DISFLUENCY SURFACE MARKERS AND COGNITIVE PROCESSING; THE CASE OF SIMULTANEOUS INTERPRETING

Myriam Piccaluga, Jean-Luc Nespoulous♦, Bernard Harmegnies**

*Laboratoire des sciences de la Parole, Académie Universitaire Wallonie-Bruxelles,
♦Laboratoire Jacques Lordat, Université de Toulouse-Le Mirail & Institut Universitaire de France
Myriam.piccaluga@umh.ac.be, nespoulo@univ-tlse2.fr, bernard.harmegnies@umh.ac.be

ABSTRACT

This paper is focused on a measurement ("Inter syllabic Interval": ISI) intended to improve the study of the chunking process in subjects performing a task of simultaneous interpreting ("SI"). ISI is introduced on the basis of a discussion of the main methodological trends in the field of SI experimental study. An exploratory experiment is carried out on a prototypical sample of 4 subjects, performing SI under several conditions (noise disturbance, speech rate disturbance, variations in the subjects' language- and task expertise). Results confirm the validity of ISI and suggest its interest for future research on spontaneous speech in various contexts.

Keywords: speech, chunk, interpreting, fluency, speaking rate

1. INTRODUCTION

Simultaneous interpreting, for the interpreter, is the task of rendering in one language a message that is simultaneously being received in another. This oral 'online' translation can be viewed, just like any other activity involving mediation between languages, as an activity that occurs spontaneously in all bilingual people, and therefore is a special case of *natural translation* [1]. After intensive training, this skill can be put to use in professional *conference interpreting* or *liaison interpreting*, and these professional services have become increasingly high-profile over the last few decades, thanks to the globalisation of working practices. In the present paper, we focus on what we shall call the *interpreting task,* regardless of the context in which it is performed, and whatever the skill level of the practitioner.

Various studies, both empirical and speculative, of cognitive processes during the interpreting task [e.g. 2 and 3] have agreed that the *simultaneous nature* of the task is more complex than the mere juxtaposition of input and output processes. For Paradis [4], the interpreting task involves successive performance of numerous sub-tasks in an iterative and interactive way, and this process ensures that the comprehension, translation and monitoring processes are maintained. Such a model clearly requires that the signal be divided into segments that are short enough to allow processing to take place. A chain of sub-tasks can thus be initiated while another chain is still in process, and several processing chains can therefore function in parallel as long as, at time ***t***, the simultaneously active sub-tasks are of different types and affect different parts of the source speech (cf. the notion of "incrementation" in Levelts's speech production model [5]).

The idea that division of the source discourse is a compulsory processing mode can be found in work by other authors, who use the concept of *'chunk'* to refer to a portion of the source speech that forms a brief unit that is nonetheless rich enough in information to be processed as a single semantic unit. This concept should be considered alongside the presence, in several models, of a memory *buffer*, which, explicitly or not, is reminiscent of the temporary storage mechanisms described in more general models of language [6].

It would be reasonable to think that the subject's ability to split into chunks the source speech is definitely an absolutely necessary (but not sufficient) condition for producing a high-quality interpretation. As a matter of fact, if the interpreter's processing proceeds without a hitch, entry chunking is regular, and this in turn leads to equally regular exit chunking. Conversely, if entry-level chunking is irregular, or if there is any disturbance at any processing level, exit-level chunking is likely to be irregular.

In this article, we explore a methodology for measuring the validity of a variable (Inter-Syllabic Interval or ISI) which is expected to be useful in assessing chunking phenomena without requiring the classical approaches whose limitations we shall show. In order to achieve such a goal, the variable in question is tested in the context of an exploratory experimental study which aims to investigate its measurement capability. In that experiment, we observed the productions of subjects involved in simultaneous interpreting tasks. Four bilingual subjects with various degrees of expertise in interpreting and various degrees of mastery of French and Spanish have been recorded while interpreting conferences given in both languages. The source conferences have moreover been disturbed by changes both in speech rate and in auditory quality.

2. QUANTIFICATION

2.1. Classical approaches

The study of temporal discourse structure can, broadly speaking, follow one of two directions. The first is based on the evaluation of the relative level of the signal produced compared to the time spent speaking, assessed either by counting, or by measuring the duration of syllabic groups, following the groundbreaking work done by Goldman-Eisler [e.g. 7]. The second uses pause analysis, as pauses can be considered to mark chunk boundaries. Whatever method used to analyse speech – measuring the speech signals that compose it, focusing more on gaps, or even simultaneously measuring both these parameters – there is always the tricky problem of determining where segment boundaries fall. Thus, it is not always easy to define with certainty where one syllable ends and when the next begins: for example, if the first syllable ends with a vowel and the second syllable begins with another one; in addition, some consonant structures (particularly approximants) are problematic when one comes to demarcation. Studying pauses seems attractive, but it must be recognised that there are several methodological problems with this approach. A pause-based study of chunking, assuming that pauses mark boundaries, requires a duration threshold: only pauses that “betray” cognitive activity (i.e. at a higher level of processing than phonetic production *per se*) can be taken into account. Therefore, pauses need to be divided into two sub-groups. The most usual criterion under use is duration, which, for some, must be above 250ms [e.g. 7]. This figure is, however, not the only one proposed in the literature, and this is probably due to differences in the methods used by various authors, but it might also be due to the cognitive nature of pauses. Quite interestingly, recent work demonstrates that there are significant differences between various types of pauses [8, 9]. Given this, the binary taxonomy underlying a decision-making system that is based on just one distinguishing measurement, causes epistemological and methodological problems. It can be seen that analysing a phenomenon (*chunks*) on the basis of a *non-event* (a break in the phonation stream), quite apart from its epistemological limitations, can cause technical problems. Moreover, if we attempt to grasp automatically the chunks borders, we soon encounter another threshold problem, this time related to intensity. If the signals are measured in anything but a completely silent environment, the silent intervals risk being contaminated by environmental noise, which would invalidate the whole procedure.

2.2. An innovative approach

The variable under study (ISI) is obtained using time differentiation of syllable intensity peaks [10]. For each peak, the time interval separating it from the preceding syllabic peak is calculated. For a corpus of N syllables, a chronologically ordered list of N–1 inter-syllabic periods is obtained. ISI graphs show a more or less regular alternating pattern of peaks and troughs. The peaks correspond to long ISIs, which are characterised by a momentary reduction in speed delivery; the troughs correspond to a period of faster delivery speed. When the interpreting process is carried out without problems, changes in ISI show a somewhat regular alternation of peaks (chunk borders) and troughs. When the *subject expertise / task difficulty* ratio decreases, the interpreting process may locally stumble, causing numerous (and/or long) pauses and vowel lengthening, and therefore provoking periods of high and frequent ISI peaks. In this study, we aim at systematizing the just mentioned observations drawn from earlier research. For this purpose, we use an experimental design allowing to control in various ways variations of the *subject expertise /task difficulty* ratio, and therefore to assess the validity of the ISI measurements.

3. EXPERIMENTAL STUDY

3.1. Subjects

Four subjects, all originating from the Barcelona region and all having lived there for several decades, underwent the experiment. All have a high mastery of French and Spanish and speak both languages frequently, even though Spanish is the dominant language in the whole sample. These similarities notwithstanding, there are across-subjects differences in terms of both their expertise in interpreting and their linguistic skills (table 1).

Table 1: subjects linguistic and task expertise

	Professionals	Non-professionals
Late bilinguals	*Int1*	*NP1*
Near-natives	*Int2*	*NP2*

3.2. Source corpora

Each subject was assigned 6 interpreting tasks, each consisting of a conference originally given in the European Parliament. Each subject had to interpret 3 conferences in each of the two combinations (from *French to Spanish* and from *Spanish to French*). The effect of the conference was neutralized by means of a latin square design.

3.3. Disturbances

Disturbance was artificially introduced into each source conference. This consisted of 1. local alterations in speech rate (increase in speed by means of a reduction in duration of 80%, 70% or 60% of the original duration, without modifying the F_0 characteristics, thanks to the time compression algorithm of a CSL 4300B Kay Elemetrics station) and 2. local addition of white noise (by merging of the source conference speech file with a file containing alternated portions of silence and portions of white noise – each 1 minute long – generated by the CSL station, at levels of 0 dB, 3 dB or 6 dB re/the average level of the speech signal). The analyses presented here have been applied only on portions of the subjects utterances produced under both speed- and noise disturbance, in a two-way 3X3 design.

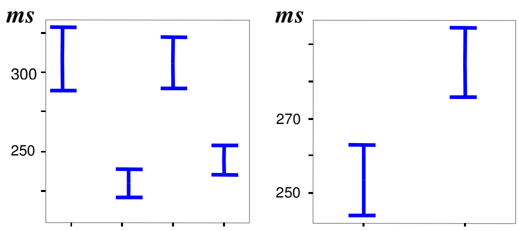
3.4. Statistical methods

We shall analyse the average ISI variation using analysis of variance, which will enable us to test the hypothetical effect of *noise disturbance*, *speed disturbance*, *linguistic combination* and *subject* considered as independent variables, taking into account the *interactions* between them (not presented here because of lack of space).

3.5. Results

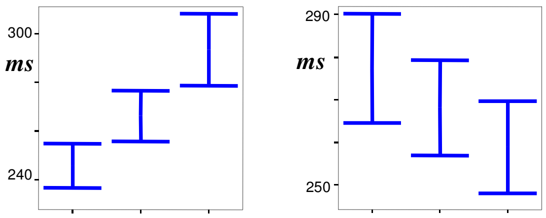
Our procedure resulted in the collection of 12,930 ISI values. The more usual ones are to be found around 170ms (median 173ms); the 50th, 90th and 95th percentiles lie within the interval [0s-1s]. The distribution nevertheless extends over almost 10s. This is probably due to the fact that when exceptionally severe incidents occur, they may produce very high ISI values (e.g. case of a subject remaining silent during several seconds because he or she lacks the appropriate words to translate in the target language), even though such incidents are very rare in our sample of highly proficient speakers. As a consequence, the distributions are not Gaussian (pronounced rightward skewness, strong kurtosis); for the purpose of applying analysis of variance to our data, we therefore applied a $\tanh^{-1}$ transform in order to normalise the distributions. To make our results easier to read, in the following figures, ISI values are nevertheless presented in their original units (milliseconds).

Figure 1: average ISI values, plus and minus 1 standard deviation (Y-axis : milli-seconds). *Left:* by speaker (from left to right: *NP1*, *NP2*, *int1*, *int2*). *Right:* by combination (from left to right: French to Spanish, Spanish to French).



As figure 1 shows, the subjects showed markedly differing performance in terms of ISI (*F=70,104, d.f.=3, α<.001).* Two distinct groups appear: *NP2* and *int2* showed values varying from 225 to 250ms, and *NP1* and *int1* showed values around 310ms. A clear distinction in terms of language expertise appears here. The ISI values also change under language combination effects. The difference, which is slight but significant (*F=216,368, d.f.=1, α<.001*), consists of a tiny increase when the target language is French, which is the least mastered language in all four subjects. As figure 2 shows, an increase in the level of noise disturbance causes an increase in ISI (*F=20,489, d.f.=2, α<.001*). This relationship is almost linear. Just as linear, however, is the significant fall (*F=19,474, d.f.=2, α<.001)* caused by an increase in the compression rate.

Figure 2: average ISI values, plus and minus 1 standard deviation (Y-axis: milliseconds). Left: by level of noise disturbance (increasing from left to right: 0, 3, 6 dB). Right: by level of speed disturbance (time compression increasing from left to right: 80%, 70%, 60% of the original duration).



4. CONCLUSIONS

From a strictly methodological point of view, it can be noted that the variable under study has proved suitable for retrospective distinguishing of groups of measurements that correspond to initial classes of the experimental groups of data (linguistic expertise, linguistic combination, noise- and speed disturbance). It therefore seems that, apart from being conceptually valuable, ISI turns out to be a valid tool for the study of the interpreting behaviour.

Observations obtained from the independent variables *subject*, *combination* and *noise disturbance* tend to show that when the *subject expertise / task difficulty* ratio is high (i.e., in a situation where cognitive constraints are minimal), ISI is low. Conversely, ISI increases when the cognitive load does (interpretation towards French, noise increase, late bilingualism). These tendencies are *temporal manifestations* in the behaviour of the subjects of the way in which they cope with these controlled *a-temporal factors*. Nevertheless, when temporal factors are manipulated, the effect is different: increasing time compression elicits lowered ISI values. Thus when the input speech rate is artificially increased, ISI decreases because the subjects are faced to the priority constraint of maintaining the overall simultaneity of the source and target discourses.

Furthermore, it should be pointed out that if our experimental study has led us to adopt a *macroscopic* view, well suited to the assessment of the measurement capacity of a numerical mechanism, however, a more detailed, *microscopic* examination of the relationship between variations in ISI and observed behaviour would be very interesting. Such a study could involve an analysis of *speech production* (e.g. phonetics, phonology, lexicon, morpho-syntax and pragmatics), *mediation* (translation) and indeed the *subjective quality* of the target speech from the point of view of the end listener (intelligibility).

Finally, we should also note that our statistical analysis led us to focus only on one aspect of the measurements: their *central tendency*. As a matter of fact, we have managed in order to have ISI variation depend upon the variations of the independent variables only, therefore maintaining the within-group variability of the data fairly constant (the group variances are quite similar, as suggested by the bar graphs in figs. 1 and 2). In other situations, where specific effects of non controlled factors could happen, a very careful examination of fluctuations in the ISI variance would nevertheless be very useful, since it would reveal unpredicted sources of variance. We do think that such treatment is likely to be fruitful in analysing other types of speakers with disfluency, be they observed in normal or pathological conditions.

5. REFERENCES

[1] De Groot, A.M.B. 1997. The cognitive study of translation and interpretation. Three approaches. In: Danks, J.H., Shreve, G.M, Fountain, S.B, McBeath, M.K (eds), *Cognitive processes in translation and interpreting*. Thousand Oaks: Sage Publications, 25-56.

[2] Gerver, D. 1976. Empirical studies of simultaneous interpretation: a review and a model. In Brislin R.W, (ed.), *Translation: Applications and Research*. New York: Gardner Press, 165-207.

[3] Moser, B. 1978. Simultaneous interpretation: a hypothetical model and its practical application. In Gerver, D., Sinaiko, W. (eds.), *Language interpretation and communication*, New York and London: Plenum Press, 353-368.

[4] Paradis, M. 1994.Toward a neurolinguistic theory of simultaneous translation: the framework. *International Journal of Psycholinguistics*, 10, 3 (29), 319-335.

[5] Levelt, W.J.M. 1991. Speaking: from intention to articulation. Cambridge: MT Press.

[6] Baddeley, A. 1997. *Human Memory Theory and Practice (Revised Edition)*, Boston: Allyn & Bacon.

[7] Goldman-Eisler, F. 1972. Segmentation of input in simultaneous translation, *Journal of Psycholinguistic Research*, 1, 2, 127-140.

[8] Campione E., Véronis, J. 2004.Pauses et hésitations en français spontané, *Actes des 25èmes Journées d'Etudes sur la Parole (JEP 2004)*, 109-112.

[9] Candéa, M. 2000. *Contribution à l'étude des pauses silencieuses et des phénomènes dits "d'hésitation" en français oral spontané.* Ph D.Thesis, Université Paris III.

[10] M. Piccaluga, *Approches psycholinguistiques de l'interprétation*. Ph D thesis, Université de Mons-Hainaut, 2004.