

Journal Pre-proof

To What Extent does Content Selection affect Surface Realization in the context of Headline Generation?

Cristina Barros, Marta Vicente, Elena Lloret

PII: S0885-2308(20)30112-1
DOI: <https://doi.org/10.1016/j.csl.2020.101179>
Reference: YCSLA 101179



To appear in: *Computer Speech & Language*

Received date: 13 March 2020
Revised date: 12 September 2020
Accepted date: 19 November 2020

Please cite this article as: Cristina Barros, Marta Vicente, Elena Lloret, To What Extent does Content Selection affect Surface Realization in the context of Headline Generation?, *Computer Speech & Language* (2020), doi: <https://doi.org/10.1016/j.csl.2020.101179>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.

- A NLG approach is analyzed for the task headline generation
- Several content selection strategies are analyzed as macroplanning stage
- An adapted version of HanaNLG is used for the surface realization stage
- Coherent and structured headlines not present in the source news are obtained
- HanaNLG-PLM headlines were among the top preferred in the human evaluation

Journal Pre-proof

To What Extent does Content Selection affect Surface Realization in the context of Headline Generation?

Cristina Barros*, Marta Vicente, Elena Lloret

Department of Software and Computing Systems, University of Alicante, Apdo. de correos 99, E-03080 Alicante, Spain

Abstract

Headline generation is a task where the most important information of a news article is condensed and embodied into a single short sentence. This task is normally addressed by summarization techniques, ideally combining extractive and abstractive methods together with sentence compression or fusion techniques. Although Natural Language Generation (NLG) techniques have not been directly exploited for headline generation, they may provide better mechanisms than summarization techniques to paraphrase the information of a text. Therefore, this paper analyzes and evaluates the effectiveness of NLG techniques for generating headlines. In NLG, both content selection and surface realization are equally important—there is no point in generating text without knowing the topic. Considering this premise, we therefore take HanaNLG—a hybrid surface realization approach—as a basis, and we analyze the effect in the generated text when different content selection strategies are integrated at macroplanning stage. The experiments conducted show that, despite not using any sophisticated summarization method, the proposed approach provided the following benefits: i) it generated a coherent, linguistically structured headline; ii) it obtained results on standard datasets (i.e., DUC 2003 and DUC 2004) that were comparable to several competitive systems, in terms of the content of the generated headline; and, iii) the headlines generated by the whole ap-

*Corresponding author

Email address: cbarros@ua.es (Cristina Barros)

proach (PLM-HanaNLG) were preferred by human assessors compared to those generated by the best performing system in DUC 2003.

Keywords: Natural Language Generation, Headline Generation, Positional Language Models, Factored Language Models, Content Selection, Abstractive Summarization

1. Introduction

An articles headline is one of the most important parts of a piece of news because it represents the main idea or the essence of the article condensed into a sentence or phrase. This fact explains why the headline generation task, whose goal is to automatically construct a headline that describes the content of a news article, is normally addressed as a summarization task.

More specifically, the headline generation task can be addressed from two common summarization strategies: the extractive approach, which identifies the most important sentence in the text and extracts it verbatim, or alternatively, the abstractive approach, which paraphrases the key information from the body of the news.

Producing a headline using an extractive summarization method has its drawbacks and may not be the most suitable approach since selecting as a representative summary a verbatim sentence from the article can lead to ignoring important facts reported from other events included. By contrast, an abstractive summarization approach would be more appropriate as it scans and paraphrases the key information of the text, combining in a single sentence or phrase information present in different sections of the document. Therefore, the latter approach can result in headlines that are more coherent and cohesive, much like professional journalists would do. Given the nature of abstract approaches, our hypothesis in the present work states that this type of approach could significantly benefit from Natural Language Generation techniques to actually create and infer new information, not expressed literally in the document.

Natural Language Generation (NLG) aims to automatically produce text

- guage Generation (ENLG), Association for Computational Linguistics, 57–60, 2015.
- [27] H. Gonçalo Oliveira, A Survey on Intelligent Poetry Generation: Languages, Features, Techniques, Reutilisation and Evaluation, in: Proceedings of the 10th International Conference on Natural Language Generation, Association for Computational Linguistics, 11–20, 2017.
- [28] V. Gupta, G. S. Lehal, Article: Named Entity Recognition for Punjabi Language Text Summarization, *International Journal of Computer Applications* 33 (3) (2011) 28–32.
- [29] Y. Gutiérrez Vázquez, A. Fernández Orquín, A. Montoyo Guijarro, S. Vázquez Pérez, Integración de recursos semánticos basados en WordNet, *Procesamiento del Lenguaje Natural* 47 (2011) 161–168.
- [30] A. Isard, C. Brockmann, J. Oberlander, Individuality and Alignment in Generated Dialogues, in: Proceedings of the INLG, Association for Computational Linguistics, 25–32, 2006.
- [31] A. Isard, J. Knox, Automatic Generation of Student Report Cards, in: Proceedings of the 9th International Natural Language Generation conference, Association for Computational Linguistics, 207–211, 2016.
- [32] R. Kiros, Y. Zhu, R. R. Salakhutdinov, R. Zemel, R. Urtasun, A. Torralba, S. Fidler, Skip-thought vectors, in: *Advances in neural information processing systems*, 3294–3302, 2015.
- [33] A. Lavie, M. J. Denkowski, The METEOR metric for automatic evaluation of machine translation, *Machine translation* 23 (2-3) (2009) 105–115.
- [34] C.-Y. Lin, ROUGE: A Package for Automatic Evaluation of Summaries, in: *Text Summarization Branches Out: Proceedings of the Association for Computational Linguistics Workshop*, 74–81, 2004.

- [35] E. Lloret, L. Plaza, A. Aker, The challenging task of summary evaluation: an overview, *Lang. Resour. Evaluation* 52 (1) (2018) 101–148.
- 960 [36] I. Macdonald, A. Siddharthan, Summarising News Stories for Children, in: *Proceedings of the 9th International Natural Language Generation conference*, Association for Computational Linguistics, 1–10, 2016.
- [37] S. Narayan, C. Gardent, Unsupervised Sentence Simplification Using Deep Semantics, in: *Proceedings of the 9th International Natural Language Generation conference*, Association for Computational Linguistics, 111–120,
965 2016.
- [38] A. Nenkova, L. Vanderwende, K. McKeown, A Compositional Context Sensitive Multi-document Summarizer: Exploring the Factors That Influence Summarization, in: *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*,
970 *SIGIR '06*, ACM, New York, NY, USA, ISBN 1-59593-369-7, 573–580, 2006.
- [39] P. Over, H. Dang, D. Harman, DUC in Context, *Inf. Process. Manage.* 43 (6) (2007) 1506–1520, ISSN 0306-4573.
- [40] L. Padró, E. Stanilovsky, FreeLing 3.0: Towards Wider Multilinguality, in: *Proceedings of the Eight International Conference on Language Resources and Evaluation*,
975 2473–2479, 2012.
- [41] K. Papineni, S. Roukos, T. Ward, W.-J. Zhu, Bleu: a Method for Automatic Evaluation of Machine Translation, in: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Philadelphia, Pennsylvania, USA,
980 311–318, URL <https://www.aclweb.org/anthology/P02-1040> , 2002.
- [42] S. Pita Fernández, Determinación del tamaño muestral, *CAD ATEN PRIMARIA* 1996 3 (1996) 138–14.
- [43] H. Pottker, News and its communicative quality: the inverted pyramid when and why did it appear?, *Journalism Studies* 4 (2003) 501–511.
985

- [44] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, Language Models are Unsupervised Multitask Learners .
- [45] R. Řehůřek, P. Sojka, Software Framework for Topic Modelling with Large Corpora, in: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, ELRA, Valletta, Malta, 45–50, 2010. 990
- [46] E. Reiter, A. Belz, An Investigation into the Validity of Some Metrics for Automatically Evaluating Natural Language Generation Systems, *Comput. Linguist.* 35 (4) (2009) 529–558.
- [47] E. Reiter, R. Dale, Building Natural Language Generation Systems, Cambridge University Press, 2000. 995
- [48] V. Rus, M. Lintean, A Comparison of Greedy and Optimal Assessment of Natural Language Student Input Using Word-to-Word Similarity Metrics, in: Proceedings of the Seventh Workshop on Building Educational Applications Using NLP, Association for Computational Linguistics, Montréal, Canada, 157–162, URL <https://www.aclweb.org/anthology/W12-2018> , 2012. 1000
- [49] A. M. Rush, S. Chopra, J. Weston, A Neural Attention Model for Abstractive Sentence Summarization, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 379–389, 2015. 1005
- [50] K. K. Schuler, Verbnet: A Broad-coverage, Comprehensive Verb Lexicon, Ph.D. thesis, 2005.
- [51] S. Sharma, L. E. Asri, H. Schulz, J. Zumer, Relevance of unsupervised metrics in task-oriented dialogue for evaluating natural language generation, arXiv preprint arXiv:1706.09799 . 1010
- [52] A. Stolcke, SRILM - An Extensible Language Modeling Toolkit, in: Proceedings International Conference on Spoken Language Processing, vol 2., 901–904, 2002.

- [53] R. Sun, Y. Zhang, M. Zhang, D. Ji, Event-Driven Headline Generation, in:
1015 Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, 462–472, 2015.
- [54] S. Takase, J. Suzuki, N. Okazaki, T. Hirao, M. Nagata, Neural Headline
1020 Generation on Abstract Meaning Representation, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 1054–1059, 2016.
- [55] J. Tan, X. Wan, J. Xiao, From Neural Sentence Summarization to Headline
1025 Generation: A Coarse-to-fine Approach, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI'17, AAAI Press, 4109–4115, 2017.
- [56] M. Vicente, C. Barros, E. Lloret, Statistical language modelling for automatic story generation, *Journal of Intelligent & Fuzzy Systems* 34 (5) (2018) 3069–3079.
- 1030 [57] M. Vicente, E. Lloret, Analysing Positional Language Models for Natural Language Generation, in: Proceedings of the 8th Language and Technology Conference, 357–361, 2017.
- [58] D. Zajic, B. J. Dorr, R. Schwartz, BBN/UMD at DUC-2004: Topiary, in:
1035 Proceedings of the North American Chapter of the Association for Computational Linguistics Workshop on Document Understanding, Association for Computational Linguistics, 112–119, 2004.
- [59] H. Zhang, H. Xu, S. Bai, B. Wang, X. Cheng, Experiments in TREC 2004 Novelty Track at CAS-ICT, in: Proceedings of the 13th Text Retrieval Conference (TREC), 2004.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof