Procesamiento del Lenguaje Natural, Revista nº 58, marzo de 2017, pp. 45-52 recibido 16-11-2016 revisado 12-01-2017 aceptado 08-02-2017

A Multilingual Multi-domain Data-to-Text Natural Language Generation Approach *

Un enfoque multilingüe y multidominio de datos-a-texto para la generación de lenguaje natural

Cristina Barros and Elena Lloret
Department of Software and Computing Systems, University of Alicante
Apdo. de Correos 99 E-03080, Alicante, Spain
{cbarros,elloret}@dlsi.ua.es

Resumen: La investigación en enfoques multidominio innovadores y flexibles puede ser un paso significativo en el área de Generación del Lenguaje Natural. En este sentido, el objetivo de este artículo es presentar un enfoque estadístico centrado en la fase de realización. Este enfoque permite la generación de oraciones que cumplan un propósito dado por una "característica semilla" de entrada, la cual se encargará de guiar el proceso de generación. Este enfoque ha sido probado en el ámbito de generar automáticamente oraciones que expresan opiniones para reseñas de películas y, además, el enfoque también ha sido probrado en el ámbito de generación del lenguaje para tecnologías de apoyo a problemas relacionados con el lenguaje. Dados los resultados obtenidos, este enfoque es capaz de generar oraciones para dos dominios diferentes con un rendimiento similar en dos idiomas diferentes, obteniendo buenos resultados y cumpliendo los requisitos especificados para cada dominio.
Palabras clave: Generación de lenguaje natural, "característica semilla", modelos de lenguaje factorizados, realización

Abstract: Research in innovative and flexible multi-domain approaches may be a significant step forward in the area of Natural Language Generation. In light of this, the aim of this paper is to present a statistical approach focused on the surface realisation stage. This approach allows the generation of sentences oriented to meet the purpose given by an specific input seed feature, that will guide all the generation process. Our approach was tested to automatically generate opinionated sentences in the domain of movie reviews and was also tested in the domain of Natural Language Generation for assistive technologies. Based on the results obtained, the approach has proved to be able to generate sentences in two different domains with similar performance and for two different languages, obtaining good results and fulfilling the requirements specified for each domain, which opens the door to be applied in new domains and applications.
Keywords: Natural language generation, seed feature, factored language models, surface realisation

1 Introduction

Currently, with the advance of the technology and the increase of the available content, human-computer communication and interaction needs to be as sound, precise and natural as possible (Jacko, 2012).

Much of this information can be given in a non-textual form, being difficult to interpret by humans. Sensor information and data obtained from electronic medical devices or visual numeric ratings and symbols (like stars) with little information about their scoring origins, are clear examples of these kind of information.

For example, in Figure 1, two types of

* This research work has been funded by the University of Alicante, Generalitat Valenciana, Spanish Government and the European Commission through the projects PROMETEOII/2014/001, TIN2015-65100-R and TIN2015-65136-C2-2-R, respectively.

ISSN 1135-5948

movie reviews can be seen, one with only numeric ratings and the other with numeric ratings and some text. These movie reviews differ on the quantity of information given, where a user has more information to make decisions and to know on what basis the movie was scored with the presence of text in the second movie review.

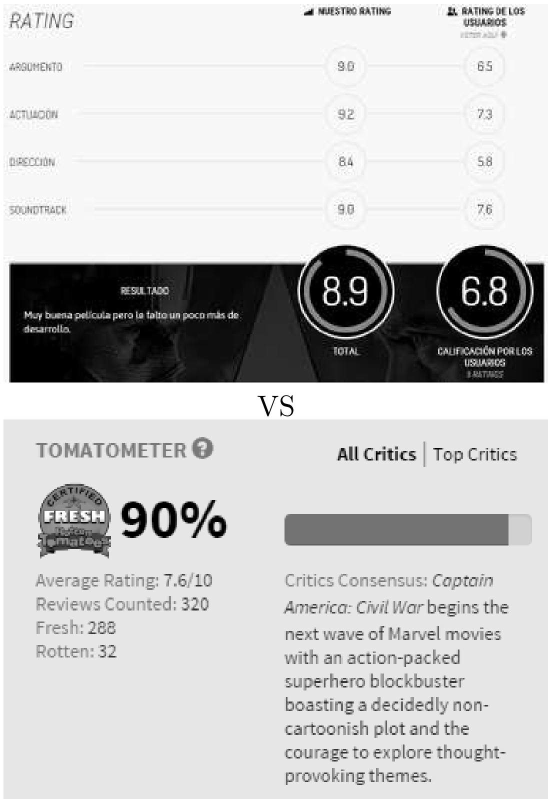

Figure 1: Example of two different types of movie reviews

The area of Natural Language Generation (NLG) aims to automatically develop techniques to produce human utterances, that can be materialised through text or speech. In these terms, NLG techniques can be useful to be used together with non-linguistic elements for generating texts to explain for example the symbols or ratings mentioned above, or another kind of data difficult to interpret such as the one obtained from sensors, among other applications.

In this research area, the development of versatile NLG approaches is still a challenge. Existing NLG systems are designed for very specific domains (Ramos-Soto et al., 2015) and languages (Ballesteros et al., 2015), as well as for particular predefined purposes (Ge et al., 2015), where the cost of adapting these systems can be very high. The research of flexible, multi-domain and multilingual techniques would be a breakthrough in the NLG area.

Towards the advance of such a big challenge, the objective of this paper is to present an almost-fully language independent statistical data-to-text NLG approach that can generate text for different domains, thanks to the concept of an input seed feature which guides all the generation process. Within our scope, this seed feature can be seen as an abstract object (e.g., a rating, a sentiment, a polarity, a phoneme) that will determine how the final sentence will be in relation to its vocabulary or the word categories that this new sentence must contain. We tested our approach in the context of two different domains, that will be explained in section 4, for the English and Spanish languages, in order to show its appropriateness to different non-related scenarios.

2 Related Work

The task of NLG comprises a wide range of subtasks which extend from an action planning until its execution. Therefore, starting from non-linguistic data or text, there are many decisions to be made such as the structure of the message and its content, the rhetorical structure at several levels, the syntactic structure and the correct words choice or the final text arrangement (Bateman and Zoch, 2003). These subtasks can be grouped into a pipeline of three broad stages: document planning, microplanning and surface realisation (Reiter and Dale, 2000). In the document planning stage, the system must decide what information should be included in the text and how to organise it into a coherent structure, leading to a document plan. From this document plan, in the microplanning stage, a discourse plan will be generated, where appropriated words and references will be chosen supplying them with a linguistic structure. Finally, the surface realisation stage generates the final text with the concrete information and structure selected.

This NLG process is commonly addressed from either statistical and knowledge based approaches, where the former are based on the calculus of the probability of certain words to appear together; and the latter resort to linguistic techniques in order to generate text. The main difference between these two approaches is that statistical approaches are more flexible than knowledge based ones in terms of language and domain.

Traditionally, statistical approaches have been based on Language Models (LM), whose probabilities are extracted from a text. Due to this, these approaches are highly adaptable to different domains and languages. Factored language models (FLM) are an extension of LM proposed in Bilmes and Kirchhoff (2003) which permit a greater flexibility and adaptability. In this model, a word is viewed as a vector of k factors such that $w \equiv \{f^1, f^2, \ldots, f^K\}$, where these factors can be anything, including the Part-Of-Speech (POS) tag, stem or any other lexical, syntactic or semantic feature. Once a set of factors is selected, the main objective of a FLM is to create a statistical model $P(f|f_1, \ldots, f_N)$ where the prediction of a feature f is based on N parents $\{f_1, \ldots, f_N\}$. These models have been widely employed in several areas of Computational Linguistics, mainly in machine translation (Crego and Yvon, 2010). Furthermore they have been used to a lesser degree in NLG, such in the BAGEL system (Mairesse and Young, 2014), where FLM are used to predict the semantic structure of the sentence to generate, or in Novais and Paraboni (2012) where FLM are used to rank sentences in Portuguese.

Moreover, there are several approaches focused on the generation of reviews such as the one presented in Gerani et al. (2014), where an abstractive summarisation for products reviews is generated taking advance of their discourse structure. However, to the best of our knowledge, there is no previous research work focused on generating opinionated sentences employing FLM, and, furthermore, with the restriction of having words related with a concrete seed input features (a specific polarity in our case). In addition, our approach is also novel in the sense that it can be applied to different domains and language with minimal adaption.

3 *A Flexible Multi-Domain Natural Language Generation Approach*

We propose a statistical approach focused on the surface realisation stage and based on over-generation and ranking techniques employing FLMs.

This technique allows the approach to be almost-fully language independent since it is necessary to adapt some resources (e.g., semantic features) for the language-specific part. This input seed feature concept introduced will permit us to make the generated text flexible regarding its domain and purpose.

This approach first generates several sentences which then will be ranked as will be explained below.

3.1 Generation

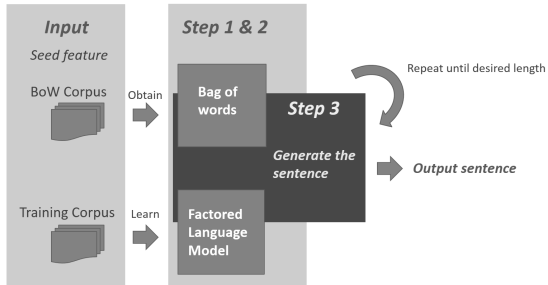

Figure 2: Our proposed approach

For a specific input seed feature (e.g., “positive” polarity), multiple sentences are generated, taking into account: i) a training corpus, ii) a corpus from where a bag of words is obtained (BoW corpus), and iii) the seed feature. The generation approach consists of three major steps, as can be seen in Figure 2:

1. **Step 1: Generate the language model**. A FLM is firstly trained over a corpus (i.e., the training corpus, a collection of texts from where the FLM is trained) in order to obtain the probabilities of the factors of appearing together.
2. **Step 2: Generate the bag of words**. A bag of words containing words related with the input seed feature and their frequency is obtained from the BoW corpus (i.e., a different collection of texts from where the bag of words is gathered). For instance, in the case that we want to generate a sentence with positive polarity, the bag of words could include words such as “great”, “good”, “outstanding”, “excellent”, etc.
3. **Step 3: Generate the sentence**. Then, a sentence is generated based on the FLM and the bag of words previously obtained. The generation algorithm follows an iterative process that will finish when the desired length of

the sentence or a full stop are reached. This will allow us to decide the length of the sentence depending on the final application (e.g., a tweet or a sentence to be integrated in a long review) In this iterative generation process, starting in the first iteration from the token start of the sentence, the following words are selected according to the highest probabilities from the FLM, prioritising the selection of words from the bag of words. In this manner, the process guarantee that the generated sentence will contain the maximum number of words related with the input seed feature.

3.2 Ranking

When several sentences are generated for a specific seed feature, the aim of this stage is to decide which one would be finally selected. Only one sentence is selected during this stage. The ranking is performed in order to select one correct sentence based on its probability and the number of words related with the seed feature. In the case that the ranking was not applied, several sentences would be generated, and the user will have to manually select the one that s/he prefers. The sentence probability is computed by the chain rule where the probability of a sentence can be calculated as the product of the probability of all the words: $P(w_1, w_2...w_n) = \prod_{i=1}^{n} P(w_i|w_1, w_2...w_{i-1})$.

The probability of a word is then calculated, as it is suggested in Isard, Brockmann, and Oberlander (2006), such the linear combination of FLMs, where a weight λ_i was assigned for each of them: $P(f_i|f_{i-2}^{i-1}) = \lambda_1 P_1(f_i|f_{i-2}^{i-1})^{1/n} + \cdots + \lambda_n P_n(f_i|f_{i-2}^{i-1})^{1/n}$, where f the selected factors from the different FLMs employed, being the total sum of the weights 1. The final selected sentence would be the one containing the maximum number of words related to the seed feature and which probability is above the average.

4 Domains

We primarily focused our experiments in the domain of generation sentences with the positive and negative polarity in the context of movie reviews. Our final application would be to provide supporting sentences to visual or numeric ratings, so that reviews could be complemented with more information, thus becoming more informative. Furthermore, in order to verify the flexibility and multi-domain of this approach, we also tested this approach in the context of NLG or assistive technologies as it will be explained in section 4.2.

4.1 Opinionated NLG

Within our first domain, the experimentation was focused on the generation of opinionated sentences with a specific polarity (positive or negative), using this polarity as the input seed feature. The main objective is to create meaningful sentences containing words with a specific polarity.

A large portion of the web is dedicated to sites where people express their opinions (such as TripAdvisor[1] or RottenTomataoes[2]), so the generation of this kind of polarity sentences could serve this type of platforms to generate sentences from visual numeric ratings (like stars). So, in a first instance, we focused the generation on the context of movie reviews, where an illustrative example of the sentences we want to generate can be seen in Figure 3. The generation of this kind of sentences can be very useful when an user uses Webpages as the one shown in section 1, where in the review there are only symbols or numbers without any type of explicative or informative associated text.

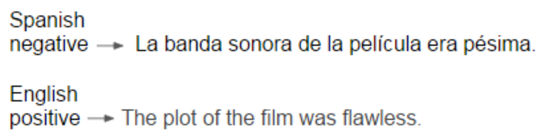

Figure 3: Illustrative example of opinionated NLG sentences (*Translation: The film's soundtrack was awful*)

Given the context seen above, we have employed the Spanish Movie Reviews corpus[3] and the Sentiment Polarity Dataset (Pang and Lee, 2004) as our corpora for Spanish and English, respectively. The approach was tested with the positive and negative polarities using the ML-SentiCon (Cruz et al., 2014) files and the polarity words from (Liu, Hu, and Cheng, 2005) to identify the polarity of a word in Spanish and English, respectively.

[1] https://www.tripadvisor.es/
[2] https://www.rottentomatoes.com/
[3] http://www.lsi.us.es/ fermin/corpusCine.zip

4.2 NLG for assistive technologies

This was a completely different scenario that was used to test and verify the flexibility of our proposed NLG approach. Within our second domain, the experimentation was focused on story generation to help children with dyslalia, a disorder in phoneme articulation. Based on this domain, a phoneme is selected as the seed feature, where the main objective is to generate meaningful sentences containing the maximum number of words in the sentences related to that concrete phoneme. This type of sentences can be useful in dyslalia speech therapies in order to reinforce the phoneme pronunciation through reading and repeating words (Rvachew, Rafaat, and Martin, 1999).

Some illustrative examples of an input phoneme and generated sentences meeting the requirements can be seen in Figure 4.

Spanish
/e/ → **É**ras**e** qu**e** s**e** **e**ra, un hombr**e** llamado **E**st**e**ban.

English
/e/ → My fr**ie**nd Fr**e**d s**ai**d that m**a**ny people eat br**ea**d.

Figure 4: Illustrative example of NLG sentences for assistive technologies (*Translation: Once upon a time, a man named Esteban*)

Therefore, for this scenario, the employed approach is the same as in the first domain specified in section 4.1, where the only difference lies in the seed feature (in this domain, a phoneme) and the corpus used. Consequently, a collection of 158 Hans Christian Andersen tales in two languages (English and Spanish) was chosen as corpora, being the vocabulary contained in it suitable for a young audience. In addition, the approach was tested with all the English and Spanish phonemes.

5 *Experiments*

With the domains previously mentioned, we conducted an experiment where, using phonemes and polarity as the seed feature in each domain, we automatically generated sentences in Spanish and English. From these generated sentences, a raking was performed over those ended by a full stop according to the linear combination explained in section 3.2. With this experimentation, we wanted to test to what extent the generated sentences were classified as positive and negative (in the case of the first domain) and if they have words with an specific phoneme (in the case of the second domain).

During the experimentation, we used several tools that will be further explained.

Each file of the corpus previously described was processed with Freeling (Padró and Stanilovsky, 2012) in order to obtain information about the selected factors of the FLM. In our case, these factors were the word itself (token), the POS-tag and the lemma. Freeling is a language analyser at a lexical, syntactic and semantic level that works for multiple languages, including English and Spanish.

In order to evaluate the polarity of the sentences, we employed the sentiment analysis classifier described in Fernández et al. (2013).

Finally, we trained the FLM with SRILM (Stolcke, 2002), a software which allows building and applying statistical language models, which also includes an implementation of factored language models.

Taking into account the different factors, Spanish and English sentences were automatically generated using trigram FLM with LEMMA+POSTAG (which proved to works better than other configurations), and subsequently these sentences were ranked with a linear combination of three FLM, as explained in section 3: $P(w_i) = \lambda_1 P(f_i|f_{i-2}, f_{i-1}) + \lambda_2 P(f_i|p_{i-2}, p_{i-1}) + \lambda_3 P(p_i|f_{i-2}, f_{i-1})$, where f can be can be either a lemma and a word, p refers to a POS tag, and λ_i are set $\lambda_1 = 0.25$, $\lambda_2 = 0.25$ and $\lambda_3 = 0.5$. These values were empirically determined.

6 *Evaluation and Results*

The evaluation of NLG approaches are difficult since there is not a an unique good output (gold-standard) as in other Computational Linguistic fields. In addition,there is no automatic manner to discern the meaningfulness of a given generated text or sentence in an automatic manner. In view of the above, the manual evaluation is the most currently type of assessment used in NLG (Resnik and Lin, 2010). On this basis, we performed a manual evaluation of the generated sentences in order to verify the meaningfulness of the automatic generated sentences.

This manual evaluation was performed by three different evaluators considering a sentence meaningful when: i) the sentence

Surface Realisation Domain		Meaningful generated sentences	Newly meaningful sent. (not in corpus)	Meaningful sent. with seed features
EN	Opinionated sentences for reviews	100%	50%	50%
	Assistive technologies	95%	70%	82.5%
ES	Opinionated sentences for reviews	100%	100%	100%
	Assistive technologies	88.89%	40.74%	88.89%

Table 1: Comparative table of the two domains

is meaningful by itself, ii) the sentence becomes meaningful by adding some punctuation marks, and iii) the sentence becomes meaningful by adding a preposition that usually follows the main verb. In order to measure the agreement between the evaluators the kappa statistic (Randolph, 2008) was employed, obtaining a very good agreement in both domains (an overall agreement of 1 for the opinionated NLG domain in both, English and Spanish; and an overall agreement of 0.83 for the assistive technologies domain in English and an overall agreement of 0.78 in Spanish).

On the other hand, these sentences were automatic evaluated to discern if they met the objective for each domain. This was carried out evaluating the polarity of the sentences with the sentiment analysis classifier mentioned before, in the first domain; and calculating the percentage of words containing the phonemes regarding the total length of the sentence in the second domain.

Table 1 shows the results of the approach once the whole NLG approach is applied (over-generation and ranking), where multiple sentences were generated for a concrete seed feature and subsequently ranked in order to obtain only one sentence for that seed feature. The statistics of the table were calculated based on the total number of selected sentences once the ranking was employed (being the maximum of sentences to generate the two polarities, one sentence for the positive polarity and one for the negative in the first domain; and the total number of phonemes in each language, being 44 phonemes for English and 27 phonemes for Spanish, in the second domain).

As it can be seen in the table, good results were obtained in the meaningful generated sentences in both domains, being almost the half of them not explicitly included in the corpus. Furthermore, we also obtained good results on those meaningful sentences containing words related with the seed feature, fulfilling the characteristics specified in Section 4 in both domains. We checked this using the sentiment analysis classifier mentioned in section 5 in the first domain, where we found that the polarity obtained for the generated sentence was the right polarity we specified as input. In the second we performed a manual evaluation of the words with the phonemes, where the sentences contained an average of 3 words out of 8, that was the average length of the sentences, with the specific phoneme in both languages.

Examples of the generated sentences in English and Spanish are shown in Figure 5.

Opinionated NLG
Polarity: Positive **Sent:** *The good work in this respect.*
Polarity: Negative **Sent:** *The acting be horrible .*
Polarity: Negative **Sent:** *Su falta de imaginación. Trans: Their lack of imagination.*

NLG for assistive technologies
Phoneme: /m/ **Sent:** *My mother be asleep.*
Phoneme: /b/ **Sent:** *I be bear in the book of fairy tale.*
Phoneme: /k/ **Sent:** *Cantar el canción popular. Trans: Sing the popular song.*

Figure 5: Example generated sentences

In view of these results, this approach obtains similar performance in the generation of sentences in both domains for English and Spanish, where the flexibility of the proposed method is demonstrated. The main problem of our approach though is that, due to the use of lemmas as factors, the words in the generated sentences are not inflected, and, in some cases, this affects the readability of the sentence. We are investigating possible methods to tackle this issue, such as the definition of rules, or the definition of a model to automatically learn the inflections.

7 Conclusions and Future work

In this research work, a multilingual and multi-domain statistical NLG approach which relies on an input seed feature to generate a sentence was presented. This approach allows the generation of sentences oriented to meet the purpose given by an specific seed input feature. This approach was first tested for generating opinionated sentences, where the input seed feature was the desired polarity of the sentence. This type of sentences may be useful for reviews generation based on ratings to support and provide evidences for the numeric values or symbols. Then, to verify that the same approach could be applied to other domains and scenarios, it was applied to the generation of sentences that can be useful in several speech therapies, having a phoneme as the seed feature.

Through the experimentation conducted, the approach has proved to be able to generate sentences in two different domains with similar performance and for two different languages, obtaining good results and fulfilling the requirements specified for each domain.

Although the obtained results are good, we need to add more syntactic and semantic information in order to guarantee the generation of meaningful sentences in all the cases. Consequently, in the future we will study different factors to be included in the FLM, and also, we will analyse to what extent the inclusion of deep learning techniques or word embedding-based method may be beneficial to the approach.

In the short term, we would like to improve the readability of the sentences, a well as to widen and conduct a more exhaustive evaluation of the generated sentences using crowdsourcing platforms.

Furthermore, there are three issues to be improved and research as the next steps. As mentioned before, the inflection of the words of the generated sentences is one the issues to be further investigated. In this respect, we first need to research in the types of transformations that can be applied to the words. For example, we could employ dictionaries containing the inflections and variants of the words,which could be combined with some kind of grammar or structure in order to finally obtain a infected sentence. An example of how could be the inflections of the example generated sentences seen above is shown in Figure 6.

Opinionated NLG
Original Sent: *The good work in this respect.*
Inflected Sent: *The good work was done in this respect.*

NLG for assistive technologies
Original Sent: *My mother be asleep.*
Inflected Sent: *My mother was asleep.*

Figure 6: Example inflections of the generated sentences

On the other hand, another issue that can be further researched is the generation of several sentences with cohesion between them in order to build a larger text. This sentences would need to have related topics to ensure the text coherence. This goal could be achieved, in a first approach, by including in the sentences the same subject or by taking as the subject of the sentence the direct object of the previous sentence.

In addition, for the story generation domain, it could be interesting if the seed feature could be composed, for example phoneme+polarity or phoneme+sentiment in order to generate stories with sentiments.In this case, we would need to adapt the input of the approach to have multiple seed features. For example, if the text to be generated has to help people feeling depressed, it would be necessary to generate an optimistic and happy text, so in this context, the seed feature would be for instance the concept *optimistic*, and, the words selected during the generation process would be related with this concept (e.g. the words *cheerful*, *joy* or *favorable* are related with the optimistic concept). In order to obtain the words related with a concept, lexicons or synsets such as WordNet-Affect (Strapparava and Valitutti, 2004) or word embedding techniques could be employed.

References

Ballesteros, M., B. Bohnet, S. Mille, and L. Wanner. 2015. Data-driven sentence generation with non-isomorphic trees. In *Proceedings of the NAACL*, pages 387–397. Association for Computational Linguistics.

Bateman, J. and M. Zoch. 2003. *Natural Language Generation*. Oxford University Press.

Bilmes, J. A. and K. Kirchhoff. 2003. Factored language models and generalized parallel backoff. In *Proceedings of the Human Language Technology Conference (HLT/NAACL)*, pages 4–6. Association for Computational Linguistics.

Crego, J. M. and F. Yvon. 2010. Factored bilingual n-gram language models for statistical machine translation. *Machine Translation*, 24(2):159–175.

Cruz, F. L., J. A. Troyano, B. Pontes, and F. J. Ortega. 2014. Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems with Applications*, 41(13):5984 – 5994.

Fernández, J., Y. Gutiérrez, J. M. Gómez, P. Martínez-Barco, A. Montoyo, and R. Muñoz. 2013. Sentiment analysis of spanish tweets using a ranking algorithm and skipgrams. *Proc. of the TASS workshop at XXIX Conference of the Spanish Society for Natural Language Processing (SEPLN 2013)*, pages 133–142.

Ge, T., W. Pei, H. Ji, S. Li, B. Chang, and Z. Sui. 2015. Bring you to the past: Automatic generation of topically relevant event chronicles. In *ACL*, pages 575–585. Association for Computational Linguistics.

Gerani, S., Y. Mehdad, G. Carenini, R. T. Ng, and B. Nejat. 2014. Abstractive summarization of product reviews using discourse structure. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1602–1613. Association for Computational Linguistics.

Isard, A., C. Brockmann, and J. Oberlander. 2006. Individuality and alignment in generated dialogues. In *Proceedings of the INLG*, pages 25–32. Association for Computational Linguistics.

Jacko, J. A. 2012. *Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Third Edition*. CRC Press, Inc.

Liu, B., M. Hu, and J. Cheng. 2005. Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web*, pages 342–351. ACM.

Mairesse, F. and S. Young. 2014. Stochastic language generation in dialogue using factored language models. *Comput. Linguist.*, 40(4):763–799.

Novais, E. M. and I. Paraboni. 2012. Portuguese text generation using factored language models. *Journal of the Brazilian Computer Society*, 19(2):135–146.

Padró, L. and E. Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality. In *Proceedings of the 8ht International Conference on Language Resources and Evaluation*. European Language Resources Association.

Pang, B. and L. Lee. 2004. A sentimental education: Sentiment analysis using subjectivity. In *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics*, pages 271–278. Association for Computational Linguistics.

Ramos-Soto, A., A. J. Bugarín, S. Barro, and J. Taboada. 2015. Linguistic descriptions for automatic generation of textual short-term weather forecasts on real prediction data. *IEEE Transactions on Fuzzy Systems*, 23(1):44–57.

Randolph, J. J. 2008. Online kappa calculator [computer software]. retrieved from http://justus.randolph.name/kappa.

Reiter, E. and R. Dale. 2000. *Building Natural Language Generation Systems*. Cambridge University Press.

Resnik, P. and J. Lin, 2010. *Evaluation of NLP Systems*, pages 271–295. Wiley-Blackwell.

Rvachew, S., S. Rafaat, and M. Martin. 1999. Stimulability, speech perception skills, and the treatment of phonological disorders. *American Journal of Speech-Language Pathology*, 8(1):33–43.

Stolcke, A. 2002. Srilm - an extensible language modeling toolkit. In *Proceedings International Conference on Spoken Language Processing, vol 2.*, pages 901–904.

Strapparava, C. and A. Valitutti. 2004. Wordnet affect: an affective extension of wordnet. In *LREC*, volume 4, pages 1083–1086.