

Online Translation Pricing Issues

Claire Larssonneur



Claire Larssonneur
University Paris 8
Claire.larssonneur@univ-
paris8.fr;
ORCID:
[0000-0002-5129-5844](https://orcid.org/0000-0002-5129-5844)

Abstract

Digital technologies such as translation platforms, crowdsourcing and neural machine translation disrupt the economics of translation. Benchmarking the pricing policies of nine global language services firms uncovers a shift towards online business models that contribute to reshaping the traditional volume-based content-oriented model of translation towards a range of linguistic services focused on user experience.

Keywords: online translation, economics, pricing, freemium, language services, user experience, content.

Resum

Les tecnologies digitals, com les plataformes de traducció, el crowdsourcing i la traducció automàtica neuronal han creat disrupció en l'economia de la traducció. L'anàlisi de les polítiques de preus de nou empreses de serveis lingüístics d'abast mundial revela un canvi cap a models en línia que contribueixen a reformar el model de traducció tradicional basat en el volum i orientat al contingut per oferir un catàleg de serveis lingüístics centrat en l'experiència d'usuari.

Paraules clau: Traducció en línia; economia; tarifació; serveis de semipagament; serveis lingüístics; experiència d'usuari; contingut.

Resumen

Las tecnologías digitales como las plataformas de traducción, el crowdsourcing y la traducción automática neuronal disrupten la economía de la traducción. El análisis de las políticas de precios de nueve empresas de servicios lingüísticos de envergadura mundial muestra un cambio hacia modelos de negocio en línea que contribuyen a reformar el modelo de traducción basado en el volumen y orientado al contenido para ofrecer un catálogo de servicios lingüísticos centrado en la experiencia de usuario.

Palabras clave: Traducción en línea; economía; tarificación; servicios de semipago; servicios lingüísticos; experiencia de usuario; contenido.



1. Introduction

The sustained growth experienced by the global translation market over the last twenty years is staggering: Marczak (2018) lists annual growth rates ranging from 6 to 10% over the 2008-2015 period, with revenues estimated at US\$ 43 billion in 2017 (DePalma et al., 2017). This is consistent with the increasing mobility and multicultural development of our societies at large, but also with technological advances. Vashee (2013:126) points out that much of this growth is due to our overwhelming reliance on digital means of communication: these have led to new forms of expression (tweets, posts) and new services (web and mobile localization, multilingual data mining and consulting), expanding the range of translation-related activities. Beyond facilitating access to content, the web is also a market place and a production factory. The translation industry is thus possibly facing disruption on several fronts: aside from sheer growth and new forms of content, there is the advent of neural machine translation (NMT), which is deemed to yield far better results than statistical machine translation (Castilho et al., 2017). On the production side and from a market perspective, there are blurred lines between users, producers and prosumers, for example in fansubbing or translation crowdsourcing.

These trends plead for an extended conception of disruption: indeed, Clayton Christensen's approach (1997), focused on technologies that impact the lower end of the market, may apply to free NMT. However, I prefer a broader definition of disruption as a change of perspective brought upon by a reversal of priorities, expectations or practices (Dru, 2016). Following Hayek's intuition that price is valuable "not only to balance supply and demand, but also to transmit critical information throughout the economy in a remarkably parsimonious way" (cited in McAfee and Brynjolfsson, 2017: 237), I view online pricing strategies as one indicator of a possible disruption of the translation industry.

And so I focused during February 2018 on the benchmarking of online translation and language services offered by nine major firms, collecting information from their websites and presented here in alphabetical order: (in an alphabetical order, at the website links where not otherwise specified in the text) Alphatrad (www.alphatrad.com), Capita (www.capitatrading.com/translation-services/), Datawords (<https://datawords.fr/>), Lionbridge (www.lionbridge.com/en-gb), Omnicien (<https://omnicien.com>), SDL (www.sdl.com), STAR (<https://www.star-group.net/en/home.html>), Systran (www.systran.fr, www.systransoft.com) and Textmaster (<https://eu.textmaster.com/>). Some of those companies, such as Systran and Omnicien, focus on providing software solutions; others like STAR and SDL provide both software solutions and translation services. Datawords, Capita and Alphatrad position themselves on a range of language and marketing services. Finally, Lionbridge and Textmaster rely on translation crowdsourcing. The sample was chosen to represent the variety of online commercial offers by large companies branding themselves as global, which represent only a section of the translation market but the most visible and in all likelihood early adopters of new business models.

2. Translation business models are evolving online

In the traditional translation business model, which is still widely used by small agencies and freelance translators, pricing is based on volume as measured either by the number of words, characters or standardized page units. Most customers expect a reliable and accurate translation and most transactions are stand-alone affairs, operating within the frame of a dedicated quote. Even though many online transactions follow this model of one-off services, the technical and commercial possibilities of the internet have enabled the emergence of specific online business models, some of which may apply to translation. McAfee and Brynjolfsson point out most online transactions share a few common traits. Digital goods are described as “free, perfect, instant” (2017:137): they seemingly cost near to nothing to distribute, all copies are absolutely identical and they can be delivered in one click. Music tracks, newspaper articles or free online translated content are all examples of those digital goods. Google Translate’s 500 million users have indeed become accustomed to getting online translations that are free and instant, though usually to this date less satisfying than a human translation (Turovsky, 2016). By contrast traditional translation is far from instant, comes at a price and its quest for perfection is elusive.

How the product is delivered also matters. Since changing the format and packaging of digital goods is extremely easy, a number of online e-commerce strategies rely on “unbundling and rebundling” (McAfee & Brynjolfsson, 2017: 144-145). In the music industry, unbundling corresponded to the shift from albums on CDs to single tracks sold on platforms; rebundling occurred when customers started buying lists (Spotify) or subscribing to monthly offers (Apple Music). The same process affects translation, at least in global firms. For example, crowdsourced translation is based on the principle of unbundling content into a multitude of smaller units. On the other hand, when SDL advertises their all-inclusive “global content operation model”, or when Systran offers both to customize NMT software and to host it on their server, they rebundle.

Finally, online business models combine direct and indirect sources of profit: data generated by online user behaviour and the consumption of digital goods are extremely valuable for marketing purposes, to the extent that free online services have been called “fake free goods” (Paillard, 2013: 124, my translation). VoiceTra, a Japanese voice recognition and translation application, can be downloaded free of charge but its privacy disclaimer states (2015): “Speech and text data input into ‘VoiceTra,’ their translation results, terminal information (including unique ID), communication history, application operation history, and user location information will be acquired. The acquired information will be used for research on speech translation as well as for commercial purposes.” Linguistic data are thus extremely valuable to improve the automated processing of language. Amazon’s Mechanical Turk platform details among its promotional case studies the translation of a very large corpus of social media from Arabic to English for the explicit purpose of training the machine translation system of DARPA. They further claim that “DARPA was able to translate 1.5 million words in two months, for a cost of around \$0.03 per word, one tenth the cost of using professional translators” (Mechanical Turk, n.d.).

The specificities of online commerce have one paradoxically conspicuous effect on the economy of translation: its price range has become largely opaque and polarized. The only company in the sample that publicized its rates catered to the lower-end of the market: Textmaster starts at €0.006 per word for gist translation, to be contrasted with a mean of €0.15 per word on the French/English pair recorded by the Société Française des Traducteurs in its 2015 survey of the translation market (SFT, 2015: 37-38). On the other hand, the NMT solution sold by Systran to the French Autorité de Contrôle Prudentiel et de Résolution (a compliance institution) amounted to circa €200,000 (Fournier-Outters, 2018). The scarcity of information and the wide discrepancies of prices threaten trade transparency: one could also read it as symptomatic of a profound upheaval of the market.

3. The specificities of digital markets: freemium, price modulation and fragmentation

Bomsel (2010: 198-200) identifies freemium, price modulation and supply-side fragmentation as characteristic of the net economy and I would argue they apply to translation. Freemium is a pricing strategy where the product or service is provided free of charge but the customer must pay for additional features or services. Free online translation tools of a greatly improved quality have recently been made available to the general public. Google Translate, released in 2006, was upgraded as Google Neural Machine Translation in November 2016. In August 2017, the German company DeepL released its NMT tool, boasting it could translate over one million words in one second (DeepL, “Press release --- DeepL Translator”). Google operates on a hundred language pairs, DeepL on seven Western languages but with even better results (Larousserie & Leloup, 2017, Wu et al., 2016). In the economics of free, the product or “loss leader” is a trigger for other sales or sources of revenue: APIs for DeepL, marketing information for Google. In addition, delivering a free software or service may lead to maintenance contracts or updates and customization packages, a practice which corresponds to the business model of open source software. Omniscien (n.d.-a) for instance sells “a highly scalable, flexible and high-quality machine translation system that fully integrates with a wider range of systems”. Once the settings are in place, translation will be automated (and thus “free”) for the customer but training, maintenance and interoperability come at a cost. Systran has developed an automated website localisation platform, SYSTRANLinks, available for free when the volume of traffic is low (less than 6,000 pages hosted by them viewed per year) but if you need more traffic, they market three other billable editions: standard, pro and enterprise. The standard edition is offered in three versions for a yearly subscription ranging from €475 to €1,425. Free delivery and price modulation go hand in hand within online business models.

In the net economy, the traditional model of pricing per output (in our case the volume of text to be translated) gives way to a logic of long-term engagement with the customer through leases, licenses and subscriptions, a model used by SDL, Star and Systran. The emphasis is shifted from product to usability: the one quantitative variable

that really matters online is the volume of traffic, measured via clicks, the number of pages viewed or the number of transactions completed. For instance, LUIS, a natural language understanding programme developed by Microsoft to be integrated in apps and bots, is free up until 10,000 transactions per month but billed at a rate of \$1.5 to \$5.5 per 1,000 transactions above that mark (Microsoft Azure, n.d.).

Another avenue explored by online translation firms is the fragmentation of tasks through a crowdsourcing model. In this model, there are no contracts or quotes; instead assignments, which may include a variety of tasks such as translation and copy-editing or vetting, are posted on a platform; price is set by the buyers instead of the sellers. Those willing to accept that price will contribute and the platform will charge the client a fee corresponding to a percentage of the transaction's amount. Billing is usually per task or per hour and not per volume, translation tasks included. Lionbridge (2013) also distinguishes between “microtasks, complex business tasks, complex business programmes”, a typical example of supply-side fragmentation. Work must be completed within certain time frameworks, such as no more than one hour for 200 words (Garcia, 2015). Contributors aren't necessarily trained translators or editors, they can be just native speakers of the target language or come under the vague heading of “linguists and reviewers” as advertised by Capita. Rather than relying on skilled specialists, these firms hire non-specialised contributors and can only do so because they redefine quality, fragment work and modulate price. For instance, the selling point of the STAR Webcheck software is that it allows translation to be revised at the local branches of the firm by “local representatives or external employees” (“STAR WebCheck – Online Translation Reviewing”): here translation is only required to be fit-for-purpose, not necessarily publishable. Similarly, in the DARPA project on Arabic social media, the quality or consistency of the translation of each post is immaterial: what counts is the volume of linguistic data processed to further train machine learning or to feed data mining. It appears that, beyond their impact on the economics of translation, digital tools redefine what is marketed as language services (McDonough Dolmaya, 2012).

4 Redefining language services

The one prominent feature on the websites of the nine companies under investigation is the relabelling of their services. Systran's telling motto is “Beyond Language” and it sells “corporate language management”; SDL brands itself as delivering “global enablement services” which echoes Lionbridge's definition of the three fields of work for which it recruits crowdsourcing workers: “data servicing, global testing and language services” (Surveys Wonk, 2017). The fact that these larger firms emphasize their use of machine translation is of course consistent with Vashee's (2013:130) observation: “MT systems make the most economic sense when used to translate content that offers business value but whose sheer volume is so extensive that the traditional translate-edit-proofread (TEP) process would be prohibitively expensive or time-consuming.” But even for smaller volumes and smaller firms, translation is never the only focus of those online translation firms: it is systematically associated with localisation and

usually also with a range of marketing and data mining services. A typical description of their services would be the creation, translation, organisation and delivery of content. Datawords (n.d.) positions itself on “e-multicultural technologies” which includes localisation, social media optimization, multicultural consulting, web analytics and web marketing and multilingual platform design. New concepts emerge such as “e-discovery” (internet documentary search) and “digital forensics” (data analytics). Most of these services will be billed per hour as consulting fees or included within subscription packages.

Embedding translation into a range of language services is consistent with a shift of focus from text to user experience. Catering to the specific needs of the client, focusing on the context in which translations will be used, facilitating access and process, offering added value, though they were always commercially important features, become essential when translation costs are significantly reduced. Offering full customization and integrated solutions, for example within client software systems and workflows, is one way of implementing this user experience policy. Systran hosts clients’ websites and Omniscien (n.d.-b) has built a “single platform for language processing, machine translation and machine learning”, all of which is much beyond the reach of freelance translators in terms of vocational training or computing power. At the other end of production, providing transcreation or intercultural consulting is another way of focusing on user experience.

The notions of text, message or even information then cease to be central. This is consistent with the whole idea that language, relabelled as “agile content” (Lizuka, “Agile content requires agile translation”) is meant to be processed first and foremost by machines, one of the challenges of the AI revolution (Knight, 2016). To do so, it needs to come as standardized and exploitable content that can be easily transformed into data. Keywords, tags, metadata are the foundation of what Bomsel labelled “an economy of words” (2010: 26, my translation). Current research in NMT for instance focuses on matching voice, text and image for further marketing purposes (Facebook needs automatic recognition and tagging of visual posts) or better interaction with digital assistants and the internet of things (Xu et al., 2015). The accuracy or consistency of translation may not be such relevant features when the goal is to streamline large volumes of user-generated content into a data mining tool.

Bibliography

- Bomsel, O. (2010). *L’Economie Immatérielle*. Paris: Gallimard NRF Essais.
- Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). “Is Neural Machine Translation the New State of the Art?”, *Prague Bulletin of Mathematical Linguistics* 108, pp.109-120. DOI: 10.1515/pralin-2017-0013.
- Christensen, C. (1997). *The Innovator’s Dilemma*. New York: Harper Business.
- Cronin, M. (2013). *Translation in the Digital Age*. London: Routledge.

- Datawords (n.d.). "e-multiculturalism" <<https://datawords.co.uk/company/e-multiculturalism>>. Page consulted on date: 30.06.18.
- DeepL "Press release --- DeepL Translator" <<https://www.deepl.com/press.html>>. Last updated: 29.08.17. Page consulted on date: 30.06.18.
- De Palma, D., Stewart, R., Lommel, A. and Pielmeier, H. (2017) "The Language Services Market: 2017".
<<http://www.common sense advisory.com/AbstractView/tabid/74/ArticleID/39815/Title/TheLanguageServicesMarket2017/Default.aspx>>. Page consulted on date: 15..03.18.
- Desjardins, R. (2016). Translation and Social Media: In Theory, in Training and in Professional Practice. London: Palgrave Macmillan.
- Dru, J.-M. (2016). New : 15 approches disruptives de l'innovation. Montreuil: Pearson France.
- Enriquez Raido, V. (2016). "Translators as adaptive experts in a flat world: From Globalization 1.0 to Globalization 4.0?", IJOC: International Journal of Translation Vol. 10, pp. 970-988.
- Faes, F. (2017). "Translate.com exposes highly sensitive information in massive privacy breach" <<https://slator.com/technology/translate-com-exposes-highly-sensitive-information-massive-privacy-breach/>> Last updated: 07.09.17. Page consulted on date: 14.10.17.
- Fournier-Outters, T. (2018). "La traduction automatique à l'ACPR", Presentation at the 27th meeting of the Groupe interministériel sur la traduction, Ministère de l'Économie et des Finances, 4th of June 2018.
- Garcia, I. (2015). "Cloud Marketplaces: procurement of translators in the age of social media", JosTrans 23. pp. 18-38
- Google "Privacy and terms". <<https://policies.google.com/terms#toc-content>>. Last updated: 25.10.17. Page consulted on date: 30.06.18.
- Lizuka, I. "Agile content requires agile translation" <<https://blog.sdl.com/digital-experience/agile-content-translation/>>. Last updated: 08.02.18. Page consulted on date: 20.06.18.
- Knight, W. (2016). "AI's Language Problem".
<<https://www.technologyreview.com/s/602094/ais-language-problem/>>. Last updated: 09.08.16. Page consulted date: 05.02.18.
- Larousserie D. & Leloup, D. (2017). "Quel est le meilleur service de traduction en ligne?" <http://www.lemonde.fr/pixels/article/2017/08/29/quel-est-le-meilleur-service-de-traduction-en-ligne_5177956_4408996.html>. Last updated: 28.08.17. Page consulted on date: 20.06.18.
- Lionbridge (2013). The complete guide to business process crowdsourcing: Changing the way work gets done.
<http://info.lionbridge.com/rs/lionbridge/images/BusinessProcessCrowdsourcing_eBook_3.pdf>. Consulted date: 30.06.18.

- Marczak, M. (2018). "Translation Pedagogy in the Digital Age", Angles #7 <<http://angles.edel.univ-poitiers.fr/index.php?id=1556>>
- McAfee, A. & Brynjolfsson, E. (2017). *Machine, Platform, Crowd: Harnessing our Digital Future*, New York: W.W. Norton.
- McDonough Dolmaya, J. (2012). "Analyzing the Crowdsourcing Model and Its Impact on Public Perceptions of Translation". *The Translator* 18 (2), pp. 167–91.
- Mechanical Turks (n.d.). "DARPA language challenge"
<https://requester.mturk.com/case_studies/cs/darpa> Page consulted on date: 30.06.18.
- Microsoft Azure (n.d.). "Cognitive Services pricing—Language Understanding (LUIS)"
<<https://azure.microsoft.com/en-us/pricing/details/cognitive-services/language-understanding-intelligent-services/>> Page consulted date: 20.06.18.
- Omniscien (n.d.-a). "Government and intelligence section"
<<https://omniscien.com/market-solutions/>>. Page consulted on date: 30.06.18.
- Omniscien (n.d.-b). "Media Studio" <<https://omniscien.com/media-studio/>>. Page consulted on date: 30.06.18.
- Paillard L. (2013). *La Gratuité intellectuelle. Pour une véritable révolution numérique*. Lyon: Parangon.
- SFT (2016). "Enquête 2015 sur les pratiques professionnelles des métiers de la traduction" <<https://www.sft.fr/chiffres-clefs.html>>
- STAR (2014) "STAR WebCheck – Online Translation Reviewing" Webcheck".
<<https://www.star-group.net/tr/products/star-webcheck.html> > Last updated: 30.09.14.
Page consulted on date: 30.06.18.
- Surveys Wonk. "Virtual Bee Review: Work From Home Data Entry".
<<https://surveyswonk.com/virtual-bee-review-work-home-data-entry/>> Last updated: 03.06.17. Page consulted on date: 30.06.18.
- Turovsky, B. "Ten years of Google Translate"
<<https://blog.google/products/translate/ten-years-of-google-translate/>> Last updated: 28.04.16. Page consulted on date: 28.02.18.
- Vashee, K. (2013). "Understanding the economics of machine translation", *Translation Spaces*, Vol. 2 pp. 125–149. DOI: 10.1075/ts.2.07vas
- VoiceTra (2015). "Privacy policy" <<http://voicetra.nict.go.jp/en/privacy.html>> Page consulted on date: 30.06.18.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... Dean, J. (2016). "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation", *Computation and Language*. <arXiv:1609.08144>
- Xu, K., Lei, J., Kirost, R., Cho, K., Courville, A., Salakhutdinov, R., ... Bengio, Y. (2015). "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention", in Bach, B. and Blei, D. (eds.) *Proceedings of the 32nd International Conference on*

Machine Learning. Lille, France, JMLR: W&CP volume 37.
<<http://proceedings.mlr.press/v37/xuc15.pdf>>