

Automated Simultaneous Interpretation: Hints of a Cognitive Framework for Machine Translation

Rafael E. Banchs

Institute for Infocomm Research
1 Fusionopolis Way, #21-01, Singapore 138632
rembanchs@i2r.a-star.edu.sg

Abstract

This discussion paper presents and analyses the main conceptual differences and similarities between the human task of simultaneous interpretation and the statistical approach to machine translation. A psycho-cognitive model of the simultaneous interpretation process is reviewed and compared with the phrase-based statistical machine translation approach. Some interesting differences are identified and their possible implications on machine translation methods are discussed. Finally, the most relevant research problems related to them are identified.

1. Introduction

Nowadays, translation has become an important element of daily life. Indeed, the emergence of modern information and communication technologies and the resulting globalization phenomenon are continuously boosting the need for translation services and applications. Within the context of professional translation, three different types of human translation tasks can be identified:

Document translation. This task refers to the situation in which the professional translator is required to generate a target-language-version of a given source document. In this kind of situations, full understanding of the source material is required and full generation of the target must be accomplished. In general, free translations are acceptable, no specific time constraints are imposed, and the best translation quality is expected.

Consecutive interpretation. This task refers to the situation in which the professional translator is required to mediate the communication between

two persons that speaks different languages. The basic communication protocol in this case is based in a turn-taking strategy, in which interlocutors must speak one at a time when they are given the right to speak. In this kind of situations, full understanding of the source material is required and full generation of the target must be ideally accomplished, while a ‘shared’ time constrains exists.

Simultaneous interpretation. This task refers to the situation in which the professional translator is required to translate on-the-fly what other person is saying in a different language. In this case no turn-taking is allowed as the translator is expected to produce the translated speech while the main speaker continues speaking. In these situations, full understanding and full generation is not mandatory, as the interpreter must keep the main speaker’s pace because ‘concurrent’ time constraints exist.

Current machine translation technologies have been theoretically and empirically designed under assumptions related to the first and second categories defined above. As far as we know, only few attempts have been done to apply machine translation to the specific problem of simultaneous interpretation. Indeed, previous research in this area can be traced back to the Vermobil¹ project, as well as to work from Kitano (1991) and Furuse and Iida (1996), who proposed the use of incremental translation. Later on, Mima *et al.* (1998) developed the idea of example-based incremental transfer.

The main objective of this discussion paper is to highlight the differences and similarities between the human task of simultaneous interpretation and statistical machine translation aiming at proposing

¹ <https://en.wikipedia.org/wiki/Verbmobil>

a research agenda for problems related to automated simultaneous interpretation. The rest of this discussion paper is structured as follows. First, in section 2, a recently proposed psycho-cognitive model of human simultaneous interpretation is presented along with its possible implications on machine translation. Then, in section 3, a cognitive framework for machine translation based on the described psycho-cognitive model is proposed.

2. Comparative Analysis

Although the translation process might slightly vary from person to person depending on a wide variety of factors, according to recently proposed psychological models of memory and attention (Padilla-Benítez and Bajo 1998), during a simultaneous interpretation task five different subtasks are conducted by the human brain: listening, segmentation, translation, reordering, and utterance production.

2.1 A Model of Simultaneous Interpretation

According to the aforementioned process, a human interpreter segments the input utterance into meaning units, which constitute basic semantic units that can be represented and manipulated at the cognitive level (Oleron and Nanpon 1965). An interesting, and also curious, fact about these meaning units is that the average human brain is able to process 7 ± 2 of such units at a time (Miller 1956). This seems to be indicating some sort of cognitive buffer size which happens to be independent of the language.

In parallel to the segmentation process, each meaning unit is translated by taking into account the surrounding 7 ± 2 unit context, and after having translated several units reordering and post-edition procedures are performed; producing, in this way, the output utterance. All these parallel processes are continuously operating while the input utterance is being received and the output utterance is being emitted; this last one with a corresponding latency of some few words.

2.2 Main Differences with SMT

Based on the information above, some important observations can be derived.

First, notice that the processes of translation and reordering are conducted sequentially, in the sense that reordering and utterance production are con-

ducted after some meaning units have been translated. So, differently from the SMT framework, in which translation and reordering are performed simultaneously during decoding; in the human interpretation case, reordering is performed after translation. This means that reordering decisions do not affect unit selection.

Second, unit selection is made by taking into account a 7 ± 2 meaning unit context, which includes both preceding and subsequent semantic information. As each meaning unit is composed of several words, the considered contexts in this case are larger than the ones considered in SMT, as well as they span over subsequent words.

No complex search strategy seems to be applied by the human interpreters. Indeed, the search strategy seems to be much simpler than in the case of SMT decoding. By decomposing the decoding task into two separated processes: translation and reordering, a simpler search strategy can be utilized.

2.3 Automated Simultaneous Interpretation

The three basic observations described above have very important implications on the way state-of-the-art SMT operates, and on the possible avenues of research for adapting this kind of systems to the specific task of simultaneous interpretation. These implications are described below.

The translation strategy is indeed very simple: a 1-to-1 mapping between meaning units which is context dependent. In the ideal case, the context and the source meaning unit must almost uniquely define the corresponding target meaning unit.

Although meaning units are not clearly defined by psychologists, they can be thought of as a set of optimal units for information representation and transference, which admit translation. According to this, meaning units should not be either generated or deleted during translation (i.e. for a given input, the corresponding translation must have the same number of meaning units).

Although reordering continues to be a NP-complete problem, the search space can be strongly reduced as meaning unit mapping should be able to produce a much reduced set of candidate translation units. In the ideal scenario, stacks of size 1 would be produced and reordering can be reduced to a simple permutation strategy of meaning units rather than words.

Source context plays a very important role in the human simultaneous interpretation framework, and

human translation in general. This means that both semantics and pragmatics have a preponderant role in the process of translation production, which moves a step ahead from current state-of-the-art technologies that make a very limited use of source context information.

Based on these important implications and the corresponding observations they were derived from, in the following section, we will attempt to define a research agenda for the problem under consideration. In such a research agenda, we define the main challenges and subtasks that must be addressed for successfully applying current state-of-the-art machine translation technologies to the problem of automatic simultaneous interpretation.

3. A Cognitive Framework for SMT

By taking into account the psycho-cognitive model for simultaneous interpretation described in the previous section, we can propose a SMT framework for automated simultaneous interpretation. In such an approach, the following four basic subtasks must be considered:

Segmentation. This subtask is responsible for segmenting the input data stream into meaning units which admit translation. Such meaning units must be minimal in the sense that not subunits can be contained into them, and must be maximal in the sense that, given a semantic context, translatability for every unit is guaranteed.

Translation or target unit selection. This subtask is responsible for selecting (or generating) the most appropriate target meaning unit for translating each input meaning unit within the current block of data under consideration. Target unit selection must be done by taking into account both the source unit to be translated and its context.

Reordering. This subtask is responsible for generating appropriate reordering for target units. Notice that this kind of reordering accounts only for long reordering (chunk reordering), as short reordering (word reordering) has been already accounted for in the unit selection stage. Reordering decisions in this task must be mainly done based on target language information.

Post-edition. This subtask is responsible for concatenating the reordered target units. This subtask must deal with two specific types of problems: boundary overlapping, where consecutive units are to be merged by resolving possible word overlap-

ping; and boundary gaps, where consecutive units are to be concatenated by filling-in possible gaps.

The proposed strategy allows for concurrently generating a target output stream while a source input stream is being received. The four aforementioned subtasks are to be pipelined so they sequentially process a given block of data. However, they are concurrently operating over a buffered stream of data. As a logical consequence of the pipeline, the overall system exhibits some latency, which must be in the range between 5 to 15 words if the automatic system is intended to mimic human simultaneous interpretation (Padilla-Benítez and Bajo 1998).

3.1 Source Input Segmentation

The main challenges of this subtask include the definition and operationalization of meaning units, as well as the definition of contextual boundaries. According to the previous discussion, a meaning unit must satisfy the following constraints:

Informative. The meaning unit must constitute a self-contained and elementary unit of information, which should be understandable within its context but without the need for specific informational elements from the surrounding units.

Translatable. The meaning unit must have an equivalent representation in the target language. If the complete source message is to be transmitted through the translation process, all individual source units must have a corresponding target unit. Source and target meaning units in a parallel sentence pair should admit one-to-one alignments.

Minimal. A meaning unit should not contain sub-units that are coherently both informative and translatable.

Although there is not a clear definition on what psychologists refer to as a meaning unit, the above described properties make it clear their utility as basic elements for information transmission and understanding. From these properties and the specific characteristics of the problem under consideration, a pragmatic definition for meaning units can be grounded on a translation optimality criterion, such as a unique segmentation which minimizes the translation effort and maximizes the translation quality.

Research work on this specific subtask must be supported by and would certainly benefit from recent research in the areas of collocation extraction,

multiword expression identification, name entity recognition and shallow parsing.

A recent study (Williams et al. 2013) explored the task of meaning unit segmentation by human annotators in languages such as English and Chinese. The result of this study suggested that no optimal solution seems to exist for this problem. Although any random segmentation is definitively not acceptable, it seems to be some preferential but variable trends on how humans perform meaning unit segmentation.

3.2 Target Unit Selection

One of the most interesting observations derived from the psycho-cognitive model is that the translation of a given meaning unit seems to be uniquely determined by its surrounding context. This fact allows for a theoretical justification on decoupling the problem of unit selection from the problem of reordering (long reordering, actually), as the problem of target unit selection becomes independent from the target language structure, which can be dealt with afterwards by means of reordering strategies that should only use target language information. The main properties that are desirable for target unit selection according to the proposed methodology are as follows:

Completeness. One source meaning unit should be translated by means of one target meaning unit which conveys an equivalent amount and type of information.

Unambiguousness. The available context information must be used to resolve ambiguity problems at this stage. According to this, the possible options for a target meaning unit, given a source meaning unit, should be restricted to a very small set of equivalent meaning units.

Different from state-of-the-art SMT, in which a large number of possible translations are considered and filtered by means of frequency-based criteria, the main objective of the proposed target unit selection approach is to fully use the context information to resolve ambiguity and produce a restricted set of candidate units. This reduction of target unit candidates should guarantee a better overall lexical selection as well as reduce the computational complexity of the decoding process.

Research work on this specific subtask of target unit selection can be supported by and would certainly benefit from recent research in the areas of word sense disambiguation, distributional seman-

tics, syntax-based machine translation, and cross-language information retrieval.

Significant effort on using source-context information to improve target unit selection has been reported during the last few years for both domain adaptation and lexical semantics (Carpuat and Wu 2005, España-Bonet et al. 2009, Haque et al. 2010, Banchs and Costa-jussà 2011)

3.3 Target Unit Reordering

Reordering is probably one of the most important challenges in SMT research, as well as it is the key factor responsible for decoding being an NP-complete problem (Zaslavskiy et al. 2009). The significant reduction of candidate hypothesis resulting from the use of source context information during target unit selection certainly reduces the computational burden for reordering. According to this, a more exhaustive search of the solution space can be afforded, which should also help improving the quality of the resulting translations.

Notice that, in this specific subtask, we are dealing with chunk-based reordering rather than word-based reordering. Indeed, word-reordering is assumed to be accounted for during the phase of target unit selection, in which each target meaning unit should already incorporate the corresponding word-reordering that is required to convey the desired meaning.

Valuable research related to this area includes lexicalized reordering as well as class-based reordering, dependency parsing and syntax based machine translation (Li et al. 2007, Nagata et al 2006, Zhang et al. 2007, Costa-jussa and Fonollosa 2006, Wang et al. 2007).

3.4 Output Post-Editon

The final subtask in our proposed framework has to do with the specific problem of merging the resulting sequence of output meaning units. After meaning unit selection and reordering, a simple concatenation strategy does not guarantee a fluid and grammatically correct output utterance.

Specific problems at the boundaries of the consecutive meaning units can be expected to occur, which can be basically grouped into two categories: boundary overlapping and boundary gaps. In this sense, the post-edition subtask can be thought of as a smoothing procedure for making the concatenation of consecutive meaning units more fluid and grammatical.

Research work on this specific subtask of output post-edition can be supported by and would certainly benefit from recent research in areas such as language modeling, syntactic- and semantic-based grammatical correction, and paraphrasing.

4. Conclusions

This paper discussed the main conceptual differences and similarities between the human task of simultaneous interpretation and the statistical approach to machine translation. A psycho-cognitive model of the simultaneous interpretation process was reviewed and compared with the statistical machine translation approach. Based on this, a research agenda for cognitive-based automated simultaneous interpretation has been discussed.

The most important application of automated simultaneous interpretation would be in the area of speech-to-speech translation; and more specifically in the case computer-mediated cross-language dialogue or discourse.

References

- Rafael E. Banchs and Marta R. Costa-jussa. 2011. A semantic feature for statistical machine translation. In Proceedings of the Fifth Workshop on Syntax, Semantics and Structure in Statistical Translation, SSST-5, pages 126–134.
- Marine Carpuat and Dekai Wu. 2005. Word sense disambiguation vs. statistical machine translation. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL '05, pages 387–394, Stroudsburg, PA, USA.
- Boxing Chen, Roland Kuhn, and George Foster. 2013. Vector space model for adaptation in statistical machine translation. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1285–1293, Sofia, Bulgaria, August.
- Marta R. Costa-jussà and José A. R. Fonollosa. (2006) Statistical machine reordering}. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, p.70--76
- Cristina España Bonet, Jesus Gimenez, and Lluís Marquez. 2009. Discriminative phrase-based models for arabic machine translation. Transactions on Asian Language and Information Processing, 8(4):15:1–15:20, December.
- Osamu Furuse and Hitoshi Iida. 1996. Incremental Translation Utilizing Constituent Boundary Patterns. In *Proceedings of Coling'96*, pages 544–549
- Hiroaki Kitano. 1991. Φ DM-Dialog: An Experimental Speech-to-Speech Dialog Translation System. *Computer* 24(6):36–50.
- C.H. Li, D. Zhang, M. Li, M. Zhou, M. Li, and Y. Guan. 2007. A probabilistic approach to syntax-based reordering for statistical machine translation. In Proc. of ACL, pages 720–727, June
- George A. Miller. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity of Processing Information. *The Psychological Review*, Vol 63:81–97.
- Hideki Mima, Hitoshi Iida and Osamu Furuse. 1998. Simultaneous interpretation utilizing example-based incremental transfer. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics*, pages 855–861
- M. Nagata, K. Saito, K. Yamamoto, and K. Ohashi. 2006. A clustered global phrase reordering model for statistical machine translation. In Proc. of ACL, page 720
- Pierre Oleron and Hubert Nanpon. 1965. Recherches sur la traduction simultanée. *Journal de Psychologie Normale et Pathologique*, 62(1):73–94.
- Presentación Padilla Benítez and Teresa Bajo. 1998. Hacia un modelo de memoria y atención en interpretación simultánea. *Quaderns. Revista de traducció*, 2:107–117.
- C. Wang, M. Collins, and P. Koehn. 2007. Chinese syntactic reordering for statistical machine translation. In Proc. of EMNLP, pages 737–745.
- Jennifer Williams, Rafael E. Banchs and Haizhou Li. 2013. Meaning unit segmentation in English and Chinese: a new approach to discourse phenomena. In Proceedings of Workshop on Discourse in Machine Translation (DiscoMT), ACL 2013.
- Mikhail Zaslavskiy, Marc Dymetman and Nicola Cancedda. 2009. Phrase-Based Statistical Machine Translation as a Travelling Salesman Problem. In Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, pages 333–341, Suntec, Singapore.
- Y. Zhang, R. Zens, and H. Ney. 2007. Improved chunklevel reordering for statistical machine translation. In Proc. of IWSLT, pages 21–28, Trento, Italy, October.