Introducing a Corpus of Human-Authored Dialogue Summaries in Portuguese

Norton Trevisan Roman
School of Arts, Sciences and Humanities
University of São Paulo
São Paulo – SP, Brazil
norton@usp.br

Paul Piwek
Centre for Research in Computing
The Open University
Milton Keynes, UK
p.piwek@open.ac.uk

Ariadne M. B. Rizzoni Carvalho
Institute of Computing
University of Campinas
Campinas – SP, Brazil
ariadne@ic.unicamp.br

Alexandre Rossi Alvares
School of Arts, Sciences and Humanities
University of São Paulo
São Paulo – SP, Brazil
alexandre.alvaresl@usp.br

Abstract

In this paper, we introduce a corpus of human-authored dialogue summaries collected through a web-experiment. The corpus features (i) one of the few existing corpora of written dialogue summaries; (ii) the only corpus available for dialogue summaries in Portuguese; and (iii) the only available corpus of summaries produced for dialogues whose participants' politeness alignment was systematically varied. Comprising 1,808 human-authored summaries, produced by 452 summarisers, for four different dialogues, this is, to the best of our knowledge, the largest individual corpus available for dialogue summaries, with the highest number of participants involved.

1 Introduction

As an important part of current mainstream research on automatic summarisation, corpora are used for a vast range of applications, from the construction of tutoring systems (*e.g.*, (Callaway et al., 2005)) to abstract production from extracts (*e.g.*, (Hasler, 2007)), to multi-document summarisation (*e.g.*, (Atkinson and Munoz, 2013)). Still, most corpora are available in English only, which may have an impact on the performance of automatic summarisation methods when applied to other languages (de Loupy et al., 2010). Also, there seems to be a preference for newswire (*e.g.*, (Amini, 2000; Copeck and Szpakowicz, 2004; Hasler, 2007; de Loupy et al., 2010)) and aca-

demic texts summaries (*e.g.*, (Teufel and Moens, 1997)), with fewer sources available for dialogue summaries, and those available mostly restricted to spoken dialogues (*e.g.*, (Murray et al., 2005; Carletta et al., 2006; Liu and Liu, 2008)).

In this paper we introduce a corpus of human-authored dialogue summaries, which we have released for use by the research community.[1] The corpus comprises 1,808 summaries, produced by 452 summarisers, for four different dialogues (each summariser produced a summary for each dialogue). To the best of our knowledge, this is the largest individual corpus available for dialogue summaries, with the largest number of participants involved. Collected through a web-experiment, where participants had to summarise a set of written dialogues, the corpus has the additional characteristic of being written in Portuguese (a language spoken by over 200 million people[2], if one accounts only for Brazil and Portugal), thereby helping reduce the dearth of corpora for written dialogue summaries in languages other than English.

Additionally, source dialogues were carefully chosen so they portray interactions with different degree of politeness, as measured in an experiment carried out by Roman et al. (2006b). Resulting summaries may therefore be used for a range of different tasks, such as (i) automatic dialogue summarisation, especially in Portuguese; (ii) studies

---

[1]At www.each.usp.br/norton/resdial/index_ing.html
[2]http://www.ibge.gov.br/home/estatistica/populacao/censo2010/default.shtm
http://censos.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=censos2011_apresentacao

on reports of emotion in dialogue; and (iii) investigation of other properties of the language used in dialogue summaries, such as most frequent typing errors (which could be helpful in, for example, spelling correction systems). We intend our release of the corpus to the research community to lead to its use as set out above and, possibly, in many further ways.

The rest of this paper is organised as follows. Section 2 describes some of the currently available corpora, presenting their size, resulting documents, and set of summarisers. Section 3 introduces our corpus, along with the methodology followed during its construction. In Section 4 we present some examples of the documents that make the corpus, along with their codification. Finally, in Section 5 we present our conclusions and directions for future work.

2 Related Work

In the search for corpora of human-authored summaries, many strategies have been adopted along the years. One of the first ones (which is still in use) was to rely on already available datasets, such as the abstracts delivered with scientific papers and textbook chapters (*e.g.*, (Teufel and Moens, 1997; Silber and McCoy, 2002; Hasler, 2007)). With the growth of the information exchange through the Internet, yet another source for raw material has emerged: online newswire documents (*e.g.*, (Amini, 2000; Jing, 2002; Copeck and Szpakowicz, 2004; de Loupy et al., 2010)), in particular those that come with a summary by their editor.

However abundant, such sources have the drawback of being quite generic, making it harder for the researcher to control different phenomena. Alternative sources include summarising e-mail threads (*e.g.*, (Rambow et al., 2004)), line graphs (*e.g.*, (Greenbacker et al., 2011)) and dialogues (*e.g.*, (Murray et al., 2005; Carletta et al., 2006; Liu and Liu, 2008)). As for this last source, there seems to be no available corpus of summaries of written dialogues. The aforementioned corpora consist of transcriptions of naturally occurring spoken dialogues, which may differ from written scripted dialogue (for example, for films, plays and adverts), as a result of the way they are produced. Scripted dialogues are an important genre in their own right, which merits academic study and has a range of applications as a result of their wide use in the entertainment, education and information presentation industries.

Apart from the source data type, size is another important feature that influences the usage of corpora. Current corpora sizes may vary from as few as 15 summaries (*e.g.*, (Jing and McKeown, 1999)) to as many as 1,000 summaries (*e.g.*, (Amini, 2000)), and up to 9,086 summaries (*e.g.*, (Copeck and Szpakowicz, 2004)), if one includes collections of corpora (in this case, gathered from four Document Understanding Conferences – DUC). Along with the size of a corpus, yet another feature to be taken into account is the number of participants that produced it, since a small number of summarisers may lead to a sample that is not representative for the phenomenon to be measured. On this account, current corpora vary from a single summariser (*e.g.*, (Hasler, 2007)) to as many as 202 (*e.g.*, (Teufel and Moens, 1997)).

Our corpus is distinctive from all these in that it consists of a total of 1,808 human-produce dialogue summaries (to our knowledge, the largest collection of summaries produced in a single initiative), authored by 452 different summarisers (again, according to our knowledge, the largest amount of summarisers reported in the literature). A further distinctive property of our corpus is that it is entirely in Portuguese, which adds to the very few existing initiatives for languages other than English (*e.g.*, (de Loupy et al., 2010; Saggion and Szasz, 2012)).

Finally, to the best of our knowledge, the current corpus is the only summary corpus whose source was chosen so as to present instances of dialogues in which the politeness of the dialogue participants varied systematically, as determined by our choice of source dialogues (see Section 3). This allows researchers to examine how politeness in dialogue is reported when the dialogue is summarised. In the next Section, we describe our corpus in more detail. We explain how we selected the source dialogues, along with the instructions presented to summarisers.

3 Data Collection

The first problem we faced, when trying to build a corpus of dialogues with different degrees of politeness for the interlocutors, was that of where to find dialogues that might fulfil this requirement. Since most available corpora are built from meeting transcriptions, and the only alternative cor-

pus that is available was automatically generated (see (Roman et al., 2006a)), we decided to go for human-authored (that is, scripted) dialogues. We then turned to film dialogues, given their availability through the web and the richness of situations they portray.

Once the source of dialogues was settled, we started to collect them from movie scripts and transcripts over the web. We collected a total of 16 dialogues, from 10 movies, which portray a customer-seller interaction.[3] This kind of interaction was chosen because (i) it delivers a situation where people would have an idea about what would be proper behaviour by the dialogue participants; and (ii) it allowed for any resulting conclusions on this subject to be compared to the existing corpus of machine-generated dialogues described in (Roman et al., 2006a), which also consists of customer-seller interactions. Dialogues were collected regardless of other features, such as genre, for example.

Given the scarcity of movie scripts and transcripts in Portuguese at the time of data collection, specially when considering the aforementioned requirements, the original materials were exclusively in English. Summarisers, on the other hand, were native speakers of Brazilian Portuguese. To overcome this mismatch, the dialogues were translated to Portuguese by one of the researchers. They were then presented to 153 subjects, in a web-experiment reported in (Roman et al., 2006b), where participants were asked to classify them according to one out of five categories on a Likert scale, ranging from "very impolite" to "very polite". The purpose of the study was to measure "first-order politeness" (Watts, 2003) (also called politeness$_1$ (Eelen, 2001)), that is, people's own interpretation of politeness (or, conversely, impoliteness). Of the original 153 participants, 89 finished the experiment, as a result of the precautions taken to avoid drop-out in the critical phase (*i.e.* the classification proper).

Finally, four dialogues were chosen from that experiment, where either one party was impolite, or both were polite (as in the experiment described in (Roman et al., 2006a)). The selected dialogues were those where the distribution of classifications was more skewed towards the positive or negative end of the scale. Although the dialogues varied

considerably in size, being 54, 61, 125 and 320 words long, respectively, no statistically significant difference (t = 0.9307, p = 0.5228) was found between the dialogue length and its classification as polite or impolite.

3.1 Dialogue Summarisation – Building the Corpus

The four dialogues selected from the experiment described in (Roman et al., 2006a; Roman et al., 2006b) were presented, in a different web-experiment, to a set of 1,385 volunteers, recruited by e-mail from all students in a Brazilian university (see (Roman et al., 2005) for details). These participants were assigned a restriction (either their summary should be no longer than 10% of the number of words in the source dialogue, or they were free to write down as much as they felt like) and a viewpoint (either customer, vendor, or an observer), under which they should write the summary. These limits were arbitrarily chosen so as to frame the summarisers' choice when forced to produce a very short summary, compared to what they would do should they be given no constraint at all, in particular when it comes to the reporting of more subjective material, such as the behaviour demonstrated by the dialogue participants, of which politeness is the prototypical case. In the sequence, participants were asked to produce a summary for each of the dialogues, under the assigned point of view and size limit.

Even though the original dialogues were in English, both classification and summarisation tasks were carried out with their Portuguese version. This, in turn, helps reducing the effects of any loss in the original content of the dialogues, by linking each summary to its source dialogue's Portuguese version, instead of its original English content. Also, participants were free to chose their own summarising style, that is, they were not asked to specifically produce abstractive or extractive summaries (we are currently studying the data to find out what summarisation styles they actually adopted). Finally, in order to keep the data as bias-free as possible, the whole experiment was designed so that participants that summarised the dialogues were different from those who classified them (*cf.* (Roman et al., 2006b)).

The experiment followed the guidelines suggested in (Roman et al., 2006b), by presenting the participants with a good number of initial web-

---

[3]Dialogues were adapted so that proper names and contextual information referring to visual elements of the scene were removed.

pages, as a way to induce those that were more susceptible to giving up the experiment to drop out before the critical phase began (*i.e.* before they were asked to produce any summary). These measures seem to have worked since, of the original 1,385 participants who started the experiment, 598 finished it. However bad that may sound, drop-out concentrated in the pre-summarisation phase, where 658 participants abandoned the experiment, resulting in a set of 652 who started the critical phase (for a more comprehensive description of the technical details involved in this kind of experiment see (Roman et al., 2005)).

Drop-out rates at each step in the summarisation process are shown in Figure 1. At first, participants were shown a web page introducing the research (*Pres* in the figure), but without giving away much information about it. The number 1,385 indicates that, out of all participants that saw the web page, 1,385 decided to move on to the next page. In the next page (*Reg* in the figure), participants had to give some personal details. At this point, a total of 860 (*i.e.* a 38% reduction in the original set) filled in the form and decided to proceed with the experiment.

In the next pages, drop-out begins to slow down. At the Log-in page, 750 (from the 860 that registered for the experiment, *i.e.* a further 13% reduction) logged in the system. These participants were then shown a web page, saying a little more about the research, but with no mention of its real intent. Out of the 750 that logged in the experiment, another 23 gave it up (*i.e.* a 3% reduction). As a result, a total of 727 participants did actually see the first dialogue to be summarised, that is, they entered the critical phase of the experiment, of which 652 submitted their first summary (a 10% decrease).

The next three pages correspond to the submission of summaries for the remaining dialogues (*D2* to *D4* in the figure). Across this set, we lost a further 7%, leaving us with 604 participants who submitted all summaries (for a total drop-out rate of around 17% at the critical phase). In the sequence, participants were prompted to classify the dialogues about their politeness (so as to verify if their perception on the dialogues matched that of the classification experiment). At this step, another four were lost. Finally, they were asked about whether they recognised any of the dialogues (*Rec* in the figure), in which page we lost

another couple of participants, ending up with 598.

The reason for moving both questions to the end of the experiment was to avoid giving the participants any information that might affect their decision on what to include in the summary. In this case, asking them about the politeness of dialogue participants right after each summary could have the participants focus on this facet of the interaction. Along the same lines, asking them whether they recognised the summarised dialogue would potentially have them effectively try to do it, which in turn might lead to false positives, whereby participants think they recognise some dialogue just because they are paying more attention to it.

Although the adopted measures succeeded in moving drop-out away from the experiment proper, it might be the case that drop-out occurred in a systematic way, in which case the experimental results could be themselves compromised (Reips, 2002). Figures 2 and 3 show the results of our analysis on drop-out according to the participants' gender, knowledge area, educational attainment and age, for all participants that provided that information. Amongst all these variables, only educational attainment was found to be related to drop-out in this experiment ($\chi^2 = 6.8327$, p<0.0090), in that postgraduate students tended to drop out less often than undergraduate students (perhaps due to a better comprehension of the experimental dynamics in general). No differences were observed for the remaining variables.[4]

Since we were dealing with movie dialogues, some participants recognised the specific movies. These participants may have included information in their summary that went beyond the information that was present in the dialogue itself. For this reason, out of the 598 participants who finished the experiment, we removed the data from all 136 participants who indicated that they were already familiar with some of the dialogues, along with the single participant who did not provide such information. An analysis of the remaining data set led us to discard a further nine from the 461 remaining participants, resulting in a total of 452. Out of these nine, three were non-native speakers of Portuguese; two produced incomplete data sets, by leaving one or more summary empty; and four produced nonsense, by typing random charac-

_____
[4]$\chi^2 = 2.0074$, p = 0.1565, for gender; $\chi^2 = 0.2966$, p = 0.8622, for area of knowledge; and $\chi^2 = 2.6390$, p = 0.7554, for age.
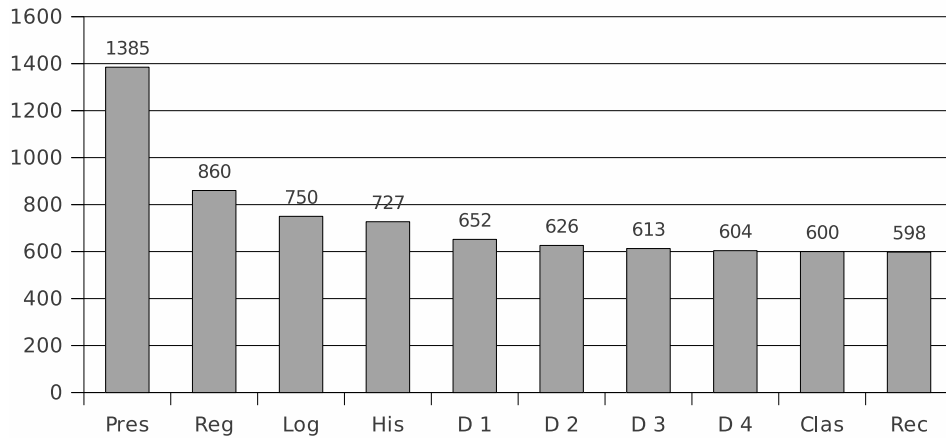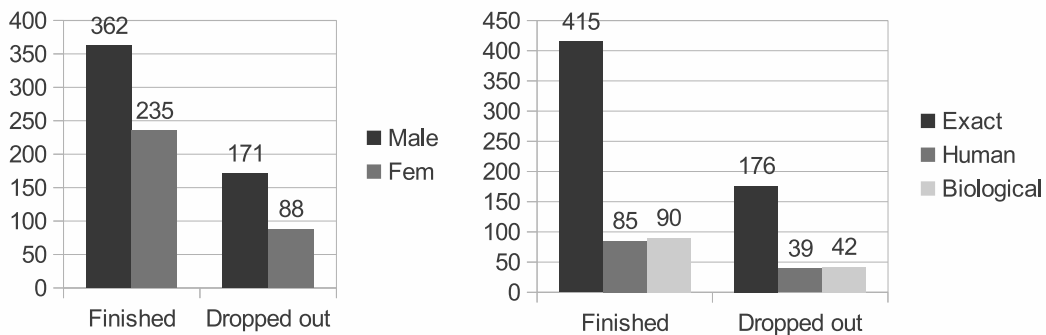
Figure 1: Number of participants at each webpage.

Figure 2: Dropout according to gender and knowledge area.

ters in the summary. All these correspond to mere 1.95% of the 461 participants, which adds to the trustworthiness of the data set.

Another source of bias in the experiment would be having an unbalanced number of participants recognise the dialogues, when compared to the 452 who did not recognise any of them. In this case, we found no statistically significant difference, between the participants who recognised any of the dialogues and those who did not, for the variables gender (χ^2 = 0.3656, p = 0.5454) and knowledge area (χ^2 = 3.4705, p = 0.1764). As for the remaining variables, once again, educational attainment showed a statistically significant difference, although borderline (χ^2 = 3.8726, p = 0.0491), whereby postgraduate students seem to have recognised the movies more often. Somewhat related to this finding is the statistically significant difference also found for the variable age (χ^2 = 23.8249, p = 0.0002), in which participants between 20-25 years old seem to have recognised proportionally less frequently the dialogues. Both results might be actually due to the participants'

life experience, whereby the older they are, the higher the odds that they are both postgraduate students and have seen the movie before. Figure 4 shows the numbers for both variables.

After filtering out the data from the participants who recognised the dialogues and from the nine with problematic data, the resulting corpus comprised 1,808 human-made summaries, produced by 452 different participants, where each participant generated four different summaries, one for each dialogue. Due to the random distribution of participants amongst the experimental categories, out of the 1,808 summaries, 896 were produced by the group with no size restrictions, whereas the remaining 912 should be no longer than 10% of the number of words of their source dialogue. Finally, the entire corpus has a total of 62,858 words (mean of 34.7 words per summary), with 11,512 (mean of 12.6 per summary) in the 10% restriction set, and 51,346 (mean of 57.3 per summary) in the set with no size restriction at all.

Of the 452 participants, 270 (59.7%) were male and 181 (40%) female, with one abstention to the
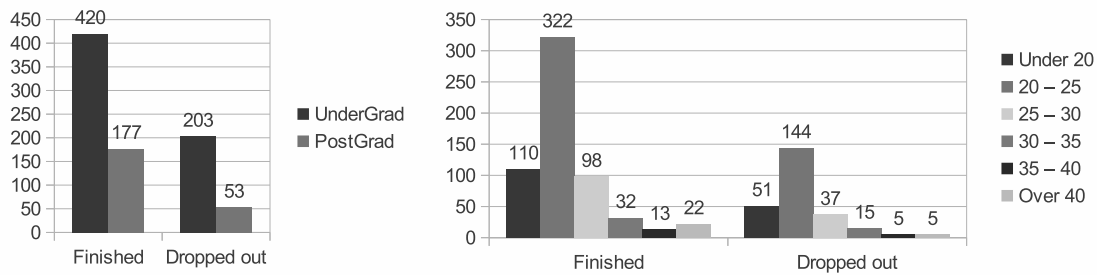
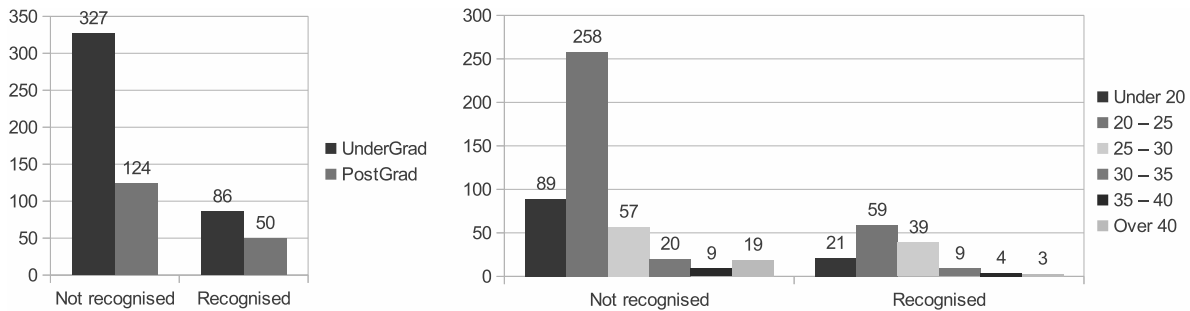Figure 3: Dropout according to educational attainment and age.

Figure 4: Distributions of participants that recognised the dialogues and those that did not.

question, 327 (72.3%) were undergraduate students, whereas 124 (27.4%) were postgraduate (and one abstention), with 322 (71.2%) pertaining to the exact sciences, 62 (13.7%) to the social sciences, other 62 to the biological sciences, and six abstentions. Ages varied from under 20 to over 40, distributed as shown in Figure 5.

Finally, regarding possible differences between the way people classified the dialogues' interaction (as reported in (Roman et al., 2006b)) and the way summarisers perceived it (in our experiment), we found no statistically significant difference[5] between both experiments, for any of the dialogues, with respect to whether participants perceived the dialogues as portraying a polite, neutral or impolite interaction. This is an indication that summarisers had understood the dialogues the same way as did those that classified them in the first experiment.

4 Corpus Delivery

The corpus is stored as a set of text files (UTF-8 encoded), in a single folder, where each file corresponds to a single summary. Within each file, data are represented using an XML compliant for-

mat[6], making them more independent of the process that created them (Müller and Strube, 2006; O'Donnell, 2008). Dialogue summaries are delivered as plain text, that is, with no further annotation added to them, so that future annotations can be made in a stand-off manner, whereby annotation and annotated data are kept in different XML files, with some link between them (Ide and Brew, 2000). Figure 6 illustrates a sample summary in the corpus.[7]

As can be seen in the figure, along with the summary, the XML includes its identification code ("R0001") and the identification of the corpus in which the summary is inserted (in this case, "C2"). There are also tags for the identification of the dialogue used to create the summary ("D1"), along with the identification of the corpus holding that dialogue (*i.e.* "Script2"). Given that summaries were produced under a viewpoint and possibly with a size constraint, both values are also recorded in their XML, followed by the summariser that produced this summary.

---

[5]χ^2 = 2.0926, p = 0.3512, for the first dialogue, χ^2 = 0.1038, p = 0.9494, for the second, χ^2 = 3.4405, p = 0.1790, for the third and χ^2 = 3.4225, p = 0.1806, for the fourth one.

[6]For a detailed description of the adopted XML codification, we refer the interested reader to (Roman, 2013).

[7]Main text may be translated as "The client in the pub wants the waitress Carol to serve him. That is not possible, because she is being replaced, since she would be better off with getting a job closer to her home. The client does not understand it at all, and he is ready to pay whatever it takes to get Carol to serve him".

Figure 5: Distribution of participants according to their age.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<plainDocument>
  <info type="id" value="R0001" />
  <info type="corpus" value="C2" />
  <info type="source" value="D1" />
  <info type="source-corpus"
        value="Script2" />
  <info type="viewpoint"
        value="attendant" />
  <info type="constraint" value="free"/>
  <info type="summariser"
        value="a30c92004183430935" />
  <text>O Cliente da lanchonete quer que
        a garçonete Carol atenda-o. Isso
        não é possivel pois ela está
        sendo substituida já que seria
        melhor ela arrumar um emprego
        mais perto da casa dela. O
        Cliente não entende de forma
        alguma, e está disposto a pagar
        o que for necessário
        para que a Carol o atenda.
  </text>
</plainDocument>
```
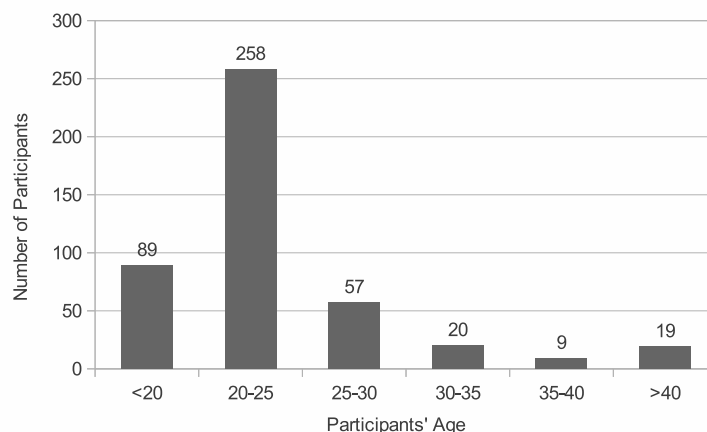
Figure 6: Codification of a plain summary.

Within this scheme, source dialogues are kept in a different folder, codified along the same lines as the corpus of summaries. Figure 7 shows a sample dialogue, adapted from the "As Good as it Gets" movie script.[8] Although following the same codification style, the stored information is different in this set. In this case, each file (and hence each

source-dialogue) has, apart from its identification code and corpus identification, the identification of the source type ("Movie Script"), the movie title and the translator of the dialogue (a necessary step, since the summaries are in Portuguese whereas the script is in English). Finally, the politeness alignment of the dialogue, as determined by the majority of participants, both in the classification experiment carried out by Roman et al. (2006b) and ours, is also added to the summary, respectively, in the "classified-politeness" and "perceived-politeness" fields.

Inside each corpus folder, there is also a subfolder named "participants", which stores all the information regarding who was responsible for the production of that corpus. In the corpus of summaries, it corresponds to the characterisation of the human summarisers, while in the set of dialogues, it corresponds to the single person that translated them. Whatever the folder, the information about each participant is kept in separate files, one per participant, as with the corpus itself.

Figure 8 shows an example of such a file, in which we keep information about the participant's identification code (within the corpus), gender, area of knowledge, educational attainment, age and Brazilian State of origin. The last two tags in the figure refer to the time the participant registered and the time s/he actually logged in to carry out the experiment. Finally, we would like to emphasise that no information is kept that could be used to directly identify any of the participants. We only report on information that is useful for statistical purposes and to characterise the sample.

---

[8]The adapted dialogue is: "In a pub. Dialogue between a client and the waitress:
Waitress: How may I serve you?
Client: No. No. Get Carol.
Waitress: I'm filling in. I don't know if she's coming back. It might be better for her to get a job closer to home.
Client: What are you trying to do to me?
Waitress: What do you mean?
Client: Listen, elephant girl, call her or something... just let her do my one meal here. I'll pay whatever. I'll wait. Do it!!!"

```xml
<?xml version="1.0" encoding="UTF-8"?>
<plainDocument>
  <info type="id" value="D1" />
  <info type="corpus" value="Script2"/>
  <info type="source-type"
        value="Movie Script" />
  <info type="title"
        value="As Good as it Gets" />
  <info type="translator" value="t01"/>
  <info type="classified-politeness"
        value="very impolite" />
  <info type="perceived-politeness"
        value="very impolite" />
  <text>
    Em uma lanchonete. Diálogo entre um
    cliente e a garçonete.
      Garçonete:  Pois não.
      Cliente:    Não, não, vá chamar a
                  Carol.
      Garçonete:  Eu to substituindo
                  ela. Não sei se ela
                  vai voltar. Talvez
                  seja melhor ela
                  arrumar um emprego
                  mais perto da casa
                  dela.
      Cliente:    O que você tá tentando
                  fazer comigo?
      Garçonete:  Como assim?
      Cliente:    Escuta aqui, ô
                  elefanta, vá chamar
                  ela... só peça que ela
                  prepare minha refei-
                  ção. Eu pago o que
                  for. Eu espero. Vá!!!
  </text>
</plainDocument>
```

Figure 7: Codification of a source dialogue.

5 Conclusion

In this paper, we introduced a corpus of human-authored dialogue summaries. Collected through a web experiment, this is, to the best of our knowledge, the largest corpus available for dialogue summaries, with the highest number of participants involved. Amongst its main characteristics, are (i) it is one of the few existing corpora of dialogue summaries and, to our knowledge, the only one produced from written dialogues, as opposed to audio transcriptions; (ii) it is the only corpus available for dialogue summaries in Portuguese; and (iii) it is the only available corpus of summaries produced for dialogues whose participants' politeness alignment was systematically varied.

Amongst other possibilities, this corpus may serve as the basis for a range of projects, from studies in generation-based summarization (or its evaluation) to sentence compression, to research on the influence the dialogue participants' politeness has on the production of summaries for such

```xml
<?xml version="1.0" encoding="UTF-8"?>
<participant>
  <info type="id"
        value="a30c92004183430935" />
  <info type="gender" value="m" />
  <info type="age" value="20-25" />
  <info type="area"
        value="exact sciences" />
  <info type="degree"
        value="undergraduate" />
  <info type="State of Origin"
        value="SP" />
  <info type="registration"
        value="Friday,1,October,2004.
        21h:0m:28s" />
  <info type="log-in"
        value="Friday,1,October,2004.
        21h:0m:58s"/>
</participant>
```

Figure 8: XML describing a summariser in the corpus.

dialogues. Since the dialogue summaries were directly typed in by the summarisers, more generic studies into language use can also be carried out, such as studies on spelling error frequencies, for example. As for future research, we intend to explore in more depth some of the topics described above.

Acknowledgements

References

Massih-Reza Amini. 2000. Interactive learning for text summarization. In *Proceedings of the PKDD'2000 Workshop on Machine Learning and Textual Information Access*, Lyon, France, September 13-16.

John Atkinson and Ricardo Munoz. 2013. Rhetorics-based multi-document summarization. *Expert Systems with Applications*, 40:4346–4352.

Charles Callaway, Myroslava O. Dzikovska, Johanna D. Moore, David Reitter, and Claus Zinn. 2005. D11: Corpus collection and specification. Technical report, The LeActiveMath Consortium, May.

Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa

Kronenthal, Guillaume Lathoud, Mike Lincoln, Agnes Lisowska, Iain McCowan, Wilfried Post, Dennis Reidsma, and Pierre Wellner. 2006. The ami meeting corpus: A pre-announcement. In *Proceedings of the Second international Workshop on Machine Learning for Multimodal Interaction (MLMI'05)*, pages 28–39, Edinburgh, UK, July 11-13.

Terry Copeck and Stan Szpakowicz. 2004. Vocabulary usage in newswire summaries. In *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pages 19–26, Barcelona, Spain, July 25-26.

Claude de Loupy, Marie Guégan, Christelle Ayache, Somara Seng, and Juan-Manuel Torres Moreno. 2010. A french human reference corpus for multi-document summarization and sentence compression. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May, 19-21.

Gino Eelen. 2001. *A Critique of Politeness Theories*. St. Jerome, Manchester, England.

Charles F. Greenbacker, Sandra Carberry, and Kathleen F. McCoy. 2011. A corpus of human-written summaries of line graphs. In *Proceedings of the UCNLG+Eval: Language Generation and Evaluation Workshop (UCNLG+EVAL '11)*, pages 23–27, Edinburgh, UK, July 31.

Laura Hasler. 2007. From extracts to abstracts: Human summary production operations for computer-aided summarisation. In *Proceedings of the RANLP 2007 Workshop on Computer-Aided Language Processing (CALP)*, pages 11–18, Borovets, Bulgaria, 30 September.

Nancy Ide and Chris Brew. 2000. Requirements, tools, and architectures for annotated corpora. In *Proceedings of Data Architectures and Software Support for Large Corpora*, pages 1–5, Paris, France. European Language Resources Association.

Hongyan Jing and Kathleen R. McKeown. 1999. The decomposition of human-written summary sentences. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '99)*, Berkeley, USA, August 15-19.

Hongyan Jing. 2002. Using hidden markov modeling to decompose human-written summaries. *Computational Linguistics*, 28(4):527–543, December.

Fei Liu and Yang Liu. 2008. What are meeting summaries? an analysis of human extractive summaries in meeting corpus. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 80–83, Columbus, Ohio, USA, June 19–20.

Christoph Müller and Michael Strube. 2006. Multi-level annotation of linguistic data with MMAX2.

In Sabine Braun, Kurt Kohn, and Joybrato Mukherjee, editors, *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, pages 197–214. Peter Lang, Frankfurt a.M., Germany.

Gabriel Murray, Steve Renals, and Jean Carletta. 2005. Extractive summarization of meeting recordings. In *Proceedings of the 9th European Conference on Speech Communication and Technology (Interspeech'2005-Eurospeech)*, Lisbon, Portugal, September 4-8.

Michael O'Donnell. 2008. The uam corpustool: software for corpus annotation and exploration. In *Proceedings of the XXVI Congreso de AESLA*, Almeria, Spain, April 3-5.

Owen Rambow, Lokesh Shrestha, John Chen, and Chirsty Lauridsen. 2004. Summarizing email threads. In *Proceedings Human Language Technology conference / North American chapter of the Association for Computational Linguistics annual meeting (HLT/NAACL 2004)*, Boston, USA, May 2-7.

Ulf-Dietrich Reips. 2002. Standards for internet-based experimenting. *Experimental Psychology*, 49(4):243–256.

Norton Trevisan Roman, Paul Piwek, and Ariadne Maria Brito Rizzoni Carvalho. 2005. A web-based experiment on dialogue summarisation. Technical Report IC-05-05, Computing Institute – State University of Campinas, Campinas, SP, Brazil, March.

Norton Trevisan Roman, Paul Piwek, and Ariadne Maria Brito Rizzoni Carvalho. 2006a. Politeness and bias in dialogue summarization: Two exploratory studies. In James G. Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing Attitude and Affect in Text: Theory and Applications*, volume 20 of *The Information Retrieval Series*, pages 171–185. Springer Netherlands, Dordrecht, The Netherlands, January 9. ISBN: 1-4020-4026-1.

Norton Trevisan Roman, Paul Piwek, and Ariadne Maria Brito Rizzoni Carvalho. 2006b. A web-experiment on dialogue classification. In Solange Oliveira Rezende and Antonio Carlos Roque da Silva Filho, editors, *Proceedings of the Fourth Workshop in Information and Human Language Technology (TIL'2006)*, Ribeir ao Preto, Brazil, October, 27–28. ICMC-USP.

Norton Trevisan Roman. 2013. Resdial – coding description (v.1.0). Technical Report PPgSI-001/2012, School of Arts, Sciences and Humanities – University of São Paulo, S ao Paulo, SP – Brazil, April.

Horacio Saggion and Sandra Szasz. 2012. The concisus corpus of event summaries. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May 23–25.

H. Gregory Silber and Kathleen F. McCoy. 2002. Efficiently computed lexical chains as an intermediate representation for automatic text summarization. *Computational Linguistics*, 28(4):487–496.

Simone Teufel and Marc Moens. 1997. Sentence extraction as a classification task. In *Proceedings of the ACL/EACL Workshop on Intelligent Scalable Text Summarization*, pages 58–65, Madrid, Spain, July 11th.

Richard Watts. 2003. *Politeness*. Cambridge University Press.