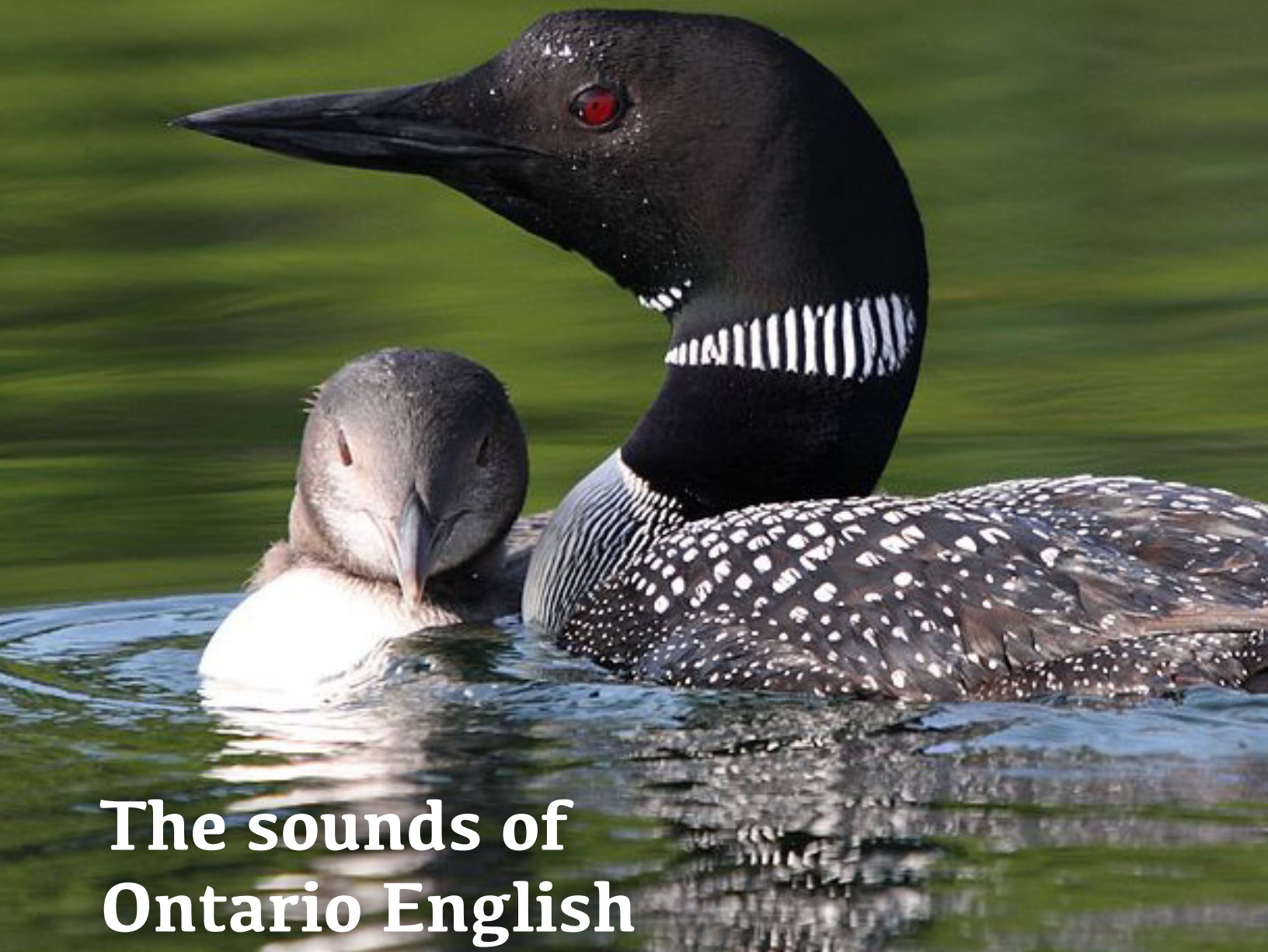
English Language Learning Magazine

CONTACT

November 2014



The sounds of Ontario English

VOLUME 40, NUMBER 4, NOVEMBER, 2014
ISSN # 0227-2938
Teachers of English as a Second Language Association of Ontario
www.teslontario.net/publication/contact-magazine

TESL Ontario
CONNECTION • COMMUNITY • EMPOWERMENT

IN THIS ISSUE

Contents

News

Articles: Phonology

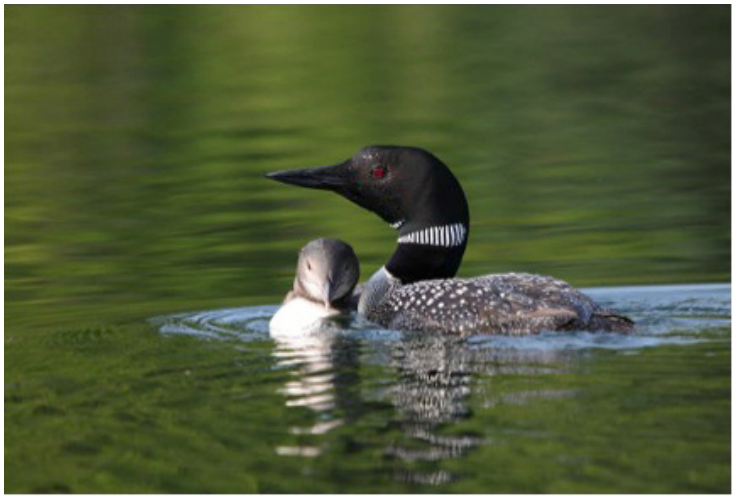
TESOL in the age of MT

More Articles

Calendar

December 3–5	The Ontario Association of Adult and Continuing Education School Board Administrators (CESBA). http://www.cesba.com/events/#Conference
January 8–11	Modern Language Association (MLA). Vancouver, BC. http://www.mla.org/convention
January 23–24	National Online Forum for English as an Additional Language (EAL) / English as a Second Language (ESL) Professionals. Online. http://www.myenglishonline.ca/for-teachers/realize/
March 1–4	Languages Canada's Annual Conference, Gatineau, QC. http://www.languagescanada.ca/en/lc-2015-conference
March 18–20	Language Testing Research Colloquium. Toronto, ON. http://ltrc2015.wildapricot.org/
March 21–24	American Association for Applied Linguistics (AAAL). Toronto, ON. http://www.aaal.org/2015conference
March 25–89	Teachers of English to Speakers of Other Languages (TESOL). Toronto, ON. http://www.tesol.org/convention2015

Please, contact us (editor@teslontario.org) to let us know about upcoming events.



EDITOR'S NOTE

In this issue, along with our usual diverse mix of articles, we have two themed sections: one on learning phonology, and one on the effects of machine translation on our profession.

When we think "phonology", we often think "accent", and I've often been told by students that they like the Canadian accent. But—which one? The pronunciation of English, as of any language, is not fixed. Rebecca Roeder introduces us to two vowel changes that are in progress in Canada right now. Two papers on the sound system in this issue relate the ability to perceive speech sounds to their acquisition. John Archibald discusses the challenges that consonant clusters present to the learner and also the strategies the learner may use to overcome them. Perception is also central to the paper by Annie Tremblay, on how perceiving prosody correctly can help learners recognize words. In the fourth sound system paper, Alene Moyer lays out for us a number of individual factors that affect the acquisition of accent by adult language learners. One noteworthy finding here is that some learners are quite happy sounding "foreign".

On the topic of machine translation, many language teachers are not happy with students using websites to translate things into or out of English. This issue includes three articles on machine translation and language teaching that should help us stop worrying and learn to love the machines. There is a good deal of potential there, the authors argue, that we may not have recognized. And language teachers are unlikely to be replaced by computers any time soon.

As usual, I'd like to thank all our contributors and strongly urge you, the reader, to become a contributor or at least let us know what you think of *Contact*, and what you'd like to see in future issues.

Brett Reynolds
editor@teslontario.org

CONTACT

Contact is published four times a year (Feb, June, August, and November) by TESL Ontario. February is our conference issue and May is our research issue. It is published for the members of TESL Ontario and is available free online to anyone.

Contact welcomes articles of general interest to association members, including announcements, reports, articles, and calls for papers.

Personnel

Editor	Brett Reynolds
EAB members	Hedy McGarrell
	David Wood
	Hanna Cabaj
Webmaster	Kevin O'Brien
Design	Yoko Reynolds

Legal

ISSN # 0227-2938

Cover image

By Cephas (Own work) [GFDL (http://www.gnu.org/copyleft/fdl.html) or CC-BY-SA-3.0-2.5-2.0-1.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons

TESL ONTARIO

TESL Ontario is a supportive community empowering educational professionals to help English language learners to thrive.

Contact TESL Ontario

TESL Ontario #405 - 27 Carlton St.
Toronto, ON M5B 1L2
Phone: 416-593-4243 or 1-800-327-4827
Fax: 416-593-0164
http://www.teslontario.net

Enquiries regarding membership or change of address should be addressed to the TESL Ontario Membership Coordinator at membership@teslontario.org.

Enquiries regarding advertising rates and reservation of advertising space should be addressed to the Office Coordinator at administration@teslontario.org.

Board

Chair	Sheila Nicholas
Vice-Chair	James Papple
Treasurer	Dan Cornish
Secretary	Irene Kosmas
Members-at-large	Cheryl Fretz
	Bhupinder Gill
	Bernice Klassen
	Jennifer MacKay
	Melita Vrakela

Executive director

Renate Tilson

Affiliates

TESL Durham, TESL Hamilton-Wentworth, TESL Kingston, TESL London, TESL Niagara, TESL North York/York Region, TESL Northern Region, TESL Ottawa, TESL Peel/Halton/Etobicoke, TESL Toronto, TESL Waterloo-Wellington, TESL Windsor

THE ROLE OF TESL IN THE AGE OF (HIGHLY ACCURATE) MT

Ana Niño*, Federico Gaspari° & Harold Somers*

* The University of Manchester (UK) & ° Dublin City University (Ireland)

Can MT undermine language learning?

Although the quality of free online machine translation (FOMT) varies greatly depending on—among other factors—the language pair, domain, text type, input style and the intended use of the output, perhaps you have come across some pretty decent online translations when wanting to check the content of a website, access a foreign blog, contribute to it, or just get the gist of what your international friends were saying on Facebook or Twitter. If you were familiar with both the source and target languages, you may have spotted some inaccuracies, or you may have been pretty impressed with the result and thought: what is the need to learn a language, now that you can have instant FOMT at any time? In this paper we explain some of the reasons why this question that (English) language learners and teachers may be asking themselves is too simplistic.

Firstly, we should not forget that FOMT systems are intended to showcase the potential of machine translation (MT), but they fall well short of what "serious" translation software used in highly specific production environments is capable of in terms of accuracy and reliability. Their attraction lies in being just one click away, available at any time, free of charge, for several language pairs and in various mobile apps. Secondly, MT cannot be compared to human translations, which remain superior in quality. MT is most commonly used in contexts with large volumes of repetitive standardized language, hence the translation quality improves if the domain is restricted or controlled. In such scenarios, the main aim is to make the translation process more productive by reducing the time and money invested; this can be achieved, for example, by pre-editing (i.e. simplifying) the source text, or by revising the raw MT output, correcting it to the level of quality that is needed by the end-users in specific situations. Thirdly, the MT quality that is desirable may differ depending on the type of text, its intended users and the circumstances under which the translation is required (its deadline, feasibility of subsequent post-editing, how many people are going to read it, for what purposes and for how long, etc.).

Hutchins (2001) distinguishes three main uses of MT (not only FOMT) in relation to the expected level of quality, namely: MT for dissemination or publication purposes (where the maximum degree of quality is required), MT for assimilation or information gathering (where less-than-perfect quality may still be acceptable), and instantaneous MT for interpersonal communication (e.g. via mobile apps such as WhatsApp Messenger, where

the objective is to get the gist of what the interlocutor writes, to reply accordingly, just to support the short-lived online interaction in real time—style, nuances, etc. are of only limited importance).

There is some common ground between MT development and language learning: for example, the closer the languages, the easier it is to achieve good results. Both MT systems and language learners need to process linguistic differences concerning for instance inflection, agreement, word reordering, false friends, homonymy, homophony, pronunciation and intonation issues (for spoken-language MT), and also cope with a variety of exceptions that make language processing and learning ever more challenging.

Just as many of these difficulties are encountered both in speaking/listening and in writing/reading in a foreign language, they are also common to written and spoken MT. The impressive breakthroughs in spoken MT over the last few years are not to be underestimated. A good example is *Skype's Voice Translator*, which supports people communicating in real time over the phone, with each caller hearing the words translated on the fly into a language that they can understand. However, these voice translation systems are just as likely to cause communication problems due to mispronunciations and homonyms, and because spoken language is often more colloquial and unpredictable than writing (e.g. false starts, interruptions, coughing, overlapping turns in conversation, use of intonation for emphasis and meaning-making with frequent contextual references), and idiomatic or slang use.

Many of the errors that students make when learning a language, which are often very much dependent on their mother tongue, can to some extent be similar to the errors produced by an MT system translating within the same language pair, although MT also produces its own characteristic errors: these clearly differ in terms of gravity and frequency from those found in human translation and in L2 speech and writing. Somers et al. (2006) conducted an experiment aimed at distinguishing MT output from low-level language students' production in order to detect use of FOMT systems by language learners (e.g. to do their homework). They came to the conclusion that "the mistakes made by MT systems are sufficiently different from those made by language learners to permit some sort of automatic detection" (Somers et al., 2006, p. 47). FOMT is therefore unlikely to constitute a threat to language learning, if texts produced by this technology can be detected and penalized accordingly. From the teaching perspective, it makes sense that language tutors learn to identify passages produced by MT using tools such as *Turnitin*, to prevent the improper (and counter-productive) use of FOMT by language students.

This is, in our view, where the use of MT, and especially FOMT, can be controversial for TESL, with plagiarism as one of its potential hazards. If students use ready-made online translations, rather than making the effort of producing the texts themselves and learn from their own mistakes, then there is little point in attending classes and being taught a language formally.

Strengths, limitations and an overview of MT system types

It is true that the quality of MT output keeps on improving at considerable speed for an ever increasing range of language pairs, but several serious limitations still remain. Unlike human translators, MT systems are generally unable to process ambiguity or to capture text-level discoursal phenomena and the interconnections between words and their abstract, implied or culturally motivated meanings. In addition, MT often fails with idioms, anaphora, proper names, culture-related words, slang, and colloquialisms. It cannot easily adjust to different registers and translate them properly—let alone handle puns and humour! For these reasons, MT output usually translates the source text too literally and can therefore be stylistically awkward. Granted, it can even be amusing if, for example, we attempt to translate a joke, figurative language, a horoscope or a dialogue from a movie. However, MT can produce surprisingly good and (under certain circumstances) usable or useful results with formal, standardized texts such as legal documents, for example, if less-than-perfect quality for gisting is sufficient, or if post-editing is an option to bring the raw MT output of uneven quality to publishable standard.

There are different types of MT systems, each with their own strengths and weaknesses. Rule-based MT (RBMT) involves using an analyzer, a parser and a transfer lexicon to link the source- and target-language structures. The rules contain morphological, syntactic and semantic information of both languages. FOMT services such as Babelfish or Apertium are based on this traditional approach, which has gradually evolved ever since MT was first demonstrated in the 1950s.

The more dominant methodology is statistical MT (SMT), which uses massive bilingual corpora of previously translated texts, with aligned sentence pairs as its main knowledge base: translations are produced by algorithms that automatically identify pairs of phrases that (most frequently) correspond to each other between the two languages. SMT uses information gleaned from hundreds of millions of documents to generate translations of similar texts using statistical techniques that calculate the probability that a target-language string of words is the translation of a particular source-language phrase. After these correspondences at the phrase level are ranked in terms of probability, a target-language model is applied to help produce fluent, correct output, filtering out unlikely phrases that are not attested in the target-language model.

This approach is based exclusively on probability, and does not require any explicit or formalized linguistic rules, only very large quantities of human-translated parallel texts. This is in principle the way in which Google Translate, Bing Translator and other statistical MT systems work: they detect translational patterns in millions of professionally translated documents and make intelligent guesses/mappings to generate new translations of unseen texts, that must be to some extent similar to the parallel corpus used for the training. This means that the system can only translate text that is in some sense similar to the text it

has already seen. So if the text it is trying to translate contains a misspelled word, or an ungrammatical sentence, it is more likely to get it wrong.

Finally, hybrid MT systems combine the strengths of the rule-based and statistical approaches, in an attempt to obtain more accurate results. These systems use an underlying statistical architecture, with the addition of explicit linguistic and translation rules that kick in to manage specific aspects of the translation process that are likely to cause mistakes if handled exclusively with a statistical approach, such as when morphologically rich languages are involved. Examples of MT systems using this hybrid method include PROMT, SYSTRAN and Asia Online.

MT in the language class

So the burning question now becomes: should you use MT in your language class, and, if so, how? Somers (2003) suggests the following uses:

- MT as bad model (Anderson, 1995 and Richmond, 1994), where language students can learn to identify the lexical and grammatical errors of MT output and correct them in their L1 or L2, in order to acquire a better grasp of the target language. This will also help avoid plagiarism by making students more aware of what FOMT can and cannot translate successfully. However, if the translation is done into the L2, especially with low-level students, it may reinforce incorrect use of the target language;
- MT as a translation-training tool for advanced students interested in acquiring new translation skills. In Lewis's (1997, p. 255) forward-looking words, "language graduates need to know what the capabilities of state-of-the-art MT are and how to evaluate its role as a practical tool in the language industry." This may also involve learning about pre- and post-editing skills, together with a range of computer-assisted translation (CAT) tools, including translation memory software, terminology managers and databases and concordancers, to name but a few;
- MT as a Computer Assisted Language Learning Tool (Somers, 2012) via programs such as *Transit Tiger*[1]. With this program language tutors can design their own guided translation exercises aiming at practising translation as an independent language learning skill into the L1 and into the L2. The interface provides hyperlinked words as hints for structures or lexical phrases that may pose problems when translating. A glossary and two model translations are also available to the students for reference purposes.

Gaspari & Somers (2007) discuss the rather widespread (mis)use of FOMT systems for lexical look-up, i.e. as if they were bilingual dictionaries. Their survey was carried out among over 100 university language and translation students in the UK, investigating their

1. Transit Tiger Software Review https://calico.org/p-184-TransIt%20Tiger%20(32000).html

usage of FOMT systems: 62.5% of the sample reported using them for single-word lookup, and the authors explain the risks associated with this practice. Some of the FOMT systems now recognize this danger, and automatically change to "dictionary mode" if a one-word translation is requested, offering a range of translations with hints as to which might be required.

Among the various instructional drawbacks of FOMT we have already mentioned the danger of plagiarism and the fact that being exposed to FOMT output into the L2 may reinforce bad language habits. These are some of the main reasons why it is understandable that language tutors tend to oppose MT: simply because the intervention of MT seems pointless or unnatural from the language teaching perspective.

In a more recent study, Niño (2009) surveyed language students and tutors about their perceptions of FOMT as a language learning resource. On the whole, students reacted positively, stating that manipulating FOMT output had enhanced their language awareness, improved comprehension, lexical retrieval and increased confidence in their target language skills. On the other hand, language tutors perceived it more negatively, arguing that they will only start incorporating FOMT in the language class in the future when translation quality is better.

The truth is that in as much as foreign language and translation students use the Internet and its various tools such as search engines, concordancers, corpora, dictionaries, term banks, and encyclopedias as means to support foreign language written production, the use of FOMT for translation and foreign language learning cannot be easily prevented or avoided. As McCarthy (2004) puts it, FOMT "is a new parameter in translation teaching and instructors and students must learn to work with it." According to him, it is unrealistic to ignore or eliminate FOMT use for students' translation assignments. He goes on to suggest some practical strategies for fair assessment, including raising students' awareness of the potential and limitations of MT: they have the ability to produce a more accurate, idiomatic and fluent translation, especially if they possess an advanced knowledge of the target language. He also recommends teaching students about the deficiencies of MT, imposing severe penalties on machine-generated mistakes, and selecting very carefully the texts to be translated, so that for example they contain figurative language and idiomatic expressions. It seems to us that these recommendations can also be implemented in language learning programmes, to ensure that writing and translation assignments are completed in a pedagogically useful way, avoiding risky shortcuts.

Conclusion

By way of conclusion we can say that, although FOMT is an omnipresent resource used by both language learners and translation students, this technology is still not very accurate and was not designed as a language-learning tool. As a result, it should not be used for translation into the L2 without a tutor's supervision, especially at lower levels of proficiency,

where students have more problems spotting incorrect language. For assimilation or communication purposes, there seems to be little pedagogic harm in using MT output into the L1, and exercises requiring students to revise MT output in their own language (Anderson 1995) may also serve a useful purpose in heightening awareness of differences between languages, and the need for natural-sounding translations; after all, being able to speak/write it is only one reason for learning a foreign language, and anyone hoping to earn a living from their knowledge of a foreign language is more likely to be employed translating *from* it into their own language.

Translation students and advanced language learners may like to explore FOMT's capabilities in combination with other online language and translation-oriented resources (e.g. concordancers, parallel and comparable corpora, dictionaries and terminology databases, among others). However, for production purposes, one should remember that MT is subject to many inaccuracies, in particular with distant language pairs, and does not cope well with informal language (as typically found in conversations or emails, for example). To achieve high quality, statistics-based FOMT depends heavily on large corpora of aligned human translations, which are currently restricted in terms of domains, text types, and language pairs.

Nothing compares to the experience of learning a new language and getting to spontaneously express oneself in it. As we know, this does not mean being 100% accurate from the beginning, but with perseverance successful language development will take place, and it will replace the tendency to rely on MT output for constant reassurance. TESL tutors should be aware of what MT is capable of, learn to spot MT errors as one form of plagiarism and to penalize it accordingly, and teach learners how to employ this technology for written production and when not to.

As for the potential threat posed by state-of-the-art MT to the future of the TESL profession, there is still uncertainty as to whether, or when, MT-enabled international communication can surpass the use of English as a global language, and this question cannot be definitively answered for quite a few more years. In David Crystal's (2003, p. 27) words: "It will be very interesting to see what happens [when MT matures as a popular communication medium] – whether the presence of a global language will eliminate the demand for world translation services, or whether the economics of automatic translation will so undercut the cost of global language learning that the latter will become otiose. It will be an interesting battle 100 years from now". Although we will not be there to witness the developments of MT and language learning of the next 100 years, we wouldn't be surprised if TESL professionals were still very much in demand in the 22nd century, and possibly well beyond that…

References

Anderson, D. (1995). Machine translation as a tool in second language learning. *CALICO Journal, 13*(1), 68–97.

Crystal, D. (2003). *English as a global language*. Cambridge: Cambridge University Press.

Gaspari, F. & Somers, H. (2007). Making a sow's ear out of a silk purse: (Mis)using online MT services as bilingual dictionaries. *Proceedings of Translating and the Computer 29*. London, United Kingdom, 29-30 November 2007. London: Aslib/IMI, (pp. 1–15).

Hutchins, J. (2001). Machine translation and human translation: In competition or in complementation? *International Journal of Translation, 13*(1-2), 5–20.

Lewis, D. (1997). Machine translation in a modern languages curriculum. *Computer Assisted Language Learning, 10*, 255–271.

McCarthy, B. (2004). Does online machine translation spell the end of take-home translation assignments? *CALL-EJ Online, 6*(1). Retrieved from http://callej.org/journal/6-1/mccarthy.html

Niño, A. (2009). Machine translation in foreign language learning: language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL, 21*, 241–258. doi:10.1017/S0958344009000172

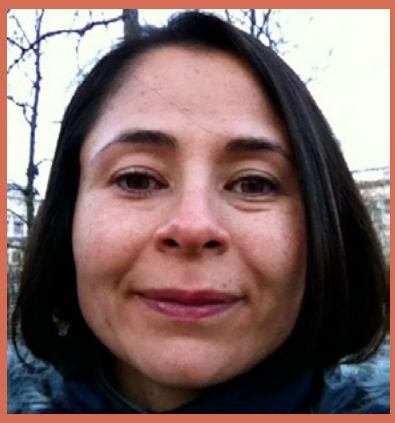
Richmond, I. M. (1994). Doing it backwards: Using translation software to teach target-language grammaticality. *Computer Assisted Language Learning*, *7*, 65–78.

Somers, H. (2003). Machine translation in the classroom. In H. Somers (Ed.), *Computers and translation: A translator's guide*. John Benjamins Publishing Company: Philadelphia, PA.

Somers, H. (2012). Computer assisted language learning and machine translation. *The Encyclopedia of Applied Linguistics*. Wiley Online Library. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0175/abstract

Somers, H., Gaspari, F. & Niño, A. (2006). Detecting inappropriate use of free online machine translation by language students – a special case of plagiarism detection. *Proceedings of the EAMT (European Association for Machine Translation) 11th Annual Conference*, 19th and 20th June 2006 in Oslo, Norway, (pp. 41–48). Retrieved from http://personalpages.manchester.ac.uk/staff/harold.somers/Babelfish.pdf

Bios



Ana Niño has a background in Applied Linguistics and holds a PhD in Machine Translation and Computer Assisted Language Learning from The University of Manchester. Since September 2006 she has been Spanish and Portuguese co-ordinator at the Language Centre, University of Manchester. Ana teaches courses in general and Medical Spanish. In the last few years she has been heavily involved in developing various e-learning resources for the Spanish VLEs. Her research interests revolve around translation, corpus linguistics and the pedagogy of teaching specialized languages, with particular emphasis on the use of emerging technologies and innovative resources for independent language learning.

Federico Gaspari has a background in translation studies and holds a PhD in machine translation from the University of Manchester (UK). After working for a few years as a lector and language tutor for Italian at the Universities of Manchester and Salford in the UK, he currently lectures in translation technology and specialized translation at the Universities of Bologna (Forlì campus) and Macerata in Italy. He is a postdoctoral researcher at Dublin City University (Ireland), where he works on international projects related to translation technology and translation quality evaluation.

Harold Somers is Professor (Emeritus) of Language Engineering at the University of Manchester, where he worked for nearly 40 years on various aspects of Machine Translation, but especially issues of usability of MT in the general community, promoting awareness and understanding of MT and other language technologies amongst language professionals. After taking early retirement in 2008, he worked for 3 years at Dublin City University, focussing on similar issues, as well as use of language technology to assist patients with limited English in doctor-patient consultations. He
is now fully retired, but still takes a keen interest in these issues.