Translation, interpreting and new technologies

Michael Carl, Sabine Braun

Abstract

The translation of written language, the translation of spoken language and interpreting have traditionally been separate fields of education and expertise, and the technologies that emulate and/or support those human activities have been developed and researched using different methodologies and by different groups of researchers. Although recent increase in synergy between these well-established fields has begun to blur the boundaries, this section will adhere to the three-fold distinction and begin by giving an overview of key concepts in relation to written-language translation and technology, including computer-assisted translation (CAT) and fully automatic machine translation (MT). This will be followed by an overview of spoken-language translation and technology, which will make a distinction between written translation products (speech-to-text translation, STT) and spoken translation products (speech-to-speech translation, SST). The key concepts of information and communications technology (ICT) supported interpreting, which is currently separate from the technological developments in written- and spoken-language translation, will be outlined in a third section and a fourth will provide an overview of current usages of translation and interpreting technologies.

1. INTRODUCTION/DEFINITIONS

The translation of written language, the translation of spoken language, and interpreting have traditionally been separate fields of education and expertise, and the technologies that emulate and/or support those human activities have been developed and researched with different methodologies and by different groups of researchers. A recent increase in synergy effects between these well-established fields has begun to blur the boundaries. However, this section will adhere to the three-fold distinction and begin by giving an overview of key concepts in relation to written-language translation and technology, including computer assisted translation (CAT) and fully automatic machine translation (MT) in section 1.1. This will be followed by an overview of spoken-language translation and technology in section 2.2, where a distinction will be made according to whether the translation product is in a written form (speech-to-text translation, STT) or in a spoken form (speech-to-speech translation, SST). The key concepts of ICT-supported interpreting, which is currently separate from the technological developments in written- and spoken-language translation, will be outlined in section 1.3, followed by an overview of current usages of translation and interpreting technologies in section 1.4.

1.1 Written-language translation and technology

There is a great range of MT systems, based on different philosophies and computer algorithms, with different advantages and disadvantages, but a feature they all share is that they are normally used as fully automatic devices, translating a source text into a target language without human intervention. In contrast to this, human translation, or for short simply translation, is an exclusively human activity without intervention by machines or collaboration between humans and machines, other than interaction with a word processor and perhaps the use of electronic dictionaries.

In between these two extremes of translating a source text into a target language, there is a plethora of tools and workbenches available that support human translators in their translation tasks (cf. http://en.wikibooks.org/wiki/CAT-Tools). Depending on the extent to which computers or humans are in the centre of the translation task, a more fine-grained distinction exists between CAT and human assisted machine translation (HAMT). Using a computer (typically a PC) to draft or format a translation is not normally considered to be a kind of CAT, even though, in a strict sense, this usage of the term would be justified. As the term is commonly used, CAT implies at least the use of electronic mono- or bilingual dictionaries, terminologies, collocation or (bi-)concordance tools, and typically translation memories. In more sophisticated versions of HAMT, an MT system would be at the core of the translation process. As the accuracy and speed of computational devices is increasing, novel forms of human-machine interaction in translation, such as interactive translation assistance, usage of multiple modalities, integration of written and spoken language, gesture and handwriting recognition, optical character recognition (OCR), speech synthesis, etc. are being explored and likely to become part of professional translation environments in the near future. These tools can—at least in principle—be combined in almost any configuration. In practice a number of workbenches exist to facilitate and support the human translation processes in various different ways.

Translation memory systems (TMs) are often used as a synonym for CAT tools. TMs do not translate by themselves. Rather, they retrieve close matches of a source-language string from a bilingual database (a so-called translation memory) and display the translation(s) associated with the retrieved segments to a translator for him/her to adjust. For this to be possible, a translation memory of aligned translations first has to be created on a segment-by-segment basis. A number of alignment tools are available to carry out this process either interactively or fully automatically. A TM also computes the similarity between the sentence to be translated and similar source-language sentences in the translation memory. The comparison is mostly based on orthographic similarity. The assumption is that similar source sentences have similar translations so that the translator can select and adapt a translation of a similar source segment. Indeed, translators are often paid by the degree of similarity: 100% identical segments are considered to require no work on the part of the translator and are therefore often not paid at all.

Given the increased quality of MT system output, the use of fully automatic translation is constantly growing. Unlike TMs, MT systems generate 'proper' translations from source texts, often based on carefully selected and tuned resources. However, depending on the expected quality of the translation product, post-editing of MT output (PEMT) is often necessary to bring the raw MT output in line with the intended purpose of the translation and to erase major translation errors and flaws that would hinder or inhibit the comprehension of the translated text. A number of MT post-editing platforms have emerged recently to facilitate this process. Like TM systems they show each source-text segment together with its MT output for post-editing (O'Brien *et al.* 2014). Post-editors usually receive a translation brief specifying the intended audience and the expected quality of the final translation product. Given the tremendous variety of resources used in an MT system—including bilingual dictionaries, phrase translations and their source-target alignments etc.—integrated MT post-editing platforms are being developed that facilitate interactive human intervention by supporting post-editors in selecting from alternative partial translations, tracing partial translations, visualizing confidence scores etc.

1.2 Spoken-language translation and technology

With regard to spoken-language translation, technological developments are still in their infancy. Automatic spoken-language translation systems are a concatenation of automatic speech recognition (ASR) and machine translation (MT) systems with an optional speech synthesis system for spoken target-language output. Transcribed speech (i.e. the output of ASR systems) differs significantly from written text. ASR output therefore requires a number of additional modifications to be suitable input for MT systems. Some of the characteristic features of spoken language, i.e. hesitations (hmm, uh, etc.), discourse markers ("well", "you know"), self -corrections ("it is- it was ..."), repetitions and

incomplete sentences produce ill-formed text that may be difficult to understand even for human readers. In addition, as ASR systems (seek to) generate a faithful transcript of the spoken words, the transcribed output also lacks segmentation and punctuation marks, making it difficult to determine when a sequence ends and is "ready" to be translated. This creates problems for MT systems which normally expect well-formed and sentence-segmented written input text. A number of operations are thus necessary to reformat and map the output of the ASR system to fit the input requirements of MT systems.

Most currently available speech translation systems operate in a consecutive fashion whereby a speaker inputs an utterance, the system processes and translates the spoken signal and outputs the translation either in written form (speech-to-text translation, STT) or spoken form (speech-to-speech translation, SST). There are only few speech translation systems that simultaneously translate unsegmented, continuous speech (Cho *et al.* 2014).

1.3 ICT-supported Interpreting

Whilst SST is still scarce, the evolution of communication and information technologies (ICT) has created ample opportunities for distance communication in real time and has led to ICT-supported human interpreting as an alternative to delivering human interpreting services onsite. On the one hand, mobile and internet telephony has facilitated conference calls with participants in two or more locations. On the other hand, videoconferencing has established itself as a tool for verbal and visual interaction in real time, including between two or more sites.

Regarding the underlying technology, telephone-based and videoconference-based/video-mediated interpreting are the two established methods of ICT-supported interpreting today. Two main uses can be distinguished on the basis of the physical or geographical distribution of the participants, including the interpreter.

One of these, remote interpreting (RI), is the use of communication technologies to gain access to an interpreter in another room, building, town, city or country. In this setting, a telephone line or videoconference link is used to connect the interpreter to the primary participants, who are together at one site. RI by telephone is often called *telephone interpreting* or *over-the-phone interpreting*. RI by videoconference is often simply called *remote interpreting* in relation to spoken-language interpreting. In sign-language interpreting, the term *video remote interpreting* has established itself. RI can be used in connection with simultaneous, consecutive and dialogue interpreting.

The second method has emerged from the demand for interpreting in telephone calls or videoconferences between parties at different sites who do not share the same language, i.e. for interpreter-mediated telephone or videoconference communication (e.g. bilingual or multilingual virtual meetings, bail hearings by video link between courts and prisons, doctor-patient phone calls or video links). In this setting, the interpreter is either co-located with one of the parties or at a separate site. The latter configuration leads to a multi-point telephone or videoconference connection. The method of interpreting required in this setting can be termed teleconference interpreting to cover both telephone and videoconference communication. However, the terms *telephone interpreting* and *videoconference interpreting* have also been used here (Braun 2015, Braun & Taylor 2012, Rosenberg 2007, Mouzourakis 2006).

Remote and teleconference interpreting have different underlying motivations but overlap to a certain extent, most notably in multi-point telephone or videoconferences. In the conference interpreting market, this combination is on the rise due to an increasing number of webinars and other events with distributed participants, and has become known as webcast interpreting. It was included as a new category in the AIIC 2012 conference interpreter survey (AIIC 2014). Webcast interpreting involves conference interpreters working in a team for remote audiences while being remote from each other rather than sharing a booth. The connection can be telephone- or videoconference-based.

1.4. Usage of translation and interpreting technologies

Figure 1 summarises the different methods of translation and interpreting.

The bottom of the pyramid reflects the market for MT and SST systems, which is known as the 'gisting' market. With the emergence of Google's translation service this market segment has grown tremendously and is by now the biggest sector of the translation market. Google started its translation service in 2002 with 5 languages, and serves 90 languages at the time of writing. Google translation services are used by more than 500 million people every month, producing more than 1 billion translations a day. Every day, Google translates more text than all human translators translate together in a whole year. Google translates web pages, tweets, blogs or communications via email or chat systems, using on-line MT systems as the translation engine. Translation quality is often far from perfect but in most cases users get access to content which would otherwise not be available to them.
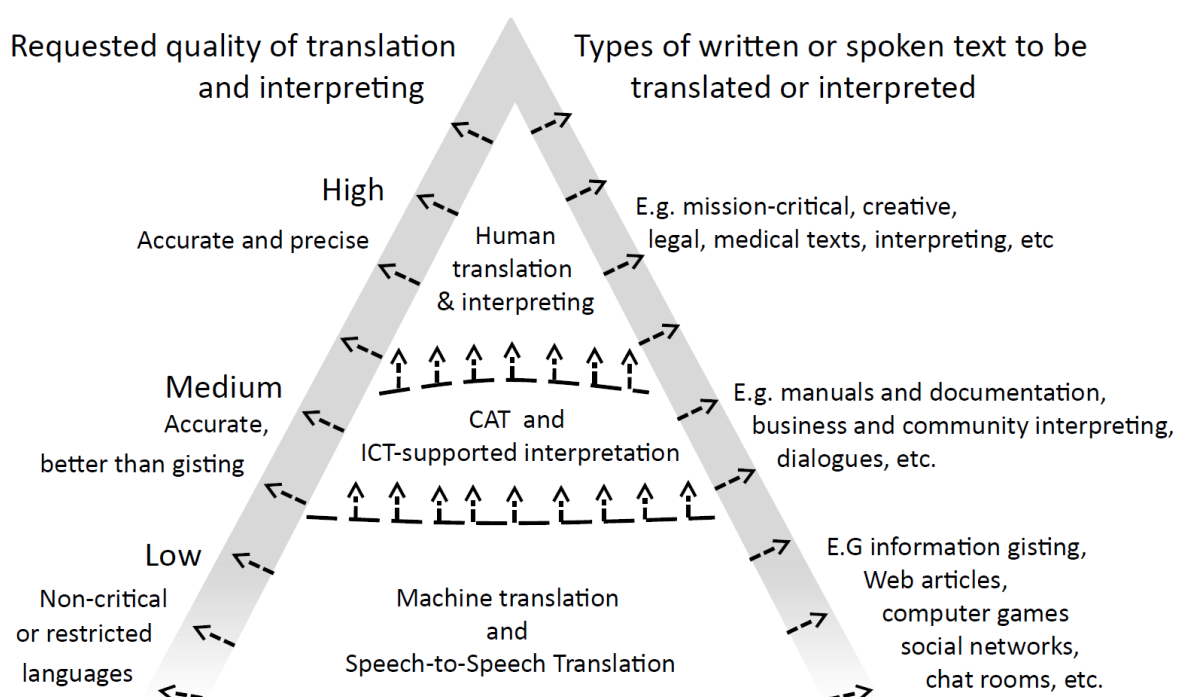

Figure 1: Information that users want to have translated (adapted from Schäler et al, 2003)

Demand for remote and teleconference interpreting varies across fields. In relation to telephone-based interpreting, an analysis of over 1000 instances by Rosenberg (2007) showed that at the time of his study, demand for remote interpreting mainly arose from migration and associated language policies, and that it was most widely used in healthcare settings while interpreting in three-way telephone conversations was more common in the business world. Variations across fields can also be identified in relation to videoconference-based interpreting. In the commercial conference market, videoconference interpreting in the form of interpreting for remote speakers during a conference is more frequent than remote interpreting but in most regions of the world both categories are outstripped by the emerging method of webcast interpreting (AIIC 2014). However, having conducted initial tests with both videoconference and remote interpreting, the European institutions, which are large users of conference interpreting services, remain mainly interested in remote interpreting (see section 3 below). In healthcare settings, remote interpreting is in high demand, but the growing trend towards tele-healthcare will also require the integration of interpreters into video calls between doctors and patients. In legal settings, both videoconference and remote interpreting have begun to establish themselves.

High-quality translation can be achieved by MT systems for restricted domains and/or controlled languages or when tuning the MT system to the type of texts to be translated. For instance, due to its

very particular domain, the Météo system (Chandioux, 1988) was a big success story in the history of MT long before the emergence of Google, providing perfect translations for weather forecasts from French into English. While the original Météo implementation was a rule-based system, later statistical implementations on the same domain provided equally good results (Gotti, 2014). However, the adaptation of the system for another domain (aviation) proved unsuccessful. Fully Automatic High Quality Translation (FAHQT) can thus be achieved for limited domains and by training MT systems to produce a particular type of texts, but for general use with a wide range of text types, it remains "a dream which will not come true in the foreseeable future" (Bar-Hillel, 1960), just as in 1960.

The requirement for reliable high-quality translation of less restricted languages can often only be met through from-scratch translation, usage of CAT tools or MT post-editing. However, as discussed above, the market for translation aides and computer-assisted translation is changing and developing at a quick pace, and many new and innovative products are emerging. In contrast to fully automated translation, the post-editor is the integral figure involved in every part during the computer-assisted translation process, to accept or reject translation suggestions, or to insert translation proposals into the target document wherever deemed appropriate.

Professional interpreting is normally required in contexts that are situated at the upper end of the triangle. Some tensions arise from the growing client-side demand for ICT-supported interpreting services, especially remote interpreting, whilst it is currently difficult to be sure whether these methods of interpreting can be associated with the same levels of accuracy and precision as onsite interpreting (Braun 2013, Roziner & Shlesinger 2010).

2. HISTORICAL PERSPECTIVES

2.1 The Beginnings of Machine Translation

In 1949, Warren Weaver became the first person to propose using computers for machine translation. After the successful decoding of German and Japanese military messages during WW-II (as generated by the Enigma machine), his idea was that translation of one language into another would be similar to the task of decoding a message. A few years later, in 1954, an MT project at Georgetown University succeeded in correctly translating 60 sentences from Russian into English, resulting in unprecedented euphoria in MT research.

It was then apparent that the process of machine translation would require a formalization of the syntax of natural language and associated automata for parsing with these grammars; this led to a highly productive period in computer science during which the foundations for programming languages and compilers were laid and pertinent theories of language were postulated in linguistics (Chomsky 1957).

However, 10 years later the Automatic Language Processing Advisory Committee (ALPAC, 1966) published a study which expressed doubt that a MT system could ever be produced, based on the assumption that understanding human language relies on information which is not present in the words which make up the message. Instead ALPAC called for investigations of which texts were suitable for automatic translation, and which were not, what tools would be helpful to support human translators, and how translation aides could be successfully integrated into the human translation workflow. In the 1980s, concrete suggestions for implementing such systems were made (Kay, 1998), although the investigation of cognitive processes in computer-assisted translation received little attention before the early decades of the twenty first century (cf. Carl *et al.* 2015).

2.2 Rule-based Machine Translation

Interest in MT was revived in the 1980s, following unprecedented development of computer hardware (PCs and cheaper storage capacity), innovative programming languages (LISP and Prolog) and modern linguistic formalisms such as Generalised Phrase Structure Grammar (GPSG) (Gazdar *et al.*

1985), Tree-Adjoining Grammar (TAG) (Joshi 1985), Lexical-Functional Grammar (LFG) (Bresnan and Kaplan 1985), Combinatory Categorial Grammar (CCG) (Steedman 1984) and Head-driven Phrase Structure Grammar (HPSG) (Pollard and Sag 1994), which were tailored to the new programming languages and suited to formalizing the structure of natural languages. A number of theoretical approaches and processing strategies for MT emerged based on these new developments, including direct MT, transfer-based MT and interlingua MT, which, at that time, relied mainly on rule-based formalisms.

The syntactic frameworks were subsequently augmented with formal representations of semantics. The idea was that the semantic expression of a natural-language text would be compositionally derived from the semantic expressions of words in the lexicon and combination rules. The semantic expressions would be devoid of the syntactic variations and vagaries of any particular language and hence two natural-language expressions representing the same "meaning" would result in the same semantic representation. This desideratum is a prerequisite for the *Interlingua-based Machine Translation* approach (cf. e.g. Eurotra), which, however, has not proved helpful in dealing with general texts at the time of writing.

2.3 Statistical Machine Translation

In the early 1990s, the availability of digitized corpora of real-world texts and translations prompted a technological challenge to process texts using tools from computational linguistics. The plan was to decompose available bilingual texts into large amounts of possible phrase translations, store them in databases and recompose new translations based on combinations of the most likely translation snippets. It soon became apparent that the tools then available for language processing could not deal with the complexities of real-world texts. For example, the tools were designed to provide all possible analyses for a given sentence which could run into thousands of analyses for even simple sentences with no indication of a preferred analysis. Consequently, the computational linguistics community turned to statistical models of language that were prevalent in information theory and speech recognition communities.

A group of researchers from IBM (Brown *et al.* 1988) re-vitalized Warren Weaver's initial idea of translation as a form of language decoding and formalized this approach in terms of a noisy channel model (Brown *et al.* 1993). Given a large amount of data and sufficient computational power, statistical models could be trained on bilingual texts and applied to decode (i.e. translate) unseen sentences. This line of research has been further developed into log-linear modelling (Och 2002) and into what is known as phrase-based statistical machine translation (PB-SMT). It is now the main MT paradigm that underlies Google translate and a number of research prototypes such as Moses (Koehn, 2007) or Thot (Ortiz-Martınez, 2014). Besides this, tools using rule-based approaches such as Systran and PROMT (see http://www.prompt.com/), developed in the pre-statistical period, have found their niche and continue to survive.

2.4 Speech-to-speech translation

Automatic Speech Recognition (ASR) dates back to the early 1950s, but the principal technological component(s) that are still used in ASR systems (i.e. the Hidden Markov Model, HMM) were introduced in the 1970s in the context of a five-year research project on speech recognition systems funded by the Defense Advanced Research Project Agency (DARPA), an agency of the U.S. Department of Defense responsible for the development of emerging technologies for use by the military (Waibel & Kai-Fu 1990).

Automatic *translation* of speech, however, is a more recent development. Research into speech-to-speech translation (SST) has attracted high levels of funding since the 1990s, including the German VERBMOBIL project (1993-2000), which received approximately €90 Million (Karger & Wahlster 2000; http://verbmobil.dfki.de/overview-us.html), the US-led C-Star initiative (1991-2004) and the European PF-Star and TC-Star projects (2002-4 and 2004-7). VERBMOBIL aimed at the development of speech-to-speech translation in restricted domains (e.g. appointment-making). More

recently, after the acquisition of Skype© by Microsoft© in 2011, Skype Translator© was released in 2014. It provides a written translation of a conversation in near-real time (for 6 languages at the time of writing) and can convert the translated output into speech. It is designed to deal with specific features of spoken language such as incomplete sentences. Whilst these projects have doubtlessly moved the topic of SST forward, at the time of writing, they could not be applied to situations in which highly accurate professional interpreting is required. Bilingual and multilingual communication in professional settings often involves complex multi-topic and multi-party interaction and may require simultaneous interpreting (i.e. a rendition while a speaker is talking) into, and out of, several languages. At the time of writing, available speech-to-speech translation systems struggled to resolve the many pragmatic intricacies of spoken interaction (from e.g. ambiguity, vagueness and differences between what is said and meant to intercultural communication problems and non-native accents/varieties of English) in a reliable manner.

2.5 ICT-supported interpreting

The first service for telephone-based interpreting was established by the Australian immigration service in 1973. In the US and in most Western European countries, such services have been offered since the 1980s and 1990s respectively (Mikkelson 2003). Although some telephone interpreting services are now being replaced by videoconference-based interpreting services, telephone-based interpreting is still a large market (Commonsense Advisory 2011). With the spread of telephone interpreting, the method has seen improvements in the technology used (e.g. dual-headset phones for clients to listen to a remotely located interpreter).

The development of video-mediated interpreting was originally driven by the interest of supranational institutions in ICT-supported interpreting as a means of optimising access to interpreters and meeting linguistic demand. The earliest experiment was organised by the UNESCO in 1976 to test the use of the Symphonie satellite. It linked the UNESCO headquarters in Paris with a conference centre in Nairobi and involved remote interpreting by telephone and video link, and interpreting in a videoconference between Paris and Nairobi. Similar experiments were organised by the UN later in the 1970s and 1980s (Viaggio 2011, Mouzourakis 1996).

From the 1990s, a series of feasibility studies of video-mediated remote interpreting (in simultaneous mode) was organised by various institutions, including the European Telecommunications Standard Institute (ETSI) in 1993 (Böcker & Anderson 1993), the European Commission in 1995, 1997 and 2000, the United Nations in 1999 and 2001, the International Telecommunications Union (ITU) in collaboration with the École de Traduction et d'Interprétation (ETI) in 1999 (Moser-Mercer 2003), the European Council in 2001, and the European Parliament in 2001 and 2004. The studies revealed a range of physiological and psychological problems which recurred in different technical conditions and which seemed to be caused by the overarching condition of remoteness (Mouzourakis 2006).

Whilst the feasibility studies cited above involved comparisons of real-life or test performances in onsite and remote interpreting, a more recent study conducted by the Fraunhofer Institute for the Interpreting Service of the European Commission (SCIC) in 2010 aimed to define minimum standards for video and audio transmission in the context of remote simultaneous interpreting. This resulted in a comprehensive list of technological recommendations (Causo 2012).

A major driving force of the spread of video-mediated interpreting in legal settings was the increasing use of videoconference technology in the court systems of many Anglo-Saxon countries since the 1990s, e.g. to link courts and prisons for pre-trial hearings (Braun & Taylor 2012, Ellis 2004, Fowler 2013). This entailed a demand for videoconference interpreting whereby the interpreter is co-located with one of the parties. Early videoconference systems, which were ISDN-based, led to problems with sound and image quality for interpreters. More recent videoconferencing systems that use high-speed Internet connections provide better audio and video quality and are more conducive to videoconference-based interpreting (Braun & Taylor 2012). Remote interpreting has been introduced more recently in courts and by the police mainly as a way of gaining timely access to interpreters and

reducing interpreter travel time and cost (e.g. Florida district courts since 2007, the Metropolitan Police in London since 2011).

The spread of videoconferencing has also promoted video-mediated interpreting services in healthcare (Locatis *et al.* 2010, Price *et al.* 2012). At the time of writing healthcare providers mainly need remote interpreting, but developments in tele-healthcare, whereby doctors make video calls to patients who are in their own home, are likely create a more diversified demand for video-mediated healthcare interpreting. At the same time, the availability of web- or cloud-based video conference services providing varying and unstable sound and image quality, and access to them on tablets and other mobile devices, raise new questions about the feasibility of video-mediated interpreting using such systems.

3. CORE ISSUES AND MAIN RESEARCH METHODS

3.1 Machine translation

One of the main difficulties in MT is the inherent ambiguity in natural languages and the fact that different languages encode information in different ways:

- Analytic languages (such as Vietnamese and Chinese) have little or no inflectional morphology, and individual words tend to consist of single morphemes. In contrast, synthetic languages such as Finnish, Turkish or Japanese make use of inflectional morphology and agglutinate several morphemes into one word. For instance, Turkish "ev-ler-iniz-den" (English: "from your houses") can be decomposed as "house-plural-your-from".
- Lexical mismatches occur when translating, for instance, English "brother" into Japanese, where a choice needs to be made between older brother (anisan) or younger brother (otouto). Similarly, English makes a distinction between "pig" and "pork" according to whether the word refers to an animal or meat, which may not be specified in all possible source language contexts.
- Categorial, conflational, structural, thematic and other translation divergences (Dorr, 1994) add to translation difficulties: For instance, Spanish "tener hambre" (have hunger; verb (have) + noun) translates into English with a different part-of-speech as "be hungry" (verb (be) + adjective). Several disconnected words, such as English "make X easier" may translate into a single word, e.g. German "X erleichtern". In some cases structural changes are required, such as translating Spanish "A entrar en B" into English "A enters B" or a change of the thematic role from Spanish "John gustar a Klaus", which is equivalent to English "Klaus likes John".
- Homonyms are words that are written in a similar way but have different meanings. For instance, English "odd" has two meanings ("uneven" in connection with numbers and "strange") but these meanings are likely to be realized differently in another language. A distinction can be made between Homographs and Homophones. The latter are words with different spelling but identical or similar pronunciation, while homographs have identical spelling but differ in pronunciation (e.g. the present tense and past tense of English "read"). Homophones formed of multiple words or phrases such as "the sky" vs. "this guy" may be difficult for ASR systems to decode correctly.

Unluckily, several of these phenomena may (and often do!) occur at the same time within a single sentence, making translation a difficult endeavour. In addition, these linguistic translation problems are complemented by cultural and contextual adaptations of the translated content.

As pointed out in section 2, the obvious limitations of analytical and rule-based methods led to the use of machine-learning methods in machine translation in the 1990s, taking advantage of the fact that "existing translations contain more solutions to more translation problems than any other available resource" (Isabelle *et al.* 1993: 205). Accordingly, much translation research in the past two decades has been corpus-based and concerned with methods to extract and make accessible from bilingual

texts the required knowledge for translation. One bottleneck for data-driven MT methods consists in a lack of a sufficient amount of exploitable bilingual texts (i.e. translations with their source texts). Due to the combinatorial complexity of human languages, millions of parallel source-target sentences are required to extract statistically reliable figures for word and phrase translation relations. In addition, in some cases the necessary translation knowledge cannot even be extracted from the texts alone, but requires 'world knowledge' including knowledge about the situation in which the source texts are produced or/and the audience for which the translations are intended to be produced. Machine translation technology therefore increasingly incorporates capabilities for domain adaptation, online learning and incremental learning.

3.2 Speech-to-speech translation

In relation to SST, one of the core issues is real-time constraints: SST systems need to be tuned to strike a balance between the size of the acoustic and language models, the search beam (i.e. how many hypotheses are considered in a search graph) and run-time behaviour. Larger search graphs and bigger models allow for more accurate output, but also require more computation time, which, given the complex processing chain ASR-MT-TTS, cannot be optimally computed in real time. A balance must be struck to translate the spoken signal within an acceptable timeframe whilst not allowing the quality of the translation to deteriorate to an unacceptable degree.

A different, but related research topic in the development of SST systems centres around removing dysfluencies (hesitations, self-corrections, repetitions, etc.) from the ASR output and tackling ungrammatical or incomplete sentences. A means of inspiration may be human interpreters who often produce condensed target texts, discarding individual words, or substituting longer phrases with shorter target-language versions, particularly in cases of very fast and/or dysfluent speech.

This research topic is also linked to the strategies used to insert punctuation marks into the ASR output, which ensures semantically consistent segments that provide an appropriate basis for generating optimal MT output.

3.3 ICT-supported interpreting

Although research topics focusing on ICT-supported interpreting are, on the whole, different from those relating to MT and SST, output quality emerges as a common denominator. Studies comparing the quality of onsite and video-mediated remote (simultaneous) interpreting in conference settings have identified few differences between the two methods, apart from earlier onset of fatigue in remote interpreting (Moser-Mercer 2003, Roziner & Shlesinger 2010). By contrast, research comparing onsite and video-mediated (consecutive) interpreting in legal settings revealed significant quality differences as well as an earlier onset of fatigue in remote interpreting (Braun 2013, Braun & Taylor 2012). In the field of healthcare, a comparison of onsite consecutive interpreting and remote simultaneous interpreting via audio connection found higher accuracy levels in the latter (Hornberger et al. 1996), although some of the differences may have resulted from the different modes of interpreting.

Many studies have gone beyond researching output quality and also elicited data on ergonomic factors and working conditions. Comprehensive studies on remote conference interpreting, for example, point to a number of psychological and physiological problems including stress, a general sense of discomfort among interpreters when working remotely, and interpreters' self-reported fatigue and dissatisfaction with their own performance (Moser-Mercer 2003, Roziner & Shlesinger 2010).

Studies in the field of healthcare and legal interpreting have also elicited user perceptions and preferences. A number of studies conducted from a healthcare management perspective has surveyed medical interpreters, physicians and patients to ascertain the efficiency of ICT-supported interpreting compared with onsite interpreting (Azarmina & Wallace 2005, Locatis et al. 2010, Price et al. 2012). This work shows that interpreters and physicians generally prefer onsite interpreting and that they prefer video-mediated to telephone-based interpreting. Notably, however, the interpreters surveyed by

Price *et al.* (2012) found all three methods satisfactory for conveying information, whilst rating the ICT-supported methods as less satisfactory for interpersonal aspects of communication due to greater difficulties in establishing a rapport with the remote participants. Patients were relatively uncritical in their judgement. Based on the self-perceptions of the participating interpreters, some of the surveys furthermore suggest that levels of accuracy in onsite and ICT-supported interpreting are similar (Azarmina & Wallace 2005), but research in the legal setting reveals discrepancies between interpreters' self-perception of their performance and objective performance analysis (Braun & Taylor 2012).

Furthermore, qualitative, observation-based research has identified changes in the communicative dynamics in videoconference-based, interpreter-mediated legal communication as well as logistical and communication problems resulting from specific participant distributions (Ellis 2004, Fowler 2013). Experimental research using simulations has explored the cognitive processes and strategies interpreters employ to address the challenges of ICT-supported interpreting (Braun 2004, 2007, 2016). This work has highlighted adaptation processes especially with regard to coordinating the communication but has also provided evidence for the increased cognitive effort required to resolve communication problems arising, for example, from listening comprehension difficulties and coordination problems in videoconferences as a possible reason for the perception that ICT-based interpreting is more tiring than onsite interpreting.

4. CURRENT DEBATES AND FUTURE DIRECTIONS

4.1 Machine translation and Post-editing

In relation to some language pairs (e.g. English - Spanish) and text types, Machine Translation output has reached a degree of quality that makes post-editing (PEMT) effective. In order to facilitate the post-editing process in the best possible way, a number of techniques are being developed and piloted to assist a human post-editor with additional automatized translation aides and real-time techniques.

During Online Learning the MT system incrementally updates its (statistical) models that are involved in the translation process. This enables the system to learn dynamically from corrections during the post-editing process and potentially leads to higher acceptance rates of MT output, since the same translation errors that appear within one text do not need to be corrected over and over again.

During Active Learning the MT re-orders sentences of the source text in a way that enables the best learning results from the modifications produced by a post-editor. The MT system selectively asks a human translator to correct a small portion of the translation so as to optimise the results of the MT system for the remaining sentences.

Interactive machine translation (IMT) assists a human translator by predicting the next piece of text that s/he is likely to input. In the process, the system takes into account all the information it has available both in the source text and in the translation already produced.

A number of browser-based and stand-alone post-editing workbenches are being implemented and piloted for PEMT. The advantages and disadvantages of translation crowdsourcing are subject to ongoing discussion. The term crowdsourcing was coined by Jeff Howe in the *Wired* magazine in 2006 to denote a "process of obtaining needed services, ideas, or content by soliciting contributions from a large group of people, and especially from an online community, rather than from traditional employees or suppliers" (Merriam-Webster).

Translation crowdsourcing is by now well established, and a number of companies have emerged using the web and web portals to realize translation projects. Facebook, for instance, had the entire site translated into French through crowdsourcing in March 2008. The advantages of translation crowdsourcing include shorter delays and lower prices compared to traditional translation. Some

companies offer also PEMT for crowdsourcing which is even more cost-effective. Major issues in translation crowdsourcing are quality control of the translation product and project management.

4.2 ICT-supported interpreting

In ICT-supported interpreting one of the points of contention is the quality of the interpreting performance. Whilst some of the discrepancies highlighted in section 3 may be due to different variables, research designs and quality measures, the different findings in relation to conference interpreting (little difference between onsite and remote interpreting) and legal interpreting (significant differences) give rise to questions about the impact of training on the adaptability of interpreters to the challenges of ICT-based interpreting.

A similar issue is the discrepancy between 'objective' and 'subjective' measures especially in studies on remote conference interpreting. Whilst these studies found little difference between onsite and remote interpreting in terms of output quality, they revealed a number of differences in the interpreters' perceptions of the two methods (Roziner & Shlesinger 2010). Mouzourakis (2006) contends that this discrepancy ultimately points to a deficit in our current understanding of the notion of remoteness and the challenges associated with it.

A further related point concerns the relative difficulty of remote interpreting (i.e. interpreter completely separated from the main parties) and teleconference interpreting (interpreter co-located with some participants). Remote interpreting is generally perceived to be more challenging. This view is also reflected in the guidance on the use of technologies in conference interpreting issued by the AIIC (2000/2012), which rejects remote interpreting whilst agreeing to teleconference interpreting (under specific circumstances).

Following the increase of videoconference interpreting in the legal sector, one of the questions arising concerns the 'best' place for the interpreter. Comparing the options—i.e. interpreter co-located with the judicial authorities vs. interpreter co-located with the minority-language speaker—Ellis (2004) and Miler-Cassino & Rybinska (2012) highlight important differences between the two configurations, e.g. concerning the rapport between interpreter and minority-language speaker.

Another point of discussion concerns access to visual information in ICT-supported interpreting and the suitability of telephone interpreting. Kelly (2008) cites a number of advantages, but Ozolins (2011) believes that her description mostly refers to the US, where the size of the market and the dominant role of Spanish have led to a level of sophistication in terms of technology use and logistics that is unlikely to be found in many other countries.

This debate is linked to the interpreters' working conditions. Ko (2006) and Lee (2007), for example, argue that the generally high levels of dissatisfaction associated with telephone interpreting partly stem from poor working conditions, including low remuneration, rather than from the use of the technology as such. There is also a debate as to whether ICT-supported interpreting, due to its potential challenges, should command higher fees than onsite interpreting.

One of the most pressing questions for future research into ICT-supported interpreting is to resolve apparent discrepancies in current research findings (see Section 3). Moser-Mercer (2005) and Mouzourakis (2006) suggest that the condition of remoteness or the lack of "presence" may be the most likely common denominator for the problems with remote interpreting. The concept of "presence" and its effects will require a substantial amount of further research.

Furthermore, Moser-Mercer (2005) has raised questions about the ways in which different groups of interpreters adapt to remote interpreting, and the reasons for variations in this process, arguing that experienced interpreters may find it difficult to adapt to the conditions of remote interpreting because they rely on automated processes, whilst novice interpreters, especially when they are subjected to

new methods of interpreting during their training, may have a greater potential for adaptation. Braun (2004, 2007) reveals a number of limits to successful adaptation in video-mediated interpreting which lead to a reduction in performance. Roziner & Shlesinger (2010) argue that the maintenance of the performance quality in ICT-supported interpreting comes at a price, i.e. that interpreters put more effort into the interpreting task than they do in other settings and may suffer post-work exhaustion. The issue of adaptation also requires further investigation.

A related consideration is how the physical separation of all participants and their perception of the situation via technical channels affect aspects such as the processing of information, the communicative behaviour of the primary participants and the communicative dynamic. Moser-Mercer (2005) outlines problems with multi-sensory integration in videoconferences, which she believes prevent interpreters from processing the information and building mental representations of the situation in the usual way. Licoppe & Verdier (2013) suggest that distributed courtrooms change the dynamic of the communication and lead to fragmentation of the communication. The sources and implications of this kind of fragmentation are not very well understood and warrant further study.

Given the speed with which communication technologies develop and spread, the future is likely to bring an increase and diversification of teleconference and remote interpreting. The latest developments which are likely to be relevant for remote interpreting fall into two categories: (i) high-end solutions such as videoconferencing systems (HD and 3D "tele-presence" or "immersive" systems) and the merger of videoconferencing with 3D virtual reality technology to create "augmented reality" communication solutions and (ii) low-end solutions such as web-based videoconferencing services which were originally developed for the home market (e.g. Skype), and video calls using mobile devices and apps. It will be important to investigate how the virtual spaces that these technologies create are able to support the development of 'presence' and the dynamic of the communication.

6. IMPLICATIONS FOR PRACTICE

Translation and interpreting practice has changed tremendously over recent decades. The changes have been brought about by a combination of technological innovation and societal change, especially increased mobility and demand for translation and interpreting in a globalised world.

On the positive side, translators/interpreters have more choices and opportunities to create translations, offer their services, network with colleagues etc. On the negative side, the introduction of technological tools has often been linked to a deterioration of working conditions and remuneration. Agencies argue, for example, that translators' remuneration can be reduced when translation technology is used to support the production of a translation.

Similarly, the rise of ICT-supported interpreting goes hand in hand with the idea of interpreters being available "at the push of a button" and with an undue simplification of the complexity of interpreting. The introduction of ICT-supported interpreting has thus sparked debate and has raised questions of feasibility and working conditions, but it has also been linked to the efficiency of service provision and the sustainability of the interpreting profession.

Any form of ICT-supported interpreting should be supported by the best possible equipment and connection. Interpreters should be involved in the planning and implementation stages, and an incremental introduction of new technology is recommendable. The room layout, positioning of equipment and seating arrangements for the interpreter and for the other parties need to be considered carefully. Kelly (2008) also highlights the importance of a quiet and undisturbed working environment for the interpreter. She refers to problems that can arise in call centres (or interpreter hubs) where interpreters may disturb each other, and problems caused by background noise when interpreters work from home.

One issue for debate is the length of interpreter-mediated encounters that involve the use of communication technology. Given that research shows a faster onset of fatigue in remote interpreting (Braun 2013, Moser-Mercer 2003), an interpreter's working turn in remote and teleconference interpreting should be shorter than in onsite interpreting.

Given the many challenges of ICT-supported interpreting, interpreters and the users of interpreting services should be trained to work in situations of remote or teleconference interpreting. The extent of the training required is not yet clear, but recent research in a legal setting suggests that short-term training may not be able to solve all problems (Braun 2016).

7. REFERENCES

AIIC (2000/2012) Guidelines for the use of new technologies in conference interpreting. *Communicate!* March-April 2000. http://www.aiic.net/ViewPage.cfm?page_id=120 (accessed 03/12/2016).

AIIC (2014) *AIIC statistics: Summary of the 2012 report*. http://aiic.net/page/6878/aiic-statistics-summary-of-the-2012-report/lang/1 (accessed (03/12/2016).

ALPAC (1966) Languages and machines: computers in translation and linguistics. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council. Washington, D.C.: National Academy of Sciences, National Research Council (Publication 1416).

Azarmina, P. and Wallace, P. (2005) *Remote interpretation in medical encounters: a systematic review*. *Journal of Telemedicine and Telecare* 11, 140-45.

Bar-Hillel, Yehoshua (1960) A Demonstration of the Nonfeasibility of Fully Automatic High Quality Translation. Appendix III of 'The present status of automatic translation of languages', Advances in Computers Vol. 1, 158-163. Reprinted in Y.Bar-Hillel (1964) Language and information. Reading/Mass.: Addis on-Wesley, 174-179. http://www.mt-archive.info/Bar-Hillel-1960-App3.pdf [accessed 11/01/2016]

Böcker, M. and Anderson, B. (1993) Remote conference interpreting using ISDN videotelephony: a requirements analysis and feasibility study. *Proceedings of the Human Factors and Ergonomics Society, 37th annual meeting*, 235-39.

Braun, S. (2004) *Kommunikation unter widrigen Umständen?* Fallstudien zu einsprachigen und gedolmetschten Videokonferenzen, Tübingen: Narr.

Braun, S. (2007) Interpreting in small-group bilingual videoconferences: Challenges and adaptation. *Interpreting* 9 (1), 21-46.

Braun, S. and Taylor, J. (Eds) (2012) *Videoconference and Remote Interpreting in Legal Proceedings*. Cambridge/Antwerp: Intersentia.

Braun, S (2013) Keep your distance? Remote interpreting in legal proceedings: A critical assessment of a growing practice. *Interpreting* 15 (2), 200-28.

Braun, S. (2016) What a micro-analytical investigation of additions and expansions in remote interpreting can tell us about interpreter's participation in a shared virtual space. *Journal of Pragmatics*, forthcoming.

Bresnan, Joan and Kaplan, Ron 1985. The mental representation of grammatical relations, MIT Press, Cambridge, MA 1985.

Brown, Della Petra, Della Pietra, and Mercer The Mathematics of Statistical Machine Translation

Causo, J.E. (2012) *Conference interpreting with information and communication technologies*. Experiences from the European Commission DG Interpretation. In S. Braun and J. Taylor (Eds), 227-32.

Chandioux, John. (1988) METEO: An Operational Translation System. Conference: Computer-Assisted Information Retrieval (Recherche d'Information et ses Applications) - RIAO 1988, 2nd International Conference, Massachusetts Institute of Technology, Cambridge, MA, USA, March 21-25, 1988. Proceedings

Chomsky, Noam (1957) Syntactic Structures. The Hague/Paris: Mouton.

Commonsense Advisory (2011) *Trends in telephone interpreting*. https://www.commonsenseadvisory.com/AbstractView.aspx?ArticleID=2102 (accessed 12/03/2016).

Dorr, Bonnie J. (1994) Machine Translation Divergences: A Formal Description and Proposed Solution. Computational Linguistics 20 (4), 597-633.

Ellis, R. (2004) *Videoconferencing in refugee hearings*. Report to the Immigration and Refugee Board Audit and Evaluation Committee. Available at http://www.irb-cisr.gc.ca/Eng/transp/ReviewEval/Pages/Video.aspx (accessed 12/03/2016).

Eunah Cho, Christian Fügen, Teresa Hermann, Kevin Kilgour, Mohammed Mediani, Christian Mohr, Jan Niehues, Kay Rottmann, Christian Saam, Sebastian Stüker, Alex Waibel (2013): A Real - World System for Simultaneous Translation of German Lectures. In: Proceedings of the 14th Annual Conference of the International Speech Communication Association (INTERSPEECH), August 25 - 29, 2013, Lyon.

Fowler, Y. (2013) Business as usual? Prison video link in the multilingual courtroom. In C. Schäffner (ed.), *Interpreting in a changing landscape. Selected papers from Critical Link 6.* Amsterdam: Benjamins, 225-248.

Gazdar, Gerald; Klein, Ewan; Pullum, Geoffrey; Sag, Ivan. Generalized Phrase Structure Grammar. Harvard University Press, Cambridge, MA, 1985.

Gotti, Fabrizio, Philippe Langlais and Guy Lapalme. 2014. Designing a machine translation system for Canadian weather warnings: A case study Natural Language Engineering / Volume 20 / Issue 03 / July 2014, pp 399 - 433

Hornberger, J., Gibson, C., Wood, W., Dequeldre C., Corso, I., Palla, B. and Bloch, D. (1996) Eliminating language barriers for non-English-speaking patients. *Medical Care* 34(8), 845-56.

http://www.dailymail.co.uk/sciencetech/article-2908490/Google-Translate-app-adds-conversation-mode-translate-SPEECH-real-time.html#ixzz3bXrCi1G2

https://www.youtube.com/watch?v=juptbCRFl4UIsabelle, Pierre, Marc Dymetman, George Foster, Jean-Marc Jutras, Elliot Macklovitch, Francois Perrault, Xiaobo Ren & Michel Simard. 1993. "Translation Analysis and Translation Automation", Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation, TMI'93 (Kyoto, July 14-16, 1993), 201-17

Kelly, N., (2008) *Telephone interpreting: A comprehensive guide to the profession*. Clevedon: Multilingual Matters.

Ko, L. (2006) The need for long-term empirical studies in remote interpreting research: A case study of telephone interpreting. *Linguistica Antverpiensia* NS5, 325-38.

Lee, J. (2007) *Telephone interpreting — seen from the interpreters' perspective. Interpreting* 2(2), pp.231-52.

Licoppe, C., and Verdier, M. (2013). Interpreting, video communication and the sequential reshaping of institutional talk in the bilingual and distributed courtroom. *International Journal of Speech, Language and the Law* 20(2), 247-276.

Locatis, C., Williamson, D., Gould-Kabler, C., Zone-Smith, L., Detzler, I., Roberson, J., Maisiak, R. and Ackerman, M. (2010) Comparing in-person, video, and telephonic medical interpretation. *Journal of General Internal Medicine* 25(4) 345-50.

Mikkelson, H. (2003) Telephone Interpreting: Boon or bane? in L. Pérez González (ed), *Speaking in tongues: language across contexts and users*. València: Universitat de València, 251–69.

Miler-Cassino, J. and Rybinska, Z. (2011). AVIDICUS comparative studies – part III: Traditional interpreting and videoconference interpreting in prosecution interviews. In Braun, S. and J. Taylor (Eds), 117-136.

Moser-Mercer, B (2005) Remote interpreting: issues of multi-sensory integration in a multilingual task. *Meta* 50(2), 727-38.

Moser-Mercer, B. (2003) Remote interpreting: assessment of human factors and performance parameters. *Communicate!* Summer 2003. http://aiic.net/ViewPage.cfm?page_id=1125 (accessed 12/03/2016).

Mouzourakis, P. (1996) Videoconferencing: techniques and challenges. *Interpreting* 1 (1), 21-38.

Mouzourakis, P. (2006) Remote interpreting: a technical perspective on recent experiments. *Interpreting* 8(1), 45-66.

Nagao, Makoto. 1984. A framework of a mechanical translation between japanese and english by analogy principle. Artificial and Human Intelligence (A. Elithorn and R. Banerji, editors). Elsevier Science Publishers. B.V. © NATO, 1984

Och F. J. and H. Ney. Discriminative training and maximum entropy models for statistical machine translation. In Proc. of ACL , 2002.

Ortiz-Martınez, Daniel and Francisco Casacuberta. 2014. The New THOT Toolkit for Fully-Automatic and Interactive Statistical Machine Translation. Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics , pages 45–48, Gothenburg, Sweden, April 26-30 2014. c 2014 Association for Computational Linguistics

Ozolins, U. (2011) Telephone interpreting: Understanding practice and identifying research needs. *Translation and Interpreting* 3(1), 33-47.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, Evan Herbst. (2007) "Moses: Open Source Toolkit for Statistical Machine Translation". Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session, Prague, Czech Republic, June 2007.
Pollard, Carl and Sag, Ivan, 1994. Head Driven Phrase Structure Grammar, University of Chicago Press, Chicago, 1994.

Price, E., Pérez-Stable, E., Nickleach, D., López, M. and Karliner, L. (2012) Interpreter perspectives of in-person, telephonic, and videoconferencing medical interpretation in clinical encounters. *Patient Education and Counseling*, 87(2), 226-32.

Rosenberg, B.A. (2007) A data driven analysis of telephone interpreting. In C. Wadensjö, B. Englund Dimitrova, and A. L. Nilsson, eds. *The Critical Link 4. Professionalisation of interpreting in the communit*y. Amsterdam: Benjamins, 65-76.

Roziner, I. and Shlesinger, M. (2010) Much ado about something remote: Stress and performance in remote interpreting. *Interpreting* 12(2), 214-47.

Schäler, Reinhard, Andy Way, Michael Carl, EBMT in a Controlled Environment, 2003/1/1 in Recent advances in example-based machine translation Pages 83-114 Publisher Springer Netherlands

Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard and Lucia Specia.2014. Post-editing of Machine Translation: Processes and Applications. Cambridge Scholars Publishing

Steedman, Mark, 1987. Combinatory Grammars and Parasitic Gaps, Natural Language and Linguistic Theory, Vol. 5, 403-439.

Viaggio, Sergio (2011) *Remote interpreting rides again*. http://aiic.net/page/3710/remote-interpreting-rides-again/lang/1 (accessed 24/01/2014).
Wahlster, Wolfgang (ed.) 2000. Verbmobil: Foundations of Speech-to-Speech Translation, Springer.

Waibel and Kai-Fu Lee, The present status of automatic translation of languages', Advances in Computers , vol.1 (1960), p.158-163. Reprinted in Y. Bar-Hillel: Language and information (Reading, Mass.: Addis on-Wesley, 1964), p.174-179.

William D.Lewis: Haitian Creole: how to build and ship an MT engine from scratch in 4 days, 17 hours, & 30 minutes. EAMT 2010: Proceedings of the 14th Annual conference of the European Association for Machine Translation, 27-28 May 2010, Proceedings ed.Viggo Hansen and François Yvon; 8pp.