How can Paidiom improve the neural machine translation of idioms?

Carlos Manuel Hidalgo-Ternero

Universidad de Málaga
IUITLM
Malaga, Spain
cmhidalgo@uma.es

Short abstract

In this paper we present research results with Paidiom, a text-preprocessing algorithm designed for 1) converting discontinuous multiword expressions (MWEs) into their continuous forms and 2) translemmatising them, i.e., converting source-text MWEs into their target-text equivalents, in order to improve the performance of current neural machine translation (NMT) systems. To test its effectiveness, an experiment with the NMT systems of VIP, Google Translate and DeepL has been carried out in the ES>EN translation direction with Verb-Noun Idiomatic Constructions (VNICs) in Spanish. The performance of Paidiom was compared to both the one of our previous algorithm (gApp) and to the manual conversion (our gold standard). In this regard, the promising results yielded by this study, the first one analysing Paidiom's performance, will shed some light on new avenues for enhancing MWE-aware NMT systems.

Extended abstract

The recent emergence of neural networks in machine translation has represented a real breakthrough, bringing forth Neural Machine Translation (NMT), which has resulted in a considerable qualitative leap compared to previous ruled-based and statistical models (Bentivogli et al., 2016; Junczys-Dowmunt et al., 2016; Wang et al., 2022).

Despite these advances, NMT systems still have an important Achilles' heel: multiword expressions (MWEs). Besides their quintessential problematic features such as syntactic anomaly, non-compositionality, diasystematic variation and ambiguity, among others, a further challenge arises for NMT: MWEs do not always consist of adjacent tokens (e.g., *You need to keep those things in mind.*), which seriously hinders their automatic detection and translation (Constant et al., 2017; Corpas Pastor, 2013; Monti et al., 2018; Ramisch & Villavicencio, 2018; Rohanian et al., 2019). To overcome the challenges that discontinuous MWEs still pose for even the most robust NMT systems (cf. Colson, 2019; Zaninello & Birch, 2020), we have designed an upgraded algorithm, called Paidiom[1], which is able not only to automatically convert discontinuous MWEs into their continuous form (analogously to our previous algorithm gApp[2]) but also to translemmatise them, i.e., to directly convert MWEs into their target-text equivalents in order to improve NMT.

To test Paidiom's effectiveness, the performance of VIP (*Voice-text integrated system for interpreters*, cf. Corpas Pastor, 2021), Google Translate and DeepL's NMT systems was examined against a total of 400 cases, comprising 100 discontinuous forms (i.e., the original texts), 100 continuous forms after gApp's conversion, 100 continuous

[1] A detailed explanation of Paidiom is available through this link: https://lexytrad.es/paidiom
[2] gApp is available through this link: http://lexytrad.es/gapp/app.php.

and translemmatised forms after Paidiom's conversion, and 100 continuous and translemmatised forms after the manual conversion (our gold standard) of Spanish Verb-Noun Idiomatic Constructions (VNICs), in the ES>EN translation direction.

Regarding the employed evaluation method, the NMTs' outputs for these different scenarios were manually assessed through an MT evaluation based on directly expressed judgements (*DEJ-based evaluation method*, cf. Chatzikoumi 2020). In this regard, for our study three professional ES/EN translators, with between 4 and 10 years of experience, were selected as annotators to directly express judgement on the translation quality of the different MT outputs employing a binary scale: 1 (good) or 0 (bad). After they submitted their judgement, final decisions on the acceptability (or not) of each specific target text were taken on a majority basis: for instance, if 2 or 3 of the translators had judged an MT output as *good*, then this output was also finally categorised as *good* for our study, and viceversa.

Against such a background, for this test set, the present experiment yielded promising global results that, on the one hand, go in line with our previous experiments with gApp (Hidalgo-Ternero, 2021 and 2023; Hidalgo-Ternero & Corpas Pastor, 2020, 2023a and 2023b; Hidalgo-Ternero, Lista & Corpas Pastor, 2023, and Hidalgo-Ternero & Zhou-Lian, 2022 and 2023)¸ along which we proved that NMT of discontinuous MWEs can overall be improved by converting them into their continuous form: in this experiment, NMT systems achieved an 8.3% accuracy in the discontinuous form vs. 13% in the continuous form after gApp, i.e., an enhancement by 4.7%. On the other hand, global results also confirmed our initial hypothesis: NMT systems can deliver a considerably better performance if MWEs are not only converted into their continuous form but also translemmatised prior to NMT. In the light of these results, MWEs' conversion into their continuous form plus their translemmatisation with Paidiom led NMT systems to achieve an overall 91.7% accuracy, attaining an analogous performance to the gold standard (92.7%). When contrasted with the original (discontinuous) version, Paidiom could, on average, improve NMT by 83.4% (84.4% with the gold standard).

In this regard, the promising results yielded by the present study invite us to continue evaluating Paidiom in further experiments in order to determine to what extent it can also improve NMT performance for other idioms, as well as for additional MWE typologies (collocations, verb-particle constructions, etc.), and in other translation directions. In addition, the present study can also constitute the basis for further research to assess the scalation of this model to other language-dependent text-preprocessing systems for the automatic translemmatisation of MWEs, with the purpose of enhancing MWE-aware NMT systems.

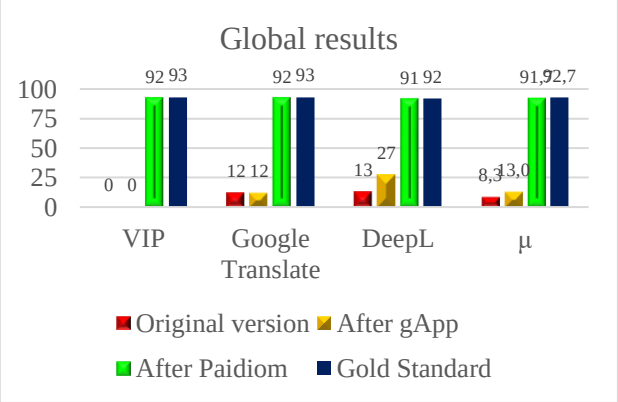

Figure 1. Global results

References

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2018). Neural versus phrase-based machine translation quality: a case study. *arXiv*

Chatzikoumi, E. (2020). How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering*, *26*(2), 137-161. doi:10.1017/S1351324919000469

Colson, J.-P. (2019). Multi-word Units in Machine Translation: why the Tip of the Iceberg Remains Problematic – and a Tentative Corpus-driven Solution. *MUMTT2019*.

Constant, M., Eryiğit, G., Monti, J., van der Plas, L., Ramisch, C., Rosner, M., & Todirascu, A. (2017). Multiword expression processing: A survey. *Computational Linguistics*, *43*(4), 1–92.

Corpas Pastor, G. (2013). Detección, descripción y contraste de las unidades fraseológicas mediante tecnologías lingüísticas. In I. Olza & E. Manero (Eds.) *Fraseopragmática. Colección Romanistik* (pp. 335-373). Frank & Timme.

Corpas Pastor, G. (2021). Technology Solutions for Interpreters: The VIP System. *Hermēneus. Revista de Traducción e Interpretación*, *23*, 91-123.

Hidalgo-Ternero, C. M. (2021). El algoritmo *ReGap* para la mejora de la traducción automática neuronal de expresiones pluriverbales discontinuas (FR>EN/ES). In G. Corpas Pastor, M. R. Bautista Zambrana & C. M. Hidalgo-Ternero (Eds.), *Sistemas fraseológicos en contraste: enfoques computacionales y de corpus* (pp. 253-270). Comares.

Hidalgo-Ternero C. M. (2023/forthcoming). A la cabeza de la traducción automática neuronal asistida por *gApp*: somatismos en VIP, DeepL y Google Translate. In G. Corpas Pastor y M. Seghiri (Eds.), *Aplicaciones didácticas de las tecnologías de la interpretación*. Comares.

Hidalgo-Ternero, C. M., & Corpas Pastor, G. (2020). Bridging the '*gApp*': improving neural machine translation systems for multiword expression detection. *Yearbook of Phraseology*, *11*, 61-80. https://doi.org/10.1515/phras-2020-0005

Hidalgo-Ternero C. M., & Corpas Pastor, G. (2023a/forthcoming). Qué se traerá *gApp* entre manos… O cómo mejorar la traducción automática neuronal de variantes somáticas (ES>EN/DE/FR/IT/PT). In Seghiri, M. & Pérez Carrasco, M. (Eds.). *Aproximación a la traducción especializada*. Peter Lang.

Hidalgo-Ternero C. M., & Corpas Pastor, G. (2023b/forthcoming). ReGap: a text preprocessing algorithm to enhance MWE-aware neural machine translation systems. In J. Monti, G. Corpas Pastor y R. Mitkov (Eds.), *Recent Advances in MWU in Machine Translation and Translation technology*. John Benjamins Publishing Company.

Hidalgo-Ternero C. M., Lista, F. & Corpas Pastor, G. (2023/under review). *gApp*-assisted NMT: how to improve the neural machine translation of discontinuous multiword expressions (IT>EN/DE). *Language Resources and Evaluation*.

Hidalgo-Ternero, C. M., & Zhou-Lian, X. (2022). Reassessing gApp: does MWE discontinuity always pose a challenge to Neural Machine Translation? In G. Corpas Pastor y R. Mitkov (eds.), *Computational and Corpus-Based Phraseology* (pp. 116–132). Springer.

Hidalgo-Ternero, C. M., & Zhou-Lian, X. (2023/under review). Minding the gApp in the ES>EN/ZH neural machine translation of discontinuous multiword expressions. *Natural Language Engineering*.

Junczys-Dowmunt, M., Dwojak, T. & Hoang, H. (2016). Is neural machine translation ready for deployment? A case study on 30 translation directions. *Arxiv*.

Monti, J., Seretan, V., Corpas Pastor, G., & Mitkov, R. (2018). Multiword units in machine translation and technology. In R. Mitkov, J. Monti, G. Corpas Pastor & V. Seretan (Eds.), *Multiword Units in Translation and Translation Technology* (pp. 1-37). John Benjamins.

Ramisch, C., & Villavicencio, A. (2018). Computational treatment of multiword expressions. In R. Mitkov (Ed.), *Oxford Handbook on Computational Linguistics* (2ª ed).

Rohanian, O., Taslimipoor, S., Kouchaki, S., An Ha, L., & Mitkov, R. (2019). Bridging the Gap: Attending to Discontinuity in Identification of Multiword Expressions. In J. Burstein, C. Doran, & T. Solorio (Eds.), *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 1*, 2692–2698.

Wang, H., Wu, H. He, Z., Huang, L. & Church, K. W. (2022) Progress in Machine Translation, *18*, 143-153. *Engineering*.

Zaninello, A., & Birch, A. (2020). Multiword expression aware neural machine translation. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara. B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis (Eds.), *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 3816–3825.