

and translation memories became mature enough to be exploited for professional translation, and henceforth translators became post-translation editors.

An important lesson from the translation business concerning potential danger for the future of lexicography is that the transition to post-editing translation was by far not easy, partially because it may have actually begun too soon (pushed by translation and localization agencies pressing to cut costs), having machine translation do yet another “over-promised and under-delivered” U-turn.

Eventually, this adoption has further progressed as the technology became more mature, and, mainly, as the translation environment for professionals has become more suited for the task of post-editing, which is very different from translating from scratch.

However, the episode had a very undesired consequence that we want to avoid in lexicography. Translators were abandoning machine translation, for both technological reasons as well as for fear of becoming jobless. These fears remain, even though the former is being improved and the latter just did not prove to be the case: the translation industry is growing and some reports describe it as one of the fastest growing businesses today.<sup>4</sup> Moreover, as the whole process gets streamlined, there are good chances that lower per-word income of translators will eventually turn into a higher per-hour rate for them.

The lessons for lexicography are straightforward: the transition to post-editing must be backed by solid technology (which we believe we have), revisited workflows (which we need to work on), and with advocacy explaining that it is not meant to steal lexicographic jobs. The shift to *post-editing lexicography* might well be a fertilizer for the falling industry, showing faster (and hence, more affordable) and more effective workflows.

A specific account comes with addressing less-resourced languages – or those that are basically not resourced at all at the moment. There are plenty of them, often geographically located in areas with growing numbers of speakers. Many of these speakers live in poverty, which nevertheless does include the possession of a smartphone. Language resources will be one of the first data needed when these societies will approach information levels of the developed world. There will be a strong business need for them but no time for twenty-years-running lexicographic projects, another reason why human efforts

need to be supplemented by automation as far as possible.

At the eLex 2017 conference we are going to present a proof of concept in the form of interconnecting Sketch Engine, a leading corpus query system, with Lexonomy, a new lightweight dictionary writing system.<sup>5</sup> We will show how a dictionary draft can be directly obtained from a reference corpus (as a One-Click Dictionary) in Sketch Engine and how it can be efficiently post-edited in Lexonomy.

The future of lexicography presents big challenges. It would be naïve not to realize that many of them pose real issues, problems and obstacles for all players in the field. However, the more so we need to look for those of them that present real opportunities – and, I believe, *post-editing lexicography* is one of them.

- 
- 5 Jakubíček M., Kovář V., Měchura M. and Rychlý P. One-Click Dictionary. In *Electronic lexicography in the 21st century, Proceedings of eLex 2017* (forthcoming).

**Sketch Engine** (SkE) started in 2004 as an academic and lexicographic product for corpus query and management and has since attracted a wide audience including translators, writers, marketers, brand naming and SEO professionals. To meet the challenge of guiding this variety of users to the functionality they need in a streamlined way, an all-new user interface is under development to not only bring in the latest Web technology but also change the way users interact with SkE. With a soft launch due in autumn 2017, users will enjoy a new friendly design which adapts to small touch screens of tablets and mobile phones. Input forms and selection screens will offer basic and advanced layouts, the former targeted at casual users without a profound knowledge of corpora or NLP and the latter serving academic and professional users. In this process, various controls have shifted to more intuitive spots, enabling the user to, for example, adjust the view on the result screen rather than make this decision beforehand as it is now, while preserving all the options and features that are currently available. Hand in hand with developing the new look and feel, SkE will become more useful to anyone in need of glossaries or dictionaries. In addition to serving as an indispensable tool for gathering data, the new link with **Lexonomy** system will enable data conversion into a lexicographic product as part of an online dictionary writing tool, which doubles as a hosting service that can produce a dictionary and have it published online instantly. Lexonomy also features an easy-to-use XML editor suitable for users with no prior knowledge to create lexicographic products complying with current standards. Embedding Lexonomy in SkE will become vital for starting brand new lexicographic projects. Users will access a corpus to identify the most frequent words and have the list pushed to Lexonomy along with part of speech tags, usage flags, example sentences, collocations, synonyms, definitions or translation, thus generating a dictionary draft for post-editing. Likewise, a subject-specific glossary can be developed analogically from a terminology list extracted from a domain corpus. This push & pull model will dramatically change the way dictionaries are built, besides its beneficial time and money saving implications.

**Ondřej Matuška**  
Lexical Computing

---

4 See, e.g., <https://gala-global.org/industry/industry-facts-and-data>.