Natural Language Generation: Recently Learned Lessons, Directions for Semantic Representation-based Approaches, and the case of Brazilian Portuguese Language

Marco Antonio Sobrevilla Cabezudo and Thiago Alexandre Salgueiro Pardo

Interinstitutional Center for Computational Linguistics (NILC)

Institute of Mathematical and Computer Sciences, University of São Paulo

São Carlos/SP, Brazil

`msobrevillac@usp.br, taspardo@icmc.usp.br`

Abstract

This paper presents a more recent literature review on Natural Language Generation. In particular, we highlight the efforts for Brazilian Portuguese in order to show the available resources and the existent approaches for this language. We also focus on the approaches for generation from semantic representations (emphasizing the Abstract Meaning Representation formalism) as well as their advantages and limitations, including possible future directions.

1 Introduction

Natural Language Generation (NLG) is a promising area in Natural Language Processing (NLP) community. NLG aims to build computer systems that may produce understandable texts in English or other human languages from some underlying non-linguistic representation of information (Reiter and Dale, 2000). Tools generated by this area are useful for other applications like Automatic Summarization, Question-Answering Systems, and others.

There are several efforts in NLG for English[1]. For example, one may see the works of Krahmer et al. (2003) and Li et al. (2018), which focused on referring expressions generation, and the work of (Gatt and Reiter, 2009), focused on developing a surface realisation tool called SimpleNLG. One may also easily find other works that tried to generate text from semantic representations (Flanigan et al., 2016; Ferreira et al., 2017; Puzikov and Gurevych, 2018b).

For Brazilian Portuguese, there are few works, some of them focused on representations like Universal Networking Language (UNL) (Nunes et al., 2002) or *Resource Description Framework* (RDF)

(Moussallem et al., 2018), and other ones that are very specific to the Referring Expression Generation (Pereira and Paraboni, 2008; Lucena et al., 2010) and Surface Realisation tasks (Oliveira and Sripada, 2014; Silva et al., 2013).

More recently, several representations have emerged in the NLP area (Gardent et al., 2017; Novikova et al., 2017; Mille et al., 2018). In particular, Abstract Meaning Representation (AMR) has gained interest from the research community (Banarescu et al., 2013). It is a semantic formalism that aims to encode the meaning of a sentence with a simple representation in the form of a directed rooted graph. This representation includes information about semantic roles, named entities, wiki entities, spatial-temporal information, and co-references, among other information.

AMR has gained attention mainly due to its simplicity to be read by humans and computers, its attempt to abstract away from syntactic idiosyncrasies (focusing only on semantic processing) and its wide use of other comprehensive linguistic resources, such as PropBank (Palmer et al., 2005) (Bos, 2016).

For English, there is a large AMR-annotated corpus that contains 39,260 AMR-annotated sentences[2], which allows deeper studies in NLG and experiments with different approaches (mainly statistical approaches). This may be evidenced in the SemEval-2017 shared-task 9 (May and Priyadarshi, 2017)[3].

For Brazilian Portuguese, Anchiêta and Pardo (2018) built the first corpus using sentences from the "The Little Prince" book. The authors took advantage of the alignment between the English and Brazilian Portuguese versions of the book to import the AMR structures from one language to

---

[1]Most of the works may be found in the main NLP publication portal at https://www.aclweb.org/anthology/

[2]Available at https://catalog.ldc.upenn.edu/LDC2017T10.

[3]Available at http://alt.qcri.org/semeval2017/task9/.

another (but also performing the necessary adaptations). They had to use the Verbo-Brasil repository (Duran et al., 2013; Duran and Aluísio, 2015), which is a PropBank-like resource for Portuguese. Nowadays, there is an effort to build a larger AMR-annotated corpus that is similar to the current one available for English.

In this context, this study presents a literature review on Natural Language Generation for Brazilian Portuguese in order to show the resources (in relation to semantic representations) that are available for Portuguese and the existent efforts in the area for this language. We focus on the NLG approaches based on semantic representations and discuss their advantages and limitations. Finally, we suggest some future directions to the area.

2 Literature Review

The literature review was based on the following research questions:

- What was the focus of the existent NLG efforts for Portuguese and which resources were used for this language?

- What challenges exist in the NLG approaches?

- What are the advantages and limitations in the approaches for NLG from semantic representations, specially Abstract Meaning Representation?

Such issues are discussed in what follows.

2.1 Natural Language Generation for Portuguese

In general, we could find few works for Portuguese (considering the existing works for English). These works focus mainly on the referring expression generation (Pereira and Paraboni, 2008; Lucena et al., 2010) and surface realization tasks (Silva et al., 2013; Oliveira and Sripada, 2014), usually restricted to specific domains and applications (like undergraduate test scoring). Nevertheless, there are some recent attempts focused on other tasks and in more general domains (Moussallem et al., 2018; Sobrevilla Cabezudo and Pardo, 2018).

Among the NLG approaches, we may highlight the use of templates (Pereira and Paraboni, 2008; Novais et al., 2010b), rules (Novais and Paraboni,

2013) and language models (LM) (Novais et al., 2010a). In general, these approaches were successful because they were focused on restricted domains. Specifically, template-based methods used basic templates to build sentences. Similarly, some basic rules involving noun and verbal phrases were defined to build sentences. Finally, LM-based methods applied a two-stage strategy to generate sentences. This strategy consisted in generating surface realization alternatives and selecting the best alternative according to the language model.

In the case of LM-based methods, we may point out that classical LMs (based on n-grams) were not suitable because it was necessary to use a large corpus to deal with sparsity of data. Sparsity is a big problem in morphologically marked languages like Portuguese. In order to solve the sparsity of the data, some works used Factored LMs, obtaining better results than the classical LMs (de Novais et al., 2011).

In relation to NLG from semantic representations for Portuguese, we may point out the work of Nunes et al. (2002) (focused on Universal Language Networking), and Moussallem et al. (2018) (focused on ontologies). Another representation was the one proposed by Mille et al. (2018) (based on Universal Dependencies), which is based on syntax instead of semantics.

In relation to NLG tools, we highlight PortNLG (Silva et al., 2013) and SimpleNLG-BP (Oliveira and Sripada, 2014) as surface realisers that were based on SimpleNLG initiative (Gatt and Reiter, 2009)[4]. Finally, other NLG works aimed to build NLP applications, e.g., for structured data visualization and human-computer interaction purposes (Pereira et al., 2012, 2015).

2.2 Natural Language Generation from Semantic Representations

Recently, the number of works on NLG from semantic representations has increased. This increase is reflected in the shared tasks WebNLG (Gardent et al., 2017), E2E Challenge (Novikova et al., 2017), Semeval Task-9 (May and Priyadarshi, 2017) and Surface Realization Shared-Task (Belz et al., 2011; Mille et al., 2018).

In general, there is a trend to apply methods based on neural networks. However, methods

---

[4]Specifically, SimpleNLG-BP was built using the French version of SimpleNLG due to the similarities between both languages.

based on templates, transformation to intermediate representations and language models have shown interesting results. It is also worthy noticing that most of these methods have been applied to English, except for the methods presented in the shared-task proposed by Mille et al. (2018).

In relation to the shared-tasks mentioned before, we point out that the one proposed by Belz et al. (2011) and Mille et al. (2018) (based on Universal Dependencies) used syntactic representations. Specifically, they presented two tracks, one focused on word reordering and inflection generation (superficial track), and other that focused on generating sentences from a deep syntactic representation that is similar to a semantic representation (deep track). Furthermore, these tasks focused on several languages in the superficial task (including Portuguese) and three languages in the deep track (English, Spanish, and French).

Among the methods used for the superficial track in these shared-tasks, we may highlight the use of rule-based methods and language models in the early years (Belz et al., 2011) and a wide application of neural models in recent years (Mille et al., 2018). In the case of the deep track, it is possible to notice that rule-based methods were applied in the first competition, and methods based on transformation to intermediate representations and based on neural models were applied in the last competition.

The results in these tasks showed that approaches based on transformation to intermediate representations obtained poor results in the automatic evaluation due to the great effort in building transformation rules for their own systems. However, they usually showed better results in human evaluations. This may be explained by the maturity of the original proposed systems. This way, although the coverage of the rules was not good, the results were good from a human point of view.

Differently from the approach mentioned before, methods based on neural models (deep learning) obtained the best results. However, some methods used data augmentation strategies to deal with data sparsity (Elder and Hokamp, 2018; Sobrevilla Cabezudo and Pardo, 2018).

One point to highlight is that the results for Portuguese were poor (compared to similar languages like Spanish). Two reasons to explain this issue are related to the amount of data for Portuguese in this task (less than English or Spanish) and the quality

of the existing models for related tasks that were used. Another point to highlight is the division of the general task into two sub-tasks: linearisation and inflection generation. Puzikov and Gurevych (2018a) pointed out that there is a strong relation between the linearisation and the inflection generation, and, thus, both sub-tasks should be treated together.

In contrast to Puzikov and Gurevych (2018a), (Elder and Hokamp, 2018) showed that incorporating syntax and morphological information into neural models did not bring significant contribution in the generation process, but incorporated more difficulty in the task.

Finally, it is important to notice the proposal of Madsack et al. (2018), which trained linearisation models using the dataset for each language independently and in a joint way, using multilingual embeddings. Although the results of this work did not present a lot of variation when used for all languages together, this work suggests that it is possible to train systems with similar languages (for example, Spanish and French) in order to take advantage of the syntax similarities and to overcome the problems of lack of data.

In relation to other used representations (Gardent et al., 2017; Novikova et al., 2017), a large number of works based on deep learning strategies were proposed, obtaining good results. However, the use of pipeline-based methods yielded promising results regarding grammar and fluency criteria in a joint evaluation (for RDF representation), but these methods (which usually use rules) obtained the worst results in the E2E Challenge.

Methods based on Statistical Machine Translation kept a reasonable performance (ranking 2nd in RDF Shared-Task), obtaining good results when evaluating the grammar. The explanation for this result comes from the ability to learn complete phrases. Thus, these methods may generate grammatically correct phrases, but with poor general fluency and dissimilarity to the target output. Finally, methods based on template obtained promising results in restricted domains, like in the E2E Challenge.

2.3 Natural Language Generation from Abstract Meaning Representation

In relation to generation methods from Abstract Meaning Representation, it was possible to highlight approaches based on machine translation

(Pourdamghani et al., 2016; Ferreira et al., 2017), on transformation to intermediate representations (Lampouras and Vlachos, 2017; Mille et al., 2017), on deep learning models (Konstas et al., 2017; Song et al., 2018), and on rule extraction (from graphs and trees) (Song et al., 2016; Flanigan et al., 2016).

Methods based on transformation into intermediate representations focused on transforming AMR graphs into simpler representations (usually dependency trees) and then using an appropriate surface realization system. Authors usually took advantage of the similarity between dependency trees and AMR graphs to map some results. However, some problems in this approach were the need to manually build transformation rules (excepting for Lampouras and Vlachos (2017), who automatically perform this) and the need of alignments between the AMR graph and intermediate representations, which could bring noise into the generation process. Overall, this approach presented poor results (compared to other approaches) in automatic evaluations[5]

Methods based on rule extraction obtained better results than the approach mentioned previously. This approach tries to learn conversion rules from AMR graphs (or trees) to the final text. First methods of this approach tried to transform the AMR graph into a tree before learning rules. As (Song et al., 2017) mentioned, these methods suffer with the loss of information (by not using graphs and being restricted to trees), due to its projective nature. Likewise, (Song et al., 2016) and (Song et al., 2017) could suffer from the same problem (ability to deal with non-projective structures) due to their nature to extract and apply the learned rules. Furthermore, these methods used some manual rules to keep the text fluency. However, these rules did not produce a statistically significant increase in the performance, when compared to learned rules.

Some problems of this approach are related to: (1) the need of alignments between AMR graph and the target sentence, as the aligners could lead to more errors (depending of the performance) in the rule extraction process; (2) the argument realization modeling (Flanigan et al., 2016; Song et al., 2016); and (3) the data sparsity in the rules, as some rules are too specific and there is a need to generalize them.

Methods based on Machine Translation usually outperformed other methods. Specifically, methods based on Statistical Machines Translation (SMT) outperformed methods based on Neural Machine Translation (NMT), which use data augmentation strategies to improve their performance (Konstas et al., 2017). In general, both SMT and NMT-based methods explored some preprocessing strategies like delexicalisation[6], compression[7] and graph linearisation[8] (Ferreira et al., 2017)

In relation to the linearisation, the proposals of Pourdamghani et al. (2016) and Ferreira et al. (2017) depended on alignments to perform linearisation. Both works point out that the way linearisation is carried out affects performance, thus, linearisation is an important preprocessing strategy in NLG. However, Konstas et al. (2017) show that linearisation is not that important in NMT-based methods, as the authors propose a data augmentation strategy, decreasing the effect of the linearisation.

In relation to compression, the dependency of alignments also occurred. Moreover, it is necessary a deep analysis to determine the usefulness of compression. On the one hand, compression contributed positively in the SMT-based methods but, on the other hand, it was harmful in NMT-based methods (Ferreira et al., 2017). It is also important to point out that both compression and linearisation processes were executed in sequence in these works. This could be harmful, as the order of execution could lead to loss of information.

Finally, according to (Ferreira et al., 2017), delexicalisation produces an increase and decrease of performance in NMT-based and SMT-based methods, respectively. An alternative to deal with data sparsity is to use copy mechanisms, which have shown performance increase in NLG methods (Song et al., 2018).

Some limitations of these methods were the alignment dependency (similar to the previous approaches) and the linearisation of long sentences. NMT-based methods could not represent or capture information for long sentences, producing un-

---

[5]Except for the work of Gruzitis et al. (2017), who incorporated the system proposed by Flanigan et al. (2016) into their pipeline.

[6]Delexicalisation aims to decrease the data sparsity by replacing some common tokens by constants.

[7]Compression tries to keep important concepts and relations in the text generation process.

[8]Linearisation tries to transform the graph into a sequence of tokens.

satisfactory results.

In order to solve these problems, methods based on neural models proposed Graph-to-Sequence architectures to better capture information from AMR graphs. This architecture showed better results than its predecessors, requiring less training data (augmented data) (Beck et al., 2018).

The main difficulty associated to deep learning is the need of large corpora to get better results. Thus, this could be hard to get for languages like Portuguese, as there are no large available corpora as there are for English.

3 Conclusions and Future Directions

This work showed a more recent literature review on NLG, specially those based on semantic representations and for Brazilian Portuguese language. As it may be seen, NLG works for Portuguese were mainly focused on Referring Expression Generation and Surface Realisation. There were a few recent works about NLG from semantic representations like ontologies or Universal Dependencies (although this last one is of syntactic nature), producing poor results.

Some resources for Portuguese were found (additional to AMR-annotated corpus), as corpora for generation from RDF (Moussallem et al., 2018) and from Universal Dependencies (Mille et al., 2018). This opens the possibility to explore the use of other resources for similar tasks in order to improve the AMR-to-Text generation. There are also corpora for languages that are relatively similar to Portuguese. Considering the proposal of Madsack et al. (2018), to learn realisations from languages that share some characteristics with Portuguese (like French or Spanish) is a reasonable alternative.

Among other strategies to deal with lack of data, it is possible to consider Unsupervised Machine Translation and back-translation strategies. The first one tries to learn without parallel corpora (these would be a corpus of AMR graphs and a corpus of sentences). This strategy has proven to be useful in this context (Lample et al., 2018a,b; Freitag and Roy, 2018). In this case, it would be necessary to extend the corpus of AMR annotations, which could represent one of the challenges. The second one aims to generate corpus in a target language (Portuguese) from other languages (as English) in order to increase the corpus size and reduce the data sparseness. In this case, it is nec-

essary to evaluate the influence of the quality of translations and how this affects the performance of the text generator.

Additionally to the above issue, there are currently large corpora for Portuguese (for example, the corpus used by Hartmann et al. (2017)), which may allow to train robust language models.

The main challenges for Portuguese are its morphologically marked nature and its high syntactic variation[9]. These challenges contribute to data sparseness. Thus, two-stage strategies might not be useful, producing an explosion in the search for the best alternative. Moreover, to treat syntactic ordering and inflection generation together could lead to the introduction of more complexity into the models. Therefore, to tackle NLG for Portuguese as two separate tasks seems to be a good alternative, reducing the complexity of the syntactic ordering and treating inflection generation as a sequence labeling problem.

Among the challenges associated to the methods found in the literature, we may highlight two: (1) the alignment dependency, and (2) the need to better understand the semantic representations (in our case, the AMR graphs) to be able to deduce how they may be syntactically and morphologically realized.

Several approaches need alignments to learn rules and ways to linearise and compress data in AMR graphs. This is a problem because there is a need to manually align AMR graphs and target sentences in order to allow the tools to learn to align by themselves and, then, to introduce these tools into some existent NLG pipeline. Thus, limitations in the aligners may lead to errors in the NLG pipeline. This problem could be bigger in NLG for Portuguese as there is limited resources, and some of these do not present alignments. To solve this, it is possible to use approaches those are not constrained by explicit graph-to-text alignments (for example, graph-to-sequence architectures). Furthermore, this could help to join all the available resources for similar tasks (i. e., corpora for other semantic representations), with no need of alignments, in a easy way and train a semantic representation-independent text generation method. However, it is necessary to measure the usefulness of this approach, comparing it with traditional methods.

---

[9]The interested reader may find an overview of Portuguese characteristics at http://www.meta-net.eu/whitepapers/volumes/portuguese.

Finally, to better understand a semantic representation (and what it means) is very important, as one may better learn the possible syntactic realisations and, therefore, to give a better clue of how sentences may be morphologically constructed. For Portuguese, there is a challenge to deal with different semantic representations. Although the concepts may be shared among different semantic representations, relations are not the same, and the decision on how to code them could generate some problems in the NLG training.

Acknowledgments

References

Rafael Anchiêta and Thiago Pardo. 2018. Towards AMR-BR: A SemBank for Brazilian Portuguese Language. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Daniel Beck, Gholamreza Haffari, and Trevor Cohn. 2018. Graph-to-sequence learning using gated graph neural networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 273–283. Association for Computational Linguistics.

Anja Belz, Michael White, Dominic Espinosa, Eric Kow, Deirdre Hogan, and Amanda Stent. 2011. The first surface realisation shared task: Overview and evaluation results. In *Proceedings of the 13th European Workshop on Natural Language Generation*, ENLG '11, pages 217–226, Stroudsburg, PA, USA. Association for Computational Linguistics.

Johan Bos. 2016. Expressive power of abstract meaning representations. *Computational Linguistics*, 42(3):527–535.

Magali Sanches Duran and Sandra M. Aluísio. 2015. Automatic generation of a lexical resource to support semantic role labeling in portuguese. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics, *SEM 2015*, pages 216–221, Denver, Colorado, USA. Association for Computational Linguistics.

Magali Sanches Duran, Jhonata Pereira Martins, and Sandra Maria Aluísio. 2013. Um repositório de verbos para a anotação de papéis semânticos disponível na web (a verb repository for semantic role labeling available in the web) [in portuguese]. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*.

Henry Elder and Chris Hokamp. 2018. Generating high-quality surface realizations using data augmentation and factored sequence models. In *Proceedings of the First Workshop on Multilingual Surface Realisation*, pages 49–53. Association for Computational Linguistics.

Thiago Castro Ferreira, Iacer Calixto, Sander Wubben, and Emiel Krahmer. 2017. Linguistic realisation as machine translation: Comparing different mt models for amr-to-text generation. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 1–10, Santiago de Compostela, Spain. Association for Computational Linguistics.

Jeffrey Flanigan, Chris Dyer, Noah A. Smith, and Jaime G. Carbonell. 2016. Generation from abstract meaning representation using tree transducers. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 731–739, San Diego California, USA. Association for Computational Linguistics.

Markus Freitag and Scott Roy. 2018. Unsupervised natural language generation with denoising autoencoders. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3922–3929, Brussels, Belgium. Association for Computational Linguistics.

Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. 2017. Creating training corpora for nlg micro-planning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 179–188. Association for Computational Linguistics.

Albert Gatt and Ehud Reiter. 2009. Simplenlg: A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 90–93, Stroudsburg, PA, USA. Association for Computational Linguistics.

Normunds Gruzitis, Didzis Gosko, and Guntis Barzdins. 2017. Rigotrio at semeval-2017 task 9: Combining machine learning and grammar engineering for amr parsing and generation. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 924–928. Association for Computational Linguistics.

Nathan Hartmann, Erick Fonseca, Christopher Shulby, Marcos Treviso, Jéssica Silva, and Sandra Aluísio. 2017. Portuguese word embeddings: Evaluating on word analogies and natural language tasks. In

Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology, pages 122–131. Sociedade Brasileira de Computação.

Ioannis Konstas, Srinivasan Iyer, Mark Yatskar, Yejin Choi, and Luke Zettlemoyer. 2017. Neural amr: Sequence-to-sequence models for parsing and generation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 146–157, Vancouver, Canada. Association for Computational Linguistics.

Emiel Krahmer, Sebastiaan Van Erk, and André Verleg. 2003. Graph-based generation of referring expressions. *Computational Linguistics*, 29(1):53–72.

Guillaume Lample, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2018a. Unsupervised machine translation using monolingual corpora only. In *International Conference on Learning Representations*.

Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2018b. Phrase-based & neural unsupervised machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5039–5049, Brussels, Belgium. Association for Computational Linguistics.

Gerasimos Lampouras and Andreas Vlachos. 2017. Sheffield at semeval-2017 task 9: Transition-based language generation from amr. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 586–591. Association for Computational Linguistics.

Xiao Li, Kees van Deemter, and Chenghua Lin. 2018. Statistical NLG for generating the content and form of referring expressions. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 482–491, Tilburg University, The Netherlands. Association for Computational Linguistics.

Diego Jesus De Lucena, Ivandré Paraboni, and Daniel Bastos Pereira. 2010. From semantic properties to surface text: the generation of domain object descriptions. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*, 14(45):48–58.

Andreas Madsack, Johanna Heininger, Nyamsuren Davaasambuu, Vitaliia Voronik, Michael Käufl, and Robert Weißgraeber. 2018. Ax semantics' submission to the surface realization shared task 2018. In *Proceedings of the First Workshop on Multilingual Surface Realisation*, pages 54–57. Association for Computational Linguistics.

Jonathan May and Jay Priyadarshi. 2017. Semeval-2017 task 9: Abstract meaning representation parsing and generation. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 536–545. Association for Computational Linguistics.

Simon Mille, Anja Belz, Bernd Bohnet, Yvette Graham, Emily Pitler, and Leo Wanner. 2018. The first multilingual surface realisation shared task (sr'18): Overview and evaluation results. In *Proceedings of the First Workshop on Multilingual Surface Realisation*, pages 1–12. Association for Computational Linguistics.

Simon Mille, Roberto Carlini, Alicia Burga, and Leo Wanner. 2017. Forge at semeval-2017 task 9: Deep sentence generation based on a sequence of graph transducers. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 920–923. Association for Computational Linguistics.

Diego Moussallem, Thiago Ferreira, Marcos Zampieri, Maria Cláudia Cavalcanti, Geraldo Xexéo, Mariana Neves, and Axel-Cyrille Ngonga Ngomo. 2018. Rdf2pt: Generating brazilian portuguese texts from rdf data. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA).

Eder Miranda de Novais, Ivandré Paraboni, and Diogo Takaki Ferreira. 2011. Highly-inflected language generation using factored language models. In *Computational Linguistics and Intelligent Text Processing*, pages 429–438, Berlin, Heidelberg. Springer Berlin Heidelberg.

Eder Miranda De Novais, Thiago Dias Tadeu, and Ivandré Paraboni. 2010a. *Improved Text Generation Using N-gram Statistics*, pages 316–325. Springer Berlin Heidelberg, Berlin, Heidelberg.

Eder Miranda De Novais and Ivandré Paraboni. 2013. Portuguese text generation using factored language models. *Journal of the Brazilian Computer Society*, 19(2):135–146.

Eder Miranda De Novais, Thiago Dias Tadeu, and Ivandré Paraboni. 2010b. Text generation for brazilian portuguese: The surface realization task. In *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*, YIWCALA '10, pages 125–131, Stroudsburg, PA, USA. Association for Computational Linguistics.

Jekaterina Novikova, Ondřej Dušek, and Verena Rieser. 2017. The e2e dataset: New challenges for end-to-end generation. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 201–206. Association for Computational Linguistics.

Maria Nunes, Graças V Nunes, Ronaldo T Martins, Lucia Rino, and Osvaldo Oliveira. 2002. The decoding system for brazilian portuguese using the universal networking language (unl).

Rodrigo De Oliveira and Somayajulu Sripada. 2014. Adapting simplenlg for brazilian portuguese realisation. In *Proceedings of the Eighth International Natural Language Generation Conference, Including Proceedings of the INLG and SIGDIAL*, pages 93–94, Philadelphia, PA, USA. Association for Computational Linguistics.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics.*, 31(1):71–106.

Daniel Bastos Pereira and Ivandré Paraboni. 2008. Statistical surface realisation of portuguese referring expressions. In *Proceedings of the 6th International Conference on Advances in Natural Language Processing*, pages 383–392, Gothenburg, Sweden.

JC Pereira, A Teixeira, and JS Pinto. 2015. Towards a Hybrid Nlg System for Data2Text in Portuguese. In *Proceedings of the 10th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6, Aveiro, Portugal. IEEE.

José Casimiro Pereira, António JS Teixeira, and Joaquim Sousa Pinto. 2012. Natural language generation in the context of multimodal interaction in portuguese. *Electrónica e Telecomunicações*, 5(4):400–409.

Nima Pourdamghani, Kevin Knight, and Ulf Hermjakob. 2016. Generating english from abstract meaning representations. In *Proceedings of the Ninth International Natural Language Generation Conference*, pages 21–25, Edinburgh, UK.

Yevgeniy Puzikov and Iryna Gurevych. 2018a. Binlin: A simple method of dependency tree linearization. In *Proceedings of the First Workshop on Multilingual Surface Realisation*, pages 13–28. Association for Computational Linguistics.

Yevgeniy Puzikov and Iryna Gurevych. 2018b. E2e nlg challenge: Neural models vs. templates. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 463–471. Association for Computational Linguistics.

Ehud Reiter and Robert Dale. 2000. *Building Natural Language Generation Systems*. Cambridge University Press, New York, NY, USA.

Douglas Fernandes Pereira Da Silva, Eder Miranda De Novais, and Ivandré Paraboni. 2013. Um sistema de realização superficial para geração de textos em português. *Revista de Informática Teórica e Aplicada*, 20(3):31–48.

Marco Antonio Sobrevilla Cabezudo and Thiago Pardo. 2018. Nilc-swornemo at the surface realization shared task: Exploring syntax-based word ordering using neural models. In *Proceedings of the First Workshop on Multilingual Surface Realisation*, pages 58–64. Association for Computational Linguistics.

Linfeng Song, Xiaochang Peng, Yue Zhang, Zhiguo Wang, and Daniel Gildea. 2017. Amr-to-text generation with synchronous node replacement grammar. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 7–13. Association for Computational Linguistics.

Linfeng Song, Yue Zhang, Xiaochang Peng, Zhiguo Wang, and Daniel Gildea. 2016. Amr-to-text generation as a traveling salesman problem. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2084–2089, Austin, Texas. Association for Computational Linguistics.

Linfeng Song, Yue Zhang, Zhiguo Wang, and Daniel Gildea. 2018. A graph-to-sequence model for amr-to-text generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 1616–1626. Association for Computational Linguistics.