

Paraphrases corpus as a user-centred approach to support translators' decision-making.

Rossana Cunha

Adriana Pagano

Fabio Alves

Federal University of Minas Gerais

Recent studies (Kenny, 2017; O'Brien et al., 2017; Castilho et al., 2017; Castilho and O'Brien, 2018) have provided evidence that suggests a growing interest in combining methods from human-computer interaction (HCI) and corpus-based translation studies. In line with this growing trend, we turn to an HCI resource to model language according to distinct language use patterns drawing on the concept of user personas (Nielsen, 2011), i.e., data-based user group models, built to guide the production of texts that can be understood by real users in various computational applications. From this perspective, our study presents preliminary results of compilation and alignment of a Brazilian monolingual corpus (Hewavitharana; Vogel, 2016; Cardon; Grabar, 2018; Mallinson et al., 2018) based on personas. We developed the personas from representations of three groups of real users -- (i) doctors, (ii) nurses and healthcare assistants, and (iii) patients. First, text parameters were defined for each persona based on Systemic Functional Theory (Halliday and Matthiessen, 2014) and texts compiled related to a common domain: diabetes self-management: (i) text retrieved from social networks in the case of patient/lay readers; (ii) text compiled from technical manuals in the case of nurses and healthcare assistants; (iii) text sampled from academic papers in the case of doctors/expert readers. We then abstracted grammar and lexical features to build a model of paraphrase detection and semantic similarity. Finally, in order to assess the comparability of the aligned instances, an expert evaluation was conducted by a Committee of domain experts. The result of this pilot study is a monolingual parallel corpus aligned with expert, specialist and lay discourse on the diabetes mellitus self-care domain. This corpus will later be used as a semantic resource (Sennrich et al., 2016; Koehn, 2018) in the architecture of a user-centred translation model.

Keywords: Translator education, human-computer interaction, user-centred design, in-domain translation, persona.

References:

- Cardon, Rémi, and Natalia Grabar. "Identification of Parallel Sentences in Comparable Monolingual Corpora from Different Registers." *Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*. 2018.
- Castilho, Sheila, et al. "Is neural machine translation the new state of the art?" *The Prague Bulletin of Mathematical Linguistics* 108.1 (2017): 109-120.
- Castilho, Sheila, and Sharon O'Brien. "Acceptability of machine-translated content: A multi-language evaluation by translators and end-users." *Linguistica Antverpiensia, New Series—Themes in Translation Studies* 16 (2018).
- Creutz, Mathias. "Open subtitles paraphrase corpus for six languages." *arXiv preprint arXiv:1809.06142* (2018).
- Halliday, Michael Alexander Kirkwood, and Christian MIM Matthiessen. *Halliday's introduction to functional grammar*. Routledge, 2014.
- Hewavitharana, Sanjika, and Stephan Vogel. "Extracting parallel phrases from comparable data for machine translation." *Natural Language Engineering* 22.4 (2016): 549-573.
- Kenny, Dorothy. "Human issues in translation technologies." *IATIS Yearbook 2015*. London and New York: Routledge. 2017.
- Koskinen, Kaisa and Minna Ruokonen. "Love letters or hate mail? Translators' technology acceptance in the light of their emotional narratives." *Human issues in translation technology*. Routledge, 2017. 26-42.
- Mallinson, Jonathan, Rico Sennrich, and Mirella Lapata. "Paraphrasing revisited with neural machine translation." *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. 2018.
- Nielsen, Lene. "Personas in co-creation and co-design." Proceedings of the 11th Human-Computer Interaction Research Symposium. 2011.
- O'Brien, Sharon, et al. "Irritating CAT tool features that matter to translators." *Hermes Journal of Language and Communication in Business* 56 (2017): 145-162.
- Suojanen, Tytti, Kaisa Koskinen, and Tiina Tuominen. *User-centered translation*. Routledge, 2015a.
- Suojanen, Tytti, Kaisa Koskinen, and Tiina Tuominen. "Usability as a focus of multiprofessional collaboration: a teaching case study on user-centered translation." *Connexions: international professional communication journal* 3.2 (2015b): 147-166.