Enriching the E2E dataset

Thiago Castro Ferreira
aiXplain, inc.
Federal University of Minas Gerais
thiagocf05@ufmg.br

Helena Vaz
Federal University of Minas Gerais
Belo Horizonte, Brazil
lenavaz13@gmail.com

Brian Davis
ADAPT Research Centre
Dublin City University
brian.davis@adaptcentre.ie

Adriana Pagano
Federal University of Minas Gerais
Belo Horizonte, Brazil
apagano@ufmg.br

Abstract

This study introduces an enriched version of the E2E dataset, one of the most popular language resources for data-to-text NLG. We extract intermediate representations for popular pipeline tasks such as discourse ordering, text structuring, lexicalization and referring expression generation, enabling researchers to rapidly develop and evaluate their data-to-text pipeline systems. The intermediate representations are extracted by aligning non-linguistic and text representations through a process called *delexicalization*, which consists in replacing input referring expressions to entities/attributes with placeholders. The enriched dataset is publicly available.[1]

1 Introduction

Data-to-text NLG is the computational task which aims to generate text from non-linguistic data (Reiter and Dale, 2000; Gatt and Krahmer, 2018). Applications of this task have become increasingly common, such as RDF-to-text (Castro Ferreira et al., 2020), AMR-to-text (Ribeiro et al., 2019), dialogue response generation (Dušek et al., 2018), robot-journalism (Rosa Teixeira et al., 2020), etc.

The growth of the field can be partially explained by increasing availability of focused data-to-text resources, such as WebNLG (Gardent et al., 2017b,a), E2E (Novikova et al., 2017; Dušek et al., 2018), ROTOWIRE (Wiseman et al., 2017), GenWiki (Jin et al., 2020) and KELM (Agarwal et al., 2021).

As with other automatic text generation fields, such as Machine Translation, significant advances in deep learning (Cho et al., 2014; Sutskever et al., 2014), along with an increasing number of data-to-text resources, have resulted in upsurge in neural end-to-end applications targeted towards data-to-text NLGk (Gardent and Narayan, 2018). Hence, given a corpus consisting of pairs between a meaning representation (MR) and its corresponding textual verbalization, a deep learning approach is usually trained in an end-to-end style, learning implicit parameters to convert the input MR into textual output. Although these approaches have shown to generate more fluent output, they also pose problems and challenges, in particular with respect to the semantic *adequacy* and overall faithfulness of the text (Wang et al., 2020). For example, some studies have shown that neural end-to-end data-to-text approaches may *hallucinate* (Rohrbach et al., 2018; Wang et al., 2018), i.e. adding information in the text which are not contained in the input data or which are untrue. This is not a trivial issue, given that accuracy of the generated output is in general considered more important than its fluency (Reiter and Belz, 2009). More importantly poor semantic adequacy is in particular unacceptable for practical applications (Dale, 2020). Furthermore, Castro Ferreira et al. (2019) has shown that traditional pipeline data-to-text systems (Reiter and Dale, 2000), which generate text from data in several explicit intermediate steps, may generalize better to new domains and in turn generate more semantically adequate text than end-to-end approaches in the context of the WebNLG corpus.

Although the current data-resources have benefited the development of end-to-end neural models, the same can not be said for pipeline systems, since these resources usually only consist of raw meaning representations and their final verbalizations. Aiming to decrease data sparsity and make data-to-text models more generalizable and generate more adequate texts, many approaches aim to extract alignments between the non-linguistic and text representations, and then use these alignments to build explicit intermediate representations for a more controllable generation process (Juraska et al., 2018; Xu et al., 2021). As examples, all

[1] https://github.com/ThiagoCF05/EnrichedE2E

Proceedings of the 14th International Conference on Natural Language Generation (INLG), pages 177–183,
Aberdeen, Scotland, UK, 20-24 September 2021.

the data-driven participating models of the E2E work by first converting the meaning representation into an intermediate template which is later realized into the final text. This is also the case in the WebNLG challenge, which makes use of the eponymous dataset.

In order to make it easy for researchers to rapidly develop and evaluate data-to-text pipeline systems, Castro Ferreira et al. (2018b) enriched the WebNLG corpus, one of the most popular data-to-text resources. The study extracts intermediate representations for popular pipeline tasks such as discourse ordering, text structuring, lexicalization and referring expression generation. Intermediate representations are automatically extracted by aligning the non-linguistic and text representations through a process called *delexicalization*, which consists of replacing in the texts referring expressions to input entities/attributes with placeholders. The same extraction process with respect intermediate representations above is applied to the recent CACAPO dataset, which is both multilingual (Dutch and English) and multi-domain, containing up to 10,000 sentences (van der Lee et al., 2020).

Highly inspired by the work of Castro Ferreira et al. (2018b) and van der Lee et al. (2020), our study aims to delexicalize and provide pipeline intermediate representations for another very popular data-to-text dataset: the E2E dataset. We believe that the enriched version of the E2E will provide another data-resource so researchers can better investigate data-driven pipeline systems, their subtasks as well as its comparison with state-of-the-art end-to-end systems.

2 The E2E Dataset

The E2E dataset is a resource initially constructed for training end-to-end, data-to-text applications in the restaurant domain. It consists of 50,602 English verbalizations to 5,751 dialog-act-based meaning representations (Novikova et al., 2017). The dataset is split into training, validation and test sets in a ratio of 76.5%, 8.5% and 15%, respectively.

An example of a pair between a meaning representation (top) and its corresponding text (middle) is depicted in Figure 1. Each meaning representation consists of 3-8 attribute-value pairs, picked from a list of 8 attributes depicted in Table 1. Verbalizations were collected through crowd-sourcing using pictures as stimuli. According to the creators, representing the inputs visually allowed the collection of more natural and informative human references phrases than depicting meaning representations (Dušek et al., 2018).

Attribute	Example Values
name	*The Punter*, *The Waterman*, ...
eatType	restaurant, pub
familyFriendly	yes / no
priceRange	cheap, high, moderate ...
food	Indian, Japanese, Chinese ...
near	*Café Rouge*, ...
area	city center, riverside ...
customerRating	low, average, high ...

Table 1: Attributes contained in a meaning representation of E2E and examples of values.

Although the crowd-workers were asked to verbalize all the information contained in the meaning representation, the creators of the corpus decided to do not penalize those who skipped some information. For this reason, the corpus may also be used to study experiments for the content selection task of pipeline data-to-text systems.

The E2E dataset differs from the WebNLG corpus, which focused on semantic variation, as it leverages higher lexical and syntactical variations, having an average of 8.1 reference verbalizations per meaning representation. The corpus is also bigger than other similar corpora such as SFRest (Wen et al., 2015), a corpus in the domain of Hotels and Restaurants with 5,192 verbalizations to 1,950 meaning representations; and Bagel (Mairesse et al., 2010), with 404 texts verbalizing 202 meaning representations.

3 Delexicalization

Following the method used by Castro Ferreira et al. (2018b) in the WebNLG corpus, we aimed to decrease the data sparsity of the corpus and to align a meaning representation with its corresponding text by *delexicalizing* the texts. The delexicalization process works by replacing the referring expressions to the values of the attributes for placeholders representing the attributes. Figure 1 shows an example of a meaning representation, the final verbalization and its delexicalized version (bottom).

The process was conducted differently for training and validation/test parts of the corpus as explained in the following sections.

3.1 Training Data

The process of delexicalizing the training data started by string matching the values of the attributes in the text and replacing them for the spe-

Attribute	Value
name	The Wrestlers
eatType	coffee shop
food	Japanese
priceRange	less than £20
area	riverside
familyFriendly	no
near	Raja Indian Cuisine

↓

Near Raja Indian Cuisine in Riverside is The Wrestlers. It is a Japanese restaurant, has reasonable prices but is not kid friendly.

↓

Near _NEAR_ in _AREA_ is _NAME_ . _NAME_ is a _FOOD_ restaurant , has _PRICERANGE_ prices but is _FAMILYFRIENDLY_ .

Figure 1: Example of the attribute-value pairs of a meaning representation (top), its corresponding verbalization (middle) and a delexicalized template annotated in this study (bottom).

cific placeholder of the attribute i.e _NAME_ or _EATTYPE_ etc. All the partial delexicalized templates were then manually reviewed and annotated by students of linguistics.

Students The students of Linguistics were recruited through a call which announce the task offering university credits in exchange. In total, 10 students volunteered to conduct the annotation.

Website In order to facilitate the annotation, the authors created a website, where, for each annotation instance, the annotators were given access to the input meaning representation, the delexicalized meaning representation, the text and the delexicalized text to be reviewed and corrected. Moreover, a checkbox was provided so the annotators could indicate any problem in the data such as wrong information or hallucination, i.e. verbalization of information which is not contained in the meaning representation.

3.2 Validation/Test Data

In order to accelerate the annotation of the validation and test sets of the corpus, we first trained a Named Entity Recognition and Classification (NERC) tool based on BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019) using the annotated training data, effectively substituting placeholders for named entities. We then replaced the referring expressions which weren't recognized by the NERC model by string matching (and substituting) the attribute values within the text. Finally, to assure the quality of the data, especially the test set, the authors manually reviewed each instance of both parts of the data.

NERC Settings Our NERC model consists of the base, cased version (`bert-base-cased`) of an English BERT encoder (Devlin et al., 2019) based on the Transformer architecture (Vaswani et al., 2017) with 12 layers, hidden dimensions of 768, 12 heads and 109M parameters in total. On top of BERT, the model has a classifier consisting of a projection layer with the `Mish` activation function (Misra, 2020) and a softmax layer. The model was trained in the train split of the enriched corpus for 20 epochs with early stop of 3 in an annotated subset of the dev split. Learning rate and batch size were set to 1e-5 and 64, respectively.

Given a text to be delexicalized, the model works by first tokenizing it and encoding the tokens in their context-sensitive embedding representations. These embeddings are then fed into the classifier head, which classifies each token. In order to know whether each token is contained within a mention of one of the 8 attributes and where each of these mentions starts and ends in terms of tokens, we used the `IOB2` format, popular in NERC applications (Ramshaw and Marcus, 1995). In total, the model classifies each token according to 17 classes[2], one that indicates whether a token is not a mention and 2 for each one of the attributes, pointing whether a mention starts (`B-`) and the remaining tokens of the mention (`I-`).

4 Explicit Intermediate Representations

Based on the alignments between the meaning representation and the text provided by the delexi-

[2] `O`, `B-FOOD`, `I-FOOD`, `B-NAME`, `I-NAME`, `B-EATTYPE`, `I-EATTYPE`, `B-FAMILYFRIENDLY`, `I-FAMILYFRIENDLY`, `B-AREA`, `I-AREA`, `B-CUSTOMERRATING`, `I-CUSTOMERRATING`, `B-PRICERANGE`, `I-PRICERANGE`, `B-NEAR` and `I-NEAR`

calization process, we can extract several explicit intermediate representations that can help to study several generation phenomenon as well as to build traditional pipeline (rule based or data driven) data-to-text systems, which may generate more adequate texts and to generalize better for new domains (Castro Ferreira et al., 2019).

Similar to Castro Ferreira et al. (2018b), we have enriched the E2E dataset with several intermediate representations about content selection, discourse ordering, text structuring, lexicalization and referring expression generation. These intermediate representations could be used to study each phenomenon as well as to develop a data-driven, pipeline data-to-text system as envisaged by Castro Ferreira et al. (2019).

Content Selection is the task of deciding which information should be verbalized. By comparing the attributes contained in a meaning representation and the presence or absence of their placeholders in the delexicalized template, we are able to automatically extract all the input content for a given verbalisation. In the example of Figure 1 for instance, we can see that the placeholder of the attribute `eatType` (e.g. _EATTYPE_) is not present in the delexicalized template, indicating that it was not selected to be verbalized in the text.

Discourse Ordering is the task of sorting the selected content in the order it should be verbalized. By looking at the order of the placeholders in the delexicalized text, we can extract this order. In Figure 1, looking at the order of the placeholders in the delexicalized template, we see that the sorted list of attributes is: `near`, `area`, `name`, `food`, `priceRange` and `familyFriendly`,

Text Structuring is the task within pipeline data-to-text systems responsible for structuring the outputs of content selection and discourse ordering into paragraphs and sentences. Using Stanza (Qi et al., 2020), we tokenized the sentences of each delexicalized template and considering their placeholders, extracted the sentence plan for each one the attributes verbalized. In Figure 1 for instance, `near`, `area`, `name` were verbalized in the first sentence, whereas `food`, `priceRange` and `familyFriendly` in the second.

Lexicalization aims to find the proper phrases and words to express the content to be included in each sentence. To obtain lexicalization templates similar to the ones used for the neural pipeline system of Castro Ferreira et al. (2019), we again used Stanza in the delexicalized templates to lemmatize determiners and verbs and extract their correct morphological inflection information. Then in these templates, determiners and verbs were replaced by their morphological inflection information and lemmas. For instance, for the delexicalized template depicted in Figure 1, the lexicalization template would be:

> Near _NEAR_ in _AREA_ VP[*Mood=Ind, Number=Sing, Person=3, Tense=Pres, VerbForm=Fin*] be _NAME_ .
> _NAME_ VP[*Mood=Ind, Number=Sing, Person=3, Tense=Pres, VerbForm=Fin*] be DT[*Definite=Ind, PronType=Art*] a _FOOD_ restaurant , VP[*Mood=Ind, Number=Sing, Person=3, Tense=Pres, VerbForm=Fin*] have _PRICERANGE_ prices but VP[*Mood=Ind, Number=Sing, Person=3, Tense=Pres, VerbForm=Fin*] be _FAMILYFRIENDLY_ .

Referring Expression Generation is the task responsible for generating the references to the entities present in the text (Castro Ferreira et al., 2018a). In our case, these entities are the attributes of the meaning representation. Following Castro Ferreira et al. (2018b), we extract the referring expression to the attributes by overlapping an original text and its delexicalized version. In Figure 1, contains examples of extracted references such as *The Wrestlers* and *It* for the `name` attribute value `The Wrestlers` in the meaning representation.

Surface Realization is responsible for taking the last decisions to convert a non-linguistic data into text. In this case, the correct morphological realisation of determiners and verbs as well as detokenizing the text. For this step in specific, we did not extract any kind of information, but refer to the extensive literature that exists on morphological inflection (McCarthy et al., 2019; Vylomova et al., 2020). These tools can be used to correctly realize our extracted lexicalization templates. Moreover, detokenization is a task already solved with high accuracy.

5 Conclusion

This work introduces the enriched version of the E2E dataset (Novikova et al., 2017; Dušek et al., 2018). Together with the enriched version of WebNLG (Castro Ferreira et al., 2018b) and CACAPO van der Lee et al. (2020), this resource will help researchers to carefully investigate particular pipeline processes in data-to-text systems in the levels of Macro- (e.g., Content Selection, Discourse Ordering and Text Structuring), Microplanning (e.g., lexicalization, aggregation and referring expression generation) and Surface Realization. In particular, we will be able to better analyse how such subtasks could obtain better performance when developed using a rule-based approach or a specific/multitask data-driven system. Moreover, in future work the community will be able to better compare pipeline and end-to-end data-to-text systems in terms of generalization as well as fluency and adequacy of their generated texts.

Acknowledgments

This research was partially funded by the Brazilian agencies CNPq, CAPES, and FAPEMIG. In particular, the researchers were supported by CNPQ grant No. 310630/2017-7, CAPES Post doctoral grant No. 88887.508597/2020-00, and FAPEMIG grant APQ-01.461-14. It was also funded by the ADAPT Centre for Digital Content Technology (`https://www.adaptcentre.ie`) which is funded under the Science Foundation Ireland Research Centres Programme (Grant 13/RC/2106_P2).

References

Oshin Agarwal, Heming Ge, Siamak Shakeri, and Rami Al-Rfou. 2021. Knowledge graph based synthetic corpus generation for knowledge-enhanced language model pre-training.

Thiago Castro Ferreira, Claire Gardent, Nikolai Ilinykh, Chris van der Lee, Simon Mille, Diego Moussallem, and Anastasia Shimorina. 2020. The 2020 bilingual, bi-directional WebNLG+ shared task: Overview and evaluation results (WebNLG+ 2020). In *Proceedings of the 3rd International Workshop on Natural Language Generation from the Semantic Web (WebNLG+)*, pages 55–76, Dublin, Ireland (Virtual). Association for Computational Linguistics.

Thiago Castro Ferreira, Chris van der Lee, Emiel van Miltenburg, and Emiel Krahmer. 2019. Neural data-to-text generation: A comparison between pipeline and end-to-end architectures. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 552–562, Hong Kong, China. Association for Computational Linguistics.

Thiago Castro Ferreira, Diego Moussallem, Ákos Kádár, Sander Wubben, and Emiel Krahmer. 2018a. NeuralREG: An end-to-end approach to referring expression generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1959–1969, Melbourne, Australia. Association for Computational Linguistics.

Thiago Castro Ferreira, Diego Moussallem, Emiel Krahmer, and Sander Wubben. 2018b. Enriching the WebNLG corpus. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 171–176, Tilburg University, The Netherlands. Association for Computational Linguistics.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.

Robert Dale. 2020. Natural language generation: The commercial state of the art in 2020. *Natural Language Engineering*, 26(4):481–487.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Ondřej Dušek, Jekaterina Novikova, and Verena Rieser. 2018. Findings of the E2E NLG challenge. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 322–328, Tilburg University, The Netherlands. Association for Computational Linguistics.

Claire Gardent and Shashi Narayan. 2018. Deep learning approaches to text production. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorial Abstracts*, pages 4–9, New Orleans, Louisiana. Association for Computational Linguistics.

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

Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. 2017b. The WebNLG challenge: Generating text from RDF data. In *Proceedings of the 10th International Conference on*

Natural Language Generation, pages 124–133, Santiago de Compostela, Spain. Association for Computational Linguistics.

Albert Gatt and Emiel Krahmer. 2018. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 61:65–170.

Zhijing Jin, Qipeng Guo, Xipeng Qiu, and Zheng Zhang. 2020. GenWiki: A dataset of 1.3 million content-sharing text and graphs for unsupervised graph-to-text generation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2398–2409, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Juraj Juraska, Panagiotis Karagiannis, Kevin Bowden, and Marilyn Walker. 2018. A deep ensemble model with slot alignment for sequence-to-sequence natural language generation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 152–162, New Orleans, Louisiana. Association for Computational Linguistics.

Chris van der Lee, Chris Emmery, Sander Wubben, and Emiel Krahmer. 2020. The CACAPO dataset: A multilingual, multi-domain dataset for neural pipeline and end-to-end data-to-text generation. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 68–79, Dublin, Ireland. Association for Computational Linguistics.

François Mairesse, Milica Gašić, Filip Jurčíček, Simon Keizer, Blaise Thomson, Kai Yu, and Steve Young. 2010. Phrase-based statistical language generation using graphical models and active learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1552–1561, Uppsala, Sweden. Association for Computational Linguistics.

Arya D. McCarthy, Ekaterina Vylomova, Shijie Wu, Chaitanya Malaviya, Lawrence Wolf-Sonkin, Garrett Nicolai, Christo Kirov, Miikka Silfverberg, Sabrina J. Mielke, Jeffrey Heinz, Ryan Cotterell, and Mans Hulden. 2019. The SIGMORPHON 2019 shared task: Morphological analysis in context and cross-lingual transfer for inflection. In *Proceedings of the 16th Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 229–244, Florence, Italy. Association for Computational Linguistics.

Diganta Misra. 2020. Mish: A Self Regularized Non-Monotonic Activation Function. In *31st British Machine Vision Conference 2020, BMVC 2020, Virtual Event, UK, September 7-10, 2020*. BMVA Press.

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, Saarbrücken, Germany. Association for Computational Linguistics.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online. Association for Computational Linguistics.

Lance Ramshaw and Mitch Marcus. 1995. Text chunking using transformation-based learning. In *Third Workshop on Very Large Corpora*.

Ehud Reiter and Anja Belz. 2009. An investigation into the validity of some metrics for automatically evaluating natural language generation systems. *Computational Linguistics*, 35(4):529–558.

Ehud Reiter and Robert Dale. 2000. *Building natural language generation systems*. Cambridge University Press, New York, NY, USA.

Leonardo F. R. Ribeiro, Claire Gardent, and Iryna Gurevych. 2019. Enhancing AMR-to-text generation with dual graph representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3183–3194, Hong Kong, China. Association for Computational Linguistics.

Anna Rohrbach, Lisa Anne Hendricks, Kaylee Burns, Trevor Darrell, and Kate Saenko. 2018. Object hallucination in image captioning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4035–4045, Brussels, Belgium. Association for Computational Linguistics.

André Luiz Rosa Teixeira, João Campos, Rossana Cunha, Thiago Castro Ferreira, Adriana Pagano, and Fabio Cozman. 2020. DaMata: A robot-journalist covering the Brazilian Amazon deforestation. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 103–106, Dublin, Ireland. Association for Computational Linguistics.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Proc. NIPS*, Montreal, CA.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, undefinedukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, page 6000–6010.

Ekaterina Vylomova, Jennifer White, Elizabeth Salesky, Sabrina J. Mielke, Shijie Wu, Edoardo Maria Ponti, Rowan Hall Maudslay, Ran

Zmigrod, Josef Valvoda, Svetlana Toldova, Francis Tyers, Elena Klyachko, Ilya Yegorov, Natalia Krizhanovsky, Paula Czarnowska, Irene Nikkarinen, Andrew Krizhanovsky, Tiago Pimentel, Lucas Torroba Hennigen, Christo Kirov, Garrett Nicolai, Adina Williams, Antonios Anastasopoulos, Hilaria Cruz, Eleanor Chodroff, Ryan Cotterell, Miikka Silfverberg, and Mans Hulden. 2020. SIGMORPHON 2020 shared task 0: Typologically diverse morphological inflection. In *Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 1–39, Online. Association for Computational Linguistics.

Qingyun Wang, Xiaoman Pan, Lifu Huang, Boliang Zhang, Zhiying Jiang, Heng Ji, and Kevin Knight. 2018. Describing a knowledge base. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 10–21, Tilburg University, The Netherlands. Association for Computational Linguistics.

Zhenyi Wang, Xiaoyang Wang, Bang An, Dong Yu, and Changyou Chen. 2020. Towards faithful neural table-to-text generation with content-matching constraints. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1072–1086, Online. Association for Computational Linguistics.

Tsung-Hsien Wen, Milica Gašić, Nikola Mrkšić, Pei-Hao Su, David Vandyke, and Steve Young. 2015. Semantically conditioned LSTM-based natural language generation for spoken dialogue systems. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1711–1721, Lisbon, Portugal. Association for Computational Linguistics.

Sam Wiseman, Stuart Shieber, and Alexander Rush. 2017. Challenges in data-to-document generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2253–2263, Copenhagen, Denmark. Association for Computational Linguistics.

Xinnuo Xu, Ondřej Dušek, Verena Rieser, and Ioannis Konstas. 2021. Agggen: Ordering and aggregating while generating.