me N], in EP and [I feel/have N when this happens], in BP. The demonstrative pronoun *this* is the topic in EP, because it is the Agent to the verb *dar*. However, in BP, the paraphrase selects a stative verb *sentir*. In the BP paraphrase, though, the support verb construction *sentia vertigens* "felt vertigo" has a resultative meaning due to the existence of an idiomatic multiword unit that has an inchoative meaning of causing something to happen (e.g., *sinto enjôo só de olhar para a comida* / *só de entrar no carro* / *só de ver a estrada* "I feel sick just by looking at the food / just by getting in the car / just by seeing the road"). So, in case of a NLP application, we would need to create a formula that can take into account not only the topic shift but also the BP verb aspectual meaning of the support verb construction *continua a dar-lhe vertigens* "continues to give him vertigo". The different support verb selection causes the disappearance of the Dative clitic in BP. The Lexical Identity (Lex) keeps existing at least partially. It is worth noting that in EP the original English aspect of continuity resulting from the use of the adverb *still* is not preserved in the BP sentence, where the notion of continuity was eliminated. We consider this more as a stylistic option than a real variety contrast between EP and BP.

In the **third example**, the variation that leaps to the eye is Lexical, with the alternation between *pequeno-almoço* / *café da manhã* "breakfast / morning coffee", well known between EP and BP. Both occur with the relative pronoun *que* "that" as antecedent that compels the proclisis of the clitic (*com que lhe preparava NP* "with which she prepared his N") and the Syntactic Identity of the paraphrases would be total were it not for the fact that BP selects the possessive *o seu* "his" postposed to the verb *preparava* "prepared", instead of the proclitic *lhe* plus the verb in EP. This alternation between *lhe*/possessive between EP and BP is widely registered and reported in grammars [11] and can also be confirmed in corpora, such as in COMPARA[9] [12]. In fact, it seems to be a constant stylistic choice in EP, to create a more complex structure with the participation of the clitic with verbs that allow it, where a possessive is perfectly acceptable.

In the **forth example**, the pair of paraphrastic units, as in the previous ones, manifests the presence of an antecedent, an adverb of negation *nunca* "never", that requires the proclisis of pronoun (PROCL), *nunca [] lhe acontecera* "it never happened to him". This clitic is necessary to complete the meaning of the verb *acontecer* "happen" in the context. However, it is not indispensable to the verb in itself provided its intransitive value. This additional information, which in EP is transmitted by the clitic, in BP is [PRO$_{NOM}$ *ele*] preceded by the PREP *com* "with, to", i.e. *com ele*, literally "with him", "to him" in the corpus.

In the **fifth example**, in EP, the clitic *lhe* occurs in a pre-verbal position without the presence of an antecedent, only a coordinating conjunction *e* "and", *e lhes conferia NP* "(and) bestowing upon them NP". In BP, the verb *agraciar*, literally, "award" / "grace", selects a direct complement in the paraphrase by the clitic *os* in enclitic position (ENCL). Again, the Syntactic Identity is undone by the occurrence of lexical elements that fill the Semantic Identity without matching the same structure of verbal predicate complements. The proclitic Dative ([PRO$_{DAT}$ *lhe*) in EP corresponds to an enclitic Accusative ([PRO$_{ACC}$ *os*) in BP.

Some remarks should be made by observing the data in Table 2. The first important observation is that it appears, just by observing the occurrences of paraphrases in this corpus, that BP seeks ways to avoid the clitic *lhe*, without necessarily breaching the rules of grammar. It is either by substituting the clitic for other elements, as in the previous examples, or by selecting other lexical items that fill the Semantic Identity (Sem). This selection often implies a total or partial change in the Syntactic Identity (Syn), as can be seen in the selected paraphrases (examples 1, 2, and 5 in the Table). The second observation concerns the verb tenses selected in the paraphrases. It seems that they are often irrelevant in terms of Semantic Identity. They belong to the mastery of style and choices of each translator, revealing also drifts of each variety. In any case, they do not interfere with information-sharing between paraphrases. Finally, it is important to remind that all contrasts presented in this paper have been found in the context of translation, which can obviously influence the choice of constructions to be used in the target language.

Figure 1 illustrates a local grammar that allows converting a EP predicate verbal construction, where the enclitic appears appears after different adverbials or the relative pronouns *que* "that" and *quem* "who(m)", into an equivalent BP construction, where the

[9] http://www.linguateca.pt/COMPARA/

clitic was elided, as in ***não lhe disse*** *que ...* > *Você* ***não contou*** *que ...* 'you haven't told him (that) ...'. The same grammar also allows the most common conversion between enclisis and proclisis as in example *entregou-**lhe*** *as chaves* > ***lhe*** *passou as chaves* 'hand over the key'.

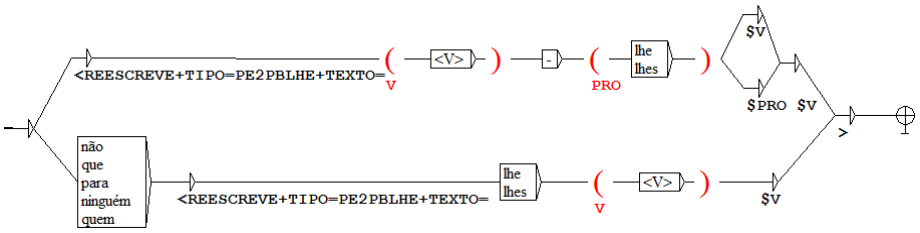

Fig. 1: Grammar to formalize the conversion of EP into BP predicate verbals with *lhe*

Figure 2 illustrates the variety adaptation capability within eSPERTo, where for a sentence written in EP, there are suggestions to rewrite it in BP. For example, for the EP sentence *Mabel Lee entregou-lhe as chaves da sala*, which in the e-PACT corresponds to the BP sentence *Mabel Lee lhe passou as chaves da sala* "Mabel Lee proceeded to hand over the key to his room", eSPERTo presents as conversion options for the verbal predicate with the enclitic *entregou-lhe* in EP, (i) the verbal predicate without the clitic, *entregou*, and (ii) the verbal predicate with the proclitic *lhe entregou*. If there is any of the following words: *não* 'not', *que* 'that', *para* 'to', *ninguém* 'nobody' or *quem* 'who(m)' (the list of words is much larger), the clitic pronoun migrates to a position before the verb, such as in *para lhe dizer* 'to tell him/her'. The grammar allows generation of the BP *não digo* from the EP *não lhe digo* and the generation of the BP *digo* and *lhe digo* from the EP *digo-lhe*. The variety adaptation capability within eSPERTo means that for a sentence written in EP, the system offers suggestions to paraphrase it in BP. In many cases, this adaptation is extremely useful when the user wants to reach an audience that speaks the variety that he/she is less familiar with.

5 Conclusions and Future Work

Variety adaptation 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, also when these are contrasting, as in the case of varieties of the same language. The placement of clitics differ considerably between BP and EP, constituting a challenge for (semi-)automated adaptation between these varieties. A clear contrast in EP–BP grammar is that displayed by the clitic pronoun *lhe*, for which we have shown the differences in syntactic behaviour. We have made a first attempt to define a typology of paraphrastic contrasts and analyzed the differing forms of expression. Some of the paraphrastic pairs indicate an approximate value, which although not assuming a full semantic correspondence, are

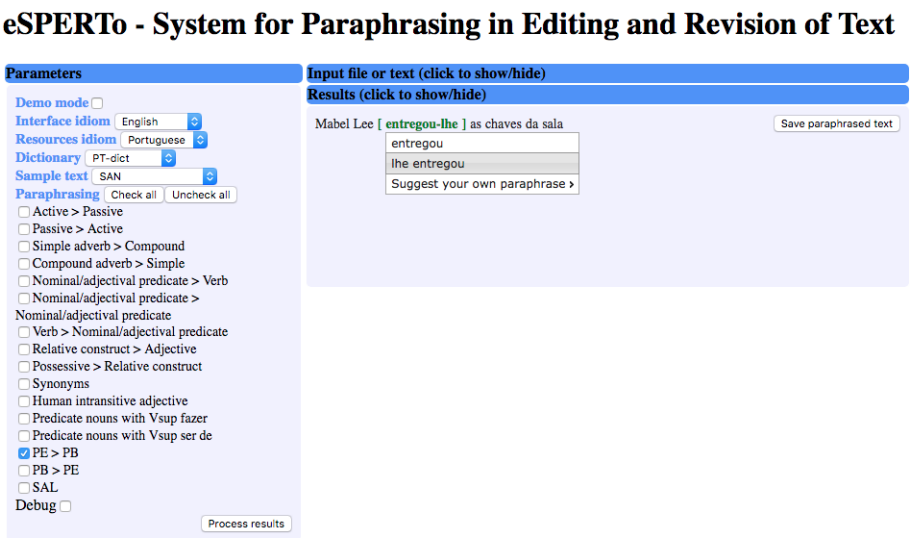

Fig. 2: Conversion of an EP predicate verbal with the clitic pronoun *lhe* into BP

extremely useful and valid in paraphrasing tasks, namely in conversion between variants. However, we do not (and cannot, given the size and characteristics of our data) distinguish between paraphrases that are possible to establish no matter the variety of Portuguese involved, and contrastive paraphrases that are "compulsory/mandatory", or strongly suggested, by the differences between the two varieties.

Our initial typology and results were achieved by analyzing a reduced subset of occurrences. In the near future, we plan to continue the alignment of paraphrastic correspondences in EP–BP sentence pairs of the existing corpus with regards to the wide variety of clitic pronouns. We plan to align the totality of the corpus since it can serve to provide a richer source of paraphrases related to the clitic phenomenon, which represents a relevant source of contrasts between the EP–BP varieties. In addition to the full corpus alignment, in order to be able to draw more meaningful conclusions on the variety contrasts involving the clitic pronoun *lhe*, it is also recommended to compare those results with larger data, namely to compare the contrasting pairs obtained with originals in EP and BP. At present, the only tool at our disposal is CLUE-Aligner, which allows to analyze two languages or language varieties simultaneously. We can search for the original sentence in English, but this is not immediately available during the alignment task. To obtain the frequencies of the different type of constructions, even if not aligned, may be relevant to gain a more accurate picture of the phenomenon, which we did not include here due to space restrictions. However, it is important to create more freely available parallel corpora for EP–BP to train and test our results in real-world paraphrasing systems, include phenomena that may only be found in other types of parallel corpora, covering not only generic texts, but also take into consideration paraphrases of different textual genres and specific or specialized domains. In addition, subtitles

could be an interesting source of corpora. The Opus project[10] contains sub-corpora of OpenSubtitles, where the Portuguese language is included.

References

1. Bagno, M.: Português ou Brasileiro: um convite à pesquisa. Parábola, São Paulo, Brasil (2001)
2. Barreiro, A., Mota, C.: e-PACT: eSPERTo Paraphrase Aligned Corpus of EN-EP/BP Translations. Tradução em Revista **1**(22), 87–102 (2017)
3. 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)
4. 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)
5. Barreiro, A., Rebelo-Arnold, I., Mota, C., Garcez, I., Baptista, J.: Automatic Paraphrasing and Normalization of Portuguese Informal into Formal Language. In: Proceedings of POP@PROPOR 2018 (2018 (in this volume))
6. Carreras, X., Chao, I.: Freeling: An open-source suite of language analyzers. In: Proceedings of the Forth Language Resources and Evaluation Conference LREC 2014, Lisbon, Portugal (2004)
7. Castilho, A.: Nova Gramática do Português Brasileiro. Contexto (2010)
8. Castilho, A.: O português do Brasil. In: Ilari, R. (ed.) Linguística Românica. pp. 237–269. Fundamentos 83, Ática (2011)
9. Castro, I.: Introdução à história do português. Colibri, Lisboa, Portugal (2011)
10. 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)
11. Cunha, C., Cintra, L.: Nova Gramática do Português Contemporâneo. Nova Fronteira (1985)
12. Frankenberg-Garcia, A., Santos, D.: Introducing COMPARA: the Portuguese-English Parallel Corpus. In: Zanettin, F., Bernardini, S., Stewart, D. (eds.) Corpora in Translator Education, pp. 71–87. St. Jerome, Manchester (2003)
13. Kato, M., Martins, A.M.: European Portuguese and Brazilian Portuguese: an overview on word order. In: The Handbook of Portuguese Linguistics. pp. 15–40. Wiley-Blackwell, Hoboken, NJ (2016)
14. Neves, M.H.M.: Gramática de usos do português. Unesp (2000)
15. Pacheco, J.C.: As construções médias do português do Brasil sob a perspectiva teórica da Morfologia Distribuída. Master's thesis, Universidade de São Paulo (2008)
16. Perini, M.A.: Modern Portuguese: a Reference Grammar. Yale Language Series, Yale University (2002)
17. Santos, D.: Os possessivos estão-me a complicar o ensino ;-) um estudo do dativo possessivo baseado em corpos. Linguística : Revista de Estudos Linguísticos da Universidade do Porto **10**, 107–130 (2015)
18. de Sousa Pereira, S.: Estudio contrastivo del régimen verbal en el Portugués de Brasil y el Español Peninsular. Ph.D. thesis, Universidade de Santiago de Compostela (2007)

[10] `http://opus.nlpl.eu/`