

Business Diversification?

I love the way some languages and cultures count. What speakers of European languages call “twenty” is rendered in the Papua New Guinea language of Mairasi—and many other languages—as “one person” (with all 20 fingers and toes combined). As someone who is not naturally mathematically inclined, my early calculating life would have been so much easier with access to such vivid practical images.

Though that is now water under the bridge, I still think I must have talked about TAUS, the Translation Automation User Society¹, at least *one person* times, so I’m not going to go into a long explanation of what TAUS is and what it does. A short explanation might be: TAUS is interested in helping its members, typically large translation buyers and large language services providers, to employ machine translation (MT) more successfully by offering the services of a think-tank and exchange forum for that particular sector of the translation world, and by exploring ways to optimize MT usage. You won’t be surprised to hear that not everyone loves everything TAUS stands for. But while I often disagree with its positions, I have appreciated engaging in dialogue with its team, including accepting invitations to participate in TAUS events, such as “Reinventing the Translation Industry,” a virtual conference held in June.²

I have also greatly appreciated that for years now TAUS has freely supplied translators with one of the best terminology tools as part of its Data Cloud ecosystem.³ And this is exactly where this story starts (but hopefully does not end).

A few weeks ago, many of us received a notification that the complete TAUS data ecosystem (the TAUS Data Cloud) was going to transition to a new platform and system (the TAUS Data Marketplace). As part of that process, the above-mentioned Data Search would be retired by October 31.



The TAUS team understood that my pleas were heartfelt, not simply as a reflection of my personal desires but also representative of many of my colleagues as well. They agreed to think about it again.

I reached out to TAUS Chief Executive Officer Jaap van der Meer and his team to find out more about the transition and to lobby for a reversal of that decision, at least concerning Data Search. The TAUS team understood that my pleas were heartfelt, not simply as a reflection of my personal desires but also representative of many of my colleagues as well. They agreed to think about it again. If they agree, they would offer continued access in the form of a legacy system, meaning it wouldn’t be updated with new data but could still be accessed at the old, or a similar, location. Let’s hope they do that.

But that’s not all we talked about. Jaap and his team also gave me an introduction to their new system and asked whether this is potentially

something interesting for translators as well.

So far, TAUS has offered credits for bilingual data that they receive from anyone, including translators, language services providers, and translation buyers. In exchange for those credits, one could download data for one’s own purposes. TAUS found that while this kind of offer might be interesting for some companies, it was, by and large, irrelevant for translators. The team hopes that this new system—a true marketplace where anyone can offer data and actually be paid any time someone else purchases that data—might be more relevant for a wider variety of stakeholders.

Let’s back up for one second, though. This whole system—which, by the

This column has two goals: to inform the community about technological advances and encourage the use and appreciation of technology among translation professionals.



way, is partly funded by the European Union—is based on legal assumptions described in a white paper that TAUS recently published in cooperation with a legal and consultancy firm.⁴ The central sentiment of the white paper might be summarized in this statement: “But, at the end of the day when the lawyers have gone home, we as professionals in the translation industry have to use our own common sense and do what’s right. We have to ask ourselves very practical questions and follow a set of simple rules to reduce regulatory risk and enhance our compliance.”

The paper examines the legal situation according to laws, how jurisdictions have responded to the use and sharing of language data and translation data (the term “translation data” is used to refer to metadata within translation memories), and actual practice across the board. While this might not appear relevant to everyone, if your clients are particularly concerned about privacy and the use of their data, it seems to be a reasonable approach.

TAUS also discovered that it’s often not particularly helpful to look only at “translation data” (i.e., the data that describes the translated bilingual

language data if you want to buy it for a certain purpose). Instead, the team developed algorithms to look more deeply into the existing language data and filter out what’s useful for a particular purpose.

So, while it will still be possible to buy bilingual data that matches certain criteria, such as “English>Romanian software localization data,” it will also be possible to have the tool look through the entire English>Romanian corpus and filter out the segments it identifies as helpful for a certain set of documents it was supplied with to find that data. Any provider of the data would be paid based on the number of segments used from their own contribution. The data will be used almost exclusively for MT training. (It could, of course, be used for translation memory purposes as well, but TAUS’s experience has taught it that it’s unlikely).


Also, any data that’s offered for purchase is cleaned. For instance, this would include the deletion of source duplicates with different targets, reflecting different stages of editing (recognize that problem??), as well as large amounts of tags being deleted, the reduction of obvious erroneous entries,

and so on. This cleaned translation memory will then be offered on the marketplace, but it’s also likely to be handed back to the original data provider as an added incentive.

When Jaap asked me whether this would be an interesting proposition for translators, I gave an answer I know many of you won’t be happy with. I said that six months ago, I wouldn’t have necessarily thought so. But now, in the midst of the crisis? Maybe.

Clearly, there has been a lot of talk about diversification among translators. There has been the realization that while specialization is really important, it might just as be important to have more than one (so you won’t be completely without work overnight if your specialization doesn’t meet the needs of a time like we’re in right now). But there’s also been a lot of talk about diversification beyond that.

I have always encouraged everyone to have some kind of professional offering beyond “just” translation. Not so much for business reasons, mind you, but more for reasons of sanity. For many of us, of course, the “business” reason now stands in the foreground.

Might data trading be an additional business for some of us? *You* tell me. TAUS will be able to add to that conversation in the months that follow the unveiling of its new system in October. 

NOTES

1. Translation Automation User Society, www.taus.net.
2. “Reinventing the Translation Industry,” <http://bit.ly/TAUS-2020>.
3. TAUS Data Cloud ecosystem, <https://data-app.taus.net>.
4. “Who Owns My Language Data?” (Translation Automation User Society), <http://bit.ly/TAUS-white-paper>.



Jost Zetzsche, CT is chair of ATA’s Translation and Interpreting Resources Committee. He is the author of *Translation Matters*, a collection of 81 essays about translators and translation technology. Contact: jzetzsche@internationalwriters.com.