Of Dragons and Speech Recognition Wizards and Apprentices

Dragoș Ciobanu
Lecturer in Translation Studies
University of Leeds
Centre for Translation Studies

ABSTRACT

Despite being used by a minority of professional translators, Automatic Speech Recognition (ASR) systems have a lot of potential for effective integration into professional workflows. A recent survey designed in the University of Leeds Centre for Translation Studies and conducted among professional translators using ASR has led to interesting findings regarding, among others, the perception of translators who use the technology on the influence of ASR on the quality of their translated output, their productivity, and the fine-tuning of workflows in order to account for the advantages and disadvantages of adopting ASR. The survey showed that the advantages outweighed the disadvantages and that the majority of professional translators are missing out by not engaging with such technologies more.

Keywords: Professional Translation, Automatic Speech Recognition, ASR, Computer-Assisted Translation tools, CAT, translator productivity, translation quality.

RESUM *(De Dragons i el reconeixement de veu, bruixots i aprenents)*

Tot i que tan sols els utilitzen una minoria de traductors professionals, els sistemes de Reconeixement Automàtic de Parla (RAP) tenen un gran potencial si s'integren de manera efectiva en els fluxos de treball professionals. Al Centre d'Estudis de Traducció de la Universitat de Leeds s'ha elaborat recentment un qüestionari dirigit a traductors professionals que utilitzen RAH. Aquest qüestionari ha donat resultats molt interessants en relació amb, entre d'altres coses, la percepció que tenen els traductors sobre l'efecte del RAP en la qualitat de la traducció, la productivitat i el flux de treball per tal d'avaluar els avantatges i inconvenients del ús d'aquesta tecnologia. El qüestionari ha mostrat que els avantatges superen els inconvenients, avantatges que la majoria dels traductors professionals s'estan perdent per no adoptar aquesta tecnologia.

Paraules clau: traducció professional, Reconeixement Automàtic de la Parla, RAP, eines de Traducció Assistida per Ordinador, productivitat del traductor, qualitat de la traducció.

RESUMEN *(De Dragones y el reconocimiento de voz, sus brujos y aprendices)*

A pesar de que tan solo los utilizan una minoría de traductores profesionales, los sistemas de Reconocimiento Automático de Habla (RAH) tienen un gran potencial si se integran de manera efectiva en los flujos de trabajo profesionales. En el Centro de Estudios de Traducción de la Universidad de Leeds se ha elaborado recientemente un cuestionario dirigido a traductores profesionales que utilizan RAH. Este cuestionario ha arrojado resultados muy interesantes en relación con, entre otros aspectos, la percepción que tienen los traductores sobre el efecto del RAH en la calidad de la traducción, la productividad y el flujo de trabajo con el fin de evaluar las ventajas e inconvenientes del uso de esta tecnología. El cuestionario ha mostrado que las ventajas superan a los inconvenientes, ventajas que la mayoría de los traductores profesionales se están perdiendo por no adoptar esta tecnología.

Palabras clave: traducción profesional, Reconocimiento Automático de Habla, RAH, herramientas de Traducción Asistida por Ordenador, productividad del traductor, calidad de la traducción.

Introducción

The quality of output in the Language Services Industry depends on a delicate and complex combination of human talent, training, experience, specialism, teamwork, optimised workflows, top-of-the-range technology and ergonomic professional set-ups. In recent decades the work of professional translators has been influenced positively by increased exposure to domain-specific knowledge and expertise through the rise of Massive Open Online Courses (MOOCs), as well as collaborative practices through the sharing of Translation Memories (TMs) and Terminology Databases (TDs). Moreover, continuously developing technologies such as Machine Translation (MT) and certain user-friendly Computer-Assisted Translation (CAT) tools are appreciated by some translators when used for quality-enhancing or productivity-enhancing reasons rather than cost-saving ones.

However, historically the situation has not been as successful in the case of Automatic Speech Recognition (ASR), which aims to take spoken language as input and transcribe it into text form verbatim. At least on health grounds, this is unfortunate. Growing pressures on productivity have led to longer working hours in front of keyboards and Visual Display Units (VDUs), incorrect posture, eye strain, discomfort and even injuries affecting wrists. These represent stress factors eventually impacting on the quality of translated output, the well-being of translators, and possibly on their longevity in the profession.

This article will report on a survey designed in 2014 in the University of Leeds Centre for Translation Studies (CTS) with a focus on the profile, motivation and practical experience of professional translators using Automatic Speech Recognition (ASR) systems in their work. The reported influence of this technology on the quality of the translators' output will be discussed alongside a comprehensive picture of the profile of these translators and their technological set-ups. I will also discuss if ASR can be a magic wand in the search for quality and the extent to which being an experienced user – a Wizard if you will – makes a difference compared to the way in which Apprentices integrate ASR in their work.

ASR and MT: two not-so-distant relatives

Although using slightly different technologies – but fairly similar approaches in collecting data and training their systems – Automatic Speech Recognition (ASR) systems represent a natural area of research and development alongside Statistical Machine Translation (SMT) in the quest for the next best thing after the Babel fish, the universal (though, sadly, fictional) translator/interpreter described in Douglas Adams' Hitchhiker's guide to the Galaxy (Adams, 1980, p. 49).

Automatic speech-to-speech (s2s) translation has been one of the main topics of linguistic and technological research and the three elements involved in such systems have been steadily developing over the years – albeit generally along separate paths. First of all, ASR, also known as speech-to-text technology, which takes spoken input and transcribes it automatically into text form – the slowest to progress and still not available for the majority of languages. Secondly, Machine Translation for text-to-text translation – nowadays users can choose between SMT, Rule-Based Machine Translation (RBMT) and Hybrid Machine Translation (HMT). Finally, text-to-speech synthesis which takes place at the end of the process, when the target text is spoken out by the computer – this was one of the first elements to be developed across several languages given its paramount importance in making operating a computer accessible to blind or visually-impaired users, as well as in the reading of interlingual subtitles in an audio-description environment.

Just as SMT was boosted by the release of the *free and open-source* MOSES platform (Koehn et al., 2007), ASR has benefitted significantly from open-source platforms such as VoxForge (*VoxForge*, n.d.), and toolkits such as CMU Sphinx (Lamere et al., 2003) or Julius (*Julius*, n.d.). Thanks to such projects, acoustic models can be crowd-sourced for languages which are considered non-commercially viable by software manufacturers. In the case of ASR provided by Microsoft in more recent versions of its Windows operating system, only the following languages are supported: American and British English, Spanish, French, German, Simplified and Traditional Chinese (Duarte, Prikladnicki, Calefato, & Lanubile, 2014). Moreover, Nuance's market-leading Dragon Naturally Speaking adds Italian, Dutch and Japanese to the list, but does not support Chinese in its desktop client version (Nuance, n.d.).

In light of the availability of ASR for some years, it is surprising that the research world has not invested much effort in investigating the impact that ASR can have on the quality of human translation output, as well as on workflows, ergonomics and productivity. By comparison, much effort has been expended in the area of MT on investigating variables such as language pair and direction, text genre, source text pre-editing, type of technology – SMT v RBMT v HMT – training data gathering, evaluation of output quality, MT post-editing, and regular workshops on these topics are being organised several times a year by one of the three chapters – for Europe, the Americas, and Asia-Pacific – of the International Association for Machine Translation (IAMT). With a few notable exceptions (Leijten and Van Waes, 2005; Leijten, Janssen and Van Waes, 2010 in Mees, Dragsted, & Jakobsen, 2013, p. 142), it seems that most of the researchers presenting at events such as the International Workshop on Spoken Language Translation (IWSLT) have directed their focus on ASR as part of the complete speech-to-speech (s2s) ideal.

This research will take a step back from this and focus on the practical application of the latest ASR technology as a possible means to increase translation output quality through integration into the translation workflow, its positive impact on productivity, the revision process, and also its extremely important influence on ergonomics and occupational health.

The physical benefits of using ASR for professional linguists

Professional translation is more often than not a sedentary activity subject to the same unfortunate consequences as other office-based roles. Repetitive Strain Injury – RSI – is one such consequence that can impact on productivity. RSI has been reported since the 1980s (Kiesler & Finholt, 1988; Pfeiffer & Heintzelman, 1997; Ratzlaff, Gillies, & Koehoorn, 2007; van Tulder, Malmivaara, & Koes, 2007) though its connection to computer work has occasionally been called into question (Mediouni et al., 2014). Other occupational risk factors for translators include neck and back pain (Paksaichol, Janwantanakul, & Lawsirirat, 2014), and eye strain (Nixon, Mazzola, Bauer, Krueger, & Spector, 2011, p. 4).

Women seem to be more likely to develop RSI (Colomb-Lippa & Wilson, 2011; Paksaichol et al., 2014), which is particularly significant in the case of our profession in which women represent – at least in Europe – over 70% of the workforce (Pym, Grin, Sfreddo, & Chan, 2012, p. 3). According to the same study, full-time professional translators spend over 4 hours in front of computers (ibid, p. 102). This is likely to increase the risk of eye-strain (Pfeiffer & Heintzelman, 1997, p. 296).

Finally, such work-based injuries and illnesses do not only affect the translators in question and the quality of their own work. They also have wider-ranging social consequences (Dembe, 2001). It is therefore quite surprising that translators do not use voice recognition systems as a matter of course. The explanation may lie in the history of speech recognition usage.

A brief (and patchy) history of using speech recognition by professional translators

A survey commissioned in 2001 by the ITI to which there were 519 responses highlighted that a "practically static minority of respondents (compared to 1998) where typists are concerned and a growing minority where speech recognition is concerned" used dictation in their professional work (Aparicio, Benis, & Cross, 2001, p. 16). Very interestingly, ever since 2001, the professional world has known that dictation is "the only productivity aid that makes a truly substantial difference to output and income, enabling experienced translators to leverage their expertise to the full" (ibid.), yet the adoption of this technology has been far from overwhelming.

In terms of the workflow used by professionals, the same report found that, on the one hand "the gross income for those using speech recognition and a typist exceed[ed] the income for those using a typist alone" (ibid.) and, on the other hand, that in 2001 "translators tend[ed] to achieve higher outputs using a typist" (ibid.). In terms of figures, out of all the survey respondents, approximately 425 respondents said they did not use dictation in their work, while approximately 40 said they used Speech Recognition, approximately 30 used a typist, and fewer than 10 used both methods.

In terms of productivity, translators who did not use ASR or a typist reported an output of approximately 3,000 words/day on average. If ASR was used, this reported output would grow to 3,100 words. If both ASR and a typist were used, the daily output would be near 3,500 words. Sending recorded tapes to a typist for transcription appeared to be the most productive method, however, with reported output nearing 4,000 words/day (ibid).

These very small differences in output could be explained in part by the accuracy of ASR technology which, in 2001, was still in its infancy. On the other hand, it is also worth noting that 54.5% of survey respondents said that translation was not their main source of income (ibid.), and therefore their motivation to set up an effective technological work environment could have been minimal at a time when Translation Memory (TM) was still used by a minority of professionals – 30%.

Ten years later, over 1,750 responses were submitted to a joint CIoL and ITI survey (CIOL & ITI, 2011, p. 4) by professional linguists who were significantly more involved with technology – 55% of respondents used CAT tools (ibid, p. 17). Nevertheless, ASR remained the tool of choice for a minority of under 10%: out of 123 users, 94% were using Dragon Naturally Speaking (ibid, p. 18).

The use of ASR by professional translators at present

So far we have seen that ASR has the strong potential to address – and indeed, in some cases it is recommended in order to prevent (Aparicio et al., 2001, p. 16) – medical conditions such as RSI, eye strain, back pain, and shoulder and neck pain. Nevertheless, this technology is still only used by a very small minority of professional translators.

A project was therefore initiated in 2014 at the University of Leeds Centre for Translation Studies (CTS) with University of Leeds Ignite funding to research two questions: first of all, why and how do professional translators use ASR at present? Secondly, to what extent are professionals making the best use of all the features of the latest ASR tools – are we seeing another case of 80% of translators using 20% of the features, as is still anecdotally reported in the case of Computer-Assisted Translation (CAT) tools? This paper will report on the findings to the first research question and only provide a brief preview of the possibilities of the latter.

An online survey (Ciobanu, 2014) was made available for one month and several communities of professional translators were contacted through three online communities: the Speech Recognition for Translators Yahoo group, the Professional Translators and Interpreters (Proz.com) LinkedIn group, and the memoQ users Yahoo group. Within that

month, 47 replies were received which is a remarkably high number given that 123 ASR users had replied to the much more visible 2011 CIoL/ITI survey. Upon closer scrutiny, 6 answers were discarded as they were either duplicates or not given by active ASR users, but by professionals who had concerns about using ASR or had used ASR but then abandoned it. Nevertheless, their comments were still saved for the qualitative analysis of the professionals' attitudes towards this technology. Moreover, the respondents were not asked whether they belonged to any professional organisations. This means that the overlap between the CIoL/ITI survey respondents and the CTS respondents cannot be established.

A range of very interesting elements emerged, starting with the profile of the ASR users. The group whose experience ranged from between 11 and 25 years appeared to be employing ASR the most (Figure 1). On the other hand, there was only 1 freelancer with less than 1 year of experience who was an ASR user and who answered the questionnaire. This in itself raises questions about the prominence given to ASR in training programmes across the world in the countries whose languages are well-supported by speech recognition engines. The University of Leeds CTS experience is that ASR has been a core component of the MA in Audiovisual Translation Studies given that live subtitling is dependent nowadays on such technologies. Moreover, from October, 2014, our newest computer lab has been equipped with Dragon Naturally Speaking Professional v.13 which gives all our MA students the chance to train on combining the currently most advanced version of the most heavily-used ASR system with subtitling, as well as CAT tools, and thus be better equipped for the professional world.
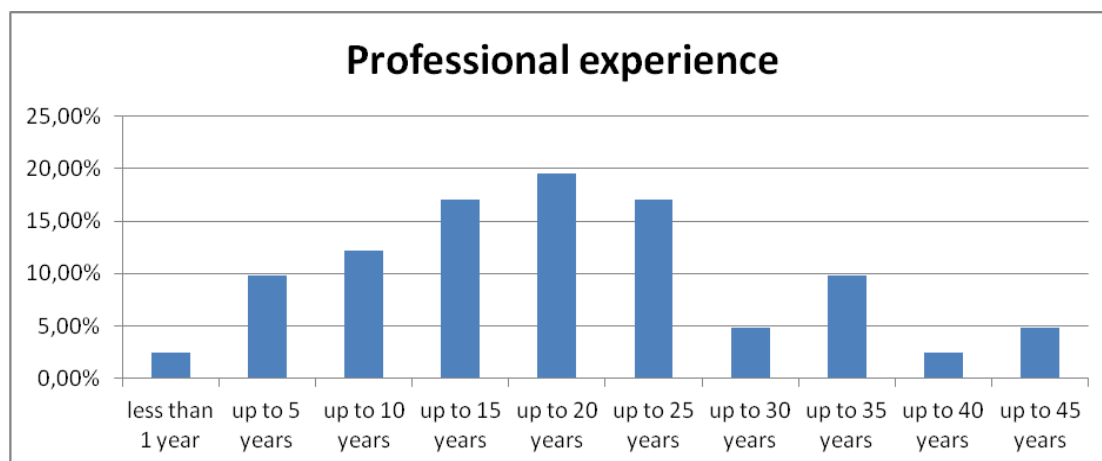

Figure 1: ASR users' professional experience

Coming back to the profile of the professionals using ASR, only 1 person reported having a disability requiring such assistive technology. The rest of the group had chosen ASR for a variety of reasons ranging from a minority of participants recovering from injuries to a majority who had adopted ASR because of improved translation quality, productivity, and a better work environment.

Moreover, for just under 81% of respondents translation was the only source of income, which suggests that full-time professional translators are becoming increasingly willing to adopt new technologies which aid both productivity and health. In terms of the operating systems on which the translators use ASR for their work, Microsoft Windows Desktop was clearly ahead, with 100% of respondents working on it; Apple iOS was mentioned by 2 translators, and Chrome OS, Mobile – Android, Mobile – Apple, and Mobile – Windows were all listed once. Finally, over 95% work with ASR in their own office, although two respondents work in shared offices.

When it came to the ASR system chosen by the professionals, Nuance's Dragon Naturally Speaking (DNS) was everyone's favourite, with only 7 respondents using another ASR tool, as well, to complement DNS (Figure 2).
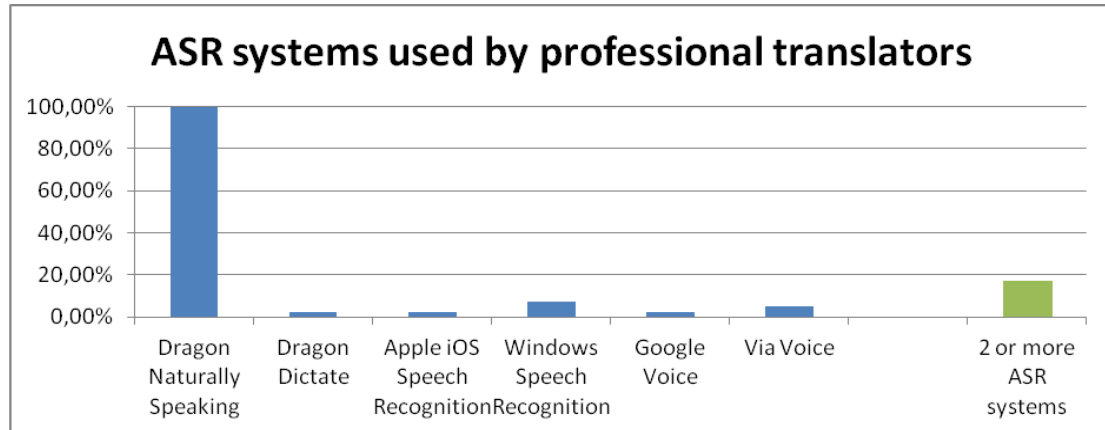

Figure 2: ASR systems used by professional translators

Regarding the languages of the respondents, there was an interesting variety of source languages, with Dutch, English, French, Galician, German, Italian, Japanese, Norwegian, Portuguese, Russian, Spanish, and Swedish being all represented. In terms of the target languages, though, using ASR into English appeared to be most frequent by far, although Dutch, French, German, Italian and Spanish were also mentioned (Figure 3).

This situation is also explained by the increased morphological and syntactic difficulties posed by those languages by comparison to English. If we are only to take the example of French, having silent endings, singular and plural form differences, and a variety of conjugations was reported by respondents to pose significant difficulties to Dragon Naturally Speaking and lead to frequent ambiguities and insertions of the wrong homophone. This in turn requires more thorough revision processes in order to ensure translation quality and is an additional challenge which not every translator feels able to deal with – particularly when it comes to correcting other translators' wrong homophones as part of a revision or – even more difficult – review process as defined by the EN15038 standard (CEN, 2006).
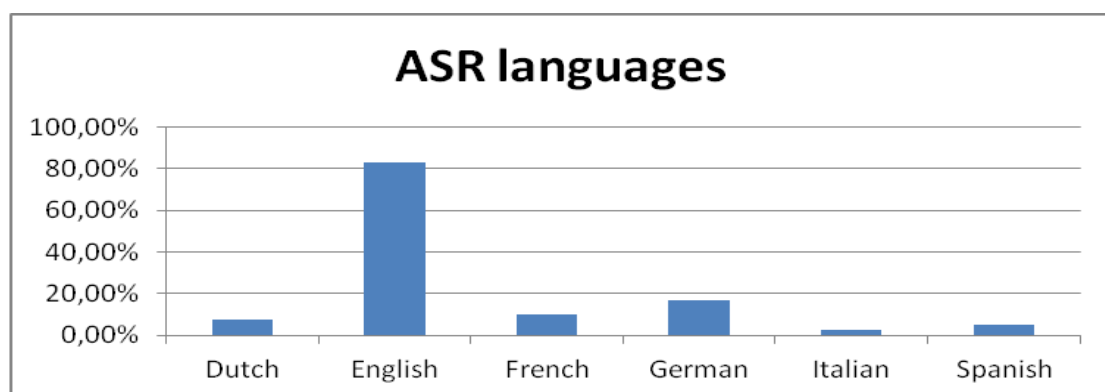

Figure 3: Languages in which ASR is used by professional translators

When taking into account the years of experience as professional translators and the years of using ASR – mainly Dragon Naturally Speaking (DNS), as this was the one tool which all the respondents had – there were a few interesting exceptions such as the heavy users, the Wizards if you will, who had managed to tame Dragon – with ratios of years of professional experience to years of DNS use looking like 5/8 (this particular translator chose this profession after having used DNS for three years in a previous capacity), 45/10, 35/15, 25/20, 15/12, 15/10, 10/7 – or the late adopters, the Apprentices who were just starting out, but were already seeing benefits – with ratios of 10/0.5, 20/0.4, or 35/0.5.

A general analysis of the average years of DNS use by category of professional experience indicates an upward trend up to 25 years of professional experience (Figure 4). The picture is not as clear for the group of even more experienced translators – the small sample size does not confirm whether they are indeed more sceptical about the advantages of ASR.

Nonetheless, despite the low amount of data, these statistics are still a good starting point for gaining an insight into the attitudes and technological set-up of some professional translators nowadays. The data suggests that the less experienced translators tend to stay away from ASR at the beginning of their careers.
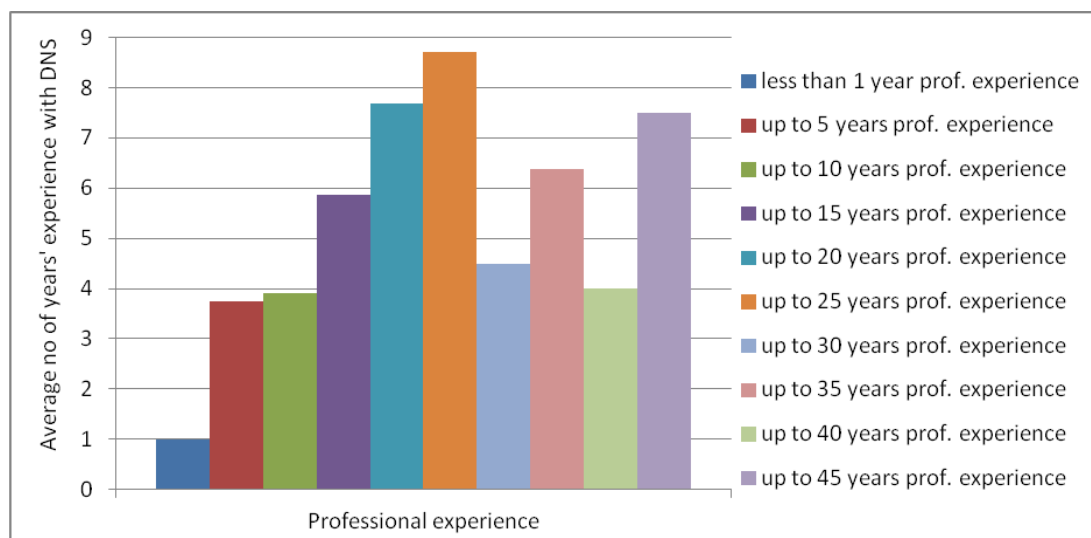

Figure 4: Years of using Dragon Naturally Speaking by years of experience as professional translators

ASR: not quite the promised magic wand and possibly a wolf in sheep's clothing?

Before looking at some of the tasks which translators listed as performing with ASR, as well as evaluating how the use of ASR is impacting on the quality of their work, a useful question to answer is how much the professionals use their speech recognition systems every day. This is particularly important given that it will be rather counterproductive to replace strained wrists, necks and backs with strained voices, as it can happen in other professions such as those belonging to the education sector (Ferracciu & Almeida, 2014).

An advantage translators have over teachers, though, is that they can invest in a good-quality microphone to use with their ASR systems – USB or Bluetooth microphones are among the best and it is worth following the manufacturers' advice regarding supported microphone headsets. This way they will not need to strain their voice as much to use the ASR commands. Talking to a computer is in some ways similar, but in many others different

from talking to humans, and it is also worth exploring voice coaching for professional translators in order to make the most of the ASR options available. For instance, speech recognition systems, just like humans, can disambiguate continuous, clearly-articulated speech much better than words enunciated one at a time. On the other hand, unlike humans, the attention of ASR systems does not need to be engaged by varying speech speed and pitch, and introducing jokes, repetitions and sound effects into one's monologues; in fact, the more monotonous and hesitation-free one's speech is, the better the ASR results are. However, explaining how to speak for ASR is beyond the scope of this article – for a comprehensive overview of such techniques, please see Romero-Fresco (2011).

With the exception of a spike in reported daily DNS usage by professionals with up to 10 years of experience – and with the proviso that more data is needed for broader generalisations – there seems to be an upward trend of working more with DNS the more established a translator is (Figure 5). This advantage of being able to relax more during work and to dictate continuously for longer without feeling fatigued was mentioned several times in the respondents' qualitative comments.
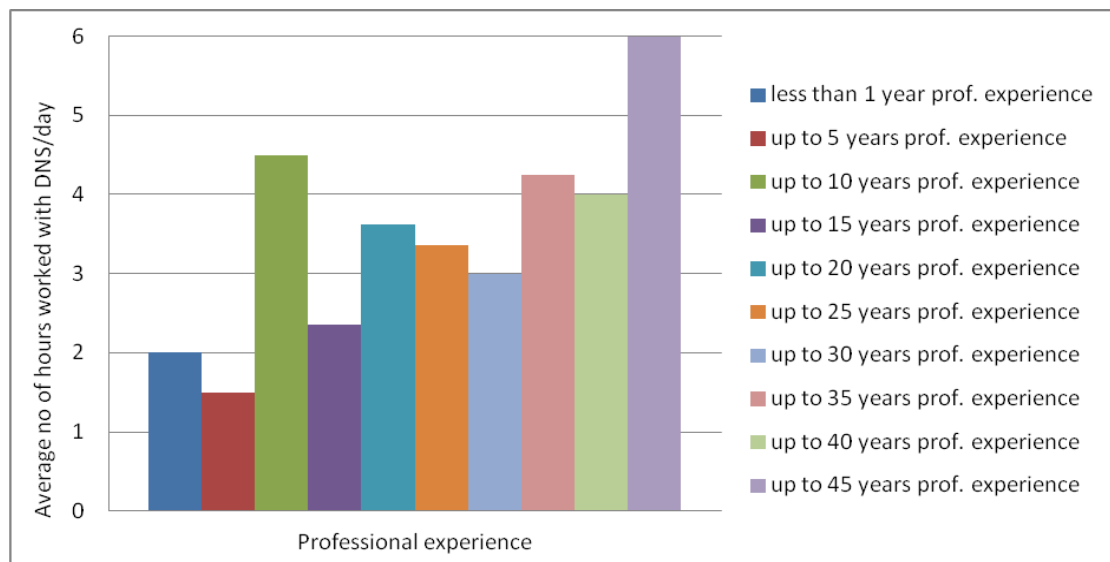

Figure 5: Hours per day of DNS use by years of experience as professional translator

Using ASR for common professional translator activities

To gain a deeper understanding of the role which ASR plays in the work of professional translators, the questionnaire asked whether this technology is being used for a range of activities, including translation, proofreading, and review as defined by the EN15038 standard (CEN, 2006), as well as transcreation and Machine Translation (MT) post-editing, within desktop or cloud-based word-processors and/or CAT tools. Other uses of ASR which are relevant for professional translators were also examined, from operating the entire computer by voice to operating individual relevant applications such as e-mail clients.

The majority of respondents stated that they used ASR for dictating translations into desktop-based word processors and CAT tools, as well as for managing e-mails. The survey makes it clear that combining speech recognition and cloud-based tools is not yet part of the professional translators' usual set-up. The same is also true of MT post-editing with ASR. This is therefore an area needing further research.

Figure 6 presents how many of the respondents used ASR with each desktop CAT tool that was mentioned. It is interesting to see that, while memoQ does not have a market share as large as SDL Studio, it appears to be used by the majority of professional translators who also use ASR in their daily work – 63.41%. Moreover, a separate analysis of the data also shows that 31.71% of respondents had used ASR with 2 or more CAT tools.

In addition, the survey gathered further information on whether the choice of CAT tool(s) installed on the professionals' machines had been influenced in any way by the integration with ASR, as well as whether there were any CAT tools installed on the respondents' machines which had not been used with ASR and why. Future articles will analyse these responses.
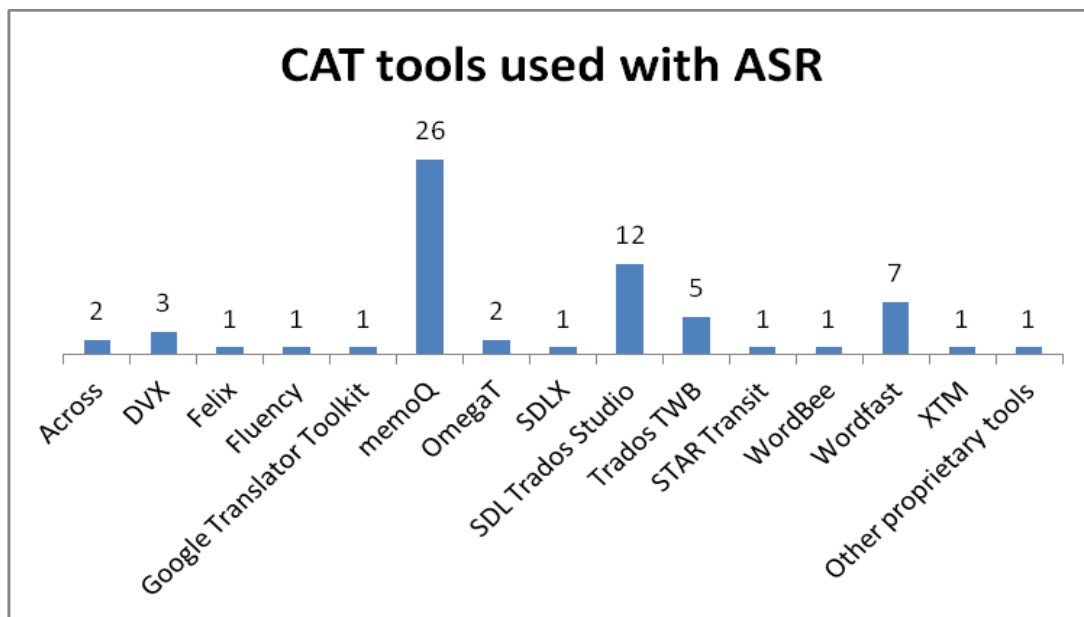

Figure 6: CAT tools used with ASR

Influence of ASR on speed and quality of translation output

Regarding the influence of ASR systems on the speed and quality of the translation output, as well as on translation workflows, the picture varies somewhat as can be seen in Figure 7. Overall, there was a general consensus within the professional experience groups that ASR does have a positive impact on productivity. When all the responses from all the professional experience groups are combined, 87.8% of respondents report between 10% and 500% increase in productivity when using ASR. The average reported productivity increase is 110.56%, though a more insightful figure in this case is the median. This is 35%. This represents the majority of productivity increases and is a much more realistic figure to expect when using ASR – at least when starting out.

Of course, there are further qualifiers for these figures – such as text genre, length of segments, and workflow set-up – which will be discussed in further articles. However, it is still encouraging to see the productivity figure increase so significantly from the approximately 3% reported in 2001 (Aparicio et al., 2001, p. 16).

The increased accuracy of ASR technology may account for most of this 10-fold+ improvement, while the reported need for high-performance computers to run this technology effectively must still be one of the biggest hurdles adopters need to overcome. There is also a certain element of myth lingering in this world of magic technology and super-users brushing

shoulders with late adopters: many times someone's negative experience with an old version of ASR software is still quoted without investigating the progress that tools such as Dragon Naturally Speaking have made especially in the case of English speech recognition.

Comments regarding sub-standard recognition quality and lengthy review processes ignore the fact that new vocabulary can be easily extracted and trained within DNS from natural speech, previously-written documents and e-mails, or even Translation Memories (TMs) by exporting the TMs and deleting the source segments from the resulting file. These individual opinions also ignore the presence of additional tools such as a text-to-speech synthesiser which can make the revision process in DNS more user-friendly and effective.
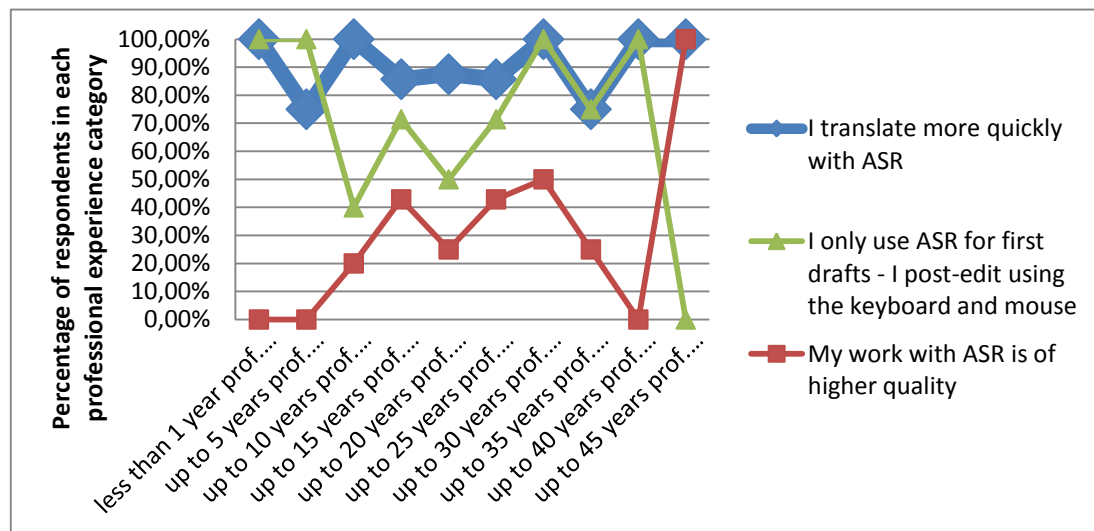

Figure 7: Impact of ASR on speed, quality and translation workflows

In terms of the impact of speech recognition on translation quality, Figure 7 is not as evocative. In fact, the picture seems to be slightly different. There appears to be a trend to rate the quality of the ASR output more highly the more experience the freelancer has with such software. At the same time, the use of ASR only for first drafts which then are post-edited by hand also appears to decrease the more experience the professional translator has with speech recognition tools such as Dragon Naturally Speaking, whose quality in the languages it supports seems to be constantly and rapidly improving – Figure 8 is representative in this respect.
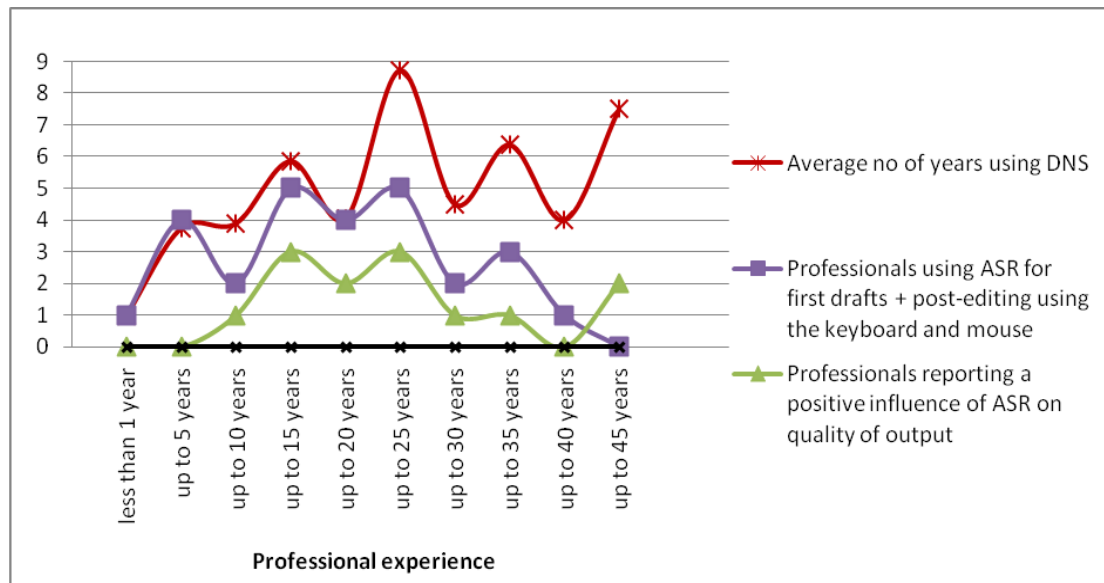
Figure 8: Relationship between DNS experience and: 1. the professional's view on translated output quality with ASR, and 2. the use of ASR only for drafting purposes

Still on the topic of quality, a few important aspects were also listed by the survey respondents, ranging from the integration of ASR and CAT tools to the impact it has on the revision process. First of all, as mentioned briefly already, just like in the case of using MT, the size of the ASR vocabulary needs to be constantly expanded for best results. In particular, speech engines need to be trained to recognise named entities – proper names – which are also a topic of MT research. Once trained, the ASR is more likely to write them correctly in the future.

Secondly, dictating the translation conflicts with the predictive typing functionality which more and more CAT tools integrate nowadays. Consistency of terminology or phrasing in the translated output can therefore be adversely affected if appropriate predictive typing suggestions come from internal databases – such as Muses in memoQ or AutoSuggest dictionaries in SDL Trados – which are not visible to the translator in the same way that TM or TD suggestions are. In addition, some translators do not find the integration between ASR and their CAT tools particularly smooth and this is why most report using their speech recognition systems just for dictating and not for any particular CAT tool-specific QA processes.

Thirdly, another very important aspect is that, in order to ensure the quality of the final translation, integrating ASR into the professionals' workflows does have a major impact on the duration and scope of the proofreading stage. While touch-typists are happy that speaking their translations addresses their fear of introducing typos while touch-typing and concentrating on the source text, a large portion of respondents did emphasise how using ASR runs the significant risk of eliminating typos only to replace them with "speakos" – as one respondent called the homophones or other lexical items introduced as a result of hesitation, ambient noise, or even hardware and software failures.

Although it is not a widely-spread practice among users of ASR, turning to text-to-speech technologies was reported as being an effective method of catching "speakos". What happens in this approach is that the computer is asked to read back to the translator his or her own translation. Hearing the translation represents a good way to spot linguistic problems in the output – after all, this methodology is not too dissimilar to what happens at present in International Organisations such as the European Space Agency where, for reasons of speed

and efficiency, the traditional Translation – Editing – Proofreading (TEP) model where translators, revisers and reviewers all use Track Changes and pass documents between them has been replaced by a Translation + Face-to-face Review model. In this process the translator has his/her translation read back to him/her by a colleague. This way the colleague is doing a monolingual target language review at the same time as the translator is doing a bilingual revision.

Of course, the approach of using a text-to-speech engine on one's computer lacks the advantage of receiving insightful comments from a human reviewer, but is still a proven way of making the revision process more effective, thus enhancing the quality of the translated output.

Interestingly, the thorough revision stage is not only needed to correct homophones and other "speakos", but also to make significant stylistic corrections, as translators reported producing more informal spoken texts whose register needed modifying when revising. This tendency to simplify, use more general language, and lower the register of the output is also reported in the literature investigating interpreting practices (Chafe and Danielewicz, 1987, p. 86 in Dragsted, Mees, & Hansen, 2011, p. 13). As more effective methods of integrating ASR, CAT tools and multilingual text-to-speech are found, research on this topic will be of great relevance to professional translators.

Conclusions

Automatic Speech Recognition systems have without doubt significant potential to make a positive contribution to the quality and ergonomics of the work of professional translators. So far few studies have looked at how these individuals use such systems in authentic professional set-ups. Moreover, while some recent studies on productivity have shown a growing interest from academia in this area, they have also suffered from certain design flaws which raise questions over the applicability of their findings (Dragsted et al., 2011; Mees et al., 2013) – e.g. the testing of ASR with students who are non-native speakers of the languages tested, and whose study set-up differs significantly from that of professional translators.

It is interesting to notice the impact that experience with ASR technology has on the use of ASR for drafting purposes only or for other tasks, too. It is also a shame that working with speech recognition remains – with few notable exceptions – the concern of Continuous Professional Development offered by professional organisations such as the ITI and YTI in the UK, rather than university training. All in all, it is high time the profession paid attention to findings going back to 2001 and belonging to the translation and applied translation studies sectors rather than the pure research one, and started investigating the latest speech recognition technologies thoroughly.

Furthermore, apart from influencing the translation drafting process and the extent to which speech recognition adopters can make the most of all the quality-enhancing functionalities offered by their CAT tools, ASR is already redefining to a significant extent some of the quality-assurance workflows. The problems which such systems introduce, together with the ones they solve, are valuable areas of Translation Process Studies research and will constitute the subject of further articles reporting on how professional translators embed ASR into their existing authentic workflows.

Future work

The second component of the research project started in the University of Leeds Centre for Translation Studies and funded through the local Ignite initiative has been investigating effective ways of embedding ASR into the work of professional translators, moving beyond simple dictation. We are still quite a long way away from freeing translators from their work

stations – 90.24% of respondents to our survey reported using ASR in combination with a keyboard and mouse, and only one participant indicated the use of ASR on its own to control the entire computer.

Advanced versions of Dragon Naturally Speaking do offer a lot of potential for hands-free work even in more complex scenarios where CAT-tool-specific keyboard combinations and functions are required, and it is certainly worth investigating the CAT features which professionals use most frequently in their work to see whether voice commands can be integrated effectively into current workflows.

Moreover, by combining CAT, ASR, and text-to-speech engines, the traditional translation workspace could become more like a consecutive interpreter's, as briefly demonstrated in a 2013 prototype (Ciobanu, 2013). This will undoubtedly have impact on translation quality and workflows, and we may start to see a change in the trend of reporting low lexical variety statistics when comparing translations to originals / source texts / start texts – depending on which terminology you prefer.

MT is also likely to have several applications in these new workspaces. Source text data either in its raw format or translated on the fly with MT has already been proposed to help increase the quality of ASR (Reddy & Rose, 2010). As the market of MT post-editing continues to grow, it will be useful to research effective ways of working with speech recognition on this task within CAT tools, in addition to the effectiveness of using such technologies for post-editing TM fuzzy matches. The task of objective data gathering will not be complete, however, without effective logging of "User Activity Data" (Moran, Saam, & Lewis, 2014) within desktop or cloud-based CAT tools, so it is important to support existing research in this area, too.

Another line of research could be based on the fact that ASR on mobile devices was mentioned in the survey responses, and is currently being investigated by other researchers (Zapata, 2014). It will therefore be useful to learn more about how professional translators integrate these platforms into their workflows, and what impact in terms of quality, ergonomics and productivity these have.

Last, but by no means least, the uneven coverage of languages by ASR systems should be one of the main topics of research and development among linguists, technologists and tools providers alike, and I am personally very much looking forward to the day when I will be able to dictate directly in Romanian. Even if the largest ASR providers still prioritise their development according to financial returns, until they are persuaded to change their position, there is still a lot of scope to harness both the wisdom and the power of the crowds to build quality acoustic models in languages not currently supported by speech recognition systems.

Acknowledgements

Author affiliation

Dr. Dragoș Ciobanu, Lecturer in Translation Studies, University of Leeds Centre for Translation Studies; e-mail: d.i.ciobanu@leeds.ac.uk

Brief author biographical notes

I am a Lecturer in Translation Studies specialising in Computer-Assisted Translation – CAT – tools, collaborative translation practices and using ASR in professional linguists' work, and I am also an instructional designer using my background in languages and technology to produce effective multilingual, multimedia online resources.

Bibilography

Adams, D. N. (1980). *The Hitchhiker's Guide to the Galaxy* (8th ed.). London and Sydney: Pan Books.

Aparicio, A., Benis, M., & Cross, G. (2001). *ITI 2001 Rates & Salaries Survey*.

CEN, E. C. for S. (2006). *European Standard EN 15038:2006. Translation Services - Service requirements*. Retrieved from http://www.babelia.pt/media/norma_en_15038.pdf

Ciobanu, D. (2013). *Demonstration of multilingual TTS, memoQ and DNS integration*. Retrieved from https://www.youtube.com/watch?v=SHRuQMrvhRY

Ciobanu, D. (2014, October). UoL Centre for Translation Studies Speech Recognition usage survey [Online survey]. Retrieved December 22, 2014, from http://goo.gl/XC7wVB

CIOL & ITI. (2011). *2011 Rates and Salaries Survey for Translators and Interpreters*.

Colomb-Lippa, D., & Wilson, S. (2011). Carpal Tunnel Syndrome. *Journal of the American Academy of Physician Assistants*, *24*(1), 49.

Dembe, A. E. (2001). The social consequences of occupational injuries and illnesses. *American Journal of Industrial Medicine*, *40*(4), 403–417.

Dragsted, B., Mees, I. M., & Hansen, I. G. (2011). Speaking your translation: students' first encounter with speech recognition technology. *The International Journal for Translation & Interpreting Research*, *3*(1), 10–43.

Duarte, T., Prikladnicki, R., Calefato, F., & Lanubile, F. (2014). Speech Recognition for Voice-Based Machine Translation. *IEEE Software*, *31*(1), 26–31.

Ferracciu, C. C. S., & Almeida, M. S. de. (2014). Occupational voice disorders in teachers and current legislation. *Revista CEFAC*, *16*(2), 628–633.

Julius. (n.d.). Retrieved from http://julius.sourceforge.jp/en_index.php

Kiesler, S., & Finholt, T. (1988). The mystery of RSI. *American Psychologist*, *43*(12), 1004–1015.

Koehn, P. et al. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions* (pp. 177–180). Stroudsburg, PA, USA: Association for Computational Linguistics.

Lamere, P., Kwok, P., Walker, W., Gouvêa, E. B., Singh, R., Raj, B., & Wolf, P. (2003). Design of the CMU sphinx-4 decoder. In *INTERSPEECH*. Citeseer.

Mediouni, Z., de Roquemaurel, A., Dumontier, C., Becour, B., Garrabe, H., Roquelaure, Y., & Descatha, A. (2014). Is Carpal Tunnel Syndrome Related to Computer Exposure at Work? A Review and Meta-Analysis: *Journal of Occupational and Environmental Medicine*, *56*(2), 204–208.

Mees, I. M., Dragsted, B., & Jakobsen, A. L. (2013). Sound effects in translation. *Target*, *25*(1), 140–154.

Moran, J., Saam, C., & Lewis, D. (2014). Towards desktop-based CAT tool instrumentation. In *Third Workshop on Post-Editing Technology and Practice* (pp. 99–112). Vancouver: AMTA.

Nixon, A. E., Mazzola, J. J., Bauer, J., Krueger, J. R., & Spector, P. E. (2011). Can work make you sick? A meta-analysis of the relationships between job stressors and physical symptoms. *Work & Stress*, *25*(1), 1–22.

Nuance. (n.d.). *Dragon Naturally Speaking*. Retrieved from http://www.nuance.com/index.htm

Paksaichol, A., Janwantanakul, P., & Lawsirirat, C. (2014). Development of a Neck Pain Risk Score for Predicting Nonspecific Neck Pain With Disability in Office Workers: A 1-Year Prospective Cohort Study. *Journal of Manipulative and Physiological Therapeutics*, *37*(7), 468 – 475.

Pfeiffer, P., & Heintzelman, M. (1997). Machines, Statues, and People: Strategies for Promoting RSI Awareness in Computing Curricula. *SIGCSE Bull.*, *29*(1), 296–300.

Pym, A., Grin, F., Sfreddo, C., & Chan, A. L. J. (2012). *The Status of the Translation Profession in the European Union* (No. DGT/2011/TST) (p. 172). Luxembourg: European Commission. Retrieved from http://ec.europa.eu/dgs/translation/publications/studies/translation_profession_en.pdf

Ratzlaff, C. R., Gillies, J. H., & Koehoorn, M. W. (2007). Work-related repetitive strain injury and leisure-time physical activity. *Arthritis Care & Research*, *57*(3), 495–500.

Reddy, A., & Rose, R. C. (2010). Integration of Statistical Models for Dictation of Document Translations in a Machine-aided Human Translation Task. *Trans. Audio, Speech and Lang. Proc.*, *18*(8), 2015–2027.

Romero-Fresco, P. (2011). *Subtitling Through Speech Recognition: Respeaking*. St. Jerome Pub.

Van Tulder, M., Malmivaara, A., & Koes, B. (2007). Repetitive strain injury. *The Lancet*, *369*(9575), 1815–1822.

VoxForge. (n.d.). Retrieved from http://www.voxforge.org/

Zapata, J. (2014). Exploring Multimodality for Translator-computer Interaction. In *Proceedings of the 16th International Conference on Multimodal Interaction* (pp. 339–343). New York, NY, USA: ACM.