

Unforeseen Consequences: Big Data and the Language Industry

Philipp B. Neubauer
Independent Researcher

1 Introduction

There are some long-term consequences of technological change that affect specific areas of social experience in ways that cannot in a direct or straightforward way be deduced from the intentions of the actors who are involved in bringing them about. For this reason, they are of considerable importance to social scientists and there is a long tradition of studying these so-called unforeseen or unintended consequences. Merton (1936) is considered to be the first to have set down systematic observations on the topic (Dietz 2004). Two key points of his observations are that unforeseen consequences need not be identified with axiologically negative effects (Merton 1936: 895) and that it need “not [be] assumed that in fact social action always involves clear-cut, explicit purpose” (ibid: 896/897). It is however safe to assume that the construction of a scenario that plausibly charts the context in which the unforeseen consequences are situated would be beneficial to their study and evaluation. This is the stated purpose of the present article. It is intended to provide some impulses for the study of unforeseen consequences of technological change – of course, our speculative/heuristic method can only produce hypotheses whose evaluation would then fall into the purview of empirical sociology and/or translation studies research, the disciplines which need to come up with designs for representative surveys – both to sociologically oriented researchers in translation studies (and particularly to those pursuing approaches based on the sociology of professions (Stichweh 2005), e. g. Diaz-Fouces and Monzó 2010; Sela-Sheffy 2011: 11) as well as to anyone interested in the broader field of technology assessment (Kalverkämper 1998: 12). This is to be achieved by charting some correlations between tendencies of the language services market and the context of industrial processes involving statistical machine translation (SMT) and post-editing (PE) within the bigger picture of the big data paradigm as it takes shape in the language industry on the one hand and the conceivable consequences this may have for the perception and economic position of translation professionals on the other hand.

Given that many of the emergent effects can be seen as “foreseen”/intended – or at least as assented to and accepted – on the part of large supply-side language industry players, there are already impressionistic

studies or personal commentaries on their impact on the translating profession (Rudavin 2009; Katan 2011) or critiques that focus on the influence of technology use on conceptions of translation equivalence and vice versa (Nogueira de Andrade Stupiello 2008). If one aims to bring the unforeseen and unintended into focus, one might look at them from the perspective of the advocates of free, libre and open source software and open access content, as this draws attention to the seeming paradox that e. g. de-professionalization might occur as a side effect of justified demands for accountability (Sandrini 2013; Mayer-Schönberger and Cukier 2013: 116), the democratic strife for access to education and freedom of information (Heylighen 2007) or simply as epiphenomena contingent on technological development. The epistemic opportunity in this regard lies in contrasting and synthesizing the perspectives of translators/post-editors and open source advocates precisely because there seems to be so little overlap between these subcultures, if one extrapolates from the current prevalence and uptake of FLOSS translation tools (García Gonzalez 2008).

Part of this synthesis will consist in arriving at a “sociological glimpse” (Diaz-Fouces/Monzó 2010: 10) which accounts for the sentiments and impressions of individual actors in the translation market. Then we will briefly expound on the ethos of open source and open access for the purpose of distinguishing, from this point of view, intended consequences from unintended/unforeseen ones. Following this, we shall introduce some more detailed observations on the technological developments driving structural change on the part of language industry suppliers:

1. Big data as a general technological trend towards the aggregation and algorithmic parsing of ever larger amounts of data; this general trend can serve as a template for interpreting developments in the translation services market by analogy.
 - a) Statistical Machine Translation (SMT), which represents the application of statistical algorithms to large repositories of translation data, e. g. such composed of translation memories (TM), on-line bitexts and parallel texts and especially the so-called open data, which public institutions disclose or release to the general public (Sandrini 2013). Another factor driving the growth of accessible translation data can be seen in the traction gained by open formats for data interchange (ibid.) which (at least in theory) facilitate the aggregation of data by ensuring its uniform structural presentation.
 - b) Post-editing (PE), by which we primarily refer to the rewriting of machine translation output in order to achieve results that are comparable to human translation, this is the subclass of “full post-

editing” (Allen 2003: 306). Within the scope of this article, this is the only relevant type as our argument depends on the commensurability with fully human (intellectual) translation. The output of PE activity can subsequently be added to the machine translation corpora used as its starting point. PE itself can be organized in the form of crowdsourcing (compare Fédération Internationale des Traducteurs (FIT) 2015) or it can be cast as a new way of professional translating, albeit one fraught with new challenges. This is reflected in the emergence of formal training courses in post-editing for which certification is available, for instance at the language service provider SDL plc. (2015a) or the industry association TAUS (2015).

Concluding the article, we will co-ordinate the insights into the technical workings of SMT/PE with the sociological glimpse obtained in the first section, which shall lead to an evaluation of the present trend in conjunction with a forecast of what there might be to come.

2 A Sociological Glimpse of the Language Industry

Here, the situation regarding the progressive automation of the workplace in general may serve as a starting point; it is noteworthy that in recent years this seems to have begun to penetrate to professions that would previously have been considered impervious to automation. According to an article published in Wired Magazine (Dormehl 2015) which quotes research by the University of Oxford conducted in 2013, approximately 47% of all jobs are predicted to be cut due to automation over the course of the next 20 years – the exact scope of the study in terms of industry and geographic scope was not amplified on; while this trend has been around since the dawn of the industrial revolution in the 19th century, its new quality seems to be that now, “white-collar professions involving a high level of training are just as likely to be displaced by software [...] because once-untouchable fields such as law and medicine include specialisms that are vulnerable to automation: medical diagnosis, the drafting of contracts and comparison of trademarks can be better carried out by a computer than by human beings” (ibid.). The researchers who published the study saw the reason for this in the fact that the subdivision of larger work processes into ever smaller series of actions, which has greatly facilitated the automation of “cognitive work”.

Although this prognosis with its more general scope does not make any specific mention of the language industry or the market for translation services, the scenario seems to resonate with some observer’s laments about the degradation in pay, prestige and working conditions that seem to prevail in

this area. Often, their blame is laid on technical innovation and/or economic developments.

Where technical innovation is concerned, the reason for the downward spiral is attributed to changes in perception regarding the translator and his or her task brought about by machine translation and translation memory technologies. One example for this is the critique articulated by Nogueira de Andrade Stupiello (2008), whose views shall be briefly summarized here. Contrary to the creed of functionalism, translators in highly automated environments are no longer seen as responsible for the semantic rendering of the target text, but are seen to be merely tasked with cosmetic changes to the semi-automatically generated output, which – as folk wisdom would have it – is already semantically complete and fully equivalent of the source. Hence, the focus is on minor flaws, details that the machine could not successfully “recover”. According to the critic, this perspective itself is not new, but follows from the tradition of translation technology and is already manifest in the conventions of translation memory use. Here, leverage is paramount even if the pre-translated segments do not fit their new context and thus any retranslation of existing matches due to textual concerns is neither desired nor remunerated. Nogueira de Andrade Stupiello (2008) thinks that the reasons for the prevalence of these attitudes can be found in the ever-shorter production cycles for translations, the need to cut cost and the “urgency of communication” under the pressures of globalization and the information age, which must eventually lead to lowered expectations regarding linguistic quality. At the end of the day, all that seems to matter is to somehow grasp the gist of a foreign language text.

Rudavin’s (2009) observations, by contrast, are formulated from a personal and practice-oriented perspective. He is concerned especially with the market situation of freelance translators, whereby the focus is less on technology assessment or the profession’s image in a stricter sense and more on the underlying structure of the language industry and its tendencies as a business sector. He observes that as such, the language industry cannot be viewed in isolation from its larger economic context and its actor’s financial incentives. In this regard, he also names “globalization” as the key driver, besides “market consolidation” and technical progress. The interrelation of the latter two is of special interest here: as global ITC networks facilitate the coordination of international multilingual projects, there emerges a market for projects which, due to time constraints, scale and the number of languages required are only manageable by the largest language service providers, actors whom Rudavin calls “translation corporations”. In some cases, these happen to be the very same corporations who also act as vendors of proprietary CAT tools that provide the workflow/process infrastructure by

which translation tasks devolve to smaller subcontracting agencies and ultimately the freelance translators. According to Rudavin, the “translation corporations” (which remain unnamed) already have a strong foothold in the market; the 30 largest vendors together are said hold a market share of 20% at an annual growth rate of 20-50%. If this tendency were to continue, a likely consequence would be the formation of an oligopoly.

3 Big Data, Open Source and Open Data

This is the initial scenario that we shall assume for the critique of the unforeseen/unintended consequences of the use of open and public data and open source technology in a for-profit translation context, since a starting hypothesis about the priorities and interests of industry actors is necessary for deducting intentions and contrasting them with the unintended/unforeseen consequences of their social actions. Before this can be attempted, there remain the enabling technological conditions to be explored.

3.1 Big Data

As it shall be seen, the big data paradigm is central to the success of the method of statistical machine translation while certain forms of openness can be seen to constitute necessary preconditions for the application of the big data paradigm to the language industry. There is hitherto no complete intensional definition of big data, however, two essential properties indicative of this state of social and technical development can be identified: on the one hand, there is a steady increase in the quantity of digital data as the digitization of ever more areas of human experience progresses; on the other, there is an emergent qualitative change of the area itself which follows the utilization of the data in its respective context. This latter is what Mayer-Schönberger and Cukier (2013: 6) assert to be the defining attribute of big data:

[D]ata has begun to accumulate to the point where something new and special is taking place. [...] The quantitative change has led to a qualitative one. The sciences like astronomy and genomics, which first experienced the explosion in the 2000s, coined the term “big data”. [...] There is no rigorous definition of big data. [...] One way to think about the issue today [...] is this: big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more.

If it is assumed that SMT (with or without downstream PE) constitutes a new mode of value creation for the language industry which has the potential to

disrupt markets and production processes, the question remains where exactly the mass (“big”) data fueling the SMT engines are sourced from and how they are exploited or ultimately monetized.

3.2 Open Source

One possibility for obtaining the mass data is to rely on open sources, whereby this statement can be confusing as the data in question need not be licensed as “open source” as in “free, libre and open source”, but need only be publicly and unrestrictedly accessible, as in “open-source intelligence” (Wikipedia contributors 2015c, Open-source intelligence) – The Open Source model (Heylighen 2007a) itself follows a principle similar to that of “communalism”, which is at work in the organization of science (Merton 1988: 680); thus, the Mertonian concepts used to describe scientific organization should be reasonably continuous with this new context. Nevertheless, such data can and does include “free and open” licensed sources in a stricter sense. According to FOLDOC (2012: Open Source), this is the intention behind Open Source as a model of software licensing and distribution:

A method and philosophy for software licensing and distribution designed to encourage use and improvement of software written by volunteers by ensuring that anyone can copy the source code and modify it freely.

This concept, which reflects a denotation of unlimited redistribution and modification, is not limited to software products, but applies to other immaterial goods as well. Insofar as a strict separation of formal language texts and digital natural language data and audiovisual material is tenable (compare Touretzky 2001), it has been designated either Open Access (Heylighen 2007) or Open Content (Gunn 2008) where it relates to the latter. Analogous to the family of open source software licenses, a few licensing models for Open Content can be distinguished from the published content itself. According to Gunn (2008), the “Creative Commons” (CC) and “Free Document” (FDL) licensing models can be cited as examples of explicitly free and open licenses for publishing. The intentions motivating Open Data initiatives which also include open translation data (compare Sandrini 2013: 33) can be seen to vary somewhat from this theme. Here one might distinguish explicitly open from public data, with the latter satisfying the criterion of de-facto open access without necessarily being meant for free redistribution and modification.

3.3 Open Data

True open data originate with the public sector and government institutions (Sandrini 2013: 33); they are often released to the general public because

public institutions can rarely do more than merely administer the data on behalf of their constituencies for want of resources and expertise (Mayer-Schönberger and Cukier 2013: 116).

An example for open translation data can be found in those published by the European Union (*ibid.*) who also hope to advance their own SMT program in this fashion. Translation data of the UN have been published in the context of the “Corpora Commons” initiative, also with the explicit aim of furthering SMT research (Gunn 2008). These two examples concern open data in the stricter sense (compare Mayer-Schönberger and Cukier 2013: 38); patents and trademarks which must by decree be published in several languages (Pariser 2011) might serve as yet another example.

The development of Google Translate, currently perhaps the most prototypical phrase-based statistical machine translation system, exemplifies the conflation of open and public data in the training of SMT engines; besides the actual open data aggregates described above, public data comprising practically all translation data of the world wide web have been leveraged for its training. Among this, there has been some with contentious legal status, as the utilization of translations from the Google Books project shows – see “Authors Guild, Inc. v. Google, Inc.” (Wikipedia contributors 2015) (Pariser 2011).

While the for-profit use of true open data is (at least in general understanding) in line with the intentions of their providers, the same treatment of merely public data constitutes a gray area at very minimum. This might also be applicable to some extent to proprietary translation data held by language service providers, provided that they meet two conditions: firstly, they need to be fungible, i. e. come in a structurally open (interchange) format (Sandrini 2013: 33) and secondly, they need to be scrambled by technical means in order to circumvent some intellectual property laws that would otherwise apply to the data in aggregate (Zetsche 2005); this at least holds inside the German jurisdiction (Cruse 2014) and shows that determining the status of such data is difficult to begin with. Once the conditions are met, these data might also be treated as public.

3.4 Distinguishing Public Data and Open Source Software

While these considerations reference the relationship of SMT and data, open source software is also directly and indirectly relevant to developments in SMT. For one, free and open source SMT software and components immediately lower the barrier for SMT research (Lopez 2008: 3), while a more indirect consequence can be discerned in the diversity of ideas, actors and projects and the flat hierarchies of open source development (Heylighen

2007) which favor rapid evolution. Even though our focus lies on data as the main driver of SMT uptake, these factors might be of interest in the assessment of any unforeseen consequences stemming from the FLOSS paradigm itself – a conceivable case in point is the use of free and open SMT systems, e.g. MOSES (2015) on the part of language service providers. Though this appears a plausible scenario, there now seems to be (to the best of my knowledge) no economically significant use of this or similar systems – however, if any such use were modeled on the patterns described here, they would qualify as cases for the study of the unforeseen effects of FLOSS products.

Considering that data is the key component, it is for now safe to neglect the impact of the actual licensing model of SMT software on the scenario to be devised. Its basis lies in the construction of the relationship between the availability of data to fuel data-driven semi-automatic production processes on the one and the structure of these processes, i. e. how language workers interface with machine output, on the other hand.

4 Machine Translation and Post-editing

Research into machine translation has been around since the advent of electronic computers in the 1940s (Ping 1998: 162). Historically, the area has seen its ups and downs, the former marked by irrational exuberance triggered by an overestimation of the impact of advances in memory capacity and computing power on machine translation capabilities, the latter by the subsequent disenchantment caused by the evaluation of the actual results delivered by predominantly rule-based historical machine translation systems (Weizenbaum 1976: 186). Such tendencies are still extant, however, the premise seems to have changed with the shift towards big data/statistical processing; here, it is plausible to assert that increasing “processor speed, random access memory size, secondary storage, and grid computing” will indeed contribute to the improved performance of machine translation systems (Lopez 2008: 3) because such performance would be based on a larger throughput of data (i. e. larger amounts parsed) to begin with.

However, this article is not intended be an in-depth review of the history, functional principles and limitations of the machine translation systems themselves; we merely draw on these to elucidate on its argument. The focus is more on current tendencies in the actual deployment of SMT systems that can be linked to both big data and open data than on their history or technical details. The following figure shows a breakdown of MT systems by the fundamental strategy used to create the semblance of a “translation”

performance on chunks of natural language input and thus a “pseudo-translation” (Torrens, cited in Wilss 1996a: 212).

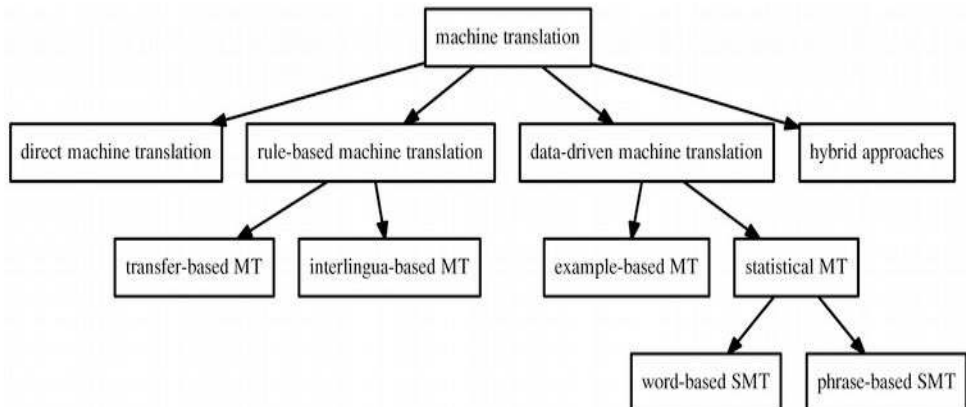


Figure 1: Taxonomy of machine translation architectures. Based on: Labaka et al. 2007; Lopez 2008; Eberle 2008; Gupta 2012; Okpor 2014.

If one completely disregards both the historical strategy of direct machine translation and any hybrid approaches there remain two fundamentally different strategies of MT, the rule-based and the data-driven. The rule-based model aims at generating a pseudo-translation by means of a pre-encoded linguistic and grammatical rule set for a generative transfer of L1 to L2. The statistical model relies on parsing large quantities of data for the probability of translation equivalence and thus constitutes the kind of technology that might benefit significantly from a quantitative hike in the available data. Here, we can discern the potential for the conversion of quantity to quality that Mayer-Schönberger and Cukier have envisioned.

4.1 Statistical Machine Translation

This potential lies in the reliance on statistical correlations between L1 and L2 renderings of chunks or phrases (in the case of the currently prevalent phrase-based SMTs, Lopez 2008: 9) rather than on explicit grammatical rules for the generation of a pseudo-translation. The linguistic material for analysis resides in parallel corpora (i. e. aligned translation data) parsed by the SMT algorithm. Unlike the rule-based model, the machine makes no attempt at emulating human interpretation or reconstructing the semantics of the source text (Ping 1998: 163-164). It does however appear to demonstrate “machine learning” (Lopez 2008: 1) in the sense described here:

A control system acts when there is a discrepancy between what it senses (sensory signal) and what it is supposed to sense or would like to sense (reference). The connections that matter are those of certain activities in the system's repertoire with the changes they provoke in certain sensory perturbations. A mechanical feedback device that replaces us in a given task is a crystallized piece of experiential learning. It is the materialization of an if-then rule that has been inductively derived from experience by the designer (Glaserfeld 1981).

What the machine "likes to sense" in this case is the larger probability of a given L1 phraseme having been translated by L2 phraseme X, as opposed to phrasemes Y, Z and so on. This figure shows what remains at the end of the mapping process:

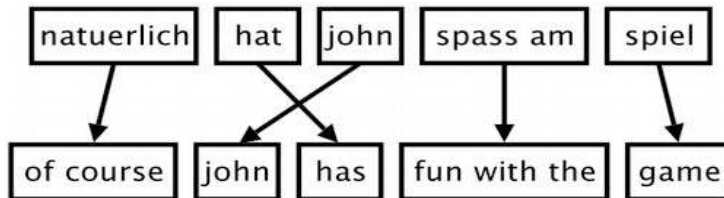


Figure 2: A phrase-based SMT model; Koehn (2010).

This however also serves to illustrate that the machine will only be capable of providing a "plausible" pseudo-translation if the search space for such probabilities is large enough, both in terms of finding positive correlations for the largest possible amount of L1 phrasemes and in terms of eliminating relatively unlikely candidate phrases; as the search space thus equals the corpus of phrase pairs "known" to the algorithm, it becomes clear why SMT performance is linked closely to corpus size and (alignment) quality (Arnold 2003: 139; Lopez 2008: 1; Labaka et al. 2007).

It also shows that the approach of so "guessing" the probability of a phrase to appear in a certain slot regardless of its semantic function is a far cry from the (always contested) idea of artificial intelligence as aiming "to simulate human intelligence as it manifests itself in the understanding of all reality, concrete or abstract, with which human beings are confronted [... b]y means of entirely automatic processes" (Wijnands 1993: 166). If one tries, for the sake of the argument, to imagine the pseudo-translation process as performed by a human, one might think of someone who is neither a speaker of L1 nor L2 in the process of assembling fragments of "fuzzy matches" from a translation memory system, guided only by their optical resemblance to

character strings which appear in L2 texts. Insofar as reading the pseudo-translation can be said to have caused someone to “understand” its intended message, this would have been a function of the database/corpus having contained very similar phrasal material, which in turn would only have been likely (read: probable) if the search space was very large indeed; this is why SMT is considered a big data application (de Palma 2013).

That this event is even possible constitutes the previously mentioned “new quality from quantity”; as recently as 12 years ago, the scarcity of data had been seen as a severe limitation of the statistical approach to machine translation (Arnold 2003: 139). Now, the increasing availability of open and public translation data have made this a non-issue, at least for some language combinations. Predictably, this increase in the volume of data has translated into better quality pseudo-translations (Scholtes 2010), to the extent that the technology has now attracted the interest of language service providers (Rex 2013) and the largest technology players (Herranz 2014) alike. Even if the quality of the output of the free (of charge) web translation offers (e. g. Google Translate) is scarcely good enough for integration into professional translation workflows, this need not be the case for proprietary engines offered by language service providers like SDL (“BeGlobal”, SDL plc. 2015b) which have been trained on well-aligned and often industry-specific input data.

4.2 Post-editing

However, to reiterate our argument, neither a large statistical search space nor a cleanly aligned MT corpus can in and of themselves grant the SMT engine the capability to translate in the sense of producing something that actually equals a human translation in form and function. It lacks the crucial element of “intelligence”, however one likes to define it (Wilss 1996b; Weizenbaum 1976: 186-187). Whether or not one believes that the original meaning of the source text can somehow be “recovered” from the phrase salad resulting from SMT or whether one asserts that it takes an act of interpretation of the pseudo-translation relative to the source in order to arrive at a semantically viable reading of any pseudo-translation that does not by chance resemble a natural language utterance (which need not bear any semantic relationship to the source language’s) is moot with regard to this statement.

To my mind, this is about the pinnacle of the “translation performance” that current systems are capable of. That the public and scientific interest in machine translation research has never completely waned despite this might be explained by venturing that linguistic utterances do not “contain” any intrinsic meaning, but that any meaning is synthesized by the recipients’ fitting

them into their experiential world. It is this act which provides considerable leeway for the benevolent interpretation of pseudo-translation as well as that of any other speech act (especially those in written language) (Berman 2013: 2-4; von Glasersfeld 1999).

If SMT technology is to be employed for the creation of value on the basis of big data, the missing ingredient needs to be added downstream, at a later stage of the production process. This stage is called post-editing (PE); it involves the use of human labor to impose potential meaning by rewriting/reordering the SMT pseudo-translation. In principle, this understanding does not significantly deviate from the definition of post-editing as the “the correction of machine translation output by human linguists/editors” (Veale and Way, cited in Allen 2003: 297). It seems likely that the literature contains many more variations on this theme.

Any of these might however be open to criticism, both from the vantage point of translation theory and from that of statistical machine translation technology itself. On the one hand, the notion of “correction” reflects the somewhat naive view of natural language criticized by Nogueira de Andrade Stupiello (2008), namely that which maintains that essential meaning (to the extent that this is believed to inhere in the source) has already been recovered by the SMTs and that the segment would only need to be polished by removing minor errors (e. g. non-agreement of suffices, superfluous or missing words and other artifacts of alignment). However, it should now have become clear that this essentially contradicts the premise of an a-semantic and non-interpretative mode of pseudo-translation generation. Insofar as a meaning is read into the signage of the segment by the post-editor or subsequent interpreter, its emergence is owed to the intervention of the person’s consciousness and their ability to interpret language within considerable tolerances – it has clearly not been actively recovered by the machine. As the term “segment” in this context suggests, the primary locus of “meaning recovery” is – in line with the prevalent design logic of current translation editor software – the micro-linguistic level of the sentence or below, where accidental matches are far more probable than on the macro-linguistic level of the complete text. Here, the chances for these to occur should be astronomically small, which is probably why the impact of SMT on texts hardly seems to feature in considerations of SMT capabilities. Granting the possibility of “lucky” selections on the segment level and minimal human intervention with the output of well-trained engines, the translation performance proper as it is perceived by the final recipient needs ultimately be enacted by the human post-editor, not the engine, which can’t (and isn’t designed to) provide it.

Having stated this, there is also the aspect of SMT economy to consider. While it is always possible to replace an inviable pseudo-translations with a completely new translation, this is certainly not the best solution in terms of leverage, considering that the post-edited output is not only there to serve the immediate need of the translation customer, but that it should ultimately return to the SMT corpus in order to enlarge its search space (i. e. the range of phrase variety covered) and so to guarantee future leverage for more plausible pseudo-translations.

“Leverage” in this sense can be understood as analogous to the use of this term in the context of translation memories, i. e. better leverage is achieved by (re-)using as many of the original SMT suggestions as possible in order to closely match similar input in the future; depending of the quality of the SMT corpus used, it is easy to see how this goal competes with that of efficiently imposing potential meaning. Incorporating both these competing goals into the PE strategy can be seen as a challenge notably absent from conventional human translation.

Hence, the capability for reconciling and balancing the human and machine demands of the task – i. e. the demand for communicative meaning and readability on the one, the demand for uniformity and future leverage on the other hand – is the distinguishing quality of post-editing when compared to translating. However, with regard to the more standard qualities demanded in commercial translation (correctness, speed, and cost), there is no question of “either ... or”; the additional challenges of post-editing simply add to the overall requirements. This translates into cumulative difficulty, as post-editing has the goal of translating more text faster. The PE additionally faces the challenge of submitting more text to QA procedures, etc. in even less time. Post-editing, which in this way differs from purely human translation both in terms of quality and of quantity, can thus appear a task that “anyone can do” (Pym 2013: 489) only at the most superficial of enquiries.

5 A Tentative Scenario for the Translation Market

To conclude this line of enquiry, it now remains to relate the aspects of underlying technology to the impressions of our “sociological glimpse”. The connecting elements are both the status of the translating profession as an income-generating factor (or, on the reverse, the decreasing rates which are a hallmark of de-professionalization) and the competition between translation workers with differing qualification profiles (compare Monzó 2011). The heart of the matter is that post-editing as an occupational activity does not seem to belong to any recognized profession which in turn would lend it the pedigree

correlated with higher remuneration (Fuchs-Heinritz et al. 1995: 521). The following statements are indicative of this observation:

- Pym (2013: 491) understands post-editing as an area associated with “technical communication” but notes that efforts at professionalizing this discipline tend to lag far behind those already undertaken for translation and interpreting;
- Allen (2003: 298-299) observed that, at least at the time of his writing, hard-and-fast criteria to certify the qualification of post-editors were lacking; recent efforts to formalize this qualification, like those already mentioned, might remedy this in the short term but will never convey the professional pedigree of a full university degree program.

Given that the self-reported status of translators in a recent study (Katan 2011: 77-78) was relatively low – respondents stated that it was largely comparable to that of a “secretary” – and that tendencies of de-professionalization are already under investigation (ibid 66) in this field, the key danger is to my mind that due to the nature of the process, crucial human capabilities are either accidentally misattributed to the SMT engines or deliberately misrepresented. If so, the likely consequence is a further erosion of the professional recognition of translators/PEs, aggravated further by clients being isolated from the translation/localization process by multiple layers of large language service provider’s corporate bureaucracies, two factors which are very likely to coincide, especially when these middle-men are vendors of language services and translation technology/SMT products at the same time.

The peril for the translation/PE practitioner lies less in falling victim to an actual deskilling, insofar as this is defined as a “reduced utilization of [... and] partial or complete devaluation of existing scholastic/academic, professional or vocational qualifications” (Fuchs-Heinritz et al. 1995: 135, my translation), as should have emerged from the present discussion. It lies in the loss of (or rather the failure to attain) the professional standing which secures expert status and monetary perks for the members of the more prototypical professions (Katan 2011: 70).

From this apparent de-professionalization results a change in the structure of competition in the market; when linguistic competence is devalued or no longer counts as a distinguishing professional qualification (Pym 2013: 489), a situation may emerge in which translation/PE professionals will have to compete against those whose qualifications are either completely different or those whose (source-)language competence might be significantly worse than is acceptable for professional translators (Katan 2011: 71). This larger competitive field may ultimately lead to further downward pressure on prices and/or

the exclusion from business opportunities of those who can't (or won't) compete under these circumstances.

This is likely to affect projects which are very demanding in terms of subject competence, e. g. specialized translations relating to law or medicine (where there perhaps might already be a possibility for semi-automatically generating the source text) as well as those where the expectations in terms of visibility and linguistic quality are very modest, e. g. "F.A.Q" sections for consumer products and the like.

This conclusion readily agrees with Rudavin's (2009) observation that subject specialists with a second language have recently been preferred over those who are (only) professional translators for complex assignments in the above fields. Add to this the observation that "[...] you often have no constant need to look at the foreign language [...] for some low-quality purposes, you may have no need to know any foreign language at all, if and when you know the subject matter very well" (Pym 2013: 489) and it should be easy to see how a combination of SMT/PE-capabilities and extant labor market tendencies might generate a synergy to that effect. This means that the growth of translation data (e. g. when already-dominant LSPs manage to appropriate large high-quality corpora for specific domains) which contributes to the recognizability/interpretability of pseudo-translations coincides with the automation of certain professions that may lead to the simultaneous "release" of a significant numbers of workers. The displacement of specialized translators by SMT-augmented multilingual specialists for the field in question would at least be a conceivable outcome. This scenario is not without a parallel in already existing situations where markets/fields of competence overlap (Katan 2011: 73); yet, the aspect of combined technological and social change holds the potential for bringing about a new, unforeseen quality in this phenomenon.

It seems even more likely when we approach the market for low-end translation services. As specialist knowledge does not matter here, there might even be a market for anonymous crowdsourcing workflows. Since the professional association *Fédération Internationale des Traducteurs* (FIT) (2015) has already published a position paper outlining the method of crowdsourcing, we will not amplify on this matter here; our assertion is the emergence of a scenario akin to that outlined for high-complexity projects, only with an aggravated tendency towards "lowest-bid market economics" (Muzii, cited in Katan 2011: 66). Translation workers will thus compete via pricing rather than competence/qualification. Between the high and the low end of the market, a visual breakdown of the projected scenario in relation to current practices might look like this:

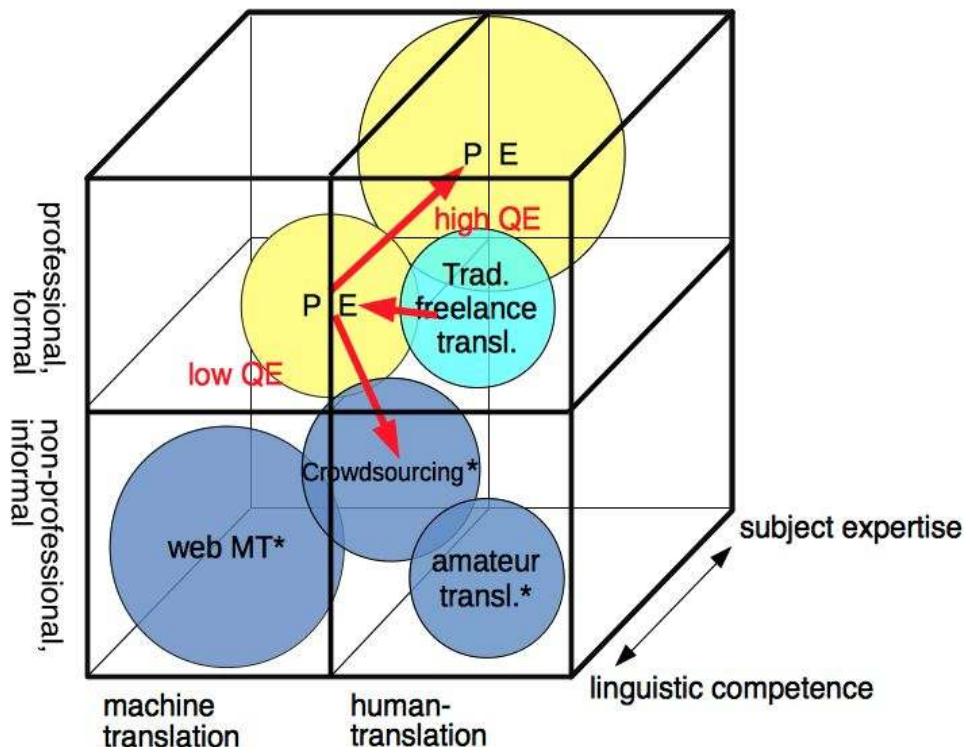


Figure 3: Intellectual translation vs. post-editing; the depth of specialized knowledge cannot be determined for activities marked with an asterisk.

For this we use a modified priority matrix with an added dimension of depth (linguistic competence vs. subject expertise). The matrix is inscribed with Venn diagrams showing any overlap between types of activities. Traditional (freelance) translating entails working a diverse portfolio of both classical translation and PE, highly specialized and general jobs, etc. It thus occupies a median position. In contrast to this, there is the noted drift towards the “back” of diagram in PE with high expectations in terms of quality (QE). Low-QE post-editing overlaps crowdsourcing in the lower right quadrant, which – due its black-box nature – may overlap with and introduce both raw machine translation from web engines and unrevised amateur human translation.

6 Outlook and Concluding Remarks

While it is conceivable that the scenario we have envisioned is likely both foreseen and intended on the part of language service providers, it is a cogent question to ask whether these consequences have been foreseen – or could have been foreseen – by any of those who have contributed to creating the basis of this economy of human/machine translation: institutional decision makers releasing open data to the public, developers of algorithms and (open source) software, academics concerned with basic research in fields like linguistics, mathematics, computer science and many more. From their vantage point, the unforeseen consequences of the growth of both open and public translation data can best be attributed to Merton's category of "chance consequences", "occasioned by the interplay of forces and circumstances which are so complex and numerous that prediction of them is quite beyond our reach" (Merton 1936: 899-900), owing to the fact that either of these endeavors seem remote from the translation services market and that there does seem to be an element of the co-incidence of a number of disparate developments involved. Nevertheless, we have managed to construct a scenario "on the ground" by identifying and connecting some of these forces and circumstances for the purpose of discussing their interplay; they are:

- the increasing automation of cognitive work,
- the role attempts at value creation through the combined use of big data resources and statistical machine learning algorithms play in this,
- the shifting expectations of translation consumers and language service providers brought about by market consolidation, globalization and the progress of certain technologies,
- the accelerated technical change through community-driven and open scientific research and software development modeled on analogous patterns,
- the economic rationalization of workflows through the combined use of human and machine resources, which gives rise to the practice of post-editing.

The most noteworthy paradox that rears its head here is that the unforeseen consequences of de-professionalization and falling proceeds from translating – even if they appear to be results of a very indirect causality – glaringly contradict the stated intention of the push to open translation data, namely to "enhance the perceived value of translation and to elevate the status of translators as a professional group" (Sandrini 2013: 33, my translation). This leaves the question of the final lesson learned from tackling the phenomenon. What the present author is paid for post-edited words is exactly half of what

the same customer is willing to pay for “new words” of a conventional human translation. If this is in any way indicative of an emergent industry trend would again need to be established by means of a representative study.

If one belongs to a group that is put a disadvantage by current developments, it is certainly tempting to feel a nostalgic longing for the “old days” of closed-off, guild-like professions and to renounce the open and collaborative mode of work which threatens to dissolve inherited privilege, even if scholars in the sociology of professions point out that the traditional professions are losing their former social and economic traction anyway (Stichweh 2005) and if one takes into account that privilege and closure in this sense have been considered an unfair advantage over laymen since the days of Adam Smith. Keen (2008) can be named as an example for this reactionary outlook on contemporary technology and culture. It seems however rather doubtful that such musings can provide any positive impulses for engaging with the present professional practice or for shaping the future of translation as a business.

They also miss the essential point. As already suggested, the true peril seems to consist in too little openness and transparency rather than too much. It would be a function of cumulative advantages – this is a concept from the sociology of science (Sismondo 2010: 39-40) which generalizes Merton’s “Matthew effect” (Merton 1968: 58; Merton 1988: 609); it might be understood as a form of positive feedback which leads to “inequalities [...that] appear to result from self-augmenting processes” (Merton 1988: 617). These effects, initially observed in scientific careers, also form a sub-category of unintended consequences (Merton 1988: 615). Apparently not limited to science, they can be observed in similar social fields, e. g. open source software development, where Heylighen (2007) observed a “rich get richer” dynamics [negatively affecting] equally valuable, competing projects [which,] because of random fluctuations or sequence effects, may fail to get the critical mass necessary to ‘take off’”. Such cumulative advantages are garnered by the “translation corporations” as a consequence of their growth and economies of scale that coincide with an environment characterized by an accelerated de-professionalization of language services in combination with a distorted perception of human/computer PE/SMT processes. Either is a consequence contingent on the big data phenomenon and some mutual interdependence can be ascribed to them.

Providing that storing larger quantities of data opens new qualitative paths for its commercial exploitation, vendors of SMT systems might start off by training their engines on open translation data and expand their reach by re-training them with data for other languages and domains as they flow back from their normal translation/PE operations. As the recognizability/ interpreta-

bility of pseudo-translations improves with rising corpus size, it will become possible for them to shunt existing customers from human translation to SMT/PE-based processes, whereby the deal can be sweetened for the consumer by passing some of the cost reductions on to them. This might create a virtuous circle (from the vendor's vantage point) as more data is funneled back into the engine, more customers are attracted and the vendor's economic clout increases. Consequently, they will find themselves in a position where they are increasingly capable of dictating (lower) translation purchasing prices and of squeezing competitors out of the market.

Any such (hypothetical) companies are practically doomed to appear as “free riders” from the vantage point of the institutions and communities that contribute technology and data in accordance with the open source ethos (Heylighen 2007): industry preferences for proprietary licensing, vendor lock-in and draconian non-disclosure agreements all but preclude any data, knowledge or technological improvement from being given back to the communities and general public. Such would be the working of a “ratchet effect” that allows the free flow of open and public resources into proprietary systems, but not the other way around.

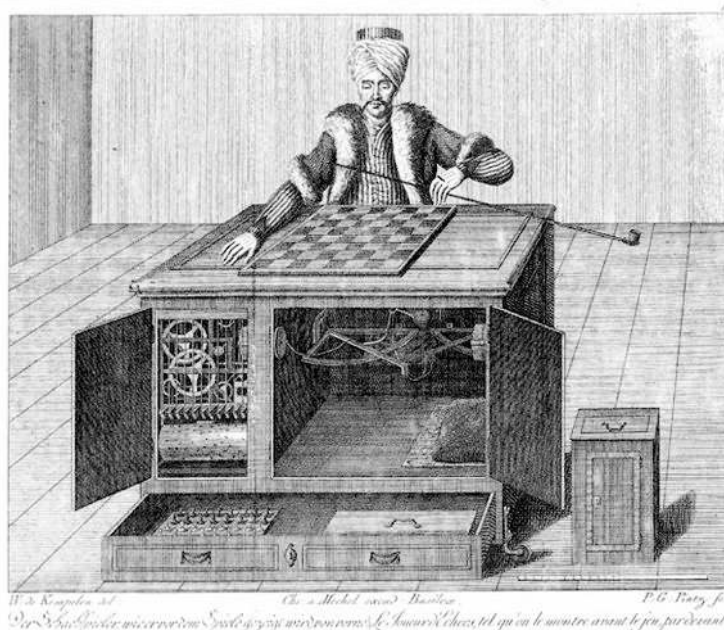


Figure 4: The “Mechanical Turk”, a 19th century make-believe chess automaton.
Source: Wikipedia contributors 2015d

Translators/post-editors would likely be affected in a different way. Here, the gap for exploitation lies in the representation of machine capabilities and their actual inability to produce more than pseudo-translations. Even if it can be assumed that no reputable language service provider would ever try to conceal this fact from their customers, downplaying it for marketing purposes would not be considered unethical by many. The human PE, the real engine of the process who ultimately bears the responsibility for the usefulness of the product – its fitness for the purpose of human communication – is blotted out from the perception of the translation consumers and thus enacts a role that begins to resemble that of the operator working in the interior of the “Turk” (Wikipedia contributors 2015d, The Turk) who helps create and maintain the illusion of an autonomously playing chess automaton by lending his or her capability to the “machine”.

Ironically, this will reinforce the impression of the “non-human, technical [...] habitus” ascribed to translating (Katan 2011: 78) and executives’ imputed opinion of translators as “human-mechanical revenue generating machines” (Rudavin 2009) – with all the perfectly foreseeable socio-economic consequences this is likely to have for the practitioners themselves.

Due to the complexity of the interplay of macro-social and technological forces that bring about similar developments, a public debate of the desirable and undesirable consequences of data-driven technologies in general is likely to benefit not only translation businesses, professional associations and translation studies as an academic discipline, but also society at large. If we fail to practice technology assessment in time, we are at peril of being overwhelmed by unforeseen consequences in the long run.

References

- Allen, J. (2003) Post-editing. In Somers, H. (ed.) *Computers and translation: a translator's guide*. Amsterdam: John Benjamins, 297-317. Available at: <https://books.google.de/books?id=a4W7IWgCqYoC> [Accessed 18 September 2015].
- Arnold, D. (2003) Why translating is difficult for computers. In H. Somers (ed.) *Computers and translation: a translator's guide*. Amsterdam: John Benjamins, 119-142. Available at: <https://books.google.de/books?id=a4W7IWgCqYoC> [Accessed 18 September 2015].
- Berman, J. J. (2013) *Principles of big data: preparing, sharing, and analyzing complex information*. Oxford: Newnes. Available at: <https://books.google.de/books?id=gEho0DI8a2kC> [Accessed 18 September 2015].
- Cruse, A. (2014) *Besitzansprüche – Urheberrecht und elektronische Datensammlungen*. *MDÜ* 14.3, 10-15.
- de Palma, D. A. (2013) *Big Data Comes to the Translation Sector*. Common Sense Advisory Blogs. Available at: <http://www.commonsenseadvisory.com/default.aspx?Contenttype=ArticleDetAD&tabID=63&Aid=3025&moduleId=390> [Accessed 18 September 2015].

- Diaz-Fouces, O. and Monzó, E. (2010) What would sociology applied to translation be like? In *MonTI* 2., 9-18. Available at: http://rua.ua.es/dspace/bitstream/10045/16432/1/MonTI_2_01.pdf [Accessed 18 September 2015].
- Dietz, H. (2004) Unbeabsichtigte Folgen – Hauptbegriff der Soziologie oder verzichtbares Konzept? *Zeitschrift für Soziologie* 33.1, 48-61. Available at: <http://www.zfs-online.org/index.php/zfs/article/viewFile/1154/691> [Accessed 18 September 2015].
- Dormehl, L. (2015) Your job automated. *Wired Magazine* (UK Edition) 01.15, 126-133. Available at: <http://www.wired.co.uk/magazine/archive/2015/01/features/your-job-automated> [Accessed 18 September 2015].
- Eberle, K. (2008) Integration von regel- und statistikbasierten Methoden in der Maschinellen Übersetzung. *Journal for Language Technology and Computational Linguistics* 23.2, 37-70. Available at: http://www.jlcl.org/2009_Heft3/kurt_eberle.pdf [Accessed 18 September 2015].
- Fédération Internationale des Traducteurs (FIT) (2015) FIT Position Statement on Crowdsourcing of Translation, Interpreting and Terminology Services. Online position paper. Available at: <http://www.fit-ift.org/wp-content/uploads/2015/04/Crowd-EN.pdf> [Accessed 18 September 2015].
- FOLDOC Contributors (2012) Free On-line Dictionary of Computing (FOLDOC). Online lexical database. Howe, D. (ed.) Available at: <http://foldoc.org/> [Accessed 18 September 2015].
- Fuchs-Heinritz, W. et al. (1995) *Lexikon zur Soziologie*, 3., völlig neu bearbeitete und erweiterte Auflage. 3rd edition. Opladen: Westdeutscher Verlag. Available at: <https://books.google.de/books?id=wSieBgAAQBAJ> [Accessed 18 September 2015].
- García Gonzalez, M. (2008) Free Software for translators: is the market ready for a change? In Diaz-Fouces, O. and García González, M. (eds.) *Traducir (con) software libre*. Granada: Comares, 3–31.
- Glaserfeld, E. von (1981) Feedback, Induction, and Epistemology. In *Applied systems and cybernetics*. Lasker, G.E. (ed.) Vol. 2. New York: Pergamon Press, 712-719. Available at: <http://www.univie.ac.at/constructivism/EvG/papers/069.pdf> [Accessed 18 September 2015].
- Glaserfeld, E. von (1999) How Do We Mean? A Constructivist Sketch of Semantics. In *Cybernetics and Human Knowing* 6.1, 9-16. Available at: <http://www.univie.ac.at/constructivism/EvG/papers/221.pdf> [Accessed 18 September 2015].
- Gunn, Allen (2008) Open Translation Tools: Disruptive Potential to Broaden Access to Knowledge. Report prepared for the Open Society Institute. Open Society Institute. Available at: <http://aspirationtech.org/files/AspirationOpenTranslationTools.pdf>.
- Gupta, S. (2012) A survey of data-driven machine translation. Available at: <http://www.cfilt.iitb.ac.in/resources/surveys/MT-Literature%20Survey-2012-Somya.pdf> [Accessed 18 September 2015].
- Herranz, M. (2014) Twitter, eBay, Facebook. Big data companies want to own machine translation. Pangeanic blog post. Available at: <http://blog.pangeanic.com/2014/08/10/twitter-ebay-facebook-big-data-companies-want-to-own-machine-translation/#> [Accessed 18 September 2015].
- Heylighen, F. (2007) Why is open access development so successful? Stigmergic organization and the economics of information. In *Open Source Jahrbuch*. Lutterbeck,

- B., Bärwolf, M. and Gehring, R.A. (eds.) *Lehmans Media*. Available at: <http://pespmc1.vub.ac.be/Papers/OpenSourceStigmergy.pdf> [Accessed 18 September 2015].
- Kalverkämper, H. (1998) 1. Fach und Fachwissen [Subject and subject knowledge]. In Hoffmann, L. and Kalverkämper, H. (eds.) *Fachsprachen – Ein internationales Handbuch zur Fachsprachenforschung [Languages for special purposes – An international handbook of special-language and terminology research]* Vol. 1. Berlin: de Gruyter, 1-24.
- Katan, D. (2011) Occupation or profession, A survey of the translators' world. In Sela-Sheffy, R. and Shlesinger, M. (eds.) *Identity and Status in the Translational Professions*. Amsterdam: John Benjamins, 67-87. Available at: <https://books.google.de/books?id=KbZxAAAAQBAJ> [Accessed 18 September 2015].
- Keen, A. (2008) *The Cult of the Amateur: How blogs, MySpace, YouTube, and the rest of today's user-generated media are destroying our economy, our culture, and our values*. New York: Crown Business. Available at: <https://books.google.de/books?id=Z59TDBx1U2UC> [Accessed 18 September 2015].
- Koehn, P. (2010) Chapter 5: Phrase-Based Models. Available at: <http://www.statmt.org/book/slides/05-phrase-based-models.pdf> [Accessed 18 September 2015].
- Labaka, G. et al. (2007) Comparing rule-based and data-driven approaches to Spanish-to-Basque machine translation. In *Proceedings of the MT Summit XI. European Association for Machine Translation*. Available at: http://doras.dcu.ie/15228/1/LabakaEtAl_summit_07.pdf [Accessed 18 September 2015].
- Lopez, A. (2008) Statistical machine translation. In *ACM Computing Surveys (CSUR) 40.3*, 8. Available at: <https://alopez.github.io/papers/survey.pdf> [Accessed 18 September 2015].
- Mayer-Schönberger, V. and Cukier, K. (2013) *Big data: A revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt. Available at: <https://books.google.de/books?id=uy4lh-WehhIC> [Accessed 18 September 2015].
- Merton, R. (1936) The Unanticipated Consequences of Purposive Social Action. In *American Sociological Review* 1.6. 894-904. Available at: <http://users.ipfw.edu/dilts/E%20306%20Readings/The%20Unanticipated%20Consequences%20of%20Purposive%20Social%20Action.pdf> [Accessed 18 September 2015].
- Merton, R. (1968) The Matthew Effect in Science. In *Science* 159.3810. Key concept source, 56-63. Available at: <http://www.garfield.library.upenn.edu/merton/matthew1.pdf>.
- Merton, R. (1988) The Matthew Effect in Science, II – Cumulative Advantage and the Symbolism of Intellectual Property. In *ISIS* 79. Key concept source, 606-623. Available at: <http://garfield.library.upenn.edu/merton/matthewii.pdf> [Accessed 18 September 2015].
- Monzó, E. (2011) Legal and translational occupations in Spain, Regulations and specialization in jurisdictional struggles. In Sela-Sheffy, R. and Shlesinger, M. (eds.) *Identity and Status in the Translational Professions*. Amsterdam: John Benjamins, 11-30. Available at: <https://books.google.de/books?id=KbZxAAAAQBAJ> [Accessed 18 September 2015].

- MOSES Project (2015) Welcome to Moses! (Statistical machine translation system). Community website. Available at: <http://www.statmt.org/moses/> [Accessed 18 September 2015].
- Nogueira de Andrade Stupiello, É. (2008) Ethical Implications of Translation Technologies. In *Translation Journal* 12.1. No longer available.
- Okpor, M.D. (2014) Machine translation approaches: issues and challenges. In *IJCSI International Journal of Computer Science Issues* 11.5, 159–165. Available at: <http://www.ijcsi.org/papers/IJCSI-11-5-2-159-165.pdf> [Accessed 18 September 2015].
- Pariser, E. (2011) *The Filter Bubble, how the new personalized web is changing what we read and how we think*. New York: Penguin. Available at: <https://books.google.de/books?id=wcalrOI1YbQC> [Accessed 18 September 2015].
- Ping, K. (1998) Machine Translation. In Baker, M.; Saldanha, G. (eds.) *Routledge Encyclopedia of Translation Studies*. 2nd Edition. London: Routledge, 162-170.
- Pym, A. (2013) Translation Skill-Sets in a Machine Translation Age. In *Meta* 58.3, 487-503.
- Rex, M. jr. (2013) Exploring the Intersection of Big Data and Machine Translation. TAUS blog post. Available at: <https://www.taus.net/think-tank/articles/translate-articles/exploring-the-intersection-of-big-data-and-machine-translation> [Accessed 18 September 2015].
- Rudavin, O. (2009) Current trends in the translation industry and what they mean to us all of us. In Baur, W. et. al (eds.) *Übersetzen in die Zukunft, Herausforderungen der Globalisierung für Übersetzer und Dolmetscher, Tagungsband der Internationalen Fachkonferenz des Bundesverbandes der Übersetzer und Dolmetscher e.V. (BDÜ)*. Vol. 32. Schriften des BDÜ. Berlin BDÜ, 69–75.
- Sandrini, P. (2013) Open Translation Data – Die gesellschaftliche Funktion der Übersetzungsdaten. In Mayer, F. and Nord, B. (eds.) *Aus Tradition in die Zukunft: Festschrift für Christiane Nord*. Berlin Frank & Timme, 27–37.
- Scholtes, J.C. (2010) Machine Translation that Works, Finally! Here is why and how... eDiscovery and Information Risk Management, Blog. Available at: <https://zylab.wordpress.com/2010/03/31/machine-translation-that-works-finally-here-is-why-and-how.../> [Accessed 18 September 2015].
- SDL plc. (2015a) Post-Editing Machine Translation Certification. Corporate website. Available at: <http://www.translationzone.com/learning/training/post-editing-machine-translation/> [Accessed 18 September 2015].
- SDL plc. (2015b) SDL BeGlobal, Cloud-based machine translation for high-volume, fast communication. Corporate website. Available at: <http://www.sdl.com/cxc/language/machine-translation/beglobal/> [Accessed 18 September 2015].
- Sela-Sheffy, R. (2011) Introduction: Identity and Status in the Translational Professions. In Sela-Sheffy, R.; Shlesinger, M. (eds.) *Identity and Status in the Translational Professions*. Amsterdam: John Benjamins, 1–9. Available at: <https://books.google.de/books?id=KbZxAAAAQBAJ> [Accessed 18 September 2015].
- Sismondo, S. (2010) *An introduction to Science and Technology Studies*. 2nd Edition. London: Wiley-Blackwell.
- Stichweh, R. (2005) *Die Soziologie der Professionen*. Working paper, Universität Bonn, Abteilung Demokratieforschung. Available at: <http://www.fiw.uni-bonn.de/>

- demokratieforschung/personen/stichweh/pdfs/38_die-soziologie-der-professionen-
2.pdf [Accessed 18 September 2015].
- TAUS (2015) Post-editing Course. Association website. Available at: <https://postedit.taus.net/post-edit/training-certification> [Accessed 18 September 2015].
- Touretzky, D. S. (2001) Viewpoint: Free speech rights for programmers. In *Communications of the ACM* 44.8. Extended online version, 23–25. doi: 10.1145/381641.381651. Available at: <http://www.cs.cmu.edu/~dst/DeCSS/Gallery/cacm-viewpoint.html> [Accessed 18 September 2015].
- Weizenbaum, J. (1976) *Computer power and human reason – from judgment to calculation*. New York: W.H. Freeman.
- Wijnands, P. (1993) Terminology vs. Artificial Intelligence. In Sonneveld, H.; Loening, K. (eds.) *Terminology – Applications in interdisciplinary communication*. Amsterdam: John Benjamins, 165-180.
- Wikipedia (2015b) Authors Guild, Inc. v. Google, Inc. In *Wikipedia, The Free Encyclopedia*. Wikipedia contributors (eds.) San Francisco, CA: Wikimedia Foundation Inc. Available at: https://en.wikipedia.org/wiki/Authors_Guild,_Inc._v._Google,_Inc [Accessed 18 September 2015].
- Wikipedia (2015c) Open-source Intelligence. In *Wikipedia, The Free Encyclopedia*. Wikipedia contributors (eds.) San Francisco, CA: Wikimedia Foundation Inc. Available at: https://en.wikipedia.org/wiki/Open-source_intelligence [Accessed 18 September 2015].
- Wikipedia (2015d) The Turk. In *Wikipedia, The Free Encyclopedia*. Wikipedia contributors (eds.) San Francisco: Wikimedia Foundation Inc. Available at: https://en.wikipedia.org/wiki/The_Turk [Accessed 18 September 2015].
- Wilss, W. (1996a) *Knowledge and Skills in Translator Behaviour*. Amsterdam: John Benjamins.
- Wilss, W. (1996b) Translation as intelligent behaviour. In Somers, H. (ed.) *Terminology, LSP, and translation: Studies in language engineering in honour of Juan C. Sager*. Amsterdam: John Benjamins, 161-168.
- Zetsche, J. (2005) *TM Marketplace White Paper, Sharing Translation Memory Data Aligned from Third-Party Documents: Legal Considerations*. Available at: <http://www.tmmarketplace.com/whitepapers/align.pdf> [Accessed 18 September 2015].