THE INTERPRETERS' NEWSLETTER

Number 13

2005

In copertina: *Sava*, photo by Alice Polenghi

CONTENTS

EDITORIAL

The present issue of the *Interpreters' Newsletter* is devoted to methodological aspects of Conference Interpreting Research (CIR), a topic put forward in the past mainly by Gile (1983, 1994, 2004) but also in other publications (Gambier *et al.* 1997; Gile *et al.* 2001; Schäffner 2004) and which has gathered momentum over the past few years reflecting a greater interest in and awareness of methodological questions. The papers included in this issue illustrate how CIR is evolving and how researchers are meeting the challenges entailed in the choice of methods and tools to be used for investigation and analysis of conference interpreting.

The contributions cover a series of areas – prosody, fluency, transcription tools, norms, personality traits, corpus linguistics, questionnaire-based surveys and statistics, with a clear predominance of fields where computer-based instruments are put to fruitful use. Interestingly, in some instances, existing tools mainly used in other disciplines have been adapted to meet the specific needs of CIR.

Barbara Ahrens describes difficulties related to the analysis of prosody in SI, from corpus recording to digitizing and transcribing the data, and comes up with possible solutions based on the use of appropriate software. In this respect, mention should be made of PRAAT, a computer program especially designed for speech analysis.

Another approach to prosody and its representation for CIR purposes is adopted by Philippe Martin, an outsider to interpreting research who illustrates the linguistic point of view in the description of prosody and analyses differences between read and impromptu speeches. Caterina Falbo, who has been collaborating with Martin, in her article describes how Winpitch has been adapted to meet the specific needs of transcribing interpreted texts and outlines the problems most frequently encountered.

Peter Mead's contribution deals with methodological aspects in the study of interpreters' fluency and adopts a practical approach to its assessment, while providing a useful point of contact with linguistic research. In particular, he concentrates on five temporal variables, three of which – speech rate, duration of pauses and mean length of run – seem to be the most relevant in the assessment of fluency.

Šárka Timarová's paper shows how corpus linguistics tools may be used in interpreting research, not only for quantitative but also for qualitative analyses.

The electronic analysis of a corpus – made up of four oral speeches in English and their interpretation into Italian Sign Language – also underlies Jane

Kellett's research aiming at producing a tri-lingual (Italian, English and Italian Sign Language) multimodal electronic glossary for interpreters. The methodology used is described in her paper in great detail.

The importance and usefulness of the translational norms paradigm in Descriptive Translation Studies and its relevance to conference interpreting are discussed by Carlo Marzocchi who stresses the need for a wider application of this notion.

Nancy Schweda Nicholson presents a study on personality characteristics of interpreter trainees. The tool used is the Myers-Briggs Type Indicator (MBTI). While admitting that language knowledge may be more important than personality type and that the MBTI is not to be viewed as a replacement for a traditional screening test, the author believes that a personality inventory might be useful to both trainers and trainees.

The present issue concludes with a sort of *querelle*. Franz Pöchhacker's paper reassesses and criticizes studies on interpreters' and users' quality expectations and preferences with the aim of consolidating and refining research practice and results and with an emphasis on statistical procedures for the analysis of survey data. Delia Chiaro and Giuseppe Nocella are among the authors discussed by Pöchhacker. Indeed, the paper they published in *Meta* is used as a starting point for methodological reflections. Chiaro and Nocella, who had read Pöchhacker's manuscript, asked us for an opportunity to reply to Pöchhacker's criticism. We have duly obliged, hopefully contributing thereby to greater transparency in the debate on methodology involving scholars and researchers.

<div align="center">Alessandra Riccardi and Maurizio Viezzi</div>

References

Gile D. (1983) "Aspects méthodologiques de l'évaluation de la qualité du travail en interprétation simultanée", *Meta* 28, 3, pp. 236-243.
Gile D. (1994) "Methodological aspects of interpretation and translation research", in *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. Ed. by S. Lambert and B. Moser-Mercer, Amsterdam/Philadelphia, John Benjamins, pp. 39-56.
Gile D. (2004) "Translation research versus interpreting research: kinship, differences and prospects for partnership", in *Translation Research and Interpreting Research. Traditions, Gaps and Synergies*. Ed by Christina Schäffner, Clevedon/Buffalo/Toronto, Multilingual Matters, pp. 10-34.

Gambier Y., Gile D. and Taylor C. (eds.) (1997) *Conference Interpreting: Current Trends in Research*, Amsterdam/Philadelphia, John Benjamins.

Gile D., V. Dam H., Dubslaff F., Martinsen B., Schjoldager A. (eds.) (2001) *Getting Started in Interpreting Research*. Amsterdam/ Philadelphia, John Benjamins.

Schäffner C. (ed.) (2004) *Translation Research and Interpreting Research. Traditions, Gaps and Synergies*, Clevedon/Buffalo/Toronto, Multilingual Matters.

ANALYSING PROSODY IN SIMULTANEOUS INTERPRETING:
DIFFICULTIES AND POSSIBLE SOLUTIONS

Barbara Ahrens
Johannes Gutenberg-University Mainz / FASK Germersheim

1.   Introduction

Prosody[1] is an integral part of orally produced texts. Firstly, it is used to structure the acoustic continuum uttered by a speaker and secondly, it is used to give prominence to those parts of the spoken text that the speaker considers to be important. Thus, prosodic elements are essential cues for the listener when processing spoken input (cf. Cutler 1983: 91). Prosody can also be an indicator of the mental-cognitive processes underlying speech production (cf. Goldman-Eisler 1958: 74).

The prosodic feature of intonation – which is defined as pitch movement due to changes of fundamental frequency (F_0) during oral speech production (cf. e.g. Cruttenden 1997[2]: 7, Günther 1999: 62; Schönherr 1997: 12, footnote 4) – has an important role to play in structuring and organizing communicative interaction. Intonation is used to indicate that the speaker will go on speaking or that further elements will follow (cf. Jin 1990: 123ff.; Selting 1995: 50ff.). Intonation has also a social function which depends on the speaker's social status or profession: e.g. priests can easily be distinguished from other professions by the way they speak (cf. Fiukowski and Ptok 1996: 670ff.).

Prosody in bilingual oral communication via an interpreter is as important as in monolingual communicative events. Prosodic elements in the source text (ST) convey meaning that is to be rendered in the target language (cf. Kade 1963: 19), and since the target text (TT) is produced orally, its prosodic features are equally important for the TT's addressee. This aspect was mentioned in early contributions on simultaneous or consecutive interpreting, but was not further developed or considered in interpreting studies for a long time.

2.   Prosody in simultaneous interpreting – the state of the art

Interpreters are professional speakers and there is no doubt that their voice and way of speaking are very important (cf. e.g. Alexieva 1990: 5; Cartellieri 1983: 213), especially in the case of simultaneous interpreting (SI) where the

---

1   Prosody comprises all suprasegmental features that depend on tonal, dynamic and durational parameters. Since these acoustic parameters can be measured objectively, they are important for computer-aided analyses.

interpreter's output is only perceived via headsets. Although most of the few authors who mentioned prosody in SI regard it to be non-marked (cf. Déjean Le Féal 1990: 155; Kirchhoff 1976: 67; Willett 1974: 103), there are also some who put forward a contrary opinion: Barik describes the way simultaneous interpreters speak as "less smooth than 'natural' speech" (Barik 1975: 294) and according to Shlesinger, "the intonational system used in simultaneous interpretation appears to be marked by a set of salient features not found in any other language use." (Shlesinger 1994: 226). For Fiukowski, the unnatural way of speaking in SI is conditioned by the linear ST comprehension and the simultaneous TT planning and production. This means that it cannot be avoided because of the simultaneous interpreting process itself (cf. Fiukowski 1986: 186). Kirchhoff (1976) too acknowledges that difficult speech processing conditions – which are characteristic for SI – can result in a less ideal TT production with hesitations, pauses, etc. (cf. Kirchhoff 1976: 67).

The few studies on prosody in SI conducted so far can be categorized according to the prosodic elements that were examined: pauses, speech rate and segmentation (cf. Alexieva 1988; Barik 1973; Gerver 1969; Goldman-Eisler 1967, 1968; Kreuzpaintner 2001; Lee 1999; Liebig 1994; Shlesinger 1994), accentuation and stress (cf. Pelz 1999; Shlesinger 1994; Williams 1995), as well as intonation and fundamental frequency (cf. Collados Aís 1998; Darò 1990; Shlesinger 1994).

3.   Difficulties to be dealt with when analysing prosody in SI

The limited number of studies dedicated to prosody in SI confirms that it has been a neglected field of scientific interest although its importance was acknowledged at the very beginning of interpreting studies (see Section 1). There are several difficulties that have to be dealt with:
- Approaches to analyses and methodology as well as definitions of prosodic phenomena are as diverse as the number of studies. This is not only the case in studies on prosody in SI, but also in studies on prosody in general (for a comprehensive overview with special emphasis to intonation, see Ahrens 2004: 75ff. and 117ff.)
- Purely auditive analyses are subjective, purely automatized speech processing is error-sensitive (cf. Schönherr 1997: 68).
- For a long time, the processing of audio and video data required very powerful computer resources and there was no user-friendly hard- and software available.
- Transcribing and analysing audio and video data is extremely time-consuming, which impairs the processing of large and representative corpora necessary for general conclusions (cf. Gile 1991: 158; Setton 1994: 183).

- Recording professional material in authentic settings is difficult and requires the consent of all parties involved, i.e. interpreters' team, speaker, conference organizer, because of the speaker's and interpreter's copyright for his/her performance (cf. AIIC – General terms of contract; Kalina 1994: 225).
- The scope and objective of the study requires certain quality standards for the recordings, such as sound quality, dual-track recordings etc. (cf. Kalina 1998: 135).
- Any transcription provides a selection of all phenomena comprised in the recordings, i.e. special attention is paid to the elements that are to be analysed (cf. Kalina 1998: 135).
- There are no generally accepted conventions of transcription for prosodic elements.
- Transcribing prosodic phenomena is difficult since they vary a lot. Nevertheless, certain patterns, e.g. falling or rising final pitch movements, can be distinguished and marked in the text (cf. Du Bois *et al*. 1993: 52).

4.   Recording the corpus

The corpus used for the study presented here comprised dual-track audio and video recordings of an English ST and three German TTs that had been made in authentic settings. The performance of three parallel booths with two professional interpreters in each of them, all working from English into German, was recorded at the Faculty of Applied Linguistics and Cultural Studies (FASK) of the University of Mainz in Germersheim during a guest lecture on an actual German-English translation job in the field of marketing communication held by a native British English speaker. All the interpreters – four women and two men – were professional interpreters who had been trained either at the FASK, University of Mainz (one woman) or at the University of Heidelberg (the other five) and were working actively and regularly as freelancers on the German private market with an average professional experience of 4.6 years. All six were native speakers of German, two of the women and the two men had English as B language, and the other two women as C language. The four women were working in Booth I and II, the two men in Booth III. In Booth I, there were the two interpreters with English as C language. After the lecture, the six interpreters filled in a questionnaire about their professional background, their preparation, their opinions on the ST and the way the ST speaker had presented it, the problems that had arisen and what they had done to solve them.

The ST speaker, whom the interpreters in the retrospective questionnaires described as "typically British", produced his speech spontaneously, using his manuscript only for short quotations and for planning how to proceed. The ST

and the way it was presented can be described as rather informal. For these reasons, the ST showed typical characteristics of spontaneous speech, such as hesitation phenomena, false starts or sudden variations in the speech rate (cf. e.g. Cruttenden 1997[2]: 174; Crystal 1969: 154; Goldman-Eisler 1958: 61).

The audio and video recordings of three parallel booths all interpreting under exactly the same external conditions – same ST, same language pair, same situation, same audience – were made in dual-track quality. For the synchronous video recordings of all three booths and the ST speaker, a fourfold splitscreen was used.

5.   Digitizing the data

The ST and the three TTs were digitized by means of *Wavelab 3.0* by Steinberg (see: http://www.steinberg.de/produkte/ps/wavelab/wavelab3/). This software allowed the synchronous digitizing of the two channels of the dual-track recordings. By using a sampling rate of 44.1 KHz (i.e. CD quality) for both channels, it was possible to obtain digitized data with optimum quality. After that, the two channels of each booth were separated and downsampled to 11.025 kHz in order to reduce the quantity of data to an amount that can be handled easily in computer-aided analyses. Due to the synchronous digitizing of the two audio channels of each booth, both channels can be aligned precisely although they were stored in different files (see Section 8).

6.   Transcribing the data

The ST and the three TTs were transcribed word by word. Although it is not very reader-friendly to use no interpunctuation, this procedure was chosen in order to avoid a misleading prosodic impression due to commas or full stops. This way of transcribing revealed that prosodic phenomena do not necessarily follow syntactically defined boundaries.

Since the analysis was to focus on pauses, segmentation into intonation units,[2] accentuation patterns and final pitch movements, these phenomena were indicated in the transcriptions. The following conventions of transcription were defined: All texts are written in small letters, syllables in capitals are stressed syllables. 1 line corresponds to 1 intonation unit (IU). "=" indicates lengthening of syllable. "\" means final falling pitch movement, "/" indicates a final rising pitch movement, "—"means final level contour, "¬" indicates a final semi-

---

2   An intonation unit (IU) is defined as a prosodic unit with a coherent F_0 contour and
       at least one pitch movement that is perceived as prominent (cf. Ahrens 2004: 111;
       Huber 1988: 71).

falling pitch movement, "—*" a final rise-level contour. "<0.96>" marks a pause with a duration of 0.96 seconds, "<A>" is a pause due to breathing, "<G>" indicates a pause with noise in the booth, "<M>" means that there is a pause during which the ST speaker is rustling his notes. "{...}" indicates paraverbal comments at the end of the intonation units they refer to, e.g. louder, faster etc. According to these conventions, the beginning of the ST reads as follows:

> *Example 1 (ST):*
> THANKS very much/ <1.29>
> ehm <1.75A> that's a rather LONG cable here and i have to be CAREful
> of— <0.39A>
> EHM— <1.69R>
> THANK you for the introDUCtion/
> THANK you for the invitation as WELL/
> i'm very pleased to BE here— <0.61A>
> ehm CAN i— <0.13>
> before i trip Over the WIRE/ <0.39A>
> can i ASK   <0.56>
> you HOW many of you hea=rd my LECture— <0.39A>
> LAST YEAR\
> THIS time LAST year— {deeper}
> was Anybody here\ <0.41A>
> is there Anybody here who was HERE a year ago\ <0.73> {faster}
> can you INdicate by raising your HANDS/ <0.69A>
> NObody\
> THAT'S very GOOD\ <0.18>

7.   Calculating speech rate

All perceived pauses were checked and measured using the speech signal of the digitized audio data and included in the transcriptions of all four texts of the corpus. In a second step, all spoken syllables were marked and counted. For each text of the corpus, the total number of syllables was divided by the total length of text (measured in seconds) in order to calculate the average speech rate (measured in syllables per second). The speech rate of the ST corresponds to what is considered a normal speech rate, i.e. an average of 5-8 syllables/second (cf. Goldman-Eisler 1961: 171), although the interpreters said that the ST had been presented very fast.

For calculating the rate of articulation, only the actual speaking time, i.e. total length of text minus total time of pauses, was used. The rate of articulation is in line with Barik's results (1973); he found that the rate of articulation of

interpreters is below that of the ST speaker (cf. Barik 1973: 257). Table 1 shows the results of all four texts.

	Total length of text (seconds)	Number of syllables	Speech rate (syllables/s)	Total time of pauses (seconds)	Rate of articulation (syllables/s)
ST	4,363.2	16,630	3.81	1,188.0	5.24
TT$_I$	4,365.6	13,971	3.20	1,347.0	4.63
TT$_{II}$	4,365.0	14,342	3.28	1,513.8	5.02
TT$_{III}$	4,363.8	13,856	3.17	1,491.0	4.82

Table 1   Speech rates and rates of articulation of the ST and TT$_{I-III}$

8.   Using PRAAT

For the analysis of the digitized audio data, *PRAAT*, a computer programme especially developed and designed for speech analysis by P. Boersma and D. Weenink at the Phonetic Sciences Department of the University of Amsterdam, was used (for further information, see: http://www.fon.hum.uva.nl/praat/). The dual-track recordings allowed the synchronization of the ST and the respective TT in *PRAAT*.

Figure 1 shows the screen view of a synchronized ST and TT$_I$ paragraph. In each analysis window, the speech signal is reflected in the upper track and the pitch and intensity contour in the lower. One can move forward or backwards in the texts by scrolling to the right or left. It is also possible to view further features in the analysis window, e.g. the spectrogram of both texts which was used for checking word and IU boundaries. Pitch is measured in Hz as indicated on the right-hand side of the analysis windows, intensity in decibels (dB) on the left-hand side. Time (in seconds) is given below the analysis windows. Like in Figure 2 below, further tiers can be defined in order to note down words, syllables, pauses, etc. in time-aligned transcriptions. Any selection of text, sound, speech signal, pitch, etc. can be stored in a time-aligned format in individual files that can be handled more easily for analysis purposes and for generating diagrams.
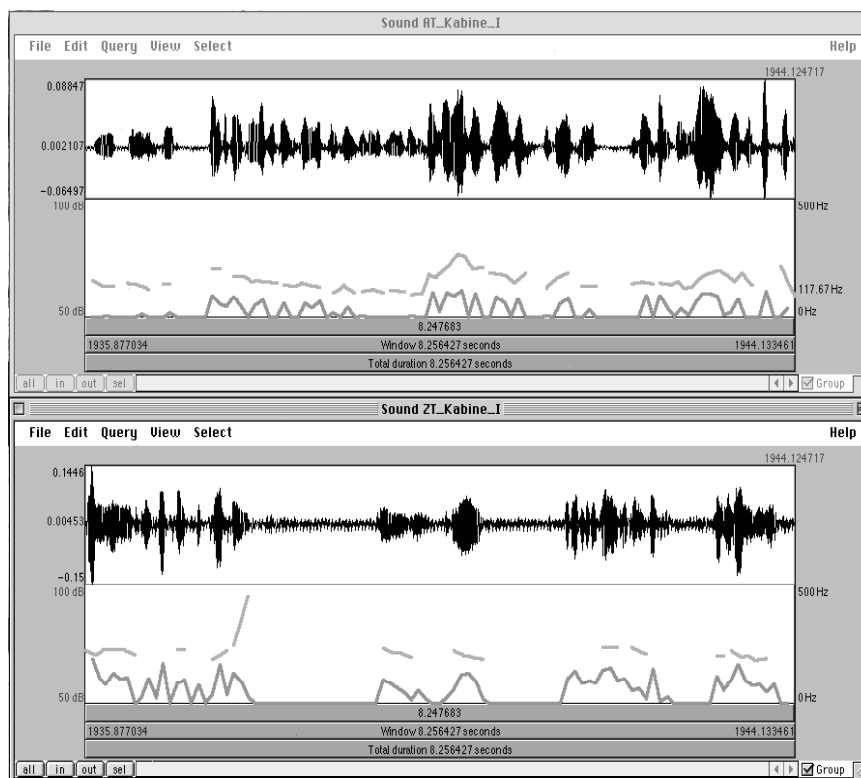
Figure 1   Screen view of a synchronized ST and TT_I paragraph in *PRAAT*

Example 2 (TT_II):
ich werde mir auch die proBLEme anschaun——*

The speech signal and the spectrogram were used to define the word boundaries in this IU, the local pitch movement in the word "proBLEme" indicated the accent in it. The characteristic rise-level F_0 contour[3] at the end of this IU can be identified clearly in Figure 2 above and Figure 3 below. Speech signal analysis confirmed these findings (see table 2).

---

3   A characteristic feature of the TTs was the interpreters' intonational singsong. In this case, the last stressed syllable in an IU showed a rising accent, i.e. the accent was carried out by a salient rising pitch movement. After the stressed syllable, the pitch remained on the frequency that had been reached by the accent until the end of the IU. Since this pattern looks like a combination of rising and level contours, it is called "rise-level" (cf. Ahrens 2004: 209ff.).
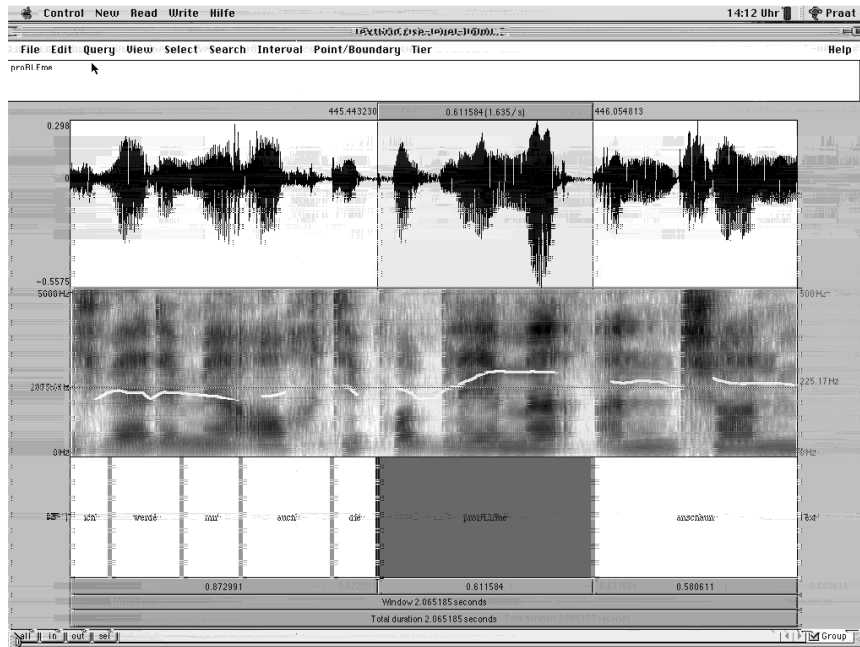
Figure 2   Screen view of example 2 – Speech signal, spectrogram, pitch and text
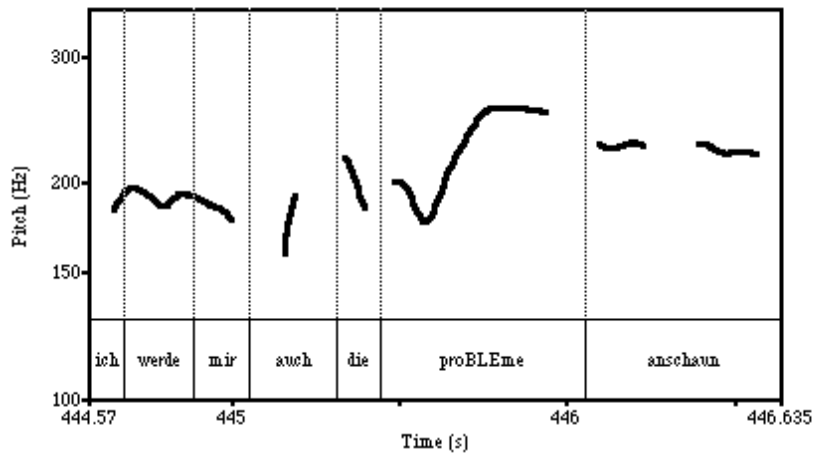
Figure 3   Example 2 - Typical rise-level contour in TT_{II}

	ich werde mir auch die proBLEme anschaun—*
Maximum F_0	253 Hz
Time Maximum F_0	445.79 s
F_0 at the end of IU	219 Hz
Mean F_0 for *level*	233 s

Table 2   Example 2 – Speech signal analysis

9.   Analysing segmentation into intonation units

In a first step, the auditively perceived pitch movements were noted down in order to reflect the segmentation into IUs and the accentuation pattern of all four texts. Then, the F_0 contour of all texts was calculated and visualized by means of *PRAAT*. The auditive results of the ST and the TTs were checked against their F_0 contour calculated by the computer. This combined two-step analysis allowed to overcome the shortcomings of purely auditive or purely computerized analyses and helped to obtain a refined picture of the intonational segmentation and the final pitch movement of each IU.

Example 3 (ST):
working with CONnotation and asSOciation\
THIS is the advertisement for ROver\ <0.28>
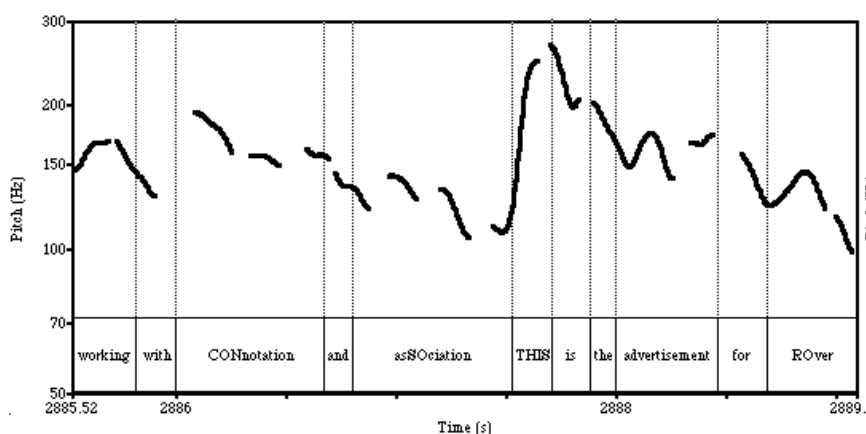
Figure 4   Sequence of two intonation units – Declination and *reset*

In Figure 4, the global falling pitch contour defined as declination (cf. Vaissière 1983: 55 ff.) as well as its *reset* in the second IU are evident. A

characteristic of the beginning of a new IU is the *reset* of the intonational contour on the first syllable of the new IU, the so called "onset syllable" (Crystal 1969: 143). By resetting, F_0 returns to the frequency level on which new IUs usually start in an utterance. Also the local pitch movements that are the reason why the words "CONnotation", "asSOciation", "THIS" and "ROver" are perceived as being prominent can be seen clearly. These prosodic phenomena were also confirmed by speech signal analysis (see Table 3).

	working with CONnotation and asSOciation	**THIS is the advertisement for ROver\ <0.28>**
Minimum F_0	103 Hz	90 Hz
Time Minimum F_0	2,887.33 s	2,889.07 s
Maximum F_0	187 Hz	279 Hz
Time Maximum F_0	2,886.11 s	2,887.71 s
Mean F_0	146 Hz	162 Hz

Table 3   Example 2 – Speech signal analysis

The continuous acoustic continuum of all three texts was divided into successive IUs. Perceivable as well as measurable boundary markers were: F_0 declination and F_0 reset, characteristic final pitch movements, such as final fall, final rise, final rise-fall, etc. (cf. e.g. Halliday 1966: 117ff., Kohler 1995[2]: 195 ff.), laryngealization, final lengthening (cf. Heuft 1999: 62) and sometimes pauses since they are not a necessary but an additional boundary marker. Very often, the end of an IU is signalized by a combination of several markers, e.g. final lengthening followed by a pause.

10.   Conclusion

Fundamental frequency is an objectively measurable parameter for analysis (cf. Gile 2003: 120). For this reason, computer-aided analysis of voice characteristics and prosody helps to gain more insight into the interplay of different prosodic phenomena and its acoustic parameters. Although computer-aided analysis is very helpful, it is always recommendable to cooperate with experts in voice and speaking skills as well as in signal processing. Nevertheless, analysing prosody remains difficult and time-consuming for the researcher. The analysing method applied to the corpus that is presented in this article is a conceptual approach. It parts from the main functions of prosody – structure and prominence – and examines how these manifest themselves in the ST and the TTs (cf. Ahrens 2004: 131ff.). In order to be able to describe prosodic characteristics of simultaneous interpreted texts, these have to be analysed in a first step as if they were monolingual, autonomous texts. In a

second step, to be examined is if salient prosodic features relate to the ST or the interpreting process itself (e.g. because of ear-voice-span, hesitations of the ST speaker, waiting for new ST input). An analysis of this kind requires digital or digitized high-quality dual-track recordings.

Hopefully, studies like the one presented here will trigger more research into the most interesting field of prosody in simultaneous interpreting. Further improvements and modifications in the analysing method are necessary and welcome in order to achieve a commonly acknowledged approach to analysing prosody which has not been reached so far.

References

Ahrens B. (2004) *Prosodie beim Simultandolmetschen*, Frankfurt, Peter Lang.

AIIC (s.t.) *Allgemeine Vertragsbedingungen für Konferenzdolmetscher*, Genf, Association Internationale des Interprètes de Conférence.

Alexieva B. (1988) "Analysis of the simultaneous interpreter's output", in P. Nekeman (ed.), *Translation, our future. Proceedings, XIth world congress of FIT*, Maastricht, Euroterm, pp. 484-488.

Alexieva B. (1990) "Creativity in simultaneous interpretation", *Babel* 36/1, pp. 1-6.

Barik H.C. (1973) "Simultaneous interpretation: Temporal and quantitative data", *Language and Speech* 16/3, pp. 237-270.

Barik H.C. (1975) "Simultaneous interpretation: Qualitative and linguistic Data", *Language and Speech* 18/3, pp. 272-297.

Cartellieri C. (1983) "The inescapable dilemma: Quality and/or quantity in interpreting", *Babel* 29/4, pp. 209-213.

Collados Aís Á. (1998) *La evaluación de la calidad en interpretación simultánea: La importancia de la comunicación no verbal*, Granada, Editorial Comares.

Cruttenden A. (1997²) *Intonation*, Cambridge, C.U.P., 1st ed. 1986.

Crystal D. (1969) *Prosodic systems and intonation in English*, Cambridge, C.U.P.

Cutler A. (1983) "Speakers' conception of the function of prosody", in A. Cutler and D.R. Ladd (eds.), *Prosody: Models and measurements*, Berlin, Springer, pp. 79-91.

Darò V. (1990) "Voice frequency in languages and simultaneous interpretation", *The Interpreters' Newsletter* 3, pp. 88-92.

Déjean Le Féal K. (1990) "Some thoughts on the evaluation of simultaneous interpretation", in D. Bowen and M. Bowen (eds.), *Interpreting – Yesterday, today and tomorrow*, Binghamton, SUNY, pp. 154-160.

Du Bois J.W., Schuetze-Coburn S., Cumming S. and Paolino D. (1993) "Outline of discourse transcription", in J.A. Edwards and M.D. Lampert (eds.), *Talking data: Transcription and coding in discourse research*, Hillsdale, Lawrence Erlbaum Associates, pp. 45-87.

Fiukowski H. (1986) "Zur Sprecherziehung für Konferenzdolmetscher", *Fremdsprachen* 30/3, pp. 186-191.

Fiukowski H. und Ptok G. (1996) "Sprechstile ausgewählter Berufsgruppen", in H. Kalverkämper und K.-D. Baumann (Hrsg.), *Fachliche Textsorten – Komponenten – Relationen – Strategien*, Tübingen, Gunter Narr, pp. 663-682.

Gerver D. (1969) "The effects of source language presentation rate on the performance of simultaneous conference interpreters", in E. Foulke (ed.), *Proceedings of the Second Louisville Conference on Rate and/or Frequency-Controlled Speech*, Louisville, Center for Rate-Controlled Recordings, University of Louisville, pp. 162-184.

Gile D. (1991) "Methodological aspects of interpretation (and translation) research", *Target* 3/2, pp. 153-174.

Gile D. (2003) "Quality assessment in conference interpreting: Methodological issues", in Á. Collados Aís, M.ª M. Fernández Sánchez y D. Gile (ed.), *La evaluación de la calidad en interpretación: Investigación. Actas del I congreso internacional sobre evaluación de la calidad en interpretación de conferencias. Almuñécar, 2001*, Granada, Editorial Comares, pp. 109-123.

Goldman-Eisler F. (1958) "Speech analysis and mental processes", *Language and Speech* 1, pp. 59-75.

Goldman-Eisler F. (1961) "The significance of changes in the rate of articulation", *Language and Speech* 4, pp. 171-174.

Goldman-Eisler F. (1967) "Sequential patterns and cognitive processes in speech", *Language and Speech* 10, pp. 122-132.

Goldman-Eisler F. (1968) *Psycholinguistics: Experiments in spontaneous speech*, London and New York, Academic Press.

Günther C. (1999) *Prosodie und Sprachproduktion*, Tübingen, Max Niemeyer.

Halliday M.A.K. (1966) "Intonation systems in English", in A. McIntosh and M.A.K. Halliday, *Patterns of language: Papers in general, descriptive and applied linguistics*, London, Longmans, pp. 111-133.

Heuft B. (1999) "Eine prominenzbasierte Methode zur Prosodieanalyse und -synthese", in W. Hess and W.Lenders (eds.), *Computer Studies in Language and Speech*, Vol. 2, Frankfurt am Main, Peter Lang.

Huber D. (1988) *Aspects of the communicative function of voice in text intonation*, Ph.D. thesis, University of Göteborg.

Jin F. (1990) *Intonation in Gesprächen. Ein Beitrag zur Methode der kontrastiven Intonationsanalyse am Beispiel des Deutschen und Französischen*, Tübingen, Max Niemeyer.

Kade O. (1963) "Der Dolmetschvorgang und die Notation. Bedeutung und Aufgaben der Notiertechnik und des Notiersystems beim konsekutiven Dolmetschen", *Fremdsprachen* 7/1, pp. 12-20.

Kalina S. (1994) "Analyzing interpreters' performance: Methods and problems", in C. Dollerup and A. Lindegaard (eds.), *Teaching translation and interpreting 2. Insights, aims, visions. Selected papers from the second "Language International" Conference, Elsinore, Denmark 4-6 June 1993*, Amsterdam-Philadelphia, John Benjamins, pp. 225-232.

Kalina S. (1998) *Strategische Prozesse beim Dolmetschen: Theoretische Grundlagen, empirische Fallstudien, didaktische Konsequenzen*, Tübingen, Gunter Narr.

Kirchhoff H. (1976) "Das Simultandolmetschen: Interdependenzen der Variablen im Dolmetschprozeß, Dolmetschmodelle und Dolmetschstrategien", in H.W. Drescher und S. Scheffzek (Hrsg.), *Theorie und Praxis des Übersetzens und Dolmetschens*, Frankfurt, Peter Lang, pp. 59-71.

Kohler K.J. (1995[2]) *Einführung in die Phonetik des Deutschen*, Berlin, Erich Schmidt, 1st ed. 1977.

Kreuzpaintner S. (2001) *Der Einsatz von prosodischen Elementen als Strategie beim Simultandolmetschen Eine Fallstudie im Sprachenpaar Spanisch-Deutsch*, Diploma thesis, University of Heidelberg.

Lee T.-H. (1999) "Speech proportion and accuracy in simultaneous interpretation from English into Korean", *Meta* 44/2, pp. 260-267.

Liebig G.-B. (1994) *Der Einsatz von Intonationsstrategien zur Optimierung des Dolmetschergebnisses beim Simultandolmetschen aus dem Englischen ins Deutsche*, Diploma thesis, University of Heidelberg.

Pelz M. (1999) *Betonung in gelesenen Konferenzansprachen und ihre Bedeutung für das Simultandolmetschen: eine akustische Analyse (Englisch-Deutsch)*, Diploma thesis, University of Vienna.

Schönherr B. (1997) *Syntax – Prosodie – nonverbale Kommunikation: Empirische Untersuchungen zur Interaktion sprachlicher und parasprachlicher Ausdrucksmittel im Gespräch*, Tübingen, Max Niemeyer.

Selting M. (1995) *Prosodie im Gespräch. Aspekte einer interaktionellen Phonologie der Konversation*, Tübingen, Max Niemeyer.

Setton R. (1994) "Experiments in the application of discourse studies to interpreter training", in C. Dollerup and A. Lindegaard (eds.), *Teaching translation and interpreting 2. Insights, aims, visions. Selected papers from the second "Language International" Conference, Elsinore, Denmark 4-6 June 1993*, Amsterdam-Philadelphia, John Benjamins, pp. 183-198.

Shlesinger M. (1994) "Intonation in the production and perception of simultaneous interpretation", in S. Lambert and B. Moser-Mercer (eds.), *Bridging the gap: Empirical research in simultaneous interpretation*, Amsterdam-Philadelphia, John Benjamins, pp. 225-236.

Vaissière J. (1983) "Language-independent prosodic features", in A. Cutler and. D.R. Ladd (eds.), *Prosody: Models and measurements*, Berlin, Springer, pp. 53-66.

Willett R. (1974) "Die Ausbildung zum Konferenzdolmetscher", in V. Kapp (Hrsg.), *Übersetzer und Dolmetscher: Theoretische Grundlagen, Ausbildung, Berufspraxis*, Heidelberg, Quelle and Meyer, pp. 87-109.

Williams S. (1995) "Observations on anomalous stress in interpreting", *The Translator* 1/1, pp. 47-64.

LINGUISTIQUE DE L'ORALITÉ :
DESCRIPTION DE LA PROSODIE ET ANALYSE INSTRUMENTALE

Philippe Martin
Université Paris 7 Denis Diderot

1.   L'écrit et l'oral

Depuis l'invention de l'écriture, le prestige de l'écrit est sans cesse réaffirmé, car, comme on le sait, « les paroles s'envolent, les écrits restent ». Rares sont les institutions qui confèrent un statut à l'oral. Pourtant aujourd'hui, l'oral peut se conserver dans le temps aussi bien que l'écrit, et peut-être est-il temps de réviser nos habitudes dans ce domaine.

En linguistique, l'idéologie de l'écrit se cache, d'une manière quelque peu insidieuse, dans le concept du « bien formé » utilisé en syntaxe générative. Est bien formée une séquence reconnue (ou produite) comme telle par tout locuteur natif de la langue considérée. Or, l'opinion du locuteur sur cette caractéristique de bonne formation, repose sur l'idéologie du convenable, de l'acceptable et renvoie donc à la connaissance scolaire du concept de « correction ». Est bien formée une séquence que l'on pourrait écrire à l'école (primaire) sans se faire réprimander par le maître, c'est-à-dire par la société. En définitive, le concept de bonne formation repose sur celle de l'écrit.

Ainsi la séquence suivante, extraite d'un corpus de français parlé (Coral Rom Project 2005), semble difficile à décoder lorsqu'elle apparaît sous sa seule forme écrite :

> *bon alors le titre déjà denrée périssable alors dans le dictionnaire ça veut dire euh c'est un fruit à consommer tout de suite ben c'est un truc à consommer tout de suite une marchandise voilà donc déjà il y a consommer qui euh consommer l'amour enfin je sais pas ça me fait ça m'a fait bizarre et euh et périssable donc ben ça voilà je vais en parler quoi donc euh moi je pense que l'amour c'est une c'est une éternelle aventure c'est-à-dire qu' on sait jamais euh sait jamais ce qui va se passer c'est comme on peut jamais dire jamais ben là voilà il y a il y a une petite citation là qui est sympa d' un gars là euh il s' appelle Jean D'ormess Jean D' Ormesson*

Notons qu'elle le serait plus encore, si les espaces entre mots graphiques étaient supprimés :

> *bonalorsletitredéjàdenréepérissablealorsdansledictionnaireçaveutdireeu hc'estunfruitàconsommertoutdesuitebenc'estuntrucàconsommertoutdesuit eunemarchandisevoilàdoncdéjàilyaconsommerquieuhconsommerl'amour*

*enfinjesaispasçamefaitçam'afaitbizarreeteuhetpérissabledoncbençavoilàj
evaisenparlerquodonceuhmoijepensequel'amourc'estunec'estuneéternelle
aventurec'estàdirequ'onsaitjamaiseuhsaitjamaiscequivasepasserc'estcom
meonpeutjamaisdirejamaisbenlàvoilàilyailyaunepetitecitationlàquiestsym
pad'ungarslàeuhils'appelleJeanD'ormessJeanD'Ormesson*

Pourtant, l'écoute de l'enregistrement de ce texte le rend tout à fait compréhensible et ce, par la présence de la prosodie, rythmant et modulant les syllabes successives, et indiquant à l'auditeur les grandes unités à décoder (groupes accentuels) et la manière de les hiérarchiser.

En fait il est facile de se rendre compte que la production spontanée de parole, dans le sens où il ne s'agit plus de texte oralisé, ne répond presque jamais aux critères de la séquence « bien formée », comme on le constate dans l'exemple précédent. Comment se fait- il alors, que de l'oral spontané fidèlement transcrit, peu ou pas compréhensible lorsqu'il est lu, devient tout à fait interprétable une fois reproduit dans sa version orale ? C'est que la prosodie fonctionne comme les espaces graphiques délimitant les mots, tout en en indiquant la hiérarchie.

Dans un petit article succinct exposant les grands principes de cette interaction texte/prosodie qui rend l'incompréhensible facilement interprétable, Deulofeu, Martin et Boulakia (2001) montrent que le discours spontané émis, se laisse aisément décrire par des séquences de macro segments, constitués chacun d'unités syntaxiques « bien formées », et en relation de parataxe (combinaison) ou de rection (dépendance) entre elles.

Parmi tous les îlots macro syntaxiques d'une séquence, l'un d'eux possède une caractéristique particulière : sa modalité peut être modifiée sans affecter les autres segments macro syntaxiques. Ainsi, dans *« ma mère son salon c'est de la moquette le sol »,* on peut changer la modalité déclarative du macro segment *« c'est de la moquette »* en interrogation *« est-ce que c'est de la moquette ? »* ou *« est-ce de la moquette ? »* sans provoquer de modifications dans les deux macro segments *« ma mère », « son salon »* ou *« le sol »*. Les trois macro segments impliqués dans cet exemple sont en relation de parataxe : ils sont simplement combinés, juxtaposés sur l'axe temporel.

La même observation peut être faite pour ce qui est de la structure prosodique. Le changement de modalité déclarative en modalité interrogative, par un contour mélodique final montant au lieu de descendant qui n'entraînerait aucune modification prosodique sur les deux premiers segments, montre l'indépendance de ceux-ci par rapport à la structure prosodique du macro segments *« c'est de la moquette »*. Le jeu d'indication de relation de rection entre macro segment indiqué par la prosodie impliquerait par contre, un changement de sens des contours prosodiques terminant ces segments.

2.   La traduction simultanée

Par prosodie de la phrase, on désigne l'ensemble des mécanismes oraux non segmentaux qui assurent l'indication de la cohésion et de la division entre unités, pour en indiquer une hiérarchie appelée structure prosodique. Dans le cas de la traduction simultanée, il s'agit de passer du décodage prosodique d'un système linguistique à un autre, tout en respectant certaines contraintes, et en particulier la contrainte respiratoire. Celle-ci impose au locuteur de positionner ses moments d'inspiration, évidemment essentiels, à des endroits spécifiques de la séquence d'unités émises, de manière d'une part à optimiser la durée des cycles d'expiration pour lesquels la génération de parole est seule possible, et à raccourcir le plus possible les cycles d'inspiration qui imposent le silence. Les segments temporels d'inspiration du cycle respiratoire ne pourront se faire, de manière privilégiée, qu'aux frontières de plus haut niveau de la structure syntaxique.

A côté de la contrainte physiologique imposée par le cycle de respiration, il existe un ensemble de caractéristiques propres au système de chaque langue, que le locuteur utilise pour indiquer la structure prosodique des énoncés. Ces caractéristiques sont constituées de marqueurs (indicateurs) qui fonctionnent de manière semblable pour des langues comme le français ou l'italien, mais en utilisant des mécanismes spécifiques. Ainsi le français utilise un mécanisme de contraste de pente mélodique à droite pour indiquer l'appartenance d'une unité prosodique à une unité plus grande, alors que l'italien fait usage d'un cadre indiqué par les contours mélodiques situés sur les syllabes accentuées (Martin 1987).

Les propriétés accentuelles des deux langues sont également différentes. Alors que le français réalise une syllabe accentuée (hors accent d'insistance ou accent emphatique) sur la dernière syllabe des groupes et mots prosodiques, l'accent lexical en italien résulte d'une dérivation impliquant le trait de longueur des syllabes du lexème composant le mot et les propriétés d'accentuabilité ou d'inaccentuabilité des suffixes éventuels qui suivent le lexème. La syllabe accentuée résultant de ce mécanisme ne se trouvant que rarement en position finale, les contours prosodiques indiquant la structure prosodique s'en trouveront affectés.

2.1.   La prosodie dans l'interprétation simultanée

Les questions qui se posent d'emblée dans l'examen des interactions prosodiques dans le cas de l'interprétation simultanée :
1. Existe-t-il des endroits temporels privilégiés pour l'élaboration de la traduction par l'interprète ?

2. Quelle est la taille des segments interprétés ? Ces segments sont-ils délimités par des catégories syntaxiques ou prosodiques particulières ?
3. La réalisation des structures prosodiques cibles suivent-elles les mécanismes prosodiques de la langue de départ ou de la langue cible, ou encore d'un autre système ?

3.   Le paradoxe du « bien formé » en interprétation simultanée

Puisque la production de parole par un locuteur ne peut se faire que pendant les temps d'expiration du cycle de respiration, la production de parole implique une stratégie complexe de planification des séquences syntaxiques à produire, de manière à faire correspondre le temps d'inspiration du cycle, évidemment nécessaire à la survie du sujet parlant, à une frontière syntaxiquement acceptable par l'auditeur dans le processus de décodage, c'est-à-dire une frontière majeure. Ainsi dans une phrase comme *Max adore Marie*, on évitera de placer un temps d'inspiration entre *Max adore* et *Marie*, cette insertion introduisant une disfluence dans la séquence produite.

L'analyse prosodique en interprétation de liaison (mais aussi dans une certaine mesure en traduction simultanée) présente un aspect semblable, impliquant en plus, pour l'interprète, le choix d'un endroit stratégique lui permettant le traitement de la séquence suivante à traduire. Autrement dit, le problème de la segmentation des unités successives traitées pendant la traduction simultanée, se pose doublement : 1) par le choix du temps d'insertion des séquences traduites ; 2) par le choix des unités syntaxiques à traiter, à la fois du point de vue de la langue de départ et de celui de la langue d'arrivée.

L'analyse des tours de parole (Martin et Yoo 2005) présente a priori quelques similarités avec celle de l'interprétation de liaison. En effet, la prise de parole dans une conversation entre deux locuteurs L1 et L2, peut se faire sans conflit (cas correspondant à la traduction séquentielle), soit avec conflit (cas présentant des analogies avec la traduction simultanée).

Dans la première situation, le locuteur L1 donne à L2 des indications qui peuvent être de nature prosodique (ou gestuelles) quant à son intention de lui céder son tour de parole (ce mécanisme peut, bien sûr, être initialisé aussi par un animateur, dans un show télévisé par exemple). De même, en traduction séquentielle, le locuteur L1 parlant dans la langue de départ, donne des indications gestuelles et/ou prosodiques au traducteur L2, pour lui laisser insérer la parole traduite.

La deuxième situation appliquée à la traduction n'est évidemment pas conflictuelle (encore que…) et le rapprochement est moins convaincant : le plus souvent L1 n'a qu'une conscience limitée de la présence d'un ou de plusieurs traducteurs effectuant une simultanée et ne cherche – normalement pas – à lui

voler son tour de parole. Il s'agit donc pour le traducteur L2, de pouvoir insérer ses séquences de production traduite aux moments opportuns, qui peuvent être signalés par des gestes (involontaires dans leur fonction, donc constituant des indices au sens de la communication linguistique) de L1, mais qui, le plus souvent, étant donné la configuration habituelle des sessions de simultanée mettant une distance géographique plus ou moins grande entre L1 et L2, sont signalés par les indices prosodiques. C'est précisément ces indices, mouvements mélodiques, changements de rythme syllabique et d'intensité, etc., qui font l'objet de l'analyse prosodique des sessions de traduction simultanée.

Cette stratégie d'insertion consiste donc essentiellement en une estimation par L2 d'une possibilité de mise à profit de l'apparition d'une pause, d'un ralentissement de rythme, etc. par le locuteur L1, pour générer la séquence préalablement traduite. A ce processus s'ajoute un choix par L2, de la grandeur et du type de séquence syntaxique à traiter. Le groupe syntaxique traduit doit être suffisamment grand, comporter assez d'unités pour ne pas provoquer trop d'interruptions dans la production du traducteur, et suffisamment petite pour permettre son analyse syntaxique (c'est-à-dire sa compréhension) sans utilisation excessive de la mémoire. On sait à ce propos que le maximum de niveaux hiérarchiques que l'on peut traiter est de 7.

A côté de l'étude des stratégies d'insertion, le processus de traduction lui-même est évidemment dominant, de par sa complexité impliquant, entre autres, les décodages syntaxiques et lexicaux de la langue de départ L1, l'accès lexical et la génération syntaxique de L2, suivie de la production orale du résultat de toutes ces opérations. La durée globale d'une opération de base, sélection d'un segment produit par L1 – compréhension, traduction et production par L2 – sera d'autant plus efficace que le traducteur aura, au cours de son apprentissage, stocké dans sa mémoire, un grand nombre de séquences toutes prêtes, comme dans le cas d'un traducteur de l'écrit utilisant les ressources d'un système de traduction assistée contenant un très vaste catalogue de séquences déjà traduites.

4.   La simultanée en production lue et spontanée

Pourquoi est-il plus facile pour un traducteur d'opérer sur un locuteur L1 qui énonce de manière spontanée plutôt que sur celui qui lit un texte ? On opterait a priori pour une hypothèse contraire : des phrases par définition bien formées, résultant d'une lecture oralisée, présenteraient toutes les caractéristiques propres à assurer un décodage lexical et syntaxique aisé, proportionnellement à la « qualité » de l'écriture et à la construction facile de la hiérarchie syntaxique.

Les praticiens de la traduction simultanée sont cependant presque tous unanimes pour affirmer le contraire, et les raisons pour trouver la simultanée à partir de l'oral spontané plus aisée, sont multiples.

Parmi les premières raisons qui viennent à l'esprit, citons :

1. les phrases écrites, quoique paradoxalement plus courtes, comparées aux phrases spontanées (Martin 2000), sont aussi plus complexes et peuvent impliquer de nombreuses parties subordonnées, imbriquées dans une structure syntaxique difficile à déchiffrer. Les phrases spontanées, par contre, présentent un grand nombre de segments courts, en relation de parataxe, c'est-à-dire simplement juxtaposés du point de vue de la structure (macro) syntaxique globale de la phrase ;

2. la lecture des phrases n'est pas toujours accompagnée d'une prosodie qui en rende la structure syntaxique plus facile à appréhender (autrement dit, les locuteurs de L1 ne sont pas nécessairement de bons lecteurs à haute voix). Du reste, l'appréhension de la hiérarchie de la phrase en lecture est un processus visuel très complexe impliquant, outre le repérage des signes de ponctuation, celui des verbes, simultanément à l'énonciation. La planification syntaxique étant imposée par le texte, le lecteur n'en a qu'une compréhension passive, au contraire de la production spontanée (qui peut cependant faire intervenir l'oralisation de macro segments tout « préparés » et ne demandant pas de génération syntaxique) ;

3. le rythme d'élocution est beaucoup plus rapide en lecture oralisée qu'en production spontanée, puisque le lecteur n'a pas besoin du temps de construction de la structure syntaxique et de l'accès lexical comme dans le cas du spontané. Il y a donc moins de temps pour le traducteur entre chaque segment traduit.

5.   Techniques d'observation et d'analyse

L'observation des données prosodiques et leur transcription constituent un problème ardu, mais l'apparition de logiciels d'analyse de la parole tels que Praat ou WinPitch facilite grandement cette entreprise. Le logiciel WinPitch (www.winpitch.com) a l'avantage d'incorporer des fonctions spécifiques pour l'analyse prosodique d'enregistrements de paroles en stéréo, particulièrement adaptées à notre étude, car elles assignent un canal d'enregistrement à la voix de la langue de départ et un autre à celle de la langue d'arrivée. Outre les nombreuses fonctions de navigation dans le fichier signal, permettant une analyse aisée de données de longue durée, ce programme affiche directement les courbes prosodiques (mélodie, durée, intensité) des deux canaux dans une même fenêtre, mais avec des couleurs différentes, ce qui permet la comparaison aisée et agréable des deux types de données simultanément.

Une autre caractéristique unique de WinPitch porte sur les nombreuses fonctions de transcription et d'alignement assisté, permettant une mise en forme rapide des données. Un lexique, un système d'interrogation des données textuelles sont également inclus. L'utilisation du programme pour l'analyse de données multimédia est possible, rendant possible l'étude des gestes accompagnant une session d'interprétation.

Si les phonéticiens sont parfois sourds à certains sons, au sens où des nuances particulières de consonnes ou de voyelles ou encore les modulations de la voix peuvent échapper à leur transcription, en revanche ils ne sont pas aveugles. Aussi la phonétique expérimentale s'est-elle développée essentiellement autour de l'analyse et de la représentation visuelle du signal de parole, analyse liée aux développements parallèles de l'électronique et de l'informatique, de manière à rendre visibles les détails parfois à peine discernables par l'oreille humaine, des sons. Ainsi a-t-on vu apparaître successivement, des spectrographes et des visualiseurs de mélodie, rendant la parole visible.

Or on sait depuis longtemps, et des domaines tels que la mécanique quantique ont particulièrement souligné cet aspect, que la réalité physique d'un phénomène n'existe qu'au travers de son observation, plus exactement à cause de l'observation qui en est faite. L'observation et partant, la description du signal de parole n'échappent pas à ce principe. Aussi, observer, décrire, modéliser des faits linguistiques en se basant sur des données obtenues autrement que par l'intercession du sujet parlant, mènent évidemment à une autre « réalité » des objets linguistiques décrits, ce qui se révèle particulièrement crucial pour les sons du langage (cf. Falbo 2005).

S'il est un objet sonore du langage singulièrement élusif, c'est bien la prosodie de la phrase, prosodie résultant d'une alchimie complexe entre fréquence laryngée, durée syllabique, intensité sonore, etc. Si des observateurs musiciens possèdent une certaine facilité pour percevoir et donc transcrire, les événements prosodiques d'une phrase, l'utilisation d'appareils de mesure par les phonéticiens ou les phonologues non musiciens s'impose, au risque d'entraîner des conséquences imprévues, du fait qu'implicitement, le point de vue adopté dans la procédure sera celui de l'appareil (ou du logiciel) de mesure.

Envisageons un exemple concret : la simple – apparemment – qualification de la syllabe d'une phrase comme proéminente ou non proéminente pose, dans une langue comme le français, d'énormes problèmes. Le problème vient du fait que l'accent phonologique de mot en français (l'accent lexical), n'est pas nécessairement réalisé, sa présence effective dépendant de facteurs tels que le nombre de syllabes non proéminentes successives, le débit de la phrase, etc. Ainsi dans *Julien adore Marie,* la réalisation d'un accent (d'une proéminence) sur la dernière syllabe de *adore* est facultative, alors qu'elle est ressentie comme

beaucoup plus nécessaire sur la syllabe finale de *avait adoré* dans *Julien avait adoré Marie*.

Tout descripteur, même pourvu d'une oreille aiguisée, ne pourra s'abstraire dans son activité, de notation des proéminences syllabiques, de sa connaissance des règles accentuelles du français ou de toute autre langue qu'il connaît. Il y aura donc, non seulement un filtrage éventuel au travers de la grille phonologique de la langue du descripteur, mais également interférence entre l'observation des données et la connaissance de règles d'attribution des phénomènes à observer.

L'affaire se complique davantage si des appareils d'observation interviennent dans le processus. Un instrument ou logiciel tel qu'un visualiseur de mélodie, donne des tracés – une valeur de la fréquence fondamentale correspondant à une estimation de la fréquence laryngée – dont les détails d'évolution ne sont pas nécessairement perçus par le sujet parlant, destinataire « normal » du signal de parole. Pire, de tels appareils donnent généralement une courbe physique « améliorée » par adoucissement des sauts, parfois détectés, de cycle à cycle laryngé, de manière à apparaître plus « agréables » à l'œil de l'observateur, au point que dans certains cas les courbes mélodiques obtenues sont bien éloignées de la réalité physique.

D'autres types d'analyse acoustique présentent chacun leurs problèmes. Ainsi le spectrographe, analogique ou numérique, est basé sur une méthode d'analyse spectrale, l'analyse de Fourier, qui par nature, est soumis au principe d'incertitude temps/fréquence : toute augmentation de la précision de mesure du temps entraîne nécessairement une perte de précision en fréquence et inversement.

Pour revenir à l'observation de la mélodie de la phrase, essentielle pour la description prosodique, l'utilisation de spectrogrammes nécessite un réglage dit à « bande étroite », pour distinguer les harmoniques des parties voisées du signal et suivre ainsi les évolutions de la fréquence laryngée, réglage qui, nécessairement, rend l'estimation des faits temporels liés à cette fréquence imprécise (ce phénomène est tout simplement dû à l'épaisseur temporelle du signal qui est nécessaire pour effectuer le calcul de la transformée de Fourier, avec une précision sur la fréquence suffisante).

Il reste que, ces précautions épistémologiques prises, tous ces instruments d'analyse acoustique sont universellement utilisés, car moins complexes à mettre en oeuvre que des instruments de mesure physiologique. En fait, visualiseurs de mélodie et spectrographes sont complémentaires, et sont du reste affichés simultanément dans des logiciels tels que Praat ou WinPitch. L'avantage d'une telle configuration vient de ce que la mesure de la fréquence fondamentale du signal de parole peut s'avérer erronée dans des parties spécifiques du signal, lors de la présence d'un bruit de fond important, d'un

filtrage excessif des basses fréquences, ou encore lors de la présence d'une autre source sonore (chevauchement de voix).

Le logiciel WinPitch est particulièrement bien adapté à l'analyse des données prosodiques de la traduction simultanée. En installant les enregistrements des voix source et traduite sur les deux canaux d'un signal stéréo (ce qui peut s'effectuer facilement à partir de deux enregistrements mono séparés), on peut faire apparaître les courbes mélodiques simultanément sur la fenêtre d'analyse, chacune dans une couleur différente, de même que les courbes oscillographiques et d'intensité. L'étude des interactions source/traduction est alors particulièrement aisée.
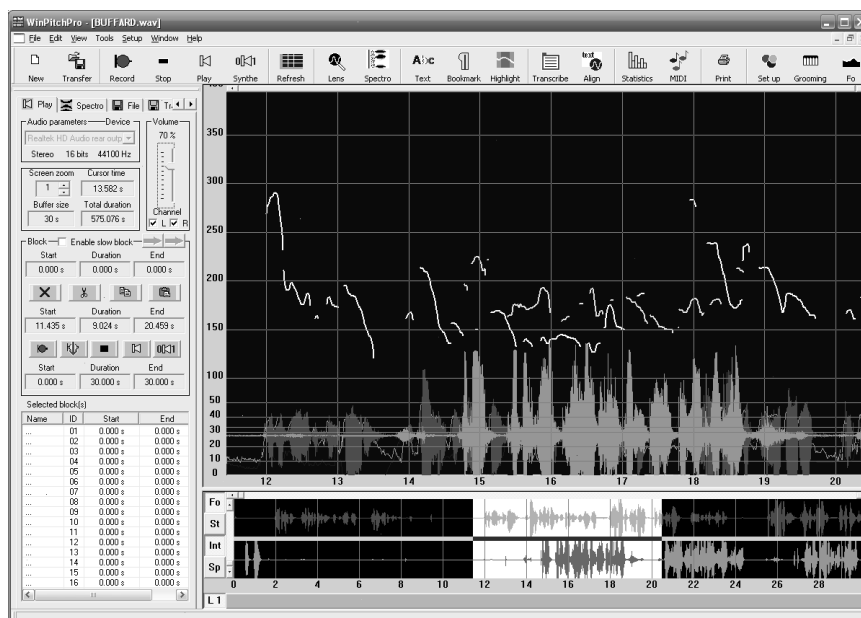
Fig. 1 Exemple de fenêtre d'analyse instrumentale d'un segment de traduction simultanée. Le logiciel WinPitchPro permet la sélection et l'analyse acoustique simultanée d'un enregistrement stéréo, dont la piste gauche est allouée à la production de L1 (langue originale), et la piste droite à la langue interprétée L2. La fenêtre d'analyse affiche les courbes d'intensité et de mélodie dans des couleurs différentes.

WinPitch permet la transcription directe (création du texte par écoute de segments de parole successifs) et l'alignement à la volée (alignement de segments successifs d'un texte préexistant). L'affichage instantané des spectro-

grammes correspondant au segment transcrit permet en outre, de démêler des voix qui se chevauchent, par l'observation sur le spectrogramme des harmoniques entremêlées, mais évoluant dans des directions différentes. De nombreuses fonctions rendent en outre aisé l'ajustement fin de la segmentation. Enfin, ce logiciel est multimodal, et fonctionne également avec des fichiers multimédia standards.

La transcription et l'alignement à la volée sont grandement facilités par la possibilité de ralentir la parole par un facteur pouvant aller jusqu'à 7 fois, sans déformation notable (le ralentissement est effectué par un des trois algorithmes disponibles, au choix de l'utilisateur : Psola, autocorrélation ou vocodeur de phase).

6.   Conclusion

L'analyse et la description des processus complexes impliqués dans la traduction simultanée font appel à tous les domaines de la linguistique : prosodie, phonologie, morphologie, syntaxe, sémantique, aussi bien pour la langue originale L1 que pour la langue interprétée L2. Tous les mécanismes cognitifs d'accès au lexique, de planification syntaxique, de décodage prosodique, etc. interviennent. Aussi, est-ce un champ d'étude passionnant qui s'ouvre dans ce domaine, qui mettra à l'épreuve les théories et modèles liés à chacun de ces domaines, pour nous en donner une meilleure compréhension et permettre une formation encore plus efficace des différents intervenants de la traduction simultanée.

Références

Coral Rom Project (2005) http://lablita.dit.unifi.it/coralrom/

Deulofeu J., Martin Ph. et Boulakia G. (2001) "Quand la prosodie bien-forme des énoncés mal formés", *Colloque Journées Prosodie 2001*, Grenoble, 10-11 octobre 2001.

Falbo C. (2005) "La transcription: une tâche paradoxale", ce volume.

Martin P. (1987) "Prosodic and Rhythmic Structures in French", *Linguistics*, Vol. 25-5, pp. 925-949.

Martin P. (2000) "L'intonation en parole spontanée", *Revue Française de Linguistique Appliquée*, Vol. IV-2, pp. 57-76.

Martin P. et Yoo H-Y. (2005) "Turn taking by contours", *Atelier Modélisation du Dialogue et Grammaire*, Université Paris 7, 9-10 mars 2005.

LA TRANSCRIPTION : UNE TÂCHE PARADOXALE

Caterina Falbo
SSLMIT Université de Trieste

1.   Introduction

La collecte d'un corpus parallèle d'interprétation implique la transcription de textes produits par l'interprète (TI) et des textes originaux (TO) correspondants. Étape obligée et apparemment facile à surmonter, la transcription se révèle bientôt parsemée d'embûches pouvant absorber une quantité énorme de temps et d'énergie. Ces difficultés apparaissent quand on se fixe l'objectif tout à fait paradoxal de recourir à la transcription : rendre visible et lisible ce qui par définition n'est qu'audible, fixer l'évanescence de l'oral, rendre l'éphémère durable.

Dans ces pages nous présentons d'abord (2.) les méthodes de transcription adoptées par la plupart des auteurs en cinquante ans de recherche sur l'interprétation. Nous essayons ensuite (3.) de mettre en évidence les problèmes posés par la transcription tout en soulignant les liens qui existent entre méthode de transcription choisie et résultats possibles, découlant de l'analyse. Nous concluons (4.) par des exemples qui illustrent concrètement les choix que le(s) transcripteur(s) est (sont) constamment appelé(s) à opérer.

2.   La transcription en littérature

Depuis les débuts de la recherche en interprétation la transcription a toujours constitué une condition nécessaire à l'analyse du TI et éventuellement à sa comparaison avec le TO correspondant.

Il est probable que les premières transcriptions ont été effectuées par les psychologues cognitivistes qui au cours des années 60-70 ont commencé à réfléchir sur l'interprétation simultanée à en étudier la qualité, certains facteurs temporels comme le décalage et le degré de simultanéité entre TO et TI, et à présenter, pour ce faire, dans leurs articles des exemples (écrits et donc transcrits) de TI (cf. Barik 1971, 1972, 1973, 1975; Gerver 1971, 1974a, 1974b, 1976; Goldman-Eisler 1967, 1972; Goldman-Eisler et Cohen 1974; Oléron et Nanpon 1964). Cependant ces chercheurs ne s'arrêtent pas, dans leurs écrits, sur le système de transcription adopté. Nous n'avons relevé qu'un seul passage dans lequel Oléron et Nanpon (1964: 75) font indirectement allusion à la transcription, lorsqu'ils affirment avoir éliminé les enregistrements d'interprétations réelles produites par des interprètes professionnels de l'UNESCO à cause de la

présence, dans les TO, "de multiples bavures (hésitations, répétitions, incor-rections) qui écartent les textes de l'organisation 'normale' de la langue et retentissent sur l'interprétation". Marianne Lederer (1981), quant à elle, présente un corpus formé par la transcription de 63 minutes d'interprétation accompagnée du TO correspondant, mais elle ne fournit aucun renseignement sur la méthode adoptée. Toutefois, en lisant les transcriptions annexées à *La traduction simultanée*, il est facile de remarquer que grâce à l'utilisation de la ponctuation l'oral est complètement assimilé à l'écrit[1].

Le même système de transposition de l'oral à l'écrit nous le retrouvons plus ou moins chez tous les auteurs qui se sont penchés sur l'investigation du produit de l'interprétation, c'est-à-dire sur le TI.

À partir des années 90 environ, nous assistons au début du questionnement, quoiqu'encore implicite, en matière de transcription. Un bon exemple nous vient de Meyer (1998: 78) qui essaie de démontrer le pouvoir 'herméneutique' de la transcription: "transcribing – following the HIAT-conventions – is more than simply writing down the linguistic surface structure". En effet le système de transcription HIAT (Heuristic Interpretative Auditory Transcription)[2] donne au chercheur la possibilité de représenter par écrit les aspects verbaux, paralinguistiques, non-verbaux et gestuels (actional) et de synchroniser la parole de l'interprète avec celle de l'orateur.

Il est aisé de reconnaître, dans le recours à ce système de transcription, le désir de représenter la totalité de l'interprétation en tant que forme particulière de l'oralité, avec ses aspects linguistiques et pragmatiques.

Ce que Meyer (1998) semble vouloir dénoncer, sans toutefois le dire explicitement, c'est que jusque là, la trascription, tout en voulant représenter l'interprétation effectuée (TI), n'en relevait en réalité que l'aspect linguistique, sous forme de simple chaîne de mots produits par l'interprète. La notation de la synchronisation entre le TI et le TO, par contre, affiche la segmentation – et donc le plan – opéré par le locuteur, qu'il soit l'orateur ou l'interprète. Il en va de même des hésitations, des pauses et des répétitions : les marquer revient à fixer des points de repère pour essayer de dégager le processus (mental) qui en est à l'origine. Le but poursuivi par Meyer (1998: 80) est de remonter aux processus cognitifs qui déterminent la performance de l'interprète; pour atteindre cet objectif il se sert de "authentic data represented in transcripts". C'est justement cette transcription conçue comme représentation authentique de l'interprétation qui doit retenir toute notre attention (cf. 2.1).

Setton (1999: 111) offre des transcriptions très soignées et très détaillées. Le texte original et le texte interprété sont enregistrés sur magnétophone à deux

---

1   Eggins et Slade (1997) adoptent le système proposé par Halliday et utilisent la ponctuation en tant que signalisation du rhytme et de l'intonation de l'oral.

2   Ce système trouve son application informatique grâce au logiciel *syncWRITER*.

pistes, synchronisés[3] et présentés de façon interlinéaire, le texte interprété transcrit apparaît au dessous de la transcription du texte original. L'auteur explique les conventions adoptées qui comprennent la notation des syllabes accentuées, de l'intonation montante/descendante, des pauses, de la respiration audible, des allongements, etc., mais, comme la plupart des chercheurs en interprétation, il n'évoque pas la procédure suivie pour transcrire, pour affronter les doutes de perception, il ne dit rien du nombre de transcripteurs.

En revanche, Pöchhacker (1994: 157-158) précise le caractère subjectif des transcriptions qu'il a personnellement effectuées et dans lesquelles on pourra retrouver, inévitablement, la perception que le transcripteur a eue du signal acoustique. L'auteur choisit une transcription orthographique pour d'évidentes raisons de lisibilité, mais aussi à cause du caractère partiel de la transcription phonétique : celle-ci en effet n'est pas non plus en mesure de représenter le rythme d'élocution ou les traits suprasegmentaux. Et même si cela était possible, le "lecteur" serait appelé, lors de la "lecture", à reconstruire la tonalité du texte. L'argumentation de Pöchhacker ne fait que réitérer la nécessité, pour le chercheur, de conjuguer le texte transcrit avec le texte oral correspondant. Dans la présentation des conventions adoptées, l'utilisation de quelques signes de ponctuation (par exemple le point final pour signaler une intonation descendante et donc la fin de phrase) révèle, à notre avis, une contamination, voire une assimilation entre l'écrit et l'oral, confirmée par l'explication donnée par l'auteur : comme plusieurs textes de son corpus existaient sous forme de manuscrits, dont les orateurs s'étaient servis lors de leurs interventions, des signes de ponctuation ont été adoptés, afin d'en respecter le registre plus écrit qu'oral[4].

La transcription retient surtout l'attention des chercheurs qui se penchent sur l'intreprétation de liaison. En général ce sont les systèmes de notation développés dans le cadre de l'analyse de la conversation qui sont adoptés. Le système mis au point par Gail Jefferson (Atkinson et Heritage 1999), ainsi que celui proposé par Traverso (1999) répondent aux exigences liées à la représentation de la conversation, c'est-à-dire à un TI produit par plusieurs locuteurs dont les énoncés se suivent les uns après les autres ou se chevauchent. Les temps des tours de parole et les pauses entre deux tours, la superposition des locuteurs ainsi que les interruptions demandant une notation rigoureuse capable de révéler les dynamiques conversationnelles en cours. A cet effet, la notation de la gestuelle ou de la direction du regard pourrait être essentielle pour le questionnement auquel on va soumettre le corpus collecté et transcrit. En outre,

---

3   L'auteur s'étend sur les difficultés de synchronisation.
4   "Da vielen Reden ein schriftliches Manuskript zugrundelag bzw. die Redner sich eher eines schriftsprachlichen Registers bedienten, wurde so weit wie möglich eine schriftnahe Gliederung (Interpuktion) verwendet" (Pöchhacker 1994: 158).

il est essentiel, comme d'ailleurs pour toute recherche sur corpus, d'utiliser des symboles présents parmi les caractères 'connus' par l'ordinateur, afin de faciliter le traitement informatique des données et éventuellement l'échange de corpus.

Nous concluons cet aperçu rapide par les mots de Cecot (2001: 73) : "The transcription of texts caused difficulties because of the absence of codified and established transcription norms". Face à cette résignation, nous répondons par un doute : peut-être existe-t-il des difficultés intrinsèques à la transcription dont le manque de conventions codifiées et univoques n'est qu'une conséquence naturelle.

3.   Le paradoxe de la transcription

Dans notre culture imprégnée d'écriture, ce paradoxe apparaît de prime abord sans raison d'être, car tout ce qui est dit nous le retrouvons très souvent et facilement écrit noir sur blanc. Nous assistons quotidiennement à ce passage du dit à l'écrit sans nous rendre compte des changements qui se glissent dans la nature même des choses "racontées". Ce que l'on oublie très souvent c'est que seuls les mots, entités physiques identifiables et identifiés, entre autres, grâce à l'écriture – puisqu'à l'oral il n'existe qu'une chaîne parlée – peuvent être fixés définitivement sur le papier ou sur un support électronique. Or, l'oralité, le parlé se compose de plusieurs éléments, dont les mots n'en représentent qu'un. Comment rendre compte, comment représenter par écrit l'intonation, le volume, l'emphase ou les allongements qui rendent les mêmes paroles si différentes les unes des autres, si elles sont prononcées plusieurs fois de suite et/ou par plusieurs locuteurs?

C'est face à ces questions que le projet de rendre le parlé, écrit, de capter et garder l'oral sous forme écrite, apparaît dans toute sa complexité.

3.1.   Les pièges de la transcription[5]

La transcription en interprétation a toujours servi les exigences des chercheurs dans la mesure où elle fournissait un texte 'sur papier' prêt à être analysé. Cette phase préalable à toute étude sur le TI cache néanmoins des difficultés, voire de véritables pièges au niveau de l'écoute ainsi qu'au niveau du traitement du texte transcrit.

La première difficulté est représentée par l'enregistrement lui-même qui, comme le disent Blanche-Benveniste et Jeanjean (1987: 93 et suiv.) poursuit

---

5   Ce paragraphe se fonde essentiellement sur les ouvrages de Blanche-Benveniste et Jeanjean (1987) et Blanche-Benveniste (1997).

exactement le même but que la transcription : arrêter l'évanescence de l'oral. Si aujourd'hui nous ne sommes plus confrontés à toute une série de problèmes techniques qui influent sur la qualité de l'enregistrement (bruits excessifs, mono vs. stéréo, etc.), il n'empêche que l'enregistrement reste une source indirecte, la directe étant l'interaction verbale réelle. Il existe toujours un décalage entre le *hic et nunc* d'un échange verbal et son enregistrement, ne serait-ce que la clarté et la qualité du son. Contrairement à ce qui se produit au cours des conversations spontanées quotidiennes, caractérisées en général par un degré d'articulation relativement bas, on a tendance à croire que les TI produits par des professionnels de l'interprétation sont caractérisés par une élocution très articulée et donc claire, pourvu que les conditions dans lesquelles l'interprète travaille le permettent (possibilité de bien entendre l'orateur, rythme d'élocution acceptable – 100-120 mots/minute –, etc.). Les exceptions toutefois ne font pas défaut. Le trac, l'émotivité qui peut entraîner le manque de contrôle de la voix et/ou un manque de concentration, rendent le TI parfois difficile à écouter, voire inaudible. Dans ces cas, la transcription devient particulièrement laborieuse parce que l'écoute est ardue et confrontée à des sons (mots) mal prononcés voire prononcés à demi[6].

Au-delà des problèmes techniques d'écoute illustrés plus haut, l'oreille peut être considérée comme le deuxième facteur de difficulté. Blanche-Benveniste et Jeanjean (1987: 102) n'ont aucune hésitation à affirmer que "l'oreille est un traître; on écoute ce qu'on s'attend à écouter". Bilger *et al.* (1997: 58) confirment cette célèbre affirmation tout en précisant que "l'oreille n'est pas un traître, elle est surtout asservie à la recherche de signification". Les connaissances préalables du transcripteur sur le sujet traité dans l'échange verbal enregistré ainsi que ses attentes sur ce qui va être dit[7], vont sans aucun doute aider le transcripteur dans son travail d'écoute-reconnaissance, la perception étant "un processus actif qui s'appuie sur la compréhension des

---

6   Bien qu'intelligible, l'interprétation simultanée du discours tenu par le frère de la Princesse Diana lors des funérailles de cette dernière, était inacceptable du point de vue des caractéristiques prosodiques de l'interprète: voix tremblante, soupirs, ton agité qui révèle un degré excessif de nervosité (Stimoli 2001). Un exemple de difficulté de décodification du dit est représenté par l'interprétation qu'a founie une professionnelle de langue maternelle française assurant l'interprétation en italien du modérateur, lors du débat présidentiel entre Jacques Chirac et François Mitterrand, passé sur Rai 3 le 28 avril 1988.

7   Il faut remarquer que le transcripteur, qui a pourtant la possibilité de revenir plusieurs fois sur le même bout d'enregistrement, procède néanmoins à la première écoute de façon linéaire, exactement comme n'importe quel écouteur. Par conséquent, ce qui n'a pas encore été entendu est assujetti aux attentes du transcripteur.

énoncés" (Blanche-Benveniste et Jeanjean 1987: 103). Il sera plus facile, par exemple, pour un transcripteur qui connaît de façon approfondie ce qui se passe sur la scène socio-politique aux Etats-Unis d'entendre, et donc de reconnaître, les noms des personnes à la une du moment[8]. Cet avantage toutefois, peut se transformer en un véritable piège pour le transcripteur en général, et pour le transcripteur-interprète en particulier. Le danger est de transcrire ce que l'on veut entendre et non pas ce que l'on écoute vraiment. On peut remédier à cet effet pervers en multipliant le nombre de transcripteurs : Blanche-Benveniste et Jeanjean (1987: 101, *multiécoute*) en recommandent au moins quatre. Chez le transcripteur-interprète qui a préparé une expérience en laboratoire pour la collecte de TI et sélectionné les TO à soumettre à interprétation, ou qui tout simplement connaît les TO, le risque d'écouter et d'entendre ce qu'il sait déjà est accru. Dans ce cas les connaissances préalables n'accélèrent pas le processus de reconnaissance du dit, mais se superposent au TI écouté et en influencent la compréhension.

Un autre piège de la transcription est représenté par la tentation du transcripteur de corriger ce qu'il entend suivant la norme de l'écrit. Très souvent, les lapsus linguae ainsi que les fautes de cohésion[9] présents à l'oral, font l'objet de corrections de la part du transcripteur. Il est aisé d'expliquer cette procédure presque automatique par les habitudes que, chez tout être humain, des siècles de formation linguistique ont façonnées vis-vis de l'écrit (Ong 1986). De même pourrait-on être tenté d'éliminer, et donc de "nettoyer" le TI de toutes les répétitions ou tentatives d'élocution (*stalling*) d'un mot. Il est entre autres très difficile, au rythme spontané d'élocution, de rendre compte de toutes les répétitions ou tentatives d'élocution. L'oreille est attirée par la version définitive du mot, du syntagme, et incapable de dénombrer les différentes tentatives qui se succèdent, en général, très rapidement.

À la lumière de tous les obstacles évoqués jusqu'ici et des remèdes visant à les surmonter – multi-écoute en tête – nous pensons avoir le droit et le devoir de

---

8   Un exemple est donné par les difficultés rencontrées par des étudiants-
     transcripteurs aux prises avec les TI correspondant au débat passé à la télé entre les
     candidats aux élections présidentielles de 1984 (Reagan-Mondale) et de 1992
     (Bush-Clinton-Perot). Lorsque les candidats faisaient allusion à des faits divers ou à
     des personnages peu connus du grand public étranger, le transcripteur était
     incapable de détecter les sons qui composaient les noms propres en questions. Seule
     une recherche successive à permis de retrouver les noms mentionnés et seulement à
     ce moment-là ils sont devenus intélligibles au(x) transcripteur(s).
9   Par fautes de cohésion nous entendons l'absence d'accord grammatical (genre et
     nombre) entre sujet et verbe, nom et adjectif, conjonction et verbe, etc. Aucune
     allusion n'est faite à l'organisation syntaxique de la langue orale par rapport à
     l'écrit (cf. Blanche-Benveniste 1997).

poser une question fondamentale. Cet interrogatif concerne l'authenticité du texte transcrit, sa correspondance avec le texte oral dont il se veut être l'image, à l'écrit. Notre réponse est qu'il serait faux de croire que la transcription puisse permettre d'aboutir à une copie authentique, absolument fidèle au texte oral. Il serait faux de croire pouvoir avoir entre les mains, écrit noir sur blanc, le texte oral lui-même. La transcription est en effet un problème de choix, de décisions à prendre au fur et à mesure que l'on écoute et que l'on transcrit. Elle est par conséquent assujettie à la subjectivité du sujet transcripteur. Avant tout, le transcripteur est appelé à choisir entre une transcription phonétique et une transcription orthographique. Il semble naturel de penser qu'une transcription phonétique puisse mieux représenter le texte oral et par conséquent lui être plus fidèle qu'une transcription orthographique. Or, il n'en est rien. S'il est vrai que la transcription phonétique fournit une représentation plus proche de la réalité 'sonore' qu'une transcription orthographique, il est tout aussi vrai que la phonétique ne peut rendre compte de "l'ensemble sémiologique d'un acte de communication" (François cité par Blanche-Benveniste et Jeanjean 1987: 120). De plus, il est illusoire de croire pouvoir avoir entre les mains le texte oral que l'on vient de transcrire. Tout interprète – au sens d'écouteur, de lecteur, voire de transcripteur – "déforme" le texte par le simple fait de l'écouter et/ou de le lire. Tout acte de réception d'un texte advient à partir de la subjectivité et de l'univers personnel du sujet. Tout cela est confirmé également par Ochs (1999: 167) qui souligne la prise de décision à laquelle, tôt ou tard, tout transcripteur est confronté:

> […] the problems of selective observation are not eliminated with the use of recording equipment. They are simply *delayed* until the moment at which the researcher sits down to transcribe the material from the audio- or video-tape. At this point, many of the classic problems just emerge.

Ce souci de fidélité authentique au texte "réel" doit forcément se conjuguer avec l'exigence de lisibilité propre à tout objet d'analyse. Et c'est là qu'encore une fois le paradoxe de la transcription ressort dans toute son ampleur:

> Transcrire de la langue parlée tient un peu du paradoxe : garder dans une représentation écrite certaines caractéristiques de l' "oralité" ; faire le "rendu" de la chose orale tout en restant dans des habitudes de lecture établies depuis longtemps pour la chose écrite … On va se trouver tiraillé entre deux exigences : la fidélité à la chose parlée et la lisibilité de son rendu par écrit. (Blanche-Benveniste et Jeanjean 1987: 115)

C'est l'exigence de faire une transcription "utilisable", lisible, qui fait privilégier la transcription orthographique, au moins dans le cadre de la recherche en interprétation[10].

Tout choix opéré par le transcripteur – ne serait-ce que la volonté de tout noter (sons, gestes, regards, intonation, etc.) – comporte un appauvrissement de la réalité communicationnelle.

Il est clair à ce point que la transcription n'est que la trace du résultat d'une élocution; une trace concrète, tangible, dans laquelle observer des phénomènes, réfléchir, étudier l'oral.

Ce n'est qu'à partir des années 50 que la linguistique a commencé à s'intéresser à la langue parlée. Très souvent l'oral a été étudié de façon plus ou moins avouée par rapport à l'écrit, ce dernier étant considéré comme le point de repère, l'étalon à partir duquel dégager les analogies et les différences. C'est à cause de cette méthode comparative, où l'un des termes constituait par définition la norme, que la langue parlée a eu droit à toute une série d'étiquettes peu attractives : imparfaite, fautive, mauvaise (cf. Blanche-Benveniste et Jeanjean 1987 : 20-28; Blanche-Benveniste 1997).

Après plusieurs décennies d'études et de réflexions orientées, entre autres, à réhabiliter la langue parlée/l'oralité en tant qu'état particulier sur le continuum écrit-écrit vs. parlé-parlé, le danger demeure de traiter la transcription comme un texte écrit. Il suffit par exemple de se pencher rapidement sur les différentes grilles proposées en littérature pour l'analyse des erreurs en interprétation (Barik 1971; Altman 1994). Ce risque ressort aussi clairement du fait que l'on oublie assez fréquemment la dimension prosodique qui caractérise les études sur les textes interprétés (Falbo 1999), ou tout simplement, du fait que les analyses sont menées exclusivement sur les aspects purement "linguistiques" (lexique, formulation syntaxique) du discours-interprète et fondées sur une comparaison avec la langue de départ. Ce danger est depuis toujours bien évident aux yeux des professionnels de l'interprétation qui refusent de consentir à l'enregistrement, à la transcription et à la publication de leurs interprétations.

Il va de soi que la transcription ne peut pas être considérée comme la forme écrite d'un texte oral. Elle n'est qu'un support matériel rappelant l'oral évanescent. Et dans une telle optique, il est alors naturel de transcrire l'intégralité du dit avec toutes les "bavures" qui le caractérisent : la nécessité d'accompagner le texte transcrit de sa dimension orale est ressentie par le chercheur comme une conditio sine qua non à toute analyse.

---

10   Il va de soi que les spécialistes de phonétique/phonologie ont des exigences bien différentes de représentation.

4.   La transcription d'un grand corpus

C'est à la lumière de cette problématique que nous nous sommes attachée à la transcription d'un grand corpus d'interprétation recueilli par Francesco Straniero Sergio (1999; 2003; Straniero Sergio et Katan 2001) à partir de 1998 et qui comprend presque toutes les interprétations (1200 environ) passées à la télévision italienne (RAI, Télé satellitaires, chaînes privées) des années 50 à nos jours. Il s'agit d'un corpus ouvert (puisqu'on ajoute au corpus toute nouvelle apparition d'interprètes à la télé), multilingue (les textes interprétés sont en langue italienne; les correspondants originaux en différentes langues étrangères), à l'intérieur duquel on peut distinguer plusieurs sous-corpus, à partir de la typologie de l'événement-objet d'interprétation : du premier pas de l'homme sur la lune, en passant par les funérailles de la Princesse Diana et de Mère Térésa, jusqu'à la guerre en Irak.

4.1.   Les instruments pour la transcription

Il y a encore peu de temps le seul moyen de conserver l'oral à côté de sa transcription, c'était de recourir à l'enregistrement. Depuis quelques années, le développement de logiciels spécifiques (Praat, Transcriber, WinPitch) permet d'informatiser le son par la création de fichiers son et de procéder à la transcription en se servant d'un seul outil : l'ordinateur. Le texte transcrit est ainsi accompagné du texte oral. Le logiciel dont nous nous servons et qui a été modifié et adapté pour répondre aux exigences propres à l'interprétation (textes assez longs, synchronisation entre texte original et texte interprété, gestion des chevauchements dans l'interprétation de liaison, par exemple) est WinPitch, mis au point par Philippe Martin, Professeur à l'Université de Paris VII. Ce logiciel offre au transcripteur la possibilité de ralentir le débit du texte sans que la voix soit déformée, de façon à mieux comprendre et mieux coordonner écoute et écriture[11]. Au fur et à mesure que l'on transcrit, le texte écrit, ainsi produit, est automatiquement aligné, c'est-à-dire 'lié', 'mis en relation permanente', avec le segment audio correspondant. Une fois l'alignement terminé, la possibilité existe de se déplacer dans le texte transcrit/oral et d'écouter en voyant ou de voir en écoutant le résultat du travail effectué. En effet, pendant l'écoute, il est possible de voir à l'écran, le texte transcrit correspondant, ainsi que l'image de la courbe mélodique (F_0). Le caractère performant de WinPitch consiste, entre autres, à gérer et à aligner des transcriptions toutes faites. Les textes en Rich Text Format peuvent être utilisés pour l'alignement avec le texte oral

---

11   Nous ne signalons ici que les aspects les plus innovants et les plus utiles au transcripteur, renvoyant à Martin (2001) pour une présentation exhaustive.

correspondant, sans devoir passer par une deuxième transcription. Évidemment toute transcription peut être corrigée, si besoin en est. En outre, il est possible de transformer un texte transcrit aligné en une 'transcription traditionnelle' répondant aux exigences de publication. Il est clair que WinPitch représente un instrument très utile qui facilite énormément la tâche onéreuse du transcripteur.

L'utilité du logiciel est apparue dans toute son évidence lors de la correction des transcriptions effectuées par un groupe d'étudiantes qui préparaient leurs mémoires de thèse. Les problèmes que nous avons eu l'occasion de relever sont de deux ordres: difficultés liées à la perception et doutes concernant la notation de certains phonèmes.

4.2.   Problèmes de perception

Dans les transcriptions effectuées par des étudiantes au moyen d'un simple magnétophone, nous avons pu constater, au moment de la correction sur WinPitch, qu'il manquait des portions de texte. Ces blancs pourraient s'expliquer par l'approche plus ou moins rigoureuse et soignée que les différentes étudiantes auraient adoptée, mais la présence du même phénomène chez toutes les étudiantes-transcripteurs semble montrer que le problème est, dans une certaine mesure, indépendant du transcripteur et inhérent au processus d'écoute/perception. Grâce à l'utilisation du ralentisseur, nous avons pu nous rendre compte de ce que les étudiantes avaient très probablement entendu, mais qu'elles avaient "oublié" de transcrire[12].

Ainsi, avons-nous été à même de rétablir les parties qui n'avaient pas été transcrites et dont nous présentons ci-dessous quelques exemples[13] :
– des tours de parole: "grazie una domanda", "sono d'accordo";
– des segments plus ou moins longs : "ecco ne abbiamo uno qui"; "no no non c'è bisogno di fare polemiche"
– des tâtonnements de l'interprète (les parties qui n'apparaissaient pas sont en majuscule) : "un UN indebitamento"; "ogni A- OGNI giorno"; "colpevole

---

12   Nous avons remarqué que très souvent, même lorsqu'on utilise le dictaphone avec la possibilité de reécouter facilement le même bout de texte plusieurs fois, le transcripteur est obligé de garder en mémoire plusieurs mots à la fois. L'effort de mémoire est renforcé mais aussi surchargé par l'opération de reécoute qui doit se conjuguer en même temps avec la tâche d'écriture. Cela parfois amène le transcripteur à confondre ce qu'il écoute et ce qu'il écrit. Evidemment il est possible de revenir sur la transcription et de la corriger, mais très souvent la fatigue et la patience dont doit faire preuve le transcripteur(-étudiant) découragent toute hypothèse de correction de la part de l'étudiant lui-même.

13   Les exemples sont tirés de la transcription du débat présidentiel entre Bush, Clinton et Perot de 1992.

CHI chi è stato"; "a ventidue ANNI ventiquattro anni"; "eh eh EH L- cosa accade"; "il DAL- il dollaro"; "e quindi non è COLPA ancora una volta il colpevole"; "questa sera con UN CON degli aspetti".

Nous avons pu également repérer des parties qui avaient été notées comme incompréhensibles :

– "in piccionaia"; "New Hampshire".

4.3. Problèmes de notation

Les problèmes de notation dont nous nous occupons ici, cachent en réalité de véritables problèmes d'interprétation (cf. Bilger *et al.* 1997). Parfois il est facile de reconnaître un phonème ou une séquence de phonèmes, mais leur notation n'est pas aisée, à cause de l'homophonie qui existe à l'intérieur de la langue et qui concerne les différentes parties grammaticales. Lorsque nous entendons [la], nous reconnaissons l'article (italien) défini féminin singulier[14]. Notre certitude disparaît lorsque nous prenons en considération le segment textuel : "per [la] l'arresto". Il est évident que les interprétations à donner à [la] peuvent être au moins deux : "per la l'arresto" ou bien "per l'a- l'arresto". Dans la première hypothèse, l'interprète "prévoyait" un substantif féminin; il renonce (autocorrection) et replanifie son énoncé; dans la deuxième hypothèse il tâtonne, il essaie de prononcer la séquence "l'arresto", mais il réussit seulement à la deuxième tentative. Cela pourrait apparaître aux yeux de ceux qui ne se sont jamais occupés de transcription comme une sorte d'élucubration sur le sexe des anges. Mais que dire devant une étude portant sur les hésitations de l'interprète? Il est évident que toute décision prise par le(s) transcripteur(s) va avoir des retombées sur les résultats de l'analyse. Dans le cas présenté ici, une écoute ralentie a révélé la présence d'une répétition de [la], ce qui pourrait nous faire trancher pour l'hésitation. Toutefois suivant l'exemple de Blanche-Benveniste et Jeanjean (1987: 143) nous avons préféré rendre compte de cette double interprétation en choisissant la *multi-transcription* et donc l'écriture "per la/l'a-la/l'a- l'arresto". Les exemples sont nombreux; nous n'en présentons que quelques uns :

– contro /la, l'a-/ /la, l'a-/ l'avvicinamento";
– "che è /al, all-/ all'interno";
– "/del, dell-/ dell'umanità";
– "/nel, nell-/ nell'aria";
– "gli iracheni /ah, ha, ha-/ sembrano hanno chiesto"; la présence d'un sujet pluriel renforce à notre avis l'interprétation "ha-"; cette hypothèse est soutenue par la présence de "sembrano" et le changement successif de

---

14 Le contexte exclut le sens "note musicale".

planification opéré par "hanno"; mais le choix le plus difficile est entre "ah" hésitation et "ha" verbe avoir;
– "aggiunta di una scadenza a questa risoluzione e:/eh non vorrei suggerirle";
– "si /eh, e-:/ ebbe"; un [e] avec allongement vocalique se confond parfaitement avec une hésitation.

5.   Conclusion

Les problèmes concernant la perception et les doutes qui surviennent lors de la notation révèlent encore une fois, si besoin en est, la subjectivité de la transcription. La réflexion sur les différentes hypothèses concernant la notation et leur évaluation, offrent un apport indéniable et indispensable pour avancer dans la voie d'une représentation de plus en plus objective de l'oral. Certes, la transcription est et reste à ce jour une image estompée de l'oralité. Ceci étant, il est aisé de comprendre que paradoxalement la transcription ne pourra jamais atteindre le but qui est en même temps sa véritable raison d'être, à savoir : effacer le caractère éphémère et évanescent de l'oralité en la transformant en écrit. S'il est vrai que l'observation des transcriptions nous permet de "voir" ce que nos oreilles sont incapables d'entendre (cf. Meyer 1998), il est tout aussi vrai que jamais le texte transcrit ne saura être considéré comme l'équivalent du texte oral.

La possibilité de rendre l'exécution de la transcription plus aisée et plus rapide, grâce à un outil comme WinPitch, ne dissipe pas la complexité du problème. Au contraire, le recours à un logiciel capable d'aligner texte transcrit et texte oral correspondant, confirme l'impossibilité d'assimiler l'oral à l'écrit en transformant le parlé en écrit et révèle, chemin faisant, toute l'autonomie et le respect que mérite l'oralité.

Bibliographie

Altman J. (1994) "Error analysis in the teaching of simultaneous interpreting: A pilot study", in *Bridging the Gap. Empirical Research in the Teaching of Simultaneous Interpreting*. Ed. by S. Lambert and B. Moser-Mercer, Amsterdam-Philadelphia, John Benjamins, pp. 25-38.
Atkinson J.M. and Heritage J. (1999) "Jefferson's transcript notation", in *The Discourse Reader*. Ed. by A. Jaworski and N. Coupland, London and New York, Routledge, pp. 158- 166.

Barik H. (1971) "A description of various types of omissions, additions and errors encountered in simultaneous interpretation", *Meta*, XVI, pp. 199-210.

Barik H. (1972) "Interpreters talk a lot among other things", *Babel*, 1/ XVIII, pp. 3-10.

Barik H. (1973) "Simultaneous interpretation: temporal and quantitative data", *Language and Speech*, 16, pp. 237-270.

Barik H. (1975) "Simultaneous interpretation: qualitative and linguistic data", *Language and Speech*, 18, pp. 272-297.

Bilger M., Blasco M., Cappeau P., Pallaud B., Sabio F., Savelli M.-J. (1997) "Transcription de l'oral et interprétation. Illustration de quelques difficultés", *Recherches sur le français parlé*, 14, pp. 57-86.

Blanche-Benveniste C. et Jeanjean C. (éd.) (1987) *Le français parlé. Transcription et édition*, Paris, INALF, Didier Erudition.

Blanche-Benveniste C. (1997) *Approches de la langue parlée en français*, Paris, Ophrys.

Cecot M. (2001) "Pauses in simultaneous interpretation", *The Interpreters' Newsletter*, 11, E.U.T, pp. 63-85.

Eggins S. and Slade D. (1997) *Analysing Casual Conversation*, Cassel, London and Washington.

Falbo C. (1999) "Interprétation : une forme particulière d'oralité", *Revue française de linguistique appliquée*, IV/2, pp. 99-112.

Gerver D. (1971) *Simultaneous and Consecutive Interpretation and Human Information Processing*, Social Science Research Council, Department of Psychology, University of Durham.

Gerver D. (1974a) "The effects of noise on the performance of simultaneous interpreters: accuracy of performance", *Acta Psychologica*, 38, pp. 159-167.

Gerver D. (1974b) "Simultaneous listening and speaking and retention of prose", *Quarterly Journal of Experimental Psychology*, 26, pp. 337-341.

Gerver D. (1976) "Empirical studies of simultaneous interpretation: a review and a model", in *Translation. Application and research*. Ed. by R.W. Brislin, New York, Gardner Press, pp. 165-207.

Gerver D. and Sinaiko H.W. (1978) *Language Interpretation and Communication*, New York and London, Plenum Press.

Goldman-Eisler F. (1967) "Sequential temporal patterns and cognitive processes in speech", *Language and Speech*, 10, pp. 122-132.

Goldman-Eisler F. (1972) "Segmentation of input in simultaneous translation", *Journal of psycholinguistic Research*, 2/1, pp. 127-140.

Goldman-Eisler F. et Cohen M. (1974) "An experimental study of interference between receptive and productive processes relating to simultaneous translation", *Language and Speech*, 17, pp. 1-10.

Lederer M. (1981) *La traduction simultanée*, Paris, Minard.

Martin P. (2001) "Une méthode d'alignement texte-parole de grands corpus", in *Oralità nella parola e nella scrittura*. A cura di M. Margarito, E. Galazzi e M. Lebhar Politi, Torino, Edizioni Libreria Cortina, pp. 305-318.

Meyer B. (1998) "What transcriptions of authentic discourse can reveal about interpreting", *Interpreting*, 3, pp. 65-83.

Pöchhacker F. (1994) *Simultandolmetschen als komplexes Handeln*, Tübingen, Gunter Narr.

Ochs E. (1999) "Transcription as theory", in *The Discourse Reader*. Ed. by A. Jaworski and N. Coupland, London and New York, Routledge, pp. 167-182.

Oléron P et Nanpon H. (1964) "Recherche sur la traduction simultanée", *Journal de psychologie normale et pathologique*, 62, pp. 73-94.

Ong W.J. (1986) *Oralità e scrittura: le tecnologie della parola*, Bologna, Il Mulino, trad. di A. Calanchi, rev. di R. Loretelli.

Setton R. (1999) *Simultaneous Interpretation. A Cognitive-Pragmatic Analysis*, Amsterdam/Philadelphia, John Benjamins.

Stimoli L. (2001) *Interpretazione simultanea di un evento mediale. I funerali di Lady Diana in Italia, Austria, Germania e Spagna*, Trieste, SSLMIT, Unpublished dissertation.

Straniero Sergio F. (1999) "The interpreter on the (talk) show", *The Translator*, 5/2, pp. 303-326.

Straniero Sergio F. (2003) "Norms and quality in media interpreting: the case of formula one press-conferences", *The Interpreters' Newsletter*, 12, pp. 135-174.

Straniero Sergio F. and Katan D. (2001) "Look who's talking", *The Translator*, 7/2, pp. 213-237.

Traverso V. (1999) *L'analyse des conversations*, Paris, Nathan.

METHODOLOGICAL ISSUES IN THE STUDY
OF INTERPRETERS' FLUENCY

Peter Mead
SSLiMIT, Forlì, University of Bologna

Two o'clock A.M. — The experiment has been tried. With what result, I
am now to describe.
Wilkie Collins, *The Moonstone* (1868)

1.   Introduction

This article illustrates a possible methodology for assessing interpreters'
fluency. The rationale for such assessment is briefly introduced, after which an
ongoing empirical study of consecutive interpreting from English to Italian is
outlined and discussed. Methods and results, which usually form two distinct
parts of a research report, are in this case presented together to illustrate how the
methodology is applied in actual practice. For this purpose, though the ultimate
aim of the study is to compare fluency in a sample of consecutive interpretations
by students and professional interpreters, only one interpretation is examined
here. An account of data for the entire sample is planned for a future article.

Assessment of interpreting is no easy task, even for experienced assessors. A
case in point is Flavia Evandri's (1998) study of how seven interpreting teachers
at Italian and Austrian universities were asked to assess Italian-to-German
interpretations by five students from the University of Bologna, who each
recorded a simultaneous and a consecutive interpretation. The teachers did not
know the students and assessed them from the recordings of their
interpretations. Lack of consistency between the various assessments indicates
considerable variability in standards and priorities from one assessor to another.
It was emblematic, for example, that there was unanimity about awarding a pass
or a fail for only three out of ten interpretations. Another interesting finding was
that almost none of the seven assessors could generally be identified as a
consistently higher (or lower) marker than others.

This lack of consensus among different assessors can, to a certain extent, be
linked to discussion of two important related issues. The first of these is whether
interpretation can be judged in isolation from a real communicative setting, with
no regard for possible interpreting strategies or how successfully they are used.
Taking the communicative setting into account involves a number of important
variables, such as delegates' ability to complement what the interpreter says
with information conveyed in handouts and slides. A second issue which can be
usefully focused on is the recognition that the assessor's judgment may differ

considerably according to whether s/he reads a transcript or (as in Evandri's study) listens to a recording of the interpretation (Gile 1999).

Even in "real" conference settings, however, surveys among delegates in different subject areas highlight a variety of opinions on the relevance of evaluation criteria such as terminological correctness and overall fidelity (Kurz 1993). This suggests that variability in assessment standards probably to a certain extent reflects lack of consensus on *what* to assess.

One possible approach is to give points for qualities such as clarity of expression, though this raises the problem that definitions of what is acceptable, comprehensible or complete are often based on "fuzzy" or subjective criteria. An alternative is to judge by default, counting errors and omissions, as in Henri Barik's (1973, 1975) early experimental work on simultaneous interpreting. The suitability of this approach is debatable, and definitions (e.g. of omissions) can again prove difficult.

Assessment need not be subject to differing opinions of what is right, wrong or missing if parameters amenable to objective measurement are taken into account – for example, duration of pauses or speech rate. Two provisos are necessary in this respect, First, such assessment should be based on instrumental measurement, since identifying features like pauses by ear alone entails the same risk of inconsistency among different assessors as focusing on more obviously subjective criteria. Second, pauses and other objectively measurable parameters may ultimately tell us little about the quality of interpretation unless content too is taken into account. However, a quantitative perspective on different features of interpreting can contribute to overall assessment of quality – for example, at a very simple level, by comparing duration of a consecutive interpretation with that of the original speech.

Provided that quantitative analysis is seen in perspective, as only a part of overall evaluation, it offers the distinct practical advantage that it is on the whole more clear-cut than assessment of content-related parameters like completeness or correctness. Admittedly, it is not always as straightforward as might seem at first sight, and will entail methodological choices – for example, choosing between syllables and words as the unit of measurement for speech rate (Pöchhacker 1993; and see section 3, below), or identifying a minimum duration below which pauses are not counted (see section 2.3.2, below). However, any such problems in quantitative analysis can generally be addressed by clear statement and consistent application of the criteria and/or methods chosen. By contrast, content-related parameters are ultimately more difficult to pin down and their evaluation may differ from one assessor to another.

Fluency lends itself to quantitative assessment through a number of indices, sometimes referred to as "temporal variables". One of these is speech rate, though this does not mean that fluency can be automatically equated with speed

– rapid speech may be formally inaccurate and/or incomprehensible. Other temporal variables make it possible to examine fluency by evidence of its absence, a perspective suggested by Erving Goffman's (1981: 172) statement of the following basic rule in public speaking:

> [speech] segments must be patched together without exceeding acceptable limits for pauses, restarts, repetitions, redirections, and other linguistically detectable faults.

Goffman considers that these features of speech reflect the efforts of reasoning and formulation which accompany linguistic production. The skill of professional speakers such as the lecturer or radio announcer is to hide these efforts and any resulting hesitations, so that no "production crisis" or "backstage considerations" (Goffman 1981: 172) will be allowed to betray moments of doubt or distraction.

It is interesting that the various "errors of performance" examined in Andrzej Kopczynski's (1981) study of interpreting quality coincide to all intents and purposes with Goffman's "linguistically detectable faults". This underlines that fluent speech production can be analysed from a similar perspective in both interpreting and public speaking. Just as Goffman argues that the lecturer's fluency will keep the curtain drawn on any production problems backstage, difficulties with any part of the interpreting process need not actually be apparent as such if the interpreter addresses them promptly and discreetly.

Against this background, professional public speaking ability tends understandably to be considered part and parcel of the interpreter's skills (Jones 1998: 40). Ingrid Kurz's (1993) survey of how different user groups and interpreters rate various features of conference interpreting is emblematic in this respect, with fluency placed fifth out of eight items in the overall ranking – ahead of correct grammatical usage, native accent and a pleasant voice. In other words, while fluency ultimately provides no guarantee of the interpreter's reliability, it is an important feature of successful interpretation.

2. Procedures

2.1. Source speech and interpretation

Information in this very brief section is intended only to outline the experimental setting in which the research was carried out, not to present the overall study sample.

The material examined here is a consecutive interpretation by a student working from English "B" into Italian "A", based on a short recorded extract

from the opening of a speech on British attitudes to Europe. The speaker was an English lecturer, addressing a non-specialist audience of Italian students.

According to the definitions of speech presentation modes used by Lehtonen (1982: 40), this was an extemporaneous delivery – planned in advance but presented freely, not read. Use of an extemporaneous speech for the experiment was preferred to manuscript delivery, which would probably have proved very difficult for beginner students like the one whose interpretation is examined here.

A transcription of the source speech, punctuated for ease of reference, is provided in Appendix 1.

The analysis of methodological issues in the following paragraphs does not examine the student's retrospective comments on silences and hesitations in the interpretation, made while listening to it on tape immediately afterwards. This aspect of the study will be touched on only very briefly in the final discussion, since it has already been examined in detail elsewhere (Mead 2002).

2.2.   Transcription and computer processing of the recorded interpretation

The recording of the interpretation was transcribed without punctuation, including all words or parts of words identified by close listening. Hesitation noises with no phonemic value were simply transcribed as "eh" or "mm", to indicate their oral or nasal character respectively. No attempt was made to indicate their duration in this initial transcription, which was intended simply as a first step towards more detailed noting of pauses.

The next step was to transfer the recording on to the hard disk of a Macintosh iMac™, using a programme for visualisation and editing of audio files (SndSampler 3.7.1™, © Alan Glenn, Midland Mi, USA). This software makes it possible to convert an acoustic signal into an oscillogram, visualising sounds as a continuous wave pattern on which any segment can be highlighted and matched with the corresponding recording. At a sampling frequency of 44 kHz, duration of different speech features can be measured in hundredths of a second. Similar programmes can be readily found for a Windows environment – for example, Adobe Audition™. Accuracy to a thousandth of a second can be achieved on some programmes, though this is necessary only for detailed phonetic study.

For the present analysis, the interpretation was divided into eleven 20-second segments and one final 6-second segment (totalling 3'46"). Each segment thus created could be visualised as a single oscillogram, so that all pauses could be identified and measured (see section 2.3.3, below). Some programmes of this kind have a zoom facility, which makes it possible to focus

on any part of the speech without first having to segment it into a sequence of smaller files.

The detailed transcription including pause data obtained from the oscillograms is reproduced below. An English gloss is provided immediately below the Italian text. This gloss is basically a word-for-word translation, with only minor adjustments where too literal a translation might prove difficult to understand. For example, in lines 8-9, the literal translation of "penso che siano necessarie le [0,18] le mie scuse" would be "I think that are necessary the the my apologies"; for purposes of clarity, this has been modified to "I think that there is need for my apologies".

For convenience of reference, lines in the transcription have been numbered on the right. All pauses are indicated in square brackets, those of at least 0.25 sec. (see section 2.3.2, below) being highlighted in bold type. Pause durations are shown in seconds. A simple indication of duration (e.g. [0.43], in l. 1) represents a silent pause. To indicate filled pauses (commonly referred to as "ums and ahs"), the duration is in each case shown alongside the corresponding vocalisation (e.g. [mm 0.51], in l. 2). A "mixed" pause, comprising an uninterrupted sequence of a silent and a filled pause, is indicated by a hyphen before or after a vocalisation (as in the first pause in l. 1). Underlined pairs of words (e.g., <u>della del</u>, in l. 6) are repetitions, in each case counted as two words (see section 2.3.1, below).

Table 1   Transcription of the interpretation, complete with English gloss

buongiorno **[- eh 1.89]** l'ultima volta che ho partecipato a un convegno è stato **[0.43]** a *hello the last time that I participated in a conference was in*	1
[mm 0.51] Bruxelles **[0.47]** dove gli interpreti mi hanno detto **[eh 0.73]** che stavo parlando * Brussels where the interpreters told me that I was speaking*	2
troppo velocemente e stavo dicendo delle stupidaggini **[0.87]** quindi nel caso ciò *too quickly and I was talking nonsense so in the event that*	3
succedesse anche questa volta vi prego di dirmelo in modo che possiamo comunicare senza *should happen this time too I ask you to tell me so that we can communicate without*	4
molti problemi **[0.83]** innanzitutto **[0.25]** vorrei iniziare con **[eh 0.43]** le mie scuse *many problems first of all I'd like to start with my apologies*	5
<u>[- eh 1.02]</u> vorrei scusarmi per il comportamento <u>della del</u> mio staff per la sua stupidità nei * I would like to apologise for the behaviour of my staff for their stupidity in*	6
confronti dell'Europa e nei vostri confronti **[1.57]** la storia che vi racconterò comunque ha *regard to Europe and towards you the story that I'll tell you however has*	7
<u>una</u> **[0.99]** <u>un</u> finale positivo **[eh - 1.68]** comunque penso che siano necessarie <u>le</u> [0.18] <u>le</u> *a positive ending nevertheless I think that there is need for*	8

mie scuse ancora una volta per quello che **[eh 0.40]** i miei concittadini hanno fatto nei *my apologies once more for what my fellow-citizens have done in*	9
vostri confronti e all'interno della **[0.33]** politica dell'Unione Europea **[- eh 1.93]** il mio *regard to you and within the politics of the European Union my*	10
[0.55] discorso **[eh 0.54]** è diviso <u>in in</u> due parti principalmente vi vorrei **[0.40]** parlare un *speech is divided into two parts mainly I'd like to talk to you a*	11
po' **[eh 0.22]** delle date più importanti che riguardano la creazione **[0.26]** dell'Unione *little about the most important dates which concern the creation of the*	12
Europea e i rapporti della Gran Bretagna nell'Unione [0.15] con l'Unione Europea **[0.36]** e *EU and the relations of Great Britain in the Union with the EU and*	13
poi **[eh 0.43]** darvi alcune spiegazioni per quanto riguarda **[0.54]** il ruolo che ha svolto *then give you some explanations regarding the role played*	14
[0.22] il mio stato all'interno dell'Unione Europea **[- eh 2.11]** il comportamento della *by my state within the European Union the behaviour of*	15
[eh 0.62] Gran Bretagna può **[eh 0.44]** sembrare un po' strano **[- eh 2.95]** forse difficile da *Great Britain may seem a little strange maybe difficult to*	16
comprendere da parte degli altri membri dell'Unione Europea soprattutto **[eh 0.36]** prima *understand by the other members of the European Union above all before*	17
che la Gran Bretagna **[eh 0.37]** diventasse **[mm 0.66]** membro dell'Unione Europea *Great Britain became member of the European Union*	18
[- eh 1.02] la prima data **[eh 0.59]** importante il primo evento di cui vi voglio parlare è il *the first important date the first event of which I want to speak to you is the*	19
discorso tenuto da Churchill a Zurigo nel millenovecentoquarantasei **[1.64]** dove [0.15] *speech given by Churchill in Zurich in nineteen forty-six where*	20
Churchill ha **[- eh 1.13]** parlato **[eh 0.47]** della situazione <u>del della</u> Gran Bretagna *Churchill spoke of the situation of of Great Britain*	21
affermato che **[eh 0.62]** la Gran Bretagna non era ancora pronta per entrare a far parte *stated that Great Britain was not yet ready to enter and become part*	22
dell'Unione Europea **[0.55]** la seconda data **[eh 0.48]** fondamentale per la creazione *of the European Union the second date fundamental for the creation*	23
dell'Unione Europea e per **[0.29]** il ruolo della Gran Bretagna all'interno dell'Unione è il *of the European Union and for the role of Great Britain within the Union is*	24
millenovecentocinquantotto in cui è stato **[0.29]** firmato il **[eh 0.26]** Trattato di Bruxelles *nineteen forty-eight with the signing of the Treaty of Brussels*	25

[0.80] ed è stato creato il **[eh 0.26]** Consiglio Europeo **[0.69]** che secondo l'Unione doveva *and the creation of the European Council which according to the Union had to*	26
svolgere un **[eh 0.47]** un ruolo **[eh 0.26]** sovrannazionale quindi di controllo *play a a supranational role thus of control*	27
sovrannazionale **[- eh 1.31]** la Gran Bretagna [0.22] si è opposta alla decisione del *[at] supranational [level] Great Britain opposed the decision of the*	28
consiglio e ha **[0.32]** affermato che **[eh 0.51]** il Consiglio Europeo dovrebbe **[0.32]** più che *council and stated that the European Council should more than*	29
altro avere un ruolo internazionale **[1.53]** il **[mm 0.48]** l'atteggiamento della Gran Bretagna *anything have an international role the the attitude of Great Britain*	30
[- eh 1.79] è **[0.69]** sembrato arrogante **[- eh - 2.15]** visto che **[ehmm 1.31]** i cittadini *seemed arrogant seeing that the English citizens*	31
inglesi e il governo **[eh 0.40]** inglese **[eh 1.31]** pensava [0.22] che l'Unione Europea aveva *and the English government thought that the European Union*	32
bisogno della **[- eh 1.02]** Gran Bretagna [0.18] per [eh 0.22] aiuti economici **[eh - 2.01]** per *needed Great Britain for economic help to*	33
[eh 0.66] risollevare la situazione **[0.36]** così negativa del **[eh 0.44]** dell'Europa *boost the situation [which was] so negative of of Europe*	34
[- eh - 2.62] infatti in tutto l'arco **[eh 0.33]** degli anni cinquanta l'economia **[0.37]** europea *indeed in the whole period of the fifties the European economy*	35
[eh 0.41] non aveva [mm 0.74] mostrato grandi miglioramenti e la situazione era piuttosto *hadn't shown great improvements and the situation was rather*	36
negativa *negative*	37

2.3.   Analysis of temporal variables

As explained above, information which reports of empirical research usually separate into distinct "methodology" and "results" sections is presented here as a single illustrative example for each stage of the study.

The following temporal variables will be examined in this way: speech rate (number of words or syllables spoken per minute), duration of pauses, phonation/time ratio (the percentage of speech time used for actual speech production, as opposed to pauses), articulation rate (number of words or syllables spoken per minute, but not counting pauses as part of speech production time) and mean length of run (the mean number of words or syllables between pauses).

Of Goffman's "linguistically detectable faults", only pauses are included in the above list. Other disfluencies such as false starts and repetitions, referred to by Goffman (1981: 172) as "restarts […] redirections", are related as much to content as to rhythm and will thus not be examined in this initial exploration of interpreters' fluency. Drawls (drawn-out vowels, often in final position, as when /i:/ becomes /i:::/ in the phrase "for me") are also excluded, mainly because their identification is to a certain extent subjective.

Several problems of content in the interpretation examined here can be readily identified from even a rapid examination of the transcribed target text – e.g., anachronistic references to the European Union, confusion between the Council of Europe and the European Council. However, these are not relevant to the type of analysis proposed.

For illustrative purposes, data on the different variables considered here will be briefly compared with those provided in studies using similar methodology for assessment of fluency in unprepared speech (see below, sections 2.3.1, 2.3.3 and 2.3.5). Speech production mode is obviously an important factor to be considered when comparing temporal variables in different speech samples – distinguishing, for example, between reading aloud, impromptu speech and consecutive or simultaneous interpretation. Even within a given production mode, a number of other factors should be borne in mind – language, register and topic being obvious examples. While isolated research efforts on temporal variables in relation to these features of speech production date back even decades (e.g. Goldman-Eisler 1967), the topic remains relatively little explored and available data are limited.

2.3.1. First temporal variable: speech rate

The expression "speech rate" is that used by Manfred Raupach (1980), whereas Richard Towell, Roger Hawkins and Nives Bazergui (1996) use "speaking rate" in their comparison of unprepared English and French oral production by twelve English students studying French at university. Luca Onnis (1999) also uses "speaking rate" in his study of English and Italian production by eight late English-Italian bilinguals (i.e. native speakers of English who have lived their adult life in Italy and acquired an excellent command of Italian).

All that is needed to calculate speech rate is a recorded speech sample and a measurement of its duration. A programme like SndSampler 3.7.1™ makes it possible to measure duration with great precision, but even a stopwatch is enough for reasonably accurate measurement of a speech sample's overall duration in minutes and seconds.

Dividing the total number of words or syllables in the speech sample by its duration in minutes or seconds gives the speech rate. Both words and syllables

were counted in the present study. The choice between the two units of measurement, together with problems of definition and method raised by use of the syllable, will be commented on below (see section 3).

The word count includes all complete words, even if these are part of a false start or repetition (as in the six cases underlined, in lines 6, 8, 11, 21 and 27 of the transcription). Incomplete words are not counted (though they would be included in a *syllable* count). It is far more convenient to use the word counter of a word processing programme than to do a visual count, though a subsequent check is necessary to identify any incomplete words which the programme will have included in the count. For the present study, this check was also used to ensure that all apostrophised forms except the definite article *l'* were counted as separate words. For example, "dell'Unione" (in l. 12 of the transcription) and "nell'Unione" (l. 13) were each treated as two words, though the automatic word count considered them as one.

The total count thus obtained for the interpretation was 394 words. To calculate speech rate in words per minute (w.p.m.), this total was divided by the interpretation's overall duration in seconds (226) and multiplied by 60. The resulting speech rate is 104.60 w.p.m. The syllable count is 880, giving a speech rate of 233.63 syllables per minute.

Like all the results presented below, this information means little out of context. One possibility is to look at it in relation to data for other individuals – either within the same sample or in other studies. For example, the mean speech rate of Onnis' (1999) eight bilinguals in Italian (99.50 w.p.m.) is fairly close to that of the student whose interpretation is examined here, while it is much lower in the study by Towell, Hawkins and Bazergui (1996) (186.92 syllables per minute in English). Comparability of speech rates in the three studies is to a certain extent limited by differences in units of measurement – Onnis counts in words, while Towell, Hawkins and Bazergui use syllables. The languages examined in the three studies also differ, as do at least two other important variables: (i) presumed level of competence (Italian is the interpreting student's native language in the present study, but the weaker of the late bilinguals' languages in Onnis' sample); (ii) experience in oral presentation (the interpretation analysed here is by a beginner student, while Onnis' subjects are teachers).

While caution is needed in comparing speech rate data across different samples, a potentially interesting alternative for future research is to look at speech rate in relation to other variables in interpretations by the same individual or group – for example, examining whether the interpreter's linguistic control is tighter or slacker at different speeds.

2.3.2. Second temporal variable: duration of pauses

The oscillograms created on the SndSampler 3.7.1™ programme (see 2.2, above) made it possible to isolate both filled pauses and silent pauses. While filled pauses are often thought of as disfluencies (e.g. Duez 1982), silent pauses can either go unnoticed or actually help the listener – for example, by holding back new information long enough for what has just been said to sink in. Silent pauses at natural syntactic breaks thus favour efficient segmentation of the incoming message by the listener – while at the same time affording the speaker an opportunity for discreet planning of what comes next (Butterworth 1980: 157; Deese 1980: 84).

Overall times for silent and filled pauses were calculated separately, while mixed sequences of silent and filled pauses (e.g. "[ ]eh[ ]") were considered as filled pauses. Thus, an initial silent pause of 0.50 sec. merging into a filled pause of 0.20 sec. and another silent pause of 0.50 sec. would be counted as a single filled pause of 1.20 sec. The reason is that there is little likelihood of each silent or filled part within the sequence being perceived as a pause in its own right. How these "mixed" pauses are transcribed has already been explained in the introductory remarks to Table 1.

There is some debate about the most appropriate minimum cut-off point for pause measurement. Towell, Hawkins and Bazergui (1996) set the threshold at 0.28 seconds, while Onnis (1999) uses 0.10 seconds. The minimum cut-off point used in the present illustration is 0.25 sec., as in a number of earlier studies (e.g. Goldman-Eisler 1958). This means that shorter pauses, though shown in the transcription, are not included in the calculation of pause duration. The upper cut-off point is less debated in the literature; 3 sec., the limit agreed on by a number of authors (Towell, Hawkins and Bazergui 1996; Onnis 1999), has been applied here – though all pauses in the interpretation analysed are well below this limit.

Once each pause had been identified and measured, total duration of silent and filled pauses for the interpretation as a whole was obtained by adding up all the individual pauses of at least 0.25 sec. (marked in the transcription, in bold type: see Table 1, above). In all, there were 68 (25 silent pauses, 43 filled pauses). Their total duration was 15.70 sec. and 40.12 sec., for silent and filled pauses respectively – in other words, almost a minute (55.82 sec.) of overall pause time.

Comparison with pause duration in other speech samples or interpretations should be based on a common denominator. In other words, total pause time (55.82 sec. in this case, 57.98 sec. in another student's interpretation of the same speech) should be examined in relation to the overall duration of each interpretation (3'46" and 3'17", respectively). Given this information, one way

of comparing pause time in the two interpretations is to calculate it in each case as a percentage of the total speech times: 24.70% and 29.43% respectively. This shows that pause time, quite similar in the two cases if simply quantified in seconds, in fact differs as a proportion of overall speaking time. Another possibility is to calculate pause duration per minute – in this case, 14.82 secs. vs. 17.66 secs. These pause times may at first sight seem high if there is no yardstick to measure them by, but they will now be considered in relation to data from other studies.

One problem in comparing pause data from different studies is that some authors consider filled pauses as non-phonemic syllables and do not include them in the calculation of pause time. If this methodology is adopted, the pause times in the two interpretations compared above are thus 15.70 sec. and 15.77 sec. Silent pause duration in the two cases thus differs very little, as reflected in the corresponding percentages: 6.95% vs. 8.00%.

Among those who include only silent pauses in calculation of pause time are Towell, Hawkins and Bazergui (1996), while Onnis (1999) takes both silent and filled pauses into account.

In both these studies, pause duration is just one parameter of fluency and does not necessarily give much information if considered in isolation from other variables such as speech rate and average length of pauses. There is also the problem that, ultimately, the researcher has no sure way of distinguishing between hesitation pauses (to allow speech planning) and functional pauses (to help the listener or create rhetorical effect).

Detailed examination of such debate, which remains at best speculative, is beyond the scope of the present study. Pause duration can nevertheless provide an interesting point of comparison as an important determinant of speech rhythm in different samples. In the next section (2.3.3), pause duration and phonation/time ratio in the present study are tentatively compared with data from the studies by Towell *et al*. and Onnis.

First, however, phonation/time ratio should be briefly explained. Since it simply gives the same information as pause time from a different perspective, how these data relate to those reported by other authors can then be discussed for the two variables together.

2.3.3. Third temporal variable: phonation/time ratio

Phonation/time ratio (PTR) is the percentage of speaking time used for phonation, or actual speech production, as opposed to pauses. It adds no real information to that provided by the calculation of pause time as a percentage of speaking time, since it is simply the calculation of the balance left when pause time is subtracted.

In the above example, pause times for the two subjects are 6.95% and 8.00% if only silent pauses are included in the calculation, or 24.70% and 29.43% if all pauses are taken into account. PTR is thus 93.05% (100 - 6.95) and 92% (100 - 8) in the first case, 75.30% (100 - 24.70) and 70.57% (100 - 29.43) in the second.

Towell, Hawkins and Bazergui (1996) identify a mean PTR of about 66% in impromptu production of English as a native language by their 12 subjects, the calculation being based on *silent* pauses of at least 0.28 sec. Though PTR in French increases (from 57% to 62%) after the students have spent several months in France, it remains slightly lower than in English. Mean PTR in the impromptu English and Italian speech of the eight late bilinguals studied by Onnis (1999) is about 65% in both languages, the calculation being based on *silent and filled* pauses of at least 0.10 sec. The PTR calculated in the present study can be tentatively compared with that in Towell, Hawkins and Bazergui (1996), the minimum cut-off point being fairly close in the two cases. If PTR in the present study is based on silent pauses alone, it is greater than 90%. This means that it is considerably higher than in the impromptu production analysed by Towell, Hawkins and Bazergui. Such a comparison can be at best tentative, for two reasons: (i) only two interpretations have been considered here, as opposed to 12 subjects in the Towell, Hawkins and Bazergui study; (ii) different languages are involved (English and French in one case, Italian in the other). It is nevertheless possible to formulate a provisional hypothesis that unprepared monolingual production, unlike consecutive interpretation, requires "on line" planning of speech content and thus involves more pausing.

This would be consistent with Daniel Gile's (1995: 89) argument that:

> L'interprète [en consécutive] connaît l'ensemble du segment de discours qu'il va interpréter avant d'en commencer la reformulation: Sur ce plan, il est parfois en meilleure situation que l'orateur, à qui il arrive de devoir improviser. (my emphasis)

It is also in line with the speech rate data discussed above (section 2.3.1), higher in consecutive interpretation than in the Towell, Hawkins and Bazergui study. Onnis' speech rate data, however, are also much higher than those reported by Towell, Hawkins and Bazergui. This underlines the need to weigh up even the most tentative conclusions against a variety of data, and also to assess how far the comparison is subject to other variables – for example, as mentioned at the end of section 2.3.1, Onnis' subjects are teachers and can thus be presumed to have greater experience of monological speech than a novice interpreting student.

2.3.4. Fourth temporal variable: articulation rate

The concept of phonation time – i.e. the time actually dedicated to speech production, calculated by removing pause time from the total duration of the speech – has already been illustrated. The question of whether pause time includes all pauses or only silent pauses has also been explained. In the interpretation analysed for the present study, what must be subtracted from the total duration of 226.00 sec. (i.e. 3'46") is either 15.70 sec. (duration of silent pauses) or 55.82 sec. (duration of silent and filled pauses combined). Phonation time will thus be 210.30 or 170.18 sec. respectively. The total word count, divided by phonation time, gives articulation rate: 108.70 w.p.m. if pause time includes only silent pauses; 134.33 w.p.m. if it includes all pauses.

Compared with pause duration and PTR (discussed in sections 2.3.2 and 2.3.3), data on articulation rate afford a different perspective on fluency. Essentially, however, the information provided is the same. Speech rate is obviously lower than articulation rate, which involves dividing the word or syllable count by only a part of the total duration, but it is interesting to see how much the two rates differ. In this example, speech rate is conspicuously lower than articulation rate only when filled pauses are included in pause time: a speech rate of 101.15 w.p.m. does not differ greatly from an articulation rate of 108.70 w.p.m. (subtracting only silent pauses from total speech production time), but is almost a third lower than an articulation rate of 134.33 w.p.m. (with filled pauses also included in pause time). This indicates that the interpreter's *filled* pauses make up an appreciably greater proportion of production time than *silent* pauses – in other words, "ums" and "ahs" are very noticeable.

The same information is, of course, given by the pause times and PTR, the only difference being that the articulation rate highlights a possible target level to measure actual speech rate against. Whether this target level can actually be taken as a realistic goal is debatable, since a possible side effect of striving to accelerate speech rate by avoiding pauses might be a clipped – and, in some languages, particularly unnatural – delivery. Limiting *filled* pauses, however, can be a relevant goal for interpreters (indeed, for all speakers) as they become more experienced and confident.

Articulation rate data will not be compared with those from other studies. This is because Onnis' methodology specifies a particularly low cut-off in pause duration ($\geq$ 0.10 sec.) for the calculation of phonation time, while Towell, Hawkins and Bazergui calculate articulation rate in syllables per second.

2.3.5. Fifth temporal variable: mean length of run

A *run* is a segment of speech uninterrupted by pauses. Mean length of run
(MLR) is sometimes included in the temporal variables through which fluency
is assessed, just as mean length of utterance is used as an index of grammatical
proficiency in investigation of language development. However, a major
methodological issue associated with the MLR is the problem of how to define a
run − in other words, is it delimited only by silent pauses (the view taken by
Towell, Hawkins and Bazergui 1996), or also by filled pauses (as in Onnis
1999)? Another methodological issue, which will be discussed in the following
section, is the unit of measurement (words or syllables). Both these questions
have obvious repercussions on comparability of data from different samples,
though the issue of definition is the more important of the two.

   In the present study, both silent and filled pauses are taken as the possible
start or finish of a run. The number of runs identified in this way is 69. Dividing
the word count of 394 by 69 gives a MLR of 5.71 words; if syllables are used,
the MLR is 880/69 = 12.75. A problem in comparing this with data in Towell,
Hawkins and Bazergui (1996) is the different definition of a run in the two
studies. If only segments between silent pauses are considered as runs, the MLR
in the present study is 394/26 = 15.15 words (or 880/26 = 33.85 syllables). The
12 individuals studied by Towell, Hawkins and Bazergui have a far lower MLR,
averaging only 7.25 syllables. It is difficult to understand why this result differs
so much in the two studies, as the difference in the languages involved would
hardly account for such an enormous gap. The hypothesis that consecutive
interpretation is in a sense less demanding than unprepared monolingual
production, because the interpreter does not have to plan speech content "on
line", has already been tentatively advanced in section 2.3.3. Even this
hypothesis, however, offers no satisfactory explanation of why MLR in the
present study is so much higher than in the sample studied by Towell, Hawkins
and Bazergui.

   Comparison with the study by Onnis (1999) does not involve the problem of
definition, since he considers that runs can start and end with either silent or
filled pauses. The problem in this case is that Onnis calculates MLR on all
segments between pauses of at least 0.10 sec. It is thus hardly surprising that the
MLR in the present study (5.71 words) proves appreciably higher than in Onnis'
sample (4.8 words in English, 3.6 words in Italian). As the pause criterion is so
different in the two studies, it makes little sense to compare data in relation to
such variables as production mode (consecutive interpreting in one case,
extemporaneous speech in the other) or language proficiency in Italian (native
command in one case, late acquisition in the other).

3.   Discussion

3.1.   Words or syllables as the unit of measurement

Franz Pöchhacker (1993) points out that, if some interpreting researchers measure speech length in syllables and others use words, there can be little basis for comparing data from different research groups. Since the average syllable count per word can differ considerably in different languages (and also in relation to other variables – e.g., sector, register, read texts vs. off-the-cuff speech), Pöchhacker suggests that the syllable is probably a better standard international unit of measurement than the word. He also reports examples of research, including his own work, in which both syllable and word counts have been obtained so that the ratio of syllables to words can be calculated. This approach has been maintained in the present study. If this practice became widespread, it would provide a good basis for more systematic study of how the conversion factor between the two units of measurement varies in different samples.

Pöchhacker (1993: 57) rightly acknowledges that syllables are not in themselves "an 'objective' yardstick of speed, let alone a measure of 'information' per time unit". An objection to measurements in syllables for comparisons across languages is raised by Onnis (1999: 87), who suggests that focusing on syllables alone can give a misleading idea of information content and that syllable counts should therefore be accompanied by ratios for converting them into words. To illustrate this point, Onnis hypothesises an extreme case of an English speech sample possibly having fewer syllables but more words than a sample of Italian.

However, argument on the respective merits of words and syllables is ultimately inconclusive – starting from the same basic observation that word/syllable ratios vary from language to language, Pöchhacker supports the syllable while Onnis prefers the word. There thus seems to be a strong case for systematically using both, as in Pöchhacker's (1993) article and in the present study. The considerable work this entails will surely be justified if it brings to light much-needed data on the conversion ratio between words and syllables.

The practice of counting syllables raises two important practical issues – whether to count the syllables indicated in dictionaries or those actually pronounced, and *how* to count them. On the first issue, it seems prudent to use a notional syllable count. Identification of how many syllables are actually pronounced depends either on the individual researcher's perception (with "top-down" processing making it difficult to distinguish between what s/he actually hears and what s/he expects to hear) or on very detailed instrumental testing of sound samples. Neither option is satisfactory – the first being unreliable, the second too demanding of time and resources. However, the approximation of a

syllable count based on the "presumed" number of syllables in a given word at least offers the practical advantage of a readily accessible, standardised methodology.

The other practical issue is whether syllables, like words, can be counted automatically on the computer. For the present study, the first step was to create a text file of the transcription and divide words manually into syllables, which the programme was then able to count as if they were words. A practical limitation of this method is that dividing words into syllables with a cursor on a computer screen would obviously prove too laborious and eye-straining for longer speech samples.

The syllable counts for the consecutive interpretation examined above have already been indicated in the sections on speech rate and mean length of run. The syllable/word ratio for the interpretation as a whole is 2.23 and, as the study progresses, it will be interesting to compare this with the ratio in interpretations by other subjects.

3.2.   Which parameters?

Five temporal variables have been examined in this study (speech rate, duration of pauses, phonation/time ratio, articulation rate, mean length of run).

For practical purposes, three of these (speech rate, duration of pauses and MLR) are probably enough. The reason is that, as explained above, PTR and articulation rate add no new information to that on pause duration – they simply offer different perspectives on the same data. In other words, pause duration indicates what proportion of speaking time the interpreter spends pausing, while PTR is the remaining proportion of speaking time. This is seen most clearly if both are expressed as percentages of total speaking time, in which case they add up to 100 (24.70% pause duration and 75.30% PTR, in the interpretation examined above). On the other hand, if pause duration were expressed in seconds per minute (in this case, 14.82 s.p.m.), no immediate relationship with a PTR of 75.30% would be apparent. Articulation rate is more complicated to calculate, but simply means how fast the interpreter speaks during phonation (in other words, when not pausing). As explained above, the potential interest of this parameter from the trainee interpreter's perspective is that it can very tentatively be taken as a theoretical speech rate to aim for by limiting pause time (see 2.3.4). It is useful for the trainee to appreciate to what extent "ums" and "ahs" can detract from fluency, though over-zealous outlawing of all pauses should not be encouraged.

Speech rate, pause duration and length of run can thus be highlighted as the most relevant of the parameters suggested above. At the same time, they can be complemented by other information not examined in this study.

3.2.1. Further analysis of pause duration

One option is to examine pause duration in greater detail. For example, Towell *et al.* (1996) and Onnis (1999) calculate the average length of pause (ALP) as a complement to overall pause duration. Examining the two parameters together makes it possible to see whether differences in pause duration from one speech sample to another are more related to the frequency of pauses or to ALP (Towell, Hawkins and Bazergui 1996). In the present study, ALP based on pauses of at least 0.25 sec. was 0.63 sec. for the 25 silent pauses and 0.93 sec. for the 43 filled pauses. While it is difficult to comment on these data in isolation, without other interpretations to provide a basis for comparison, they again underline the preponderance of filled pauses – in terms of both frequency and duration.

Another option is to examine the range of individual pause durations. For example, more than half the silent pauses identified in this study were less than half a second in duration, while only three (i.e. 12%) lasted more than a second. Of the 43 filled pauses, as many as 16 (i.e. almost 40%) lasted more than a second (including 5 – i.e. over 10% – which lasted more than 2 seconds). This information complements the message which has already emerged from the data on pause duration and frequency – i.e., that the student concerned should keep a tighter rein on filled pauses.

Detailed analysis of pause duration also suggests that a run may in practice be a very disfluent speech segment if the only landmarks used to identify it are initial and final silent pauses (as in Towell, Hawkins and Bazergui 1996). For example, the first 29 words in lines 1-3 of the transcription in Table 1 read as follows:

> buongiorno [- eh 1,89] l'ultima volta che ho partecipato a un convegno è stato [0,43] a [mm 0.51] Bruxelles [0,47] dove gli interpreti mi hanno detto [eh 0,73] che stavo parlando troppo velocemente e stavo dicendo delle stupidaggini [0,87]

Here, there are four silent pauses and two filled pauses. If runs are defined as segments between silent pauses, three can be identified here:

1. buongiorno [- eh 1,89] l'ultima volta che ho partecipato a un convegno è stato [0,43] (11 words)
2. a [mm 0.51] Bruxelles [0,47] (2 words)
3. dove gli interpreti mi hanno detto [eh 0,73] che stavo parlando troppo velocemente e stavo dicendo delle stupidaggini [0,87] (16 words).

However, if runs are considered to start and finish with either silent or filled pauses, their number doubles:

1.   buongiorno [- eh 1,89] (<u>1 word</u>)
2.   l'ultima volta che ho partecipato a un convegno è stato [0,43] (<u>10 words</u>)
3.   a [mm 0.51] (<u>1 word</u>)
4.   Bruxelles [0,47] (<u>1 word</u>)
5.   dove gli interpreti mi hanno detto [eh 0,73] (<u>6 words</u>)
6.   che stavo parlando troppo velocemente e stavo dicendo delle stupidaggini [0,87] (<u>10 words</u>).

The choice between the two definitions thus makes a considerable difference to both the number of runs (3 vs. 6) and the MLR (9.7 vs. 4.8 words). For the interpretation as a whole, the number of runs can be counted as 26 or 69 according to which definition is used, while the corresponding MLR is. 15.15 or 5.71 words. Given that long filled pauses such as those in the above example can on the whole be readily identified as disfluencies by the listener, there is a strong argument for considering them as cut-off points for segmentation into runs, not as non-phonemic syllables within runs.

3.2.2. Pause position

A final point of interest is the question of where pauses occur. Towell, Hawkins and Bazergui (1996) do not address this issue, while Onnis (1999) catalogues pause distribution in relation to syntactic position (e.g., between clauses or phrases, within phrases) but does not discuss the question in much detail.

In the present study, position of pauses was classed very simply as: (i) at sentence boundaries; (ii) at clause or phrase boundaries; (iii) just after the initial conjunction of a clause; (iv) within a clause; (v) within a phrase. The syntactic units referred to might lend themselves to criticism as having been borrowed from conventional grammatical analysis (of the written language), but they are intended only to provide a crude framework for a cursory analysis of pause distribution.

Distribution of the five classes of pause in the interpretation, shown in Appendix 2, breaks down as follows:

(i)   at sentence boundaries: 13 pauses (e.g., <u>[- eh 1,89] l'ultima volta che ho partecipato a un convegno</u>, in l. 1);

(ii) at clause or phrase boundaries: 11 pauses (e.g., <u>innanzitutto [0,25] vorrei iniziare</u>, in l. 12);

(iii) just after the initial conjunction of a clause: 7 pauses (e.g., <u>per quello che [eh 0.40] i miei concittadini hanno fatto</u>, in l. 9);

(iv) within a clause: 11 pauses (e.g., <u>prima che la Gran Bretagna [eh 0.37] diventasse</u>, in l. 17-18);

(v) within a phrase: 26 pauses (e.g., <u>iniziare con [eh 0.43] le mie scuse</u>, in l. 5).

Consistent with Goffman's advice that the public speaker should make any disfluencies as unobtrusive as possible (see above, Introduction), pauses at major syntactic boundaries (classes i and ii) make up about a third of the total. If pauses just after the first word of a clause (class iii) are also included in this category, it accounts for almost half the overall count. However, most of the pauses in the interpretation belong to classes (iv) and (v). Those in class (v), the most obtrusive, are the most frequent of all.

The breakdown for silent and filled pauses is shown below:

(i)   at sentence boundaries: 5 silent pauses, 8 filled pauses;

(ii) at clause or phrase boundaries: 6 silent pauses, 5 filled pauses;

(iii) just after the initial conjunction of a clause: 1 silent pause, 6 filled pauses;

(iv) within a clause: 2 silent pauses, 9 filled pauses;

(v) within a phrase: 11 silent pauses, 15 filled pauses.

The only classes for which silent pauses compete on more or less equal terms with filled pauses are the first two, which suggests that pauses at major syntactic boundaries are in many cases "physiological" and well controlled. The severe imbalance in favour of filled pauses for the remaining three classes indicates that the interpreter often fails to follow Goffman's advice about keeping difficulties backstage when there is no convenient syntactic break to take advantage of (if only as a breathing space). In such cases, pauses tend increasingly to be voiced as "ums" or "ahs" and betray what Goffman calls a "production crisis" (see above, Introduction).

3.2.3. Taking the interpreter's retrospective comments into account

As explained in the initial presentation of the experimental procedure (section 2.1), the present study does not include a description of the methodology for collection of interpreters' retrospective comments.

It is, however, interesting to look briefly at how information collected in this way can complement the data on pause duration and distribution. One obvious consideration is that the interpreter's perception of problems with aspects such as reading notes and coping with difficulties of reformulation can help identify possible causes of any "production crises" which mar the quality of the interpretation. Analysing the interpreter's comments side by side with the evidence of disfluencies in the interpretation can also afford insight into the interpreter's understanding of these difficulties and ability to address them. In addition, it can offer a basis for hypotheses about difficulties which, though

successfully managed, might create a "knock-on" effect in other speech segments.

One way of using the interpreter's comments is to focus on whether they indicate types of difficulty which coincide with frequent and/or prolonged pausing. To give a simple example, about a third of the interpreter's comments in this case focus on language difficulties (as opposed to problems in managing notes and/or in following the speaker's logic). These difficulties are mentioned when the interpreter focuses on phrases such as "è stato [0.43] a [mm 0.51] Bruxelles" (for "was in Brussels"), "vorrei iniziare con [eh 0,43] le mie scuse [-eh 1,02] vorrei scusarmi" (for "I have to give you a very apologetic lecture; I've got to apologize"), "ha una [0.99] un finale positivo" (for "has a happy ending") and "svolgere un [eh 0,47] un ruolo [eh 0,26] sovrannazionale quindi di controllo sovrannazionale" (for "to be in some degree supranational"). The comments indicate various types of language difficulty (doubts about whether Brussels is "Bruxelles"; problems in finding an equivalent for "apologetic" and "apologise", perceived as more grandiloquent than "scuse" and "scusarmi"; hesitation as to whether the English form "happy end[ing]", often used in Italian, is suitable here; the problem of finding an Italian equivalent for "supranational", specific to the discourse field of international relations). In this respect, having the student focus on which difficulties have been associated with comparatively long pauses can help ensure that the language points concerned are appropriately addressed.

Students should understand the importance of recognising potential problems during a consecutive interpretation well before they have to reformulate the speech segments concerned. Though it is counterproductive to focus exclusively on these at the expense of listening, note-taking and reformulation, it is better to think about possible solutions beforehand than simply to stumble across unforeseen surprises when glancing down at the notepad during reformulation. If the interpreter can avoid last-moment hesitation in assessing the possible need for strategic choices such as paraphrases or omissions, s/he can achieve an acceptable trade-off between completeness and a fluent, agreeable presentation.

4.   Conclusion

The methodology described in the previous sections is intended to provide a practical approach to fluency assessment in interpreting. While the methodology should also lend itself to research on fluency in simultaneous interpreting, the interpreter is obviously subject to different constraints in the two modes. Fluency in simultaneous is more subject to the quality of source speech delivery; in consecutive, the interpreter must achieve a good balance of careful

listening and judicious use of notes, with ability to read notes at a glance and speaking skills coming to the fore during reformulation.

This kind of study takes time, and envisaging its use for (self-)assessment during training is probably unrealistic. On the other hand, the kind of software required for research of this kind is readily available and relatively inexpensive. Studies of fluency could thus be undertaken in many different settings. This would make it possible to pool data in relation to a range of variables, offering an interesting point of contact and exchange with mainstream linguistics research. An important proviso in this respect is the need to recognise any limitations in terms of comparability because of methodological differences. This has been clearly seen, for parameters like PTR and MLR, in a number of examples analysed above.

Research on fluency in interpretation, though still in its infancy, is an exciting area of study. It offers considerable interest, not only for purposes of academic debate but also for what should surely be the ultimate goal of interpreting studies – increasingly informed insight into how trainee interpreters can be helped to negotiate the many difficulties of the learning process.

Appendix 1: Source text for the consecutive interpretation

The last time that I was in a simultaneous translation situation, it was in Brussels, at a committee of the European Union, and I heard the translator saying, 'This man is talking too fast and I think he's talking nonsense'. So if I talk too fast or if I use obscure words, please do, as suggested, interrupt in any way and we can have questions afterwards.

I have to give you a very apologetic lecture; I've got to apologize for my country for its diplomatic stupidity, I think, as I shall show, over the last 50 years in relation to Europe.

I think that the story I'm going to tell you has a happy ending, or there's a chance of it having a happy ending, but meanwhile I just have to apologize, in a sense, for what my country and its leaders have done, people of all parties, in relation to European politics over the last 50 years. Now what I'm going to do is to go very fast through the list of dates on the handout I have given you and then I'm going to try and seek explanations for what has happened in Europe. Britain has been an awkward partner in relation to Europe, before and since we became members of the European Economic Community and now the European Union.

The first date I put on my list was 1946, Winston Churchill's speech in Zurich, where he called for a United States of Europe. And then he said, but he meant that just for the Europeans, and Britain of course doesn't really quite know whether it's in Europe or not ; Europe begins at the English Channel, in a large amount of English discourse about politics. We didn't think that we

wanted, or we were not prepared to involve ourselves fully in Europe. In 1948, there was the Brussels Treaty, a defensive treaty, with the Benelux countries and France, and the Council of Europe was set up. The French wanted the Council of Europe to be in some degree supranational, but the British insisted on it being purely international. And so we were reluctant when people were trying to get on with things then.

The British at that stage, I think, had the arrogance to believe that they were still the richest country in Europe, less damaged by the war than any other Western European country, and that the wretched Europeans were trying to import our strength to compensate for their weakness. Now, of course, the tables were fairly rapidly turned. And the rate of economic growth in Western Europe in the course of the 1950's turned Britain from the top nation in Europe, in terms of economic measurements, economic success, into a relatively middle of the road nation, in terms of economic success. We were dropping down the growth league all the time. But we didn't recognize that at the beginning, when we were being difficult.

Appendix 2: Distribution of pauses in the interpretation

Pauses of at least 0.25 sec. are indicated in bold type. Five classes have been identified, according to where they occur:
(i)   *pauses between sentences are shown against the left margin, with no accompanying symbol*;
(ii) *pauses between clauses are indicated with one asterisk* (*);
(iii) *pauses just after the initial conjunction of a clause are indicated with two asterisks* (**);
(iv) *pauses within a clause are indicated with three asterisks* (***);
(v)   *pauses within a phrase are indicated with four asterisks* (****).

> buongiorno
> [- **eh 1,89**] l'ultima volta che ho partecipato a un convegno è stato ****[**0,43**] a ****[**mm 0,51**] Bruxelles *[**0,47**] dove gli interpreti mi hanno detto *[**eh 0,73**] che stavo parlando troppo velocemente e stavo dicendo delle stupidaggini
> [**0,87**] quindi nel caso ciò succedesse anche questa volta vi prego di dirmelo in modo che possiamo comunicare senza molti problemi
> [**0,83**] innanzitutto *[**0,25**] vorrei iniziare con ****[**eh 0,43**] le mie scuse
> [- **eh 1,02**] vorrei scusarmi per il comportamento della del mio staff per la sua stupidità nei confronti dell'Europa e nei vostri confronti
> [**1,57**] la storia che vi racconterò comunque ha una ****[**0,99**] un finale positivo

[eh - 1,68] comunque penso che siano necessarie le [0,18] le mie scuse ancora una volta per quello che ****[eh 0,40]** i miei concittadini hanno fatto nei vostri confronti e all'interno della ******[0,33]** politica dell'Unione Europea

[- eh 1,93] il mio ******[0,55]** discorso *****[eh 0,54]** è diviso in in due parti principalmente vi vorrei *****[0,40]** parlare un po' [eh 0,22] delle date più importanti che riguardano la creazione ******[0,26]** dell'Unione Europea e i rapporti della Gran Bretagna nell'Unione [0,15] con l'Unione Europea ***[0,36]** e poi ****[eh 0,43]** darvi alcune spiegazioni per quanto riguarda ****[0,54]** il ruolo che ha svolto [0,22] il mio stato all'interno dell'Unione Europea

[- eh 2,11] il comportamento della ******[eh 0,62]** Gran Bretagna può *****[eh 0,44]** sembrare un po' strano ***[- eh 2,95]** forse difficile da comprendere da parte degli altri membri dell'Unione Europea soprattutto ***[eh 0,36]** prima che la Gran Bretagna *****[eh 0,37]** diventasse *****[mm 0,66]** membro dell'Unione Europea

[- eh 1.02] la prima data ******[eh 0,59]** importante il primo evento di cui vi voglio parlare è il discorso tenuto da Churchill a Zurigo nel millenovecentoquarantasei ***[1,64]** dove [0,15] Churchill ha ******[- eh 1,13]** parlato *****[eh 0,47]** della situazione dell della Gran Bretagna affermato che ****[eh 0,62]** la Gran Bretagna non era ancora pronta per entrare a far parte dell'Unione Europea

[0,55] la seconda data ******[eh 0,48]** fondamentale per la creazione dell'Unione Europea e per ******[0,29]** il ruolo della Gran Bretagna all'interno dell'Unione è il millenovecentocinquantotto in cui è stato ******[0,29]** firmato il ******[eh 0,26]** Trattato di Bruxelles ***[0,80]** ed è stato creato il ******[eh 0,26]** Consiglio Europeo ***[0,69]** che secondo l'Unione doveva svolgere un ******[eh 0,47]** un ruolo ******[eh 0,26]** sovrannazionale quindi di controllo sovrannazionale

[- eh 1,31] la Gran Bretagna [0,22] si è opposta alla decisione del consiglio e ha ******[0,32]** affermato che ****[eh 0,51]** il Consiglio Europeo dovrebbe *****[0,32]** più che altro avere un ruolo internazionale

[1,53] il ******[mm 0,48]** l'atteggiamento della Gran Bretagna *****[- eh 1,79]** è ******[0,69]** sembrato arrogante ***[- eh - 2,15]** visto che ****[ehmm 1,31]** i cittadini inglesi e il governo ******[eh 0,40]** inglese *****[eh 1,31]** pensava [0,22] che l'Unione Europea aveva bisogno della ******[- eh 1,02]** Gran Bretagna [0,18] per [eh 0,22] aiuti economici ***[eh - 2,01]** per ****[eh 0,66]** risollevare la situazione ******[0,36]** così negativa del ******[eh 0,44]** dell'Europa

[- eh - 2,62] infatti in tutto l'arco *****[eh 0,33]** degli anni cinquanta l'economia ******[0,37]** europea *****[eh 0,41]** non aveva ******[mm 0,74]** mostrato grandi miglioramenti e la situazione era piuttosto negativa

References

Barik H.C. (1973) "Simultaneous interpretation: temporal and quantitative data", *Language and Speech* 16/3, pp. 237-270.

Barik H.C. (1975) "Simultaneous interpretation: qualitative and linguistic data", *Language and Speech* 18/3, pp. 272-297.

Butterworth B. (1980) "Evidence from pauses in speech", in *Language Production*, vol. 1, *Speech and Talk*. Ed. by B. Butterworth, London, Academic Press, pp. 155-176.

Butterworth B. (ed.) (1980) *Language Production*, vol. 1, *Speech and Talk*. London, Academic Press.

Dechert H.W. and Raupach M. (eds.) (1980) *Temporal Variables in Speech*, The Hague, Mouton.

Deese J. (1980) "Pauses, prosody, and the demands of production in language", in *Temporal Variables in Speech*. Ed. by H.W. Dechert and M. Raupach, The Hague, Mouton, pp. 69-84.

Duez D. (1982) "Silent and non-silent pauses in three speech styles", *Language and Speech* 25/1, pp. 11-28.

Enkvist N.E. (ed.) (1982) *Impromptu Speech: a Symposium*, Åbo, Åbo Akademi.

Evandri F. (1998) *La resa interpretativa: problemi di valutazione*, unpublished thesis, Università degli Studi di Bologna, Scuola Superiore di Lingue Moderne per Interpreti e Traduttori.

Garzone G. e Viezzi M. (eds.) (2002) *Interpreting in the 21st Century. Challenges and Opportunities*, Amsterdam/Philadelphia, John Benjamins.

Gile D. (1995) *Regards sur la Recherche en Interprétation de Conférence*, Lille, Presses Universitaires de Lille.

Gile D. (1999) "Variability in the perception of fidelity in simultaneous interpretation", *Hermes* 22, pp. 51-79.

Goffman E. (1981) *Forms of Talk*, Philadelphia, University of Pennsylvania Press.

Goldman-Eisler F. (1958) "The predictability of words in context and the length of pauses in speech", *Language and Speech* 1, pp. 226-231.

Goldman-Eisler F. (1967) "Sequential temporal patterns and cognitive processes in speech", *Language and Speech* 10, pp. 122-132.

Jones R. (1998) *Conference Interpreting Explained*, Manchester, St Jerome Press.

Kopczynski A. (1981) "Deviance in conference interpreting", in *The Mission of the Translator Today and Tomorrow. Proceedings of the 9th World Congress of the International Federation of Translators.*

Ed. by A. Kopczynski, A. Hanftwurcel, E. Karska and L. Rywin, Warsaw, Polska Agencja Interpress, pp. 399-404.

Kopczynski A., Hanftwurcel A., Karska E. and Rywin L. (eds.) (1981) *The Mission of the Translator Today and Tomorrow. Proceedings of the 9th World Congress of the International Federation of Translators*, Warsaw, Polska Agencja Interpress.

Kurz I. (1993) "Conference interpretation: expectations of different user groups", *The Interpreters' Newsletter* 5, pp. 3-16.

Lehtonen J. (1982) "Non-verbal aspects of impromptu speech", in *Impromptu Speech: a Symposium*. Ed. by N.E. Enkvist, Åbo, Åbo Akademi, pp. 33-45.

Mead P. (2002) "Exploring hesitation in consecutive interpreting: an empirical study", in *Interpreting in the 21st Century. Challenges and Opportunities*. Ed. by G. Garzone and M. Viezzi, Amsterdam/ Philadelphia, John Benjamins, pp. 75-84

Onnis L. (1999) *On Processing Language in Real Time: Fluency in Native and Non-Native Speakers*, unpublished thesis, Università degli Studi di Bologna, S.S.Li.M.I.T.

Pöchhacker F. (1993) "On the *science* of interpretation", *The Interpreters' Newsletter* 5, pp. 52-59.

Raupach M. (1980) "Temporal variables in first and second language speech production", in *Temporal Variables in Speech*. Ed. by H.W. Dechert and M. Raupach, The Hague, Mouton, pp. 263-70.

Towell R., Hawkins R. and Bazergui N. (1996) "The development of fluency in advanced learners of French", *Applied Linguistics* 17/1, pp. 84-115.

CORPUS LINGUISTICS METHODS
IN INTERPRETING RESEARCH: A CASE STUDY

Šárka Timarová
Charles University, Prague

1.   Introduction

Among the central issues in interpreting research methodology is the question of how to approach and analyse experimentally collected data. Computer-aided analysis (cf. Pöchhacker 2004: 199) and corpus-linguistic methods in particular are one possible path (Pöchhacker 2004: 202). The use of corpus managers for analysis of large data files has been proposed more than once in translation studies by Baker who also published several empirical studies with examples of such analyses (e.g. Baker 1993, 1995, 2000). A similar proposal for interpreting studies was made in Shlesinger (1998). In this paper, I would like to describe some contributions and implications of the corpus-linguistic methods for interpreting research, and show two detailed step-by-step analyses to encourage more ideas.

2.   Corpus Managers

The first and most obvious advantage of corpus managers (CM), the basic software tool, is their speed and capacity to process large amounts of data. Many previously laborious steps in data analysis can be done as, literally, one-click operations on a large number of data files. CMs also have many in-built functions. Some of those that only require pressing one button include list generation of all words found in the files, in alphabetical or frequency order, basic statistics on the total number of words (*tokens*), number of different words (*types*), number of sentences, average number of sentences per text, average number of words per sentence, number of sentences with 3, 4, 5 ... words. When searching for a particular item (a word, phrase ...), functions such as concordance (displays the item in a context of e.g. 5 preceding and 5 following words) or plotter (shows the distribution of the item throughout the text) are of great assistance. All statistics and searches can be done on a very large amount of individual files at the same time (i.e. all participant outputs). This list, already quite long, still does not cover all basic functions.

   Perhaps the major challenge for the use of CMs in interpreting research is the need for availability of the data in an electronic format. This requires that the researcher still undertake a rather laborious transcription of the audio

recordings. At the moment, there are no reliable speech-to-text converting tools for many languages. Also the transcription requires that the researcher stop and think beforehand what exactly she wishes to investigate. Common CMs were primarily developed for processing written texts. This implies that they are not able to "read" the text in other than orthographic form. The transcript of audio output cannot therefore include any extralinguistic features (marks), such as intonation rise or hesitation within a word (e.g. *presi↑dent* would not be recognised as *president*), and a sentence would not be recognised if it does not start with a capital letter and end with full stop. Similarly, any unfinished words will not be recognised as such, but rather as words in their own right (delimited by a space on each end of the letter string, e.g. *I would like to co*[me] *go home*., where the unfinished *co* for *come* would be recognised as a word per se).

CMs can indeed be of great help for quantitative analysis, but one must bear in mind that they are only tools. The possession of an oven and a cookery book does not mean one has a meal, and having the most advanced text editor still means one has to write all papers oneself. Similarly, even with a CM, the researcher must have a very clear idea of what she wants to look for and how to look for it. In the following section, I will describe two analyses with emphasis on all major decisions that had to be made throughout the process to arrive at the desired result.

3.   Sample Analyses

Participants
There were 18 participants: interpreting students who had completed their interpreting training and graduates with a maximum of 3 years of professional experience.

Materials
Two genuine recordings of conference speeches in English were used as source texts. Interpreting was recorded on common audio cassettes. A standard MS Office package was used for transcriptions (MS Word) and partial data analysis (MS Excel). WordSmith Tools[1], a corpus manager, was used for data analysis.

---

1   The corpus manager employed in these analyses was WordSmith Tools published by Oxford University Press. Very helpful tutorials and support materials are available at the author's web pages. Mike Scott's webpage can be found at http://www.liv.ac.uk/~ms2928/wordsmith/screenshots/index.htm.

Procedure
Each participant interpreted two source texts: one consecutively and one simultaneously, from English into Czech (C to A). The output was 36 recordings divided into four groups according to text and mode (text 1 consecutively, text 1 simultaneously, text 2 consecutively, text 2 simultaneously).

3.1. Analysis 1: Text Length

Rationale
As a first step in analysing differences between CI and SI, I decided to measure the length. In interpreting research, length is measured either in terms of words or syllables (word count is more frequent, but some authors have serious reservations, cf. Čeňková, 1988:101-102). As English words are generally shorter than Czech words, and as there are e.g. no articles in Czech, it seemed that a mere comparison of the ST and TT number of words would not be informative. Therefore, I decided to take both counts, words and syllables.

Procedure
The first decision had to be made at the stage of transcription. As I decided early on to use a CM for analysis, it was obvious I would transcribe the texts orthographically. For purposes of measuring the length of the output, I decided to include in the transcription everything the interpreters said, including unfinished words. For purposes of the syllable count, I also had to transcribe some abbreviations (such as *USA*) as pronounced so they might be recognised as three syllables (*u es a*). Transcribed TTs (30,000 words, over 60 printed pages!) were then uploaded into the CM. One click produced an overview of the number of words for each TT and a total for a group of TTs (grouped according to text and mode). This operation took about 10 seconds. Counting the syllables was slightly more difficult, as the CMs are not able to recognise syllables. The decision that was made[2] is a good example of how to come up with a procedure which the tool is suited for. In Czech, all syllables are centered around a vowel, with only two relatively infrequent exceptions of diphthongs. Hence, I asked the CM to find all instances of A, E, I, O, U, etc., regardless of what came before or after them (whether they were at the beginning of a word, at the end of it, preceded/followed by other letters). The total number of instances found was the desired number of syllables (I still had to discount the diphthongs, using the same method, this time looking for AU and OU combinations). The most difficult part was coming up with the procedure: the search and count itself was

---

2   For this idea I am indebted to Mirek Pošta, a colleague and a corpus-linguistics enthusiast.

again a matter of several seconds. Of course, this method will not work for every language. I am just trying to illustrate how the functions of a CM can be used, and how to adapt a research question into a workable procedure.

3.2.   Analysis 2: Lexical Density

Rationale
Lexical density is one of the key quantitative corpus parameters (Stubbs, 2002:39). The parameter is based on the fact that languages are composed of content words which are the primary carriers of meaning (nouns, adjectives, verbs, etc.) and function words (auxiliary verbs, pronouns, conjunctions, etc.). Lexical density is calculated as a ratio of the number of content words to the total number of words in a text and is expressed as a percentage, or
lexical density = 100 x number of content words/ total number of words
Lexical density is known to be higher in written texts than in spoken texts. Within the domain of spoken text, Stubbs (1996) found differences in lexical density between texts delivered in an environment with or without a direct contact with the listener. Genres where there is no feedback from the audience, such as answering machine messages or radio commentaries have a higher lexical density than genres where there is such feedback, such as public speeches or radio discussions (Stubbs, 1996:74). This raises an interesting question as to whether there would be a difference in lexical density between consecutive interpreting and simultaneous interpreting output: the consecutive interpreting environment allows for feedback and contact between the interpreter and her audience, while simultaneous interpreting does not. The following analysis describes a procedure for answering this question using a CM.

Procedure
The same small corpus of 36 samples was used. First of all, the transcriptions needed to be adjusted from the previous analysis (*u es a* back to *USA* to be counted as one word, etc.). The 36 files were uploaded to the CM and using the word list function a list of all different words (*types*) was obtained. The word list function produces a list of all words found at least once in at least one file. For each file, it will show how many times a given word appears in the file. The samples had a total of 30,000 words (*tokens* in CM terminology), but because many of them appeared more than once, there were only 3967 different words (*types*). Hence the CM reduced the total number of words the researcher needed to process by a factor of 7.5. The next step was to separate function and content words: I decided to isolate the function words, as their number is much lower than the number of content words. This step had to be done manually by going

through the list of 3967 words. By deleting the "unwanted" content words from the list, the resulting product was a list containing function words amounting to only 447 items. The list was exported into MS Excel and converted to a text file with individual words separated by a comma and a space. The result was a small text file composed solely of function words, which could be included among the "normal" files and serve as a reference file. The text file was uploaded to the CM and a new word list was generated. Clearly, this time the word list contained only function words, as there were no content words in this file.

The 36 tested files were divided into 4 groups according to mode (simultaneous, consecutive) and text (text 1, text 2). The CM generated a word list for each group, and the four word lists together with the function word list were compared: the CM produced a combined overview of all words from the corpus and their frequencies in each of the five word lists. The overview was exported again to MS Excel and ordered according to the function word list. This produces a list where the first 447 lines contained frequencies of function words, and the remaining 3520 lines with content word statistics were deleted. A total number of function words per group was calculated by adding up all frequencies, and slotted into the modified formula for lexical density, where the number of content words was expressed as the total number of words minus number of function words. As a result, four scores of lexical density (one for each group) were obtained.

While the above description may sound somewhat complicated, the actual procedure is rather straightforward. As in analysis 1, the important step is the operationalisation of the research question.

4.   Conclusion

The aim of this paper was to provide some practical examples of the use of corpus linguistics methods and its tools in interpreting research, and hopefully to encourage researchers to explore the possibilities corpus managers have to offer for data processing. While primarily designed for quantitative research, corpus managers can help with some aspects of qualitative research as well. It is my belief that corpus linguistics methodology offers valid tools for interpreting research.

References

Baker M. (1993) "Corpus Linguistics and Translation Studies. Implications and Applications", in *Text and Technology*. Ed. by M. Baker, F. Gill

and E. Tognini-Bonelli, Amsterdam-Philadelphia, Benjamins, pp. 233-250.

Baker M. (1995) "Corpora in Translation Studies: An Overview and Some Suggestions for Future Research", *Target* 7: 2, pp. 223-243.

Baker M. (2000) "Towards a Methodology for Investigating the Style of a Literary Translator", *Target* 12: 2, pp. 241-366.

Čeňková I. (1988) *Teoretické aspekty simultánního tlumočení* [Theoretical Aspects of Simultaneous Interpreting], Praha, Univerzita Karlova.

Pöchhacker F. (2004) *Introducing Interpreting Stud*ies, London and New York, Routledge.

Shlesinger M. (1998) "Corpus-based Interpreting Studies as an Offshot of Corpus-based Translation Studies", *Meta* XLIII, 4, pp. 486-493.

Stubbs M. (1996) *Text and Corpus Analysis*, Oxford, Blackwell.

Stubbs M. (2002) *Words and Phrases. Corpus Studies of Lexical Semantics*, Oxford, Blackwell.

FROM VOICE TO GESTURE: METHODOLOGICAL ASPECTS IN THE COMPILATION OF A TERMINOLOGICAL, TRI-LINGUAL, MULTIMODAL DICTIONARY

Cynthia Jane Kellett Bidoli
SSLMIT, University of Trieste

1.   Introduction

The First Forlì Conference on Interpreting Studies held in November 2000, highlighted the growing importance of community interpreting within the interpreting profession today (Garzone and Viezzi 2002: 296). Sign language interpreting falls within this context, and over the past ten years has gained greater recognition in Italy mirrored in research articles and a growing number of educational and training initiatives offered by both state and privately run institutions (see Amorini *et al*. 2000, Cameracanna and Franchi 1997, Carli *et al*. 2000, Gran and Kellett Bidoli 2000, 2001, Kellett Bidoli 2001, 2002, 2004a, 2005, forthcoming a, b, Stocchero 1995, Woll and Porcari Li Destri 1998). An Italian Sign Language (*Lingua dei Segni Italiana* - LIS[1]) course was introduced at the Advanced School of Modern Languages for Interpreters and Translators (SSLMIT) of the University of Trieste in 1998 generating curiosity and enthusiastic participation among students, as well as several interesting graduate dissertations.

Over the past two years, a seemingly unrelated investigation has been conducted by several research units throughout Italy, within a MIUR COFIN national project entitled *Intercultural Discourse in Domain-specific English* coordinated by Professor M. Gotti[2], into how and to what extent the English language influences cultural and linguistic communication in contact with Italian. Among the research groups, Trieste has been represented within the University of Turin unit investigating *Intercultural Practices and Strategies of Textual Recasting* to verify whether the production/reception of written and oral English discourse within a number of different domains leads to a propensity for cultural and linguistic intrusion from English into Italian. Italian society includes a particular 'community of practice' within its confines, that of the

---

1   Although LIS stands for *Lingua Italiana dei Segni* the Italian Deaf prefer to call it *Lingua dei Segni Italiana*. This is because the former version could imply that one is referring to a signed version of Italian rather than a sign language with its own rules of grammar adopted in a specific geographical area.

2   See: http:/www.unibg.it/cerlis/progetti.htm

Italian Deaf[3] community, which was targeted by the Turin unit as a very particular area of interlinguistic/cultural contact to study.

Contact with the English-speaking world within the Italian Deaf community almost exclusively depends on written Italian sources: translated books and articles, subtitled films, Italian websites on British or American issues and press reports from English-speaking countries found in newspapers or on TV news which at set times of the day are delivered in simultaneous sign language (translated from the Italian bulletin) during brief news broadcasts on some TV channels (Kellett Bidoli 2004 a: 129). But direct contact between the Italian Deaf and English may occur on the Internet, during English language lessons at school or university (see Ochse 2001, 2004), during study abroad on cultural exchanges (Socrates, Erasmus or Fulbright Scholarships) or at public conferences on deaf issues in the presence of English native speakers. If necessary the Deaf may resort to help through specialized teaching assistance and/or the services of professional sign-language interpreters in all those instances of direct intercultural and interlinguistic interaction with the English-speaking world.

Until recently, the nature and extent of cross-cultural encounters between English and the Italian Deaf signing community had not been investigated. To this purpose a survey was conducted in 2003 among professional Italian sign-language interpreters to determine the extent of English to LIS interpretation in Italy and discover which genres are commonly involved in order to better understand the market requirements of this specialized form of oral translation (Kellett Bidoli 2005). Data analysis revealed that interpreters with an active knowledge of English, who could if necessary mediate from English to LIS, are more numerous than expected, but interpretation is normally filtered through Italian; the source language (English) passes through Italian and is thus relayed from an aural/oral mode through headphones to the LIS interpreter who transfers the received message into a gestural/visual mode for the Deaf. Several genres emerged from the survey, the most common not within the context of community interpreting as might have been expected, but within conference interpreting in which a number of specific specialist fields were identified, and in particular the field of linguistics (conferences on various linguistic aspects of sign language).

After this first stage of investigation, four oral speeches in English (delivered by American native speakers) were subsequently selected, in the form of 115

---

3   It is an accepted convention in the literature to use "deaf" (with a lowercase "d") to refer to the audiological condition, while "Deaf" is used to refer to those deaf people who share a sign language and distinct cultural values.

minutes of video recordings taken in authentic conference settings[4]. The small corpus of speeches was composed of linguistics related topics containing 12,616 tokens of which there were 3,075 types. The speakers were video-recorded in small insets and a wide screen view of simultaneous interpretation into LIS provided a combination of multimodal parallel visual, oral/aural and gestural elements to analyse. The original video recordings in VHS were transformed into a digital corpus for electronic analysis of intercultural and interlinguistic features. Parallel corpora resulted in the form of:

– *a written transcription in English of the original spoken discourses;*
– *transcribed glosses of the signs in LIS;*
– *a written 'interpreted' version in English of the signed corpus;*
– *a written 'interpreted' version in Italian of the signed corpus.*

Detailed, contrastive, microtextual analysis was undertaken by aligning the parallel corpora to unveil intercultural and linguistic aspects of textual recasting during the mediation process from English to LIS. Alignment of the English and LIS transcriptions revealed evidence of disparity in the form of omissions and additions of information (from lexical items to whole chunks) leading to occasional instances of intercultural communicative failure through semantic misrepresentation or distortion. Detailed comparison of segments at microtextual level focussed on: word order asymmetries to detect syntactic anomalies; grammatical textual cohesion devices such as temporal succession, tense use and reference; substitution; intrusions; as well as lexical and cultural features of interest (see Kellett Bidoli forthcoming a and b). A few instances of cultural and linguistic intrusion from English were found but on the whole clear evidence emerged of awareness by the English-LIS interpreters of the need for adjustment during the mediation process to the specific linguistic and cultural traits of the target language.

The LIS corpus (composed of sign language glosses) was checked with the assistance of both a professional LIS interpreter and a deaf teacher of LIS. Several instances were found of unclear, ambiguous signing or even omission of technical phrases and lexical items related to linguistics. The perplexity and doubt experienced by the deaf expert made us wonder how much of the original

---

4   W.C. Stokoe, a paper on the evolution of sign language, presented at the First National Conference on Sign Language, *Studi, esperienze e ricerche sulla lingua dei segni in Italia*, ENS, Trieste 13-15 September 1995, published in Italian in Caselli and Corazza 1997; W.P. Isham, "Research on Interpreting with Signed Languages", C.J. Patrie, "Sequencing Instructional Materials in Interpreter Education", and B. Moser-Mercer, "The Acquisition of Interpreting Skills", all three papers presented at the International Conference "*Meeting of Sign and Voice*", University of Trieste, Trieste 12-13 December 1997, published in Gran and Kellett Bidoli 2000.

message reaches the Deaf end-user at a conference and sparked the idea that some form of didactic support could be developed for trainee interpreters to enhance their signing ability in this specialist field as well as bridge the gap between English and LIS.

During electronic analysis of the corpus, word counts, word frequencies and concordances were run of both the English and LIS (glosses) to detect lexical items related to the field in question and to determine language use and translation strategies in context. It was soon realized that this data could also be turned to good advantage to enable the compilation of the didactic support we were looking for, or rather, the compilation of a multimodal terminological data bank or glossary to be used by students. Hence, a pilot version of a trilingual terminological glossary of linguistics in English, Italian and LIS was produced in electronic format on CD-ROM to be used as a teaching aid targeted at interpreter trainees of LIS (Kellett Bidoli 2004b).

10 lexical items were initially selected to produce over 60 entries (including synonyms and cross-references) across the three languages, each accompanied by phonetic transcription in English, a definition, examples of usage in context selected from the concordances, linguistic comments and easily accessible images of signs illustrated singly or in signed sentences in context. This paper briefly discusses existing LIS dictionaries available to interpreters before passing on to methodological aspects encountered during the glossary compilation.

2.   Italian Sign Language dictionaries

Traditional dictionaries are today increasingly based on large and diverse corpora of written and spoken text as their primary data source, providing lexicographers with a limitless tool to compile up-to-date core vocabulary as well as collecting less frequently used words. Because of the three dimensional, kinetic nature of sign languages and the frequent lack of word-to-sign equivalence, there are difficulties involved in representing, transcribing or simply illustrating them 'on paper', compared to oral languages that can be represented graphically through conventional alphabets more easily. Dictionary compilation of signed languages is extremely arduous and intricate as they are composed of individual signs that convey meaning predominantly through arm and hand movements but also through simultaneous non-manual features such as posture, eye movement, gaze, head, lip and shoulder movements and much varied facial expression. Each sign is distinguished from another through four universally recognized parameters: handshape, palm orientation, movement and location. Generally, in traditional sign language paper-based dictionaries, each entry consists in a rudimentary sketch or photograph of a signer waist up, with

arrows indicating movement, and transcription graphics chosen from one of the numerous notation systems that have been devised added below. A gloss of the meaning or nearest equivalent in spoken language is offered, but to the untrained eye, the whole resembles a mix of indecipherable Roman letters, numbers and abstract symbols. The average dimension of each static illustration is approximately 4x4 cm, which leads to a serious limitation in the number of signs presented per page, which is further reflected in the overall limited volume of entries offered in most printed sign language dictionaries. To further complicate matters, just as in spoken languages, compounds exist in sign languages, composed of more than one sign representing a single referent or concept. For example, in LIS the term 'intelligent' is composed of the signs 'HEAD+YES' in quick succession, requiring a more complex graphic representation because the four parameters of each of these signs differ. The parameters in some compounds may differ so much that arm and hand positions may have to be duplicated or triplicated in staggered stages in the same sketch (Radutzky 1992: 33). Therefore, because of the combined difficulties of graphic representation and space, the average size of sign language paper-based dictionaries is restricted and hence, of generic nature (Angelini *et al.* 1991, Magarotto 1995, Radutzky 1992, Romeo 1991). They are certainly of considerable use to students learning basic sign language, but of little help to the interpreter grappling with conference papers on topics such as: *The role of bilingualism* (words and signs) *in the teaching of mathematics to deaf school children* or *Speech therapy as an aid to cognitive development in deaf infants*.

Specialist dictionaries and glossaries in LIS are lacking. The reason for this is that sign language evolves at home and in clubs where non-technical everyday 'vocabulary' is used to discuss daily events. At work, deaf people find themselves isolated in a hearing environment and are thus obliged to communicate through speech and lip-reading. They may have access to specialized terminology for their job but they rarely need to use it when signing outside the workplace. Therefore, signs do not evolve and spread rapidly through the Deaf community to describe technical language during signed 'conversation'. Indeed, standard signs may not exist in LIS for numerous technical and complex terms found in spoken Italian or English. Interpreters may be hard-pushed to find an adequate solution to express an unfamiliar Italian term by joining together existing signs or inventing a new one. Newly coined signs will only catch on and be repeated in future if they are transparent enough to convey meaning to the Deaf and if frequently used by other interpreters. Often 'technical' signs differ in their configuration from one interpreter to another causing perplexity among the Deaf, as was discovered on analysing the corpus of conference speeches. Only one dictionary of specialized nature is known to the author containing religious lexis (Puricelli *et al.* 1993).

Today the problems of graphic representation and space can be overcome thanks to computer technology and the widespread adoption of alternative media such as CD-ROMs and DVDs to provide dynamic images of signs together with superimposed written information or hypertextual links. Electronic dictionaries of this kind have started to make an appearance in Italy, such as *Dizionario mimico gestuale* (Pignotti 1997) and *Dizionario Italiano/LIS* (Piccola Cooperativa Sociale "Alba" 2003).

Advantages in using an electronic format in this particular tri-lingual spoken signed combination are evident:

– the possibility to include dynamic illustration of sign language terminology and its exemplification in context as opposed to its static representation in paper-based dictionaries;
– the speed of instant access through hyperlinks to translation equivalents and related terms, versus turning over numerous printed pages;
– unlimited space to provide definitions and examples which are normally lacking in multilingual paper-based specialized dictionaries (Bowker 2003: 159): often only headwords and their multilingual equivalents are listed;
– graphics can be varied and made more interesting through the use of colour, insets and numerous creative visual as well as acoustic devices.

Video and/or CD-ROMs are an ideal, innovative media for conveying sign languages or any didactic support materials for the training of sign language interpreters.

3.   Tri-lingual multimodal electronic glossary compilation

The basic starting point of the English-Italian-LIS glossary was a breakdown of the 12,616 tokens obtained from the linguistics corpus. As the corpus is a very small one it was not too problematic to visually scan the list to eliminate the most frequently used words in English which, as expected, were: the definite and indefinite articles, and, to, that, is, pronouns and prepositions; of no use for the purpose of this particular glossary.

Different styles of interpreter signing were apparent during observation of the videos and highlighted by the 'word count' of the LIS glosses. In one interpretation the use of the '*c'è*' sign (there is), a common LIS sign, resulted as being the most frequently used. It occurred disproportionately 113 times compared to 55, 28, and 19 in the other three interpretations. Also '*ma*' (but) was found to be the third most frequent sign used by an interpreter (28 occurrences compared to 14,13 and 13). The mouthed "*pà, pà, pà*" and accompanying hand gesture (parallel divided vertical palms) - which is a deictic marker that changes meaning according to context, often meaning 'thus', 'done this way', 'set out this way', 'so, so, so' - was the most frequently used sign in

the rendering of the Patrie speech with 51 occurrences compared to 3 in the Stokoe discourse and none in the others. However, in contrast to the other texts the Patrie text, on sequencing of interpreter teaching materials, did call for a means to transfer the idea of '*things*' (texts, exercises, skills etc.) being presented in temporal succession; not such a necessary requirement in the other interpretations, for example:

<u>"*pà, pà, pà*"</u>
MATERIAL ACTIVITY TEACHER MUST ORDER IMPORTANT.
Interpreted as: A teacher has to sequence teaching materials and activities in an orderly manner.

However, the aim of this particular glossary is to offer a selection of English technical entries within the subject field of linguistics (from the subfields of sign language and interpretation) in the conference setting and render them in Italian and LIS. Thus, almost 300 lemmata were accordingly selected from the 3,075 types and concordances run for each using *Wordsmith Tools*. It was decided to produce a pilot version of the glossary in electronic format on CD-ROM to test it before compiling a full version (Kellett Bidoli 2004b). Only 10 English lemmata out of the 300 were selected that are commonly used in discourses on spoken and/or signed interpretation and language:

chunk	**fingerspell**
classifier	**gesture**
consecutive	**interpret**
décalage	**language**
field	**negation**

which were extended to 27 items through the addition of synonyms and related terms:

chunk (verb) - chunk (noun).
classifier
consecutive (noun) - consecutive (adjective), consecutive interpretation, consecutive interpreting, consecutively.
décalage - ear-voice span, lag, lag time.
field
fingerspell - fingerspelling.
gesture (noun) - gesture (verb).
interpret - interpretation, interpreter, interpreter education, interpreter training, interpreting, interpreting booth, booth.

language
negation

The glossary was initially compiled as an 18-page Microsoft Word document before transfer into hypertextual format. An application in HTML was chosen to use a 'cross-browser' approach that permits access to the glossary through a wide choice of browsers and operative systems. Once the document in Word was transferred into hypertext the contents were split up into 87 separate HTML pages, all generated from the original 10 lemmata apart from 6 pages including the title, indices etc.

To view the CD-ROM one starts from a main menu by clicking on one of the following options:

Premessa (front matter or foreword in Italian, containing background information on the research project and its aims, followed by a bibliography)
Indice dei termini in inglese (index of English terms)
Indice dei termini in italiano (index of Italian terms)
Indice delle glosse in LIS (index of LIS glosses).

A semasiological approach, which seems to be the dominant ordering in thesauri and dictionaries containing the specialized terminology of language for special purposes (LSP) was chosen, leading to an alphabetical ordering of the three separate indices. However, during initial compilation in Word format, headwords and corresponding articles in each of the three languages were ordered vertically and alphabetically irrespective of language. It was only through colour coding that the languages could be quickly, visually identified during compilation. On the pilot CD-ROM version after each headword, the corresponding 'clickable' equivalents in Italian and LIS are also colour coded. In the following example there are three monochrome articles for the lemma **fingerspell.** Where the word IMMAGINE (image) is located, the trainee interpreter can find an icon on which to click in order to obtain a dynamic image of the correctly signed lexical item, or a fully signed version of the example provided below the definition, in order to learn correct word order sequences and collocations that more often than not differ from English or Italian:

****

INGLESE
fingerspell verb [ˈfɪŋgə spel-fɪŋgər spel] eseguire in dattilologia
eseguire in dattilologia

The use of the manual alphabet to spell out unfamiliar proper names and terms.

They fingerspelled for half an hour to practise word recognition, which is the single most difficult thing for sign language learners.

Note: Past tense and past participle: 'fingerspelled' or 'fingerspelt' (mainly in British English).

****

ITALIANO

eseguire in dattilologia sintagma verbale fingerspell eseguire in dattilologia

Utilizzare l'alfabeto manuale per indicare l'ortografia di nomi propri o di termini non familiari.

Una delle cose più difficili da imparare per gli studenti della lingua dei segni è eseguire esercizi di dattilologia.

****

LIS

eseguire in dattilologia eseguire in dattilologia fingerspell

Utilizzare l'alfabeto manuale per indicare l'ortografia di nomi propri o di termini non familiari. IMMAGINE

Una delle cose più difficili da imparare per gli studenti della lingua dei segni è eseguire esercizi di dattilologia. IMMAGINE

Nota: Nella LIS per segnare ESEGUIRE si muove il braccio per arrivare allo spazio del segno DATTILOLOGIA.

****

Each article is headed by a main lemma (originally selected from the English corpus) followed by its syntactic category in English and Italian but not in LIS, as signs are often not the equivalent of single words but may often convey concepts expressed by whole phrases in spoken language. Abbreviations were avoided given the space a CD-ROM offers. Thus all abbreviations are made explicit with no need for an explanatory list.

Phonetic data is provided through phonetic transcription of the English lemmata in the International Phonetic Alphabet, first in British English followed by any American variant on the right where applicable. Computers permit the insertion of sound clips of the correctly pronounced lexical items, which can be accessed by the user through a click of the mouse on the phonetic spelling. Though not included in the pilot sample of the glossary, it is planned to make

sound recordings available in British and American English in the full version. Phonetic information on the Italian was deemed superfluous as the end users of the glossary are intended to be native speakers of Italian: Italian, hearing, interpreter trainees.

Following on from the phonetic data are bi-lingual translation equivalents of the headword, which are distinguished by colour. Though there may seem to be no difference between the Italian and LIS equivalents, by clicking on one or the other, bi-directional access can be obtained to separate articles which have the same definitions and examples but different notes, and the addition of imagery in the case of the LIS articles. The glossary is tri-directional, in that starting from an index or entry article in any of the three languages one can access information in the other two.

Next there follows a sentential definition of the headword to conceptually describe its individual denotational meaning in the context of linguistics, in as short, simple and unambiguous manner as possible.

Exemplification in context was obtained from concordances run to show all occurrences in the corpus for each lexical item as illustrated in the extract of concordances for **field** below:

rpreters but really its, it tries to be very comprehensive of the sign language **field** and it is pretty comprehensive. So I recommend that you get it. As I menti

e of scope that is an overview of research from fields other than my particular **field**. I also used um, availability in some of the choices I made, papers that I

've cited are either from the spoken language field, or not in the interpreting **field**, linguistics or psychology. So we're really very behind in research, there

ow what could we do to use this new language? Two important developments in our **field** of interest are, first recognizing that all children communicate gesturall

search that there is little to begin with. We are all kind of beginners in this **field** of research of interpreting and the quality is uneven. However, I think th

semiotics and writes well about it. It's of course the scientific discipline or **field** of study that considers signs generally, most generally. Thus, in semiotic.

The concordances provided a wide choice of examples, often several pages long, which led to the difficult task of selecting only one or two of the most representative and interesting.

Cross-reference entries (synonyms, related terms, compounds and derivations of interest) are included as separate text pages, sometimes without a complete entry, but guide the user to a headword with a complete one. For example:

****

INGLESE
consecutive interpretation noun/ uncountable
[k↔n sekʊtɪv ɪn t prɪteΣ↔n]
interpretazione consecutiva interpretazione consecutiva
See: consecutive (noun)

****

Concordances revealed patterns of language usage, which led to the compilation of notes providing information on spelling variants and linguistic or semantic features of interest to interpreters. For example in the case of the head word **language** there are four separate observations noted:

Note:
- In Italian there are two separate terms for the above definitions: 1) *lingua* and 2) *linguaggio*. Also in LIS there two different signs;
- frequently related pre-modifiers are: *A -, American sign -, B -, British sign -, conventional -, first -, gestural -, Italian sign -, local -, political -, real -, second -, sign -, signed -, source -, speech-based -, spoken -, target -, working -*;
- frequently related verbs: *to acquire, to adopt, to check, to define, to find out about, to go between two, to hear, to interpret, to listen to, to produce, to see, to sign, to shadow, to speak, to talk about, to test, to use, to work into, to work from, to write about*;
- related expressions: *language acquisition, language competence, language experience, language field, language interpreters*.

The dynamic images provided by a deaf signer (a teacher of LIS) and inserted in the LIS text frames, were filmed with a digital camera in a naturally illuminated classroom. He wore dark cloths to contrast with the white wall behind him to highlight his signing. For each entry in LIS he signed the headword and then the full example. Each headword and example were

numbered and during filming separated by a rudimentary clapper board (small blackboard and chalk) to enable the spliced segments to be correctly positioned in the glossary. The 'clapper board' was invaluable as often several takes were made for each item to be filmed. Problems included false starts, loss of memory while signing long or more complex examples, signs that were too wide, high up or low down that exited the film frame and the author's elbow or arm that occasionally invaded the screen while juggling with the tools of the trade between each take: blackboard, chalk, duster and a large font size list of the terms and examples to be filmed. This list, contrary to plans, could not be used as a prompt during the filming, because it diverted the signer's gaze laterally instead of straight at the camera. Hence the signer's mnemonic capacity was occasionally stretched to the limit.

The methodology described above is essentially straightforward and simple and can be applied to any terminological dictionary including a signed language once a subject field has been identified and the lexis collected. Multilingual transcription of the original corpus was by far the most time consuming and arduous phase of the research (Kellett Bidoli forthcoming a).

4.   Conclusion

Interpreters and translators alike rely on general and terminological dictionaries (specialist glossaries) for their work, and trainees even more so, lacking the years of experience that build up an expert's individual, subconsciously assimilated, lexical and world knowledge. Interpreters need so-called active or production-oriented dictionaries, those that provide translation equivalents in context in languages other than one's mother tongue.

Dictionaries other than general-purpose dictionaries in the Italian-LIS language combination are lacking in the area of LSP. The technical problems involved in the compilation of paper-based LIS dictionaries, which must convey linguistic information about a gestural three-dimensional language in two-dimensional format, are being overcome by the advent of the electronic dictionary through the application of sophisticated computer technology which has opened the way to a revolutionary means of processing and representing multimodal data. Any such electronic audio-visual support in this direction would be an invaluable aid for anyone involved in sign language interpreter training, not only from Italian to LIS, the common interpreting directionality in Italy, but also from English to LIS in view of the unceasing spread of English as an international language of communication or in other spoken language combinations. Trainee interpreters need stimulating material abreast of our technological times. When in doubt as to the semantic meaning of a word or sign, when in need of correct pronunciation or when tormented by polysemic

equivalence, they need swift access to dictionaries based on the real language of native speakers and signers.

Computer corpora provide a huge quantity of written or spoken material for lexicographers to process and interpreters to use, but also an incredible speed with which to select and sort lexical items compared to the card archives of the past. Corpus data once analysed and selected for the compilation of dictionaries, thesauri and glossaries can now be stored and presented on line, on CD-ROM or DVD, which provide limitless space and technological wizardry in the form of hyperlinks, acoustic, and visual features.

Corpus analysis of interpreted discourse in specialist subject fields is expected to yield a wealth of intercultural/linguistic features, that commonly emerge during interpretation from English to LIS. It is hoped that such findings will lead to the future compilation of invaluable, terminological, didactic tools in multimodal format for the training of future interpreters of sign language.

References

Amorini G., Corazzi M.C. and Eletto G.M. (2000) "Idiomatismi e metafore nell'interpretazione in e dalla LIS", in *Viaggio nella città invisibile. Atti del 2° Convegno nazionale sulla Lingua Italiana dei Segni*. Ed. by C. Bagnara, P. Chiappino, M.P. Conte and M. Ott, Pisa, Edizioni del Cerro, pp. 181-194.

Angelini N., Borgioli R., Folchi A. and Mastromatteo M. (1991) *I primi 400 segni. Piccolo dizionario della Lingua Italiana dei Segni per comunicare con i sordi*, Firenze, La Nuova Italia.

Bowker L. (2003) "Specialized Lexicography and Specialized Dictionaries" in *A Practical Guide to Lexicography*. Ed. by P. van Sterkenburg, Amsterdam/Philadelphia, John Benjamins, pp. 154-164.

Cameracanna E. and Franchi M.L. (1997) "Considerazioni sull'interpretariato al termine di un corso per interpreti di LIS", in *LIS. Studi esperienze e ricerche sulla Lingua dei Segni in Italia. Atti del 1° Convegno Nazionale sulla Lingua dei Segni*. Ed. by M.C. Caselli and S. Corazza, Tirrenia (Pisa): Edizioni del Cerro, pp. 281-285.

Carli F., Folchi A. and. Zanchetti R. (2000) "Processi di interpretazione dei proverbi", in *Viaggio nella città invisibile. Atti del 2° Convegno nazionale sulla Lingua Italiana dei Segni*. Ed. by C. Bagnara, P. Chiappino, M.P. Conte and M. Ott, Pisa, Edizioni del Cerro, pp. 67-71.

Caselli M.C. and Corazza S. (eds.) (1997) *LIS. Studi esperienze e ricerche sulla Lingua dei Segni in Italia. Atti del 1° Convegno Nazionale sulla Lingua dei Segni*, Tirrenia (Pisa): Edizioni del Cerro.

Garzone G. and Viezzi M. (eds.) (2002) *Interpreting in the 21st Century. Challenges and Opportunities*, Selected papers from the 1st Forlì Conference on Interpreting Studies, 9-11 November 2000, Amsterdam/Philadelphia, John Benjamins.

Gran L. and Kellett Bidoli C.J. (eds.) (2000) *Signed-Language Interpretation and Training: Theoretical and Practical Aspects*, Proceedings of the International Conference "Meeting of Sign and Voice" in Trieste, December 12-13, 1997*, Trieste, Edizioni Università di Trieste.

Gran L. and Kellett Bidoli C.J. (2001): "Sign language interpretation: a newcomer to Italian university studies", in *"Languaging" in and across Human Groups, Perspectives on Difference and Asymmetry.* Ed. by G. Cortese and D. Hymes, in *Textus* Vol. XIV no. 2 Genoa, Tilgher, pp. 421-446.

Kellett Bidoli C.J. (2001) "Sign Language Interpretation: a newcomer to the interpreting forum", *The Interpreters' Newsletter* 11, DSLIT, Università degli studi di Trieste, Trieste, pp. 131-151.

Kellett Bidoli C.J. (2002) "Spoken-language and Sign-language Interpretation: are they really so different?", in *Interpreting in the 21st Century. Challenges and Opportunities*, Selected papers from the 1st Forlì Conference on Interpreting Studies, 9-11 November 2000. Ed. by G. Garzone and M. Viezzi, Amsterdam/Philadelphia, John Benjamins, pp. 171-179.

Kellett Bidoli C.J. (2004a) "Intercultural Features of English-to-Italian Sign Language Conference Interpretation: A Preliminary Study for Multimodal Corpus Analysis", in *Intercultural Discourse in Domain-specific English,* in *Textus,* Vol. XVII no. 1. Ed. by C. Candlin and M. Gotti, Genoa, Tilger, pp. 127-142.

Kellett Bidoli C.J. (2004b) *Glossario inglese - italiano - lingua italiana dei segni (LIS). La lingua dei segni e l'interpretazione: il linguaggio delle conferenze,* (progetto pilota), on CD-ROM, Turin, Piccola Cooperativa Sociale "Alba" a r.l – O.n.l.u.s.

Kellett Bidoli C.J. (2005) "Investigation into Linguistic and Cultural Mediation between the English-speaking World and the Italian Deaf Signing Community", in *Cross-cultural Encounters: Linguistic Perspectives*. Ed. by M. Bondi, N. Maxwell, Roma, Officina Edizioni.

Kellett Bidoli C.J. (Forthcoming a) "A Multimodal Approach in a Multilingual Context: Interpretation from English to Italian to Italian Sign Language". Paper presented at the 25th ICAME Conference, *Corpus linguistics: the state of the art twenty-five years on*, University of Verona, Italy 19-23 May 2004.

Kellett Bidoli C.J. (Forthcoming b) "The Linguistics Conference Setting: A Comparative Analysis of Intercultural Disparities during English to Italian Sign Language Interpretation", in *The Role of English in Institutional Settings: An Intercultural Perspective*. Ed. by G. Garzone. and C. Ilie, Bern, Peter Lang.

Magarotto C. (ed.) (1995) *Vocabolario della lingua gestuale italiana dei sordi*, Rome, Armando Editore.

Ochse E. (2001) "EFL with Adult Deaf Students: Two Cultures, Two Approaches". Ed. by G. Cortese and D. Hymes, *"Languaging" in and Across Human Groups Perspectives on Difference and Asymmetry*. In *Textus. English Studies in Italy*. 14/2, Genoa: Tilger, pp. 447-471.

Ochse E. (2004) "Language – English – Difficult//Question – You – Think – What?". Ed. by C. Candlin and M. Gotti, *Intercultural Discourse in Domain-specific English*. In *Textus. English Studies in Italy*. 17/1, Genoa, Tilgher, pp.143-158.

Piccola Cooperativa Sociale "Alba" (2003) *Dizionario Italiano/LIS*, on CD-ROM, Turin, Piccola Cooperativa Sociale "Alba" a r.l – O.n.l.u.s.

Pignotti G. (ed.) (1997) *Dizionario mimico gestuale* (CD-ROM), Ascoli Piceno, Rinascita.

Puricelli E., Marcioni M., Domini M. and Leogrande E. (1993) *Anch'io voglio comunicare. Manuale dei principali segni religiosi*, Milan, Arte Grafica 2B.

Radutzky E. (ed.) (1992) *Dizionario bilingue elementare della Lingua Italiana dei segni*, Roma, Kappa.

Romeo O. (1991) *Dizionario dei segni. La lingua dei segni in 1400 immagini*, Bologna, Zanichelli.

Stocchero I. (1995) "L'interprete come intermediario tra sordi e udenti", *Sociologia della comunicazione*, 20, pp. 61-66.

Woll B. and Porcari Li Destri G. (1998) "Higher Education interpreting", in *Proceedings of the 18th International Congress on Education of the Deaf*. Ed. by A. Weisel, Tel Aviv University, Ramot Publications.

ON NORMS AND ETHICS IN THE DISCOURSE ON INTERPRETING

Carlo Marzocchi
SSLMIT, University of Trieste

This paper deals in general terms with the way the notion of "translational norm" has been used in research on interpreting, mainly in papers by Shlesinger (1989 and 1999), Harris (1990), Schjoldager (1995), Gile (1999), and, more recently, Garzone (2002).[1] I shall then advance reasons why this notion could find wider application, or at least be made more explicit. I shall not be looking, yet, at a corpus of scholarly or non-scholarly discourse *stating* more or less implicitly what the norm is,[2] although this body of "extratextual pronouncements" is, as Shlesinger (1999) noted, an important source of evidence for the norms operating in interpreting (with some caveats concerning the gap between normative discourse and norms in practice, as recalled in Toury 1998). A recent example of how prevailing norms for a given setting, conference interpreting, can be elicited from discourse and then compared with actual behaviour is provided in the study by Diriker (2004). For the time being, however, I shall keep within the limits of a broad and admittedly abstract argument in favour of a "thicker" description of norms in interpreting, substantiated by anecdotal evidence that I hope will be deemed relevant.

## 1.	Norms as the key to variability in Descriptive Translation Studies

The main result of the strand of research known as Descriptive Translation Studies[3] may well lie in the fact that it has "discovered the complexity of translation" as a sociocultural product; in this way, translation appears as "characterised precisely by its variability" (Toury 1998:12), whereas essentialist

---

1	This paper documents a step in my ongoing research on interpreting within the framework of the EU institutions (moving from Marzocchi 1998). I am indebted to Miriam Shlesinger and Franz Pöchhacker for inspiration, for their constructive criticism on previous drafts and for drawing my attention to relevant literature. My thanks also go to the Editors for their valuable support. Of course, I remain solely responsible for the arguments put forward.

2	An interesting corpus to this end could consist in the early writings on interpreting (1930 – 1970) that were the object of a recent study by Falbo (cf. in this respect 2004: 21 ff. and *passim*).

3	Hereinafter DTS, although the acronym misleadingly suggests a monolithical school of thought.

statements as to what constitutes true or faithful translation are determined by historical contingencies and are culture- and period-bound.

Accordingly, the move towards awareness of the historical variability of the object of study can be said to have implied a radical *undefining* of translation. A-priori definitions were felt by DTS scholars to project the researcher's own assumptions on translation onto a corpus that was often culturally and historically distant; this would involve the risk of circular reasoning, as translational corpora selected according to a particular definition of translation are bound to confirm the definition (cf. Hermans 1999, ch. 4).

An instrument of the attempt to have "variability in all its facets *introduced into the notion of translation itself*" (Toury 1998:13) is the notion of norms, used as an explanatory tool to account for the diverse ways translation is historically, socially and culturally determined. Norms can be defined as regularities of translatorial behaviour, departure from which implies some form of social sanction, that in turn reflect the values shared by a social group. Historical instances of translational behaviour can then be explained in terms of

– preliminary norms, governing for example the choice of source texts, of source languages, the option to translate directly or through a relay language, or to translate into the foreign language; in other words, what Toury terms "translation policy";
– initial norms, governing a very broad orientation towards *adequacy* with respect to the source text or *acceptability* within the target culture;
– operational norms, that guide decision-making during the process of translation at macro- and micro-structural level (cf. Toury 1995, ch. 2).

Different articulations of the concept have been proposed, notably by Chesterman (1993), who suggested a distinction between norms operating at the level of the translator's role and of the relationship between ST and TT (professional norms) and norms pertaining to what is expected from a translation product to be recognized as such (expectancy norms).

2.   Norms in interpreting: early doubts and explorations

The beginning of *explicit* discussion on translational norms in interpreting can be traced back to a programmatic contribution by Miriam Shlesinger (1989) in the 'Forum' section of the inaugural issue of *Target*. Shlesinger's paper strikes the reader – this reader at least – for its methodological insight and scepticism at the same time. On the one hand it cast light on the methodological advantages, for TS as a whole, of extending the concept of norms to interpreting, in a way that was entirely in line with the systemic approach of DTS and fitted neatly within the general framework of TS as charted by Holmes in the 1970s. On the other hand Shlesinger discussed doubts as to how to elicit norms and even as to

whether norms do operate in conference interpreting, that is, as to whether interpreters' translatorial behaviour is determined by anything else than personal preference or cognitive constraints (cf. her later paper on this subject, 1999). Shlesinger's doubts were motivated by the size and scattered distribution of the profession, thought not to favour the socialization processes by which norms emerge and are passed on; however, the fact that interpreter training is (was, rather) concentrated in a few institutions could in fact lead to more consistent transmission of norms to future interpreters. On balance, this led Shlesinger to conclude that norms may well govern interpreter behaviour, despite the difficulties in eliciting them.

Other methodological considerations advanced by Shlesinger concerned the difficulties in finding and designing corpora, legal obstacles to recording performances, and the impact of monitoring on performance. Apart from these, however, a key insight in Shlesinger's paper – again, firmly within a DTS framework – is that in order to study norms in interpreting one needs to place interpreting within a 'system'. Given the interactional nature of interpreting, the system cannot be defined at the level of the ST, nor at the level of a vaguely defined receiving culture, and must therefore be conceptualised at the level of the interpreting event or setting. A thread may be seen running from Shlesinger's call for a systemic look at interpreting settings to later research, such as Pöchhacker's detailed analysis of the context of his case–study (1994), or to Alexieva's (1997) reasoned typology of interpreter–mediated events. Although Alexieva draws her analytical tools from sources other than DTS, her analysis of settings in terms of sociolinguistic parameters could be complemented by looking at the different professional and expectancy norms associated with each constellation of parameters.

A response to Shlesinger's early methodological doubts came very quickly from Harris (1990). In Harris' reply, observations from specific sectors of practice, that testify to the variability of norms in different interpreting settings are somewhat inconsistently juxtaposed to an essentialist statement as to a "fundamental norm" constituting all translatorial activity, that of acting "as a honest spokesperson". The opening statement illustrates how norms not only govern interpreters' behaviour, but are also *expected to do so* by fellow practitioners and scholars, in other words, it testifies to the 'psychological reality' of norms: Harris states that norms do in fact operate in interpreting, and that "anybody well acquainted with the activity could point to some of [them]" (1990: 115). He then moves on to name a few 'norms' organized around binary oppositions:

1. professional vs. natural interpreting (the respective norms being speaking in the first person vs. reporting speech);

2. conference interpreting (CI) vs. TV interpreting (the norms being "conventional fictions" in CI such as 30-minute turns regardless of change of speaker, vs. preference for consistency of voice, gender and prosodic features on TV);
3. his own training experience at Ottawa, and Russian tradition, vs. Western, AIIC practice when it comes to encouraging or frowning upon interpreting into the B language;
4. interpreted speech vs. written translation (production errors and calques being more acceptable in the former than in the latter).

The examples presented by Harris actually represent different levels on a continuum between habits, preferences and socially sanctioned norms, but also raise the issue of how to distinguish between the statement of a norm and its practical operation, and between stating that there is a norm and eliciting it from textual data.

The same lack of a clear distinction between habitual behaviour and a socially sanctioned norm can be found years later in a paper by Schjoldager (1995). This paper is a more articulated attempt to infer an initial norm from the way interpreters treat a particular cultural item in a corpus of Danish–English interpretation. However, the fact that the author's research project underwent changes in progress explains a certain inconsistency: the stage is set in an evaluative framework, where performances by two groups of subjects are to be compared in terms of equivalence, in a search for interpreting *quality*. The conclusions are then drawn in a descriptive framework, in terms of norms governing the choice of how to treat the cultural item while coping with processing constraints. Nevertheless, the pattern that emerged deserves further study, even though the author formulates it rather sweepingly as a generally valid norm: "[The interpreter] is allowed to say something which is *apparently unrelated to the source text* […] provided s/he can say something that is contextually plausible" (Schjoldager 1995:310, my emphasis – the normative discourse is implicit for example in "is allowed to"). In fact, apart from the limited size of the corpus, it is the lack of information on the social acceptance of this translatorial behaviour that makes me hesitate to call this a norm governing interpreting. It would have been interesting to have Schjoldager's subjects comment on their performance, or have other students or trainers assess the performance, before concluding that the subjects' behaviour was in line with a norm.

Going back to Harris's response to Shlesinger, as could be expected from a scholar coming from an entirely different background, no notion of system comes to organize relations in Harris' examples, although Harris identifies different social agents as those who actually set the norms: TV managers, conference organizers, trainers, institutions. Yet it is precisely the apparently

heterogeneous character of Harris' reply that illustrates the *undefining* potential of the notion of norms. Examples in the very short response given by Harris range from what is known as natural interpreting, performed by children in bilingual communities, through interpreting in conference and diplomatic settings, to legal and TV interpreting; the author introduces them stating that "norms will not be the same everywhere" (Harris 1990:115). In other words, the simple fact of reflecting rather loosely on norms in interpreting elicited an academic discourse that began building bridges between different interpreting settings, otherwise still marked by different social and academic prestige and by separate research paradigms, with a cognitive and process-based one prevailing in conference interpreting and a more socio-culturally oriented and discourse-based one in court and community interpreting (despite unifying efforts notably by Pöchhacker in his more recent work, and by Diriker 2004).

Gile, in a contribution to a discussion on translational norms in 1998, also saw the research potential in a norm-based approach to interpreting. He saw it from two points of view:

– as a way to open up the object of study to paradigms that had not figured prominently in his conceptual toolbox, or at least had gone unnoticed in the reception of his writings; to him, norms could be a tool to "foster more empirical research into interpreting and more interdisciplinarity, in particular with sociology and with research on written translation" (1998:99);

– as a way to have research itself undergo a relativizing scrutiny: research is seen as norm–governed behaviour and, at least to judge by the following sharp criticism, the prevailing paradigms loose any aura of intrinsic, objective adequacy: "research in the field has been increasingly governed by status-oriented norms at the expense of problem-solving. Becoming aware of these norms and their operation is important for researchers in the field" (1998:99).

3.   Norms in cognitive paradigms and in the social construction of interpreting

A shared concern in Gile's and in Shlesinger's contribution – perhaps the main concern – is the extent to which norm-governed behaviour on the part of research informants can interfere with the results of experimental or observational studies; this is a problem inasmuch as the studies were originally designed to relate behaviour to *cognitive* constraints. Gile (1998) reports an experiment on the variability of fidelity perceptions, looking at how target-speech segments were reported by participants as errors or omissions; among the results, he found that the same TT segments were by no means reported as errors or omissions by all assessors, and wondered whether they were simply

missed or whether the different appreciation reflected different fidelity norms. Similarly, the lack of correlation between the number of errors identified and the general impression of fidelity reported by the assessors led Gile to hypothesize that something else – the operation of a norm – must account for the way an overall performance is evaluated, possibly overruling the perception of individual errors at micro level.

Gile also noted that norms should be taken into consideration in studies comparing ST and TT on the basis of propositional analysis, if results are not to be distorted by deliberate, norm-based departures from literalness. In Gile's terms, norms presiding over the assessment of performance should enter into the picture in such studies because otherwise

> the metric chosen by researchers […] may measure the opposite of what it is supposed to measure. A correct identification of norms is necessary in order to calibrate the propositional or other metrics used (Gile 1998:99).

Along very similar lines, Shlesinger's (1999) concern in a study of how interpreters handle long strings of modifiers is how to "tease apart" omissions or rearrangements that are due to cognitive overload from those that reflect the norms governing what is sufficient output in simultaneous interpreting. In a move that illustrates how experimental and discoursal data can be analyzed to elicit norms, Shlesinger then surveys the literature on interpreting, in search of normative statements that could account for her subjects' behaviour. This leads her to the very sensible conclusion that interpreters abide by what could be termed a "condensation norm", that "not only condones but often encourages strategic macroprocessing", so that "not every element of every proposition in the source text needs to be reproduced as such. It is appropriate for a simultaneous interpreter to produce the underlying meaning of the proposition" (Shlesinger 1999:69). Intuitively, this is corroborated by training experience (condensation techniques figure equally prominently in the curriculum in Trieste and in recent research on strategies adopted by students, cf. Donato 2003); it is also in line with the long-standing discourse on conference interpreting as documented in professional literature and analyzed by Diriker (2004) and Jones' recent *Conference Interpreting Explained*. However, in addition to seeing the operation of norms in interpreting as a "how-to-tease-apart-from-the-cognitive" issue, the institutional and social construction of norms should also be investigated, e.g. the role of training and professional institutions in shaping and passing on this and other norms. It may also be the case that norms start their 'lifecycle' as a strategy to cope with cognitive constraints in a given situation and are then interiorized and generalized, as could have been the case with condensation or "macro-processing".

Things become more problematic, in my opinion, when professional or academic discourse elevates a particular norm to a defining trait of (segments of) the interpreting profession, i.e. a tool to maintain self-perception, as is testified by the countless statements routinely heard to the effect that translators *say it all*, word for word, whereas conference interpreters get the "real" message across (if possible in a polished, TL-idiomatic version). Diriker (2004: 32 ff.) shows how in the discourse on interpreting condensation and "macro-processing" coexist side by side with ambitious claims to true "fidelity"; this is taken to distinguish interpreting from a received idea of written translation as a word-for-word exercise. This could be dismissed as a hasty but innocuous generalization were it not for the fact that it may have been instrumental in sustaining a status and prestige gap between the different language professions. One often finds, among language professionals, evaluative discourse that attributes different norms to different settings, even within the limited field of oral translation. The "getting the message" vs. "word-for-word" argument, for example, has been invoked in discourse by opposite camps: to underscore the interpreter's sovereign intellectual command of the ST and its cultural implications as opposed to the translator's purported literalness (again, see Diriker 2004 for quotes from professionals and organizations employing interpreters), but also to suggest an entirely different hierarchy, as is illustrated by the introduction to a standard work on court interpreting:

> Court interpreters must conserve the tone of the language, the timbre of the vocabulary with a fidelity that distinguishes the *truly great literary translations*; conference interpreting is first draft translating, Court interpreting is polished translation (Gonzales *et al.* 1991: 27, my emphasis).

In a interesting search for intellectual legitimacy, we see here leading scholars marking a milestone in the academic and institutional consolidation of their field by linking it to one of the most prestigious types of language work in Western tradition, literary translation, with a discourse of, predictably, fidelity and truth.

To understand how statements like the one above are best read as instrumental to socio-professional needs, it is perhaps worth contrasting them with the landscape of norms on the field. In a study on court interpreting in Italy (Siviero 2003), normative statements were elicited in interviews with interpreters working in courts in Trieste and Rome. Some of them, mostly self-taught native speakers from immigrant communities, practising ad-hoc and with low professional status, did in fact play down their role and responsibility in the proceedings by denying precisely the operational norm of completeness (akin to what Gonzalez et al. would call "conservation"). Rather, they reported that their

task was simply to 'get the gist across', and in *plain* language for that matter, since they mostly interpreted for uneducated defendants. In a previous study, interpreters with higher levels of education and permanent status within the court in an officially bilingual region reported norms that they perceived as more demanding, for example in terms of completeness, fidelity, and role delimitation (cf. Roncalli 2001). A difference in actual behaviour was also found: interpreters surveyed in the former study often extended their role into various forms of interaction with and advocacy for defendants – without other participants objecting, for that matter. In the latter study this was not the case, although the comparison is made less relevant by the fact that the role of interpreters there was much more limited, as almost all other participants were bilingual too.

This difference in reported norms between the standard discourse on court interpreting in the USA and the one produced by some practitioners in Italy may well be explained, at least partly, by the different levels of professionalization achieved by court interpreters in the two countries; an explanation may also lie in the different legal systems: Italian proceedings are still largely paper-based rather than relying on interaction as in an adversarial system. However, the difference may also hint at a discrepancy in the discourse on norms between the grassroots and the more established segments of the profession.

Contextualized studies of conference interpreting also show a discrepancy between (assumed) norm and practice. Among the scholars in interpreting more aware of methodological issues, Pöchhacker has consistently used concepts from TS, in particular from the functionalist approaches of Vermeer and Holz-Mänttäri (cf. 1994, 1995). Some features of his corpus of interpreted speech, such as the choice of forms of address, were easily accounted for in terms of an initial norm prescribing "adaptation to target cultural conventions" in professional conference interpreting (1995: 47-49). Moving from this assumption, Pöchhacker then noted that interpreters may fail to adapt instances of culture-bound communication – such as humour – to target cultural conventions, since this would involve segmenting the source text in larger units, which is at odds with the constraints of time and linearity typical of simultaneous interpreting. Interpreters would then be left in the ironic situation of being inherently unable to comply with the very initial norm they set themselves.

Pöchhacker's solution to this paradox lay in suggesting that "cultural transfer" would not be a relevant norm in settings marked by a shared, supra-national socio–professional culture. As a corollary, Pöchhacker doubted that *Skopos*–theoretical concepts were fully applicable to interpreting. Elsewhere I suggested – although at that time not on the basis of the notion of norms, cf. Marzocchi 1998 – that a solution could also lie in pointing at possible

alternatives to adaptation to the target culture, namely at *documentary translation* as a viable initial norm for interpreting, at least in some settings. In other words, I was arguing for more caution in assuming that a fully *instrumental* initial norm is viable for conference interpreting, although it may appear frequently in professional discourse.[4]

The distinction between instrumental and documentary translation was put forward by Nord and is well known in Translation Studies (obviously, it also partly overlaps with dichotomies such as *covert/overt, dynamic/formal, domesticating/foreignizing*):

> We find two basic types of translation processes. The first aims at producing in the target language a kind of *document* of (certain aspects of) a communicative interaction in which a source-culture sender communicates with a source-culture audience via the source text under source-culture conditions. The second aims at producing in the target language an *instrument* for a new communicative interaction between the source-culture sender and a target-culture audience, using (certain aspects of) the source text as a model (Nord 1997: 50).

In a documentary strategy, SI would then no longer attempt to mask features of the original speech such as idioms, humour, intertextual reference, perhaps even forms of address by *replacing* them with supposed cultural equivalents but would resort to a visible intervention by the interpreter to mediate them.[5]

A further reason to refrain from assuming a single initial norm lies in the fact that, as already suggested by Shlesinger (1989), the interpreting setting is best construed as a system in order to study norms. But if this is the case, a broad socio-professional label like 'conference interpreting' can hardly be construed

---

4   Individual aspects of interpreting behaviour in a given setting may very well respond to a norm prescribing instrumental translation: in an interesting study that deserves a more thorough discussion, Turrini (2004) analyzed the way interpreters handle set and creative metaphors in a corpus of SI at a plenary session of the European Parliament. Turrini found a trend towards translating set metaphors non-literally, i.e. with idiomatic equivalents. It would be interesting, however, to check whether *realia*, idioms, forms of address and humour are also translated in a TL-oriented, idiomatic way or whether they sometimes lead to a non-functional output as in Pöchhacker's corpus. As Turrini rightly recalls, the EP plenary is a *sui generis* setting in view of the notorious speed of delivery, planned speeches and unshared knowledge; however, I wonder if these very features do not make it a suitable setting for research into norms, as cognitive constraints can be taken to be at their relative peak for *all* interpreters involved.

5   I have developed this argument with reference to the treatment of idioms in a more didactic paper to appear in the online *Translation Journal* (Marzocchi forthcoming b).

as single system, to which one could associate a single, default initial norm. This is obvious if one thinks of how this label actually covers (covers up, to a point) a wide variety of settings, ranging from the one-off conference serviced by highly mobile free-lance interpreters, to national or international organizations where staff interpreters serve for a whole career in a known range of meetings and for well-known customers, to settings that are in fact court proceedings, albeit at international level. The fact that the range of settings designated by 'conference interpreting' cannot be construed as a single system does not mean that some elements have no systemic, potentially normative effects across the range of settings: training institutions and professional associations come to mind as an example. But pretending that it is possible to formulate norms at the level of "conference interpreting" as a whole, and only at that level, means neglecting all the elements of each specific setting (institutional status and goal, membership, specific language policy a.o.) that enter, within the boundaries of the setting, into a systemic relationship with interpreting. The example discussed in the next section shows that it is at this level, and not at the level of a generic notion of "interpreting", that norms emerge.

4.   Norms, institutions and ethics

Both Gile and Shlesinger seem to take up the norms issue for its methodological potential, but somehow still as a by–product of studies focusing on cognitive aspects. This is of course in line with the researchers' background and interests, and is reflected in the fact that both authors organize most of their examples of norms operating in interpreting around the quantitative metaphors of *addition, omission, condensation, completeness, leaving out* etc.[6] My impression, however, is that the real potential of "norms" as a conceptual tool does not lie exclusively in the fact that it allows us to tune cognitive paradigms more finely, leading to "better calibrated metrics" in Gile's terms. A wider significance of the notion of norms lies in the fact that it evokes the issue of *ethics*. I use the term still rather tentatively to designate both the body of codes of ethics, but also, more generally, any discourse relating translatorial choices to socially shared *values*.

     To build this argument, one has to develop an insight that emerged in Shlesinger's 1999 contribution and, perhaps only implicitly so, in Gile's 1998

---

6   This is only partly the case for Gile, who lists a series of statements that he sees as
       candidate to norm-status in interpreting. One of these statements, "maximizing the
       communication impact of the speech" (1998: 99), can be read as a TT-oriented
       initial norm.

paper. From her review of normative statements in the literature on interpreting in court and community settings, Shlesinger rightly observes that norms potentially collide, especially in institutional settings. Shlesinger sees

> a potential conflict between two sets of norms: expectancy norms implicitly or explicitly projected by the court, and performance norms based on the interpreter's own perceptions of her role and of what she ought to do to fulfil it (1999: 66).

The realization of the potential conflict of norms brings in another key actor, although only mentioned in passing in Shlesinger's paper: *institutions* and the way they shape the norms interpreters are supposed to abide by.

Again, court interpreting provides interesting examples: with the development of court interpreting in the US an explicit translational norm – the *verbatim requirement,* in fact a very strict initial and operational norm in Toury's terms – was inserted into many of the codes of ethics adopted by courts and professional bodies (cf. Gonzales *et al.* 1991, Mikkelson 1996). The verbatim norm prescribes a very ST-oriented translation, including, as was mentioned above, the conservation (that is, the reproduction in the TT) of paralinguistic features, pauses, hedges, hesitations and the like. Its controversial adoption by the profession has accompanied a gain in status and the growing professionalization of the sector. At the same time, by framing interpreting within the usual conduit metaphor, the verbatim requirement seems to safeguard the different roles in the courtroom, protecting other actors from a potentially intrusive role of the interpreter as a would-be mediator or cross-cultural consultant.

An initial norm comparable to the verbatim requirement is prescribed by the Code of Ethics for Interpreters and Translators adopted in 1999 by the International Criminal Tribunal for the Former Yugoslavia, a body whose staff comprises both interpreters coming from the mainstream "conference" circuit in Europe and ad-hoc interpreters of a more "community" profile. The content of the Code ranges from fairly simple prescriptions on courtesy and timeliness, professional integrity and development, role delimitation vis-à-vis legal counsel to, on the other hand, a more sophisticated attempt to enforce complete conservation of all linguistic and paralinguistic features of the ST. The rule stems from a reference to "truth and completeness" in the crucial article 10 (note the reference to the *wording*); at the same time the Code seems to provide for some latitude and visibility for interpreters: paralinguistic clues are deemed relevant if they "facilitate the understanding", and rectifying own errors as well as asking for clarifications are explicitly prescribed practices:

1. (a) Interpreters and translators shall convey with the greatest fidelity and accuracy, and with complete neutrality, the wording used by the persons they interpret or translate.

1. (b) Interpreters shall convey the whole message, including vulgar or derogatory remarks, insults and any non-verbal clue, such as the tone of voice and emotions of the speaker, which might facilitate the understanding of their listeners.

[…]

2. (a) Interpreters and translators shall acknowledge and rectify promptly any mistake in their interpretation or translation.

2. (b) If anything is unclear, interpreters and translators shall ask for repetition, rephrasing or explanation [my emphasis throughout].

The case of the verbatim requirement is a powerful illustration of how norms and ethics can interact. In the first place, it shows how ethical discourses in different settings go into varying degrees of detail in prescribing translational norms, be they initial or operational; indeed, the Tribunal establishes a clear link between the judicial setting and ethical discourse, when in the second recital it "considers that being subject to a Code of Ethics is an integral attribute of being an interpreter and translator employed in a judicial environment". By contrast, the conference interpreting profession in Europe does not seem to have felt the need to spell out any initial norm in its code of ethics, apart from the generic appeal to faithfulness and professionalism (see Diriker 2004: 29-30 for a brief analysis of the AIIC code). Historians of the profession may wish to account for the different development.[7] Secondly, the case of the verbatim requirement illustrates how the conflict between initial norms takes place at the level of ethics. The criticism of the verbatim requirement came out of value-based stances, pertaining to what kind of interpreting and how visible an interpreter role and status is better suited to serve the rights of the defendant or of whole ethnic communities, or even out of a more abstract idea of *fair trial* and *justice*.[8]

---

7   My own very idiosyncratic explanation, partially confirmed by the material analyzed by Falbo (2004), is that early scholarly writings on conference interpreting in Europe, written by recognized professionals and providing plenty of practical advice, did in fact replace an explicit translation norm in ethical discourse; they have since shaped the the self-perception of conference interpreters in Europe in much in the same way as explicit codes of ethics have done for court interpreting and other non-conference settings in the United States and elsewhere. Of course this is an interpretive hypothesis that cannot be retrospectively falsified.

8   Similarly, in the discussion of translation within the European Institutions one finds the same narrative of conflicting norms (roughly literalist vs. functional) over-

A further, anecdotal, illustration of how the discussion on norms almost automatically implies a discussion of ethical stances can be found in the following excerpts from an article by a professional in a semi-scholarly journal; here, the comments on a colleague's conduct very quickly turn from an expert register ("inaccurate", "key term", "comparison") to a discourse of values, rights and indeed sins ("entitlement", "dirty lie"). Note that the controversy apparently starts with an issue of completeness (the colleague had omitted some embarrassing remarks, i.e. he had violated a norm prescribing completeness in sensitive settings) but then the ethical dimension prevails, to the extent that one section of the article carries the heading "Trust":

> On a fine day in Windhoek, Namibia, the President ad-libbed a few remarks, as is his wont. At a certain point the interpreter stopped translating, looked at the President, and interrupted: "Presidente, não estou entendendo", a remark that can be construed both as *Mr. President, I cannot understand what you say* or *Mr. President, I cannot understand why you are saying this*. […] The President must have heard his interpreter, since he repeated the phrase and added a second phrase, so as to build a comparison. The interpreter then provided what would usually be considered an inaccurate rendering of the first part of the comparison and omitted what would generally be considered a key term from the second. […] Now, why isn't the wilful deletion of a couple words from a presidential speech considered *a lie* [..]*? A lie is a lie*, of course, and although there may be cases where a *lie is ethically justified*, such as when a life is in danger, we should be really careful with exceptions. […] I am sure the audience wants to know what the President of Brazil said, not what his interpreter thought he ought to have said instead. I would go farther than that: the audience is *entitled* to know what the President said (Nogueira 2004, my emphasis).

It is also worth noting that Nogueira's account is sophisticated enough to reconcile a completeness norm, named here an "exact translation", with the ethical stance that he concedes as a possible line of conduct for the interpreter, i.e. to "try and prevent fights". The reconciliation is possible by stretching the interpreter's role into that of a neutral but visible intercultural consultant who is the principal of some communicative acts ("letting the parties understand"), as appears from the following excerpt (note also that the discourse is about participants having rights: "I am *entitled*"):

> [colleagues] told stories of how an interpreter prevented a fight by omitting insulting statements from the translation. Great. But, first, if

arched, for example in Wagner *et al.* 2002, by a value-based discourse on the function of institutional communication with citizens (cf. also Marzocchi 2004).

someone calls me an &$#@!!! to my face in a foreign language I believe
I'm entitled to know what he said and, second, the purpose of avoiding a
fight might have been equally well served by *providing an exact
translation* and *letting the parties understand* that the other guy was mad
as hell and meant business, too (Nogueira 2004, my emphasis).

5.   Three reasons to step up research on norms in interpreting

Up to now I have argued, with the help of a few illustrations from scholarly and
professional discourse, that norms are not only a complement to cognitive
paradigms, that norms have to do with institutions, and that they also have to do
with ethics, i.e. with the values presiding over translation. There are three
further reasons why I insist on advocating a fuller use of the notion of norms
(we might as well call for a "thicker" notion) in research on institutional and
social aspects of interpreting.

   Firstly, if we look for an ethical construct[9] that can inform interpreting
behaviour across the range of modes and settings, the lack of an explicit
translational norm in ethical discourse is a missing link, a grey area in the way
the conference interpreting profession depicts itself. Investigating ethical
discourses that include an explicit initial norm as developed in other fields of
interpreting can help the profession fill this gap – court interpreting, but also
Sign Language interpreting are obvious candidates for this exercise. Ultimately,
I would like to see norms emerge from the way the profession perceives the
communicative needs of a given setting, rather than being dictated by purely
institutional needs – as was probably the case with the verbatim requirement in
US court interpreting – or simply taken for granted with a vague reference to
fidelity, as is the case today in conference interpreting.

   Let me exemplify what I mean by the latter: elsewhere (Marzocchi
forthcoming a) I have suggested that the defining feature of the EU institutions
as a setting for language work is the complex configuration of languages in
contact that prevails in the daily life of the institutions, despite official discourse
focusing on isolated languages and identities. I have argued that relevant
translational norms should be inspired by this feature, and that interpreting in
such a setting cannot possibly be based upon the assumption that languages and
cultures are isolated and that official language and mother tongue coincide. In
other words, an initial norm should be elaborated that builds on and
acknowledges the highly stratified language repertoire skilfully exploited by
(some) speakers in that setting, as was documented in another thesis at Trieste

---

9   In fact, this has been my own (re)search as a professional prior to any scholarly
     endeavour.

(cf. Aquino 2001) on a corpus of speeches from the European Parliament plenary.

Secondly, a fresh normative elaboration by the profession could overcome some paradoxes between interpreter behaviour and norms governing language work in the same institutional setting, which I see particularly in the treatment of idioms and *realia*. Another anecdote can illustrate this point: European Court of Justice, spring 2004, novice colleagues from Poland are practising in the dummy booth. Suddenly one of the parties in the hearing uses the English expression "the proof of the pudding is in the eating". The Polish colleague, who later reported thinking "we don't say that in Polish" and "I cannot possibly talk about puddings in Polish", renders the idiomatic expression in plain language, neutralizing the idiom. Later on another participant refers to what the previous speaker has said, takes up the "pudding" image again and in a perfect instance of intertextual reference develops it, half-jokingly, for a couple of sentences. The other colleague on the microphone at the moment cannot refer back to anything similar to puddings in the previous output of her booth mate. In normative terms, and excluding for the moment cognitive constraints,[10] the first colleague seems to have been trying to adhere to a 'full cultural transfer' norm prescribing idiomaticity and proscribing a hybrid TT. For some reason she could not come up with a more literal or mediated solution, that would have provided material for intertextual reference in the TT (for example framing it as in: "as they say in English, the proof of the pudding…").

The anecdote is particularly telling since it involved trainee interpreters, who are arguably more prone to norm-compliant behaviour. This contrasts sharply with an institutional setting that frequently resorts to literalist solutions leading to a strikingly hybrid TT, especially as regards the written translation of a specific type of realia, i.e. names of institutional bodies in legal texts. The striking juxtaposition of languages begins already on the cover pages of case documents, where a uniform layout tells us what the original language was as well as the official denominations of the parties and of the referring jurisdiction in their respective languages. The body of the translations illustrates the practice of keeping institutional names in the original language, possibly adding an explanation in brackets, even when the translation between cognate languages could allow for a calque; this is now an explicit norm of written translation at the ECJ and is visible in all case documents available to interpreters. One wonders, then, how powerful the idiomaticity norm must be to prevent my trainee colleagues from producing even a slightly hybrid output, for example

---

10   The comments made by the colleague afterwards led me to explain the anecdote with a deliberate choice of idiomaticity of the TT rather than of lack of processing resources to come up with equivalents for, say, the lexical item "pudding" (cf. also Marzocchi forthcoming b).

importing what is after all a fairly transparent idiom; and this in a setting where they are surrounded by an extremely visible translation practice (considering the pile of documents that ECJ interpreters are supposed to go through to prepare for the hearing) that displays its hybrid features on every page of the documents.

Anecdotes like this one also reinforce the doubts on cultural adaptation as a viable norm for interpreting, especially in settings where language contact is so obvious and interaction so structured that intertextual reference is the rule rather than the exception. To tackle this issue, research should not be limited to operational norms (e.g. Shlesinger's "condensation" norm) and should venture into the ethically mined field of initial norms, that imply a basic option as to what translators and interpreters are there for in a given setting.

A final reason why research into norms and interpreting should be linked to ethical discourse has to do with an overtly personal and possibly misplaced dissatisfaction with the following: we 'inherited' the notion of norms from DTS scholars who had used it, not without some intellectual daring, to legitimize variability in translatorial behaviour as the object of study in TS; in the study of interpreting, however, I see the risk of it being applied in an innocuous version that does not challenge the core of our received wisdom. This was already partly the case in Harris' (1990) reply to Shlesinger. Harris first listed several examples of variability in interpreter behaviour as evidence for the operation of different norms in different settings; yet he then made his case irrelevant with a final statement to the effect that "under all circumstances" all interpreters are assumed to serve as an "honest spokesperson", and convey the "ideas and manner of the speaker" as accurately as possible.

There is some merit in an ethics centred around the notion of "honesty", in a move not unlike Nord's proposal of "loyalty" as an overarching principle that informs translatorial behaviour within her functional approach (Nord 1997, ch. 8); at least, this would be in line with the traditional discourse based on the virtue of the *fidus interpres*. However, the problem with such a statement is that a single explicit accuracy criterion – an initial norm – cannot automatically be derived from the qualification as a "honest spokesperson"; and this is precisely because of the operation of socially and historically determined norms; in other words, the same bona fide honest spokesperson will perform in (slightly?) different ways in different settings. Denying this, and subsuming all forms and settings of interpreting under the umbrella of an undefined notion of fidelity "to the ideas and manners" of the speaker, means adopting the notion of norms only superficially. It amounts to neglecting the main lesson that can be derived from its application, precisely a lesson in *undefining*.

The same could be said of a more recent contribution by Garzone (2002), that deserves a closer look. The author reviews the debate on quality in simultaneous interpreting and traces the move from early, error-based notions of

quality to more recent, context-aware approaches that raise the issue of how to model all relevant variables; she puts forward the notion of norms as the principle

> located at a sufficiently high level of generalization to explain the rationale underlying the interpreter's behaviour and choices, thus providing a basis for understanding the intrinsic quality of a given SI performance as well as the user's quality expectations (Garzone 2002: 110-111).

Garzone then identifies case- and corpus studies on interpreted speech, together with the analysis of discourse on interpreting, as the source of evidence for norms. She illustrates the socio-cultural specificity and the instability of preliminary norms by reference to the norm favouring interpretation into the mother tongue (specific to Western international organizations and increasingly challenged with, for example, the successive enlargements of the EU). Her discussion of operational norms focuses on the issue of completeness. In slight contradiction to Shlesinger (1999), Garzone states that

> One of the basic norms shared by the interpreting community is that the interpreter should give a complete rendition of the ST, which in theory would rule out omissions (Garzone 2002: 114).

Departures from this norm, which are widely documented in literature, are then seen as "repair and emergency strategies which contribute to assuring the quality of the final product" (2002: 115). A section is then dedicated to the "variability of norms and quality criteria across cultures, space and time": quality, both from the interpreter's and the user's perspective, is redefined in terms of the negotiation between ideal norm and repair strategies (the latter made necessary, according to Garzone, by the fact that "most of the time" interpreters work in an "emergency situation", 2002: 117).[11]

Although one cannot but agree with Garzone's reasoning on quality and possibly with its didactic implications, something is missing in her discussion. Garzone does not mention Toury's category of initial norms (1995: 56), those governing a broad orientation towards the source or the target system, i.e. orienting a translator's decision in solving the tension between adequacy and acceptability. I would argue that initial norms are highly relevant, as they are the ones more closely related to ethical discourse on translation.

The reason for this omission may be linked to the end of her paper. The conclusion reached by Garzone is that user surveys on quality criteria are prone to ideological distortion, because of the gap between the ideal norm and the repair strategies that users too are willing to deploy to maintain communication.

---

11   Perhaps an overestimation of objectively difficult working conditions.

A consequence of this is that "quality assurance rests exclusively on interpreters", as the only "guarantors of the intrinsic quality and fidelity of the TT to the ST"; at the same time they have to make sure the finished product is "sufficiently fluent, plausible and coherent" not to lose the user's confidence. However this "formal" criterion should not be used as a substitute for "*real* fidelity to the ST" (all quotes from 2002: 118, emphasis in the original). In my reading, Garzone thus places the ultimate responsibility back on the interpreter, so it is only to be expected that her paper should end with a discussion of ethics. However, in a slightly inconsistent ending, the author insists on a conceptual distinction between ethics and deontology that effectively *defuses* the potential for variability inherent in a norm-based approach, just as happened in the end of Harris's paper. Deontology is taken to refer to the professional commitment to provide the best possible product; ethics would then refer to the moral orientation informing the choices made by an "upright person" outside the professional sphere but possibly "interfering" with deontology. The distinction is then developed by stating that the "formal acceptability" of the TT would be prescribed by professional deontology, whereas a properly ethical stance would also prescribe "fidelity to the letter and spirit" of the ST (2002: 119).

Garzone's goal in upholding this distinction may well be to put a conceptual emergency brake againonst the relativistic drift hidden in a socially determined notion of quality; if this is the case, the goal is certainly worth sharing. Yet I doubt that it can be reached through the idea of an *intrinsic* quality of interpreting across the whole range of settings, framed in terms of an unspecified *fidelity to the letter and the spirit* of the ST, reminiscent, among other things, of Harris's "honest spokesperson". This would mean precisely that the descriptive scholar would nevertheless approach interpreting with a *built-in* norm, thereby wasting the whole intellectual effort invested in the move to a descriptive, norm-based approach in the first place.

Awareness of norms could also inform the selection of hypotheses and the way they are formulated in our own scientific rhetoric. This can be illustrated by reference to an interesting corpus-based study by Van Besien and Meuleman (2004), where they look at the way two simultaneous interpreters handle repaired and unrepaired speech errors by the speaker in a corpus of Dutch into English interpreting. Confirming their main hypothesis, their finding is that "in more than 4 out of 5 cases interpreters correct speakers' unrepaired errors and translate speakers' repairs without translating the original utterance", which is in line with experience and teaching practice in conference interpreting (but would go against apparent norms in some court settings, as we have seen). Whereas I by no means question the relevance or plausibility of the result, there may be something slightly circular in the evaluative discourse in which they frame it: the hypothesis that interpreters, being listener/client-centred, "will always try to

produce a correct and unambiguous message", i.e. they will not reproduce speech errors, is related at the outset to a distinction between "good" and "not-so-good" interpreters (2004: 65). Instances where the interpreters, seemingly without much effort, only translated the repaired utterance are repeatedly labelled as "successful translation" (2004: 77). More awareness of different norms would probably have led the authors to frame their finding, in itself a relevant one, in more detached terms.

The point in approaching interpreting as norm-governed behaviour, in conclusion, is to acknowledge and account for the fact that "fidelity to the letter and the spirit", or the difference between good and not-so-good, means different things in different settings (again, perhaps only *slightly* different things). While the interpreting profession developed in different settings in society, the translation-normative discourse that accompanied it has intertwined with ethical issues to such an extent that there is not much methodological point in adopting the concept of norms without accepting the undefining potential, the awareness of variability, and ultimately the risk of relativism that it carries with it from TS.[12] This variability extends all the way up to the choice of initial norms and to the ethical, value-based definition of the interpreter's role in each specific setting. Excluding from the scope of socially- and historically determined norms an undetermined essential *quid*, that should characterize all interpreting under all circumstances (Harris' "honest spokesperson", Garzone's ethics of "fidelity to the letter and the spirit") implies making the notion of norms a blunt conceptual tool, without much "added value", i.e. additional explanatory power, compared to less abstract and more process-oriented notions such as *techniques* and *strategies*.

References

Alexieva B. (1997) "A Typology of Interpreter–Mediated Events", *The Translator* 3:2, pp. 153-174.

Aquino R. (2001) "La percezione dei fenomeni linguistici e traduttivi in una organizzazione multilingue attraverso i dibattiti al Parlamento europeo", unpublished thesis, Trieste, SSLMIT.

---

12   A different story is of course how to tackle relativism in an academic field that needs to be at least as normative as to teach students know-how. The conceptual solution may well lie, rather than in an undefined essence of translation, in shared values that are defined outside and above translation (conflicting examples that come to mind: "intercultural understanding", "language purity", "experiencing the other", "low translation costs", "minimal effort" and many more).

Chesterman A. (1993) "From 'Is' to 'Ought': Laws, Norms and Strategies in Translation Studies", *Target* 5:1, pp. 1-20.

Diriker E. (2004) *De-/Re-contextualizing conference interpreting: interpreters in the ivory tower ?*, Amsterdam-Philadelphia, John Benjamins.

Donato V. (2003) "Strategies adopted by student interpreters in SI: a comparison between the English-Italian and the German-Italian language pairs", *The Interpreters' Newsletter* 12, pp. 101-134.

Falbo C. (2004) *La ricerca in interpretazione. 1930-1970*, Milano, Franco Angeli.

Garzone G. (2002) "Quality and norms in interpretation" in *Interpreting in the 21st Century*. Ed. by G. Garzone and M. Viezzi, Amsterdam/ Philadelphia, John Benjamins, pp. 107-119.

Gile D. (1998) "Norms in Research on Conference Interpreting: A Response to Theo Hermans and Gideon Toury", *Multilingual Matters. Current Issues in Language and Society* 5:1-2, pp. 99-106.

González R.D., Vásquez V.F. and Mikkelson H. (1991) *Fundamentals of Court Interpretation. Theory, Policy and Practice*, Durham, Carolina Academic Press.

Harris B. (1990) "Norms in Interpretation", *Target* 2:1, pp. 115-119.

Hermans T. (1999) *Translation in Systems*, Manchester, St. Jerome.

Jones R. (1998) *Conference Interpreting Explained*, Manchester, St Jerome.

Marzocchi C. (1998) "The case for an institution-specific component in interpreting research", *The Interpreters' Newsletter* 8, pp. 51-74.

Marzocchi C. (2004) "Reviews of Wagner, Bech, Martínez, *Translating for the European Union Institutions,* and Cosmai, *Tradurre per l'Unione europea. Problematiche e strategie operative*", *Across Languages and Cultures. A Multidisciplinary Journal for Translation and Interpreting Studies*, 5 (2), pp. 282-293.

Marzocchi C. (Forthcoming a) "On a contradiction in the discourse on language arrangements in EU institutions", *Across Languages and Cultures. A Multidisciplinary Journal for Translation and Interpreting Studies.*

Marzocchi C. (Forthcoming b) "On Idioms, Intertextuality, Puddings and Quantum Physics (all of them in simultaneous, please)", *Translation Journal*, 9:2, http://accurapid.com/journal/32idioms.htm.

Mikkelson H. (1996) "Community Interpreting: An Emerging Profession", *Interpreting* 1:1, pp. 125-129.

Nogueira D. (2004) "His Excellency and His Interpreter", *Translation Journal* 8:2, http://www.accurapid.com/journal, last accessed September 2004.

Nord C. (1997) *Translating as a Purposeful Activity. Functionalist Approaches Explained*, Manchester, St. Jerome.

Pöchhacker F. (1994) *Simultandolmetschen als komplexes Handeln*, Tübingen, Narr.

Pöchhacker F. (1995) "Simultaneous Interpreting: a functionalist perspective", *Hermes, Journal of Linguistics* 14, pp. 31-53.

Roncalli S. (2001) "L'interpretazione in tribunale in una provincia bilingue: il caso di Bolzano", unpublished thesis, Trieste, SSLMIT.

Schjoldager A. ([1995] 2002) "An exploratory study of translational norms in simultaneous interpreting. Methodological reflections", in *The Interpreting Studies Reader*. Ed. by F. Pöchhacker and M. Shlesinger, London/New York, Routledge, pp. 301-311.

Shlesinger M. (1989) "Extending the Theory of Translation to Interpretation: Norms as a Case in Point", *Target* 1:1, pp. 111-115.

Shlesinger M. (1999) "Norms, strategies and constraints: how do we tell them apart?", in *Anovar/Anosar: Estudios de Traducción e Interpretación*. Ed. by A.A. Lugris. y A.F. Ocampo, Vigo, Universidade de Vigo, pp. 65-77.

Siviero F. (2003) "Interpretazione in tribunale: problemi organizzativi ed elaborazioni deontologiche", unpublished thesis, Trieste, SSLMIT.

Turrini C. (2004) "L'interpretazione del linguaggio non letterale in IS al Parlamento europeo", in *Linguistica e interpretazione*. A cura di G. Bersani Berselli, G. Mack e D. Zorzi, Bologna: CLUEB, pp. 301-311.

Toury G. (1995) *Descriptive Translation Studies and Beyond*, Amsterdam/ Philadelphia, John Benjamins.

Toury G. (1998) "A Handful of Paragraphs on 'Translation' and 'Norms'", in *Translation and Norms*. Ed. by C. Schäffner, Clevedon, Multilingual Matters, pp.10-32.

Van Besien F. and Meuleman C. (2004) "Dealing with Speakers' Errors and Speakers' Repairs in Simultaneous Interpretation", *The Translator* 10:1, pp. 59-81.

Wagner E., Bech J. and Martínez M. (2002) *Translating for the European Union Institutions*, Manchester, St. Jerome.

**PERSONALITY CHARACTERISTICS OF INTERPRETER TRAINEES:
THE MYERS-BRIGGS TYPE INDICATOR (MBTI)**

Nancy Schweda Nicholson
University of Delaware

1.   Introduction

People have been interested in personality for thousands of years. The Chinese as well as the ancient Greeks, Aristotle and Hippocrates, developed various systems and labels which enabled them to identify and define basic personality traits.

   In more recent times, Jung created his own way of looking at personality via "personality types" based on individuals' preferences for functioning in both the personal and professional arenas of everyday life (Jung 1923, 1971). For the past 50 years, the Japanese have been gathering data to support "Theory B", a system used to classify personality based on blood type (Nomi and Besher 1983). Theory B has become so popular in Japan that virtually every stratum of society has been affected by it, from advertising to the development of managerial strategies. Other models include "Spectral Theory", which uses the seven colors of the spectrum as a basis for identifying personality characteristics, and VALS (Value and Life Styles), created in the 1970s by the California-based SRI International. VALS has become very popular in the business world (Oldenburg 1988). The 'communication value orientation model' was developed by Casse (1981). Praendex Incorporated has produced a "Performance Requirement Options" (PRO) worksheet which asks respondents to indicate what they believe are important "behavioral requirements" for any given job (PRO 1989). A list of 90 possibilities is offered. These include items such as "maintaining complete, accurate records", "making major decisions independently", and "being a patient, sympathetic listener" (PRO 1989: 1-3). The general idea is that individuals' qualifications and strengths can be matched to the requirements for a particular position. The Five-Factor Model (FFM) of personality situates distinctive and restricted traits within an umbrella grouping of five basic categories: Extraversion, Neuroticism, Agreeableness, Openness to Experience and Conscientiousness (Digman 1990; McCrae and John 1992). There is a growing body of literature on the use of the FFM to assess personality traits cross-culturally (see, for example, McCrae and Allik 2002; Hampson 2000; Saucier, Hampson and Goldberg 2000; Williams, Satterwhite and Saiz 1998. Also see Mohan 2000 for general cross-cultural studies of personality, identity, and factors such as anxiety, stress and neuroticism). Saucier, Hampson and Goldberg (2000) consider whether or not the basic dimensions employed to

describe personality are generalizable across a wide variety of cultures and languages. These include Germanic, Slavic, Romance and non-Indo-European tongues. Block, writing about personality and affect, states that the psychology of personality especially looks at how persons "perceive, respond to, and understand their respective worlds as they seek to establish adaptive life modes" (2002: xii).

The Myers-Briggs Type Indicator (MBTI), developed in the 1950s by a mother-daughter team and based on Jung's theory of psychological type, plays an influential role in personality evaluation in the United States (Bayne 1995; Keirsey 1998; Myers 1987, 1980, 1962; Myers, McCaulley, Quenk and Hammer 1998; Quenk 2000, 1993; Tieger 1995). The MBTI identifies our preferences for (1) interacting with others, (2) gathering information, (3) making decisions about what we experience, and (4) controlling ourselves and the world around us. There is no one "perfect" or "ideal" personality type. All individuals exhibit certain preferences and "preferences are not a matter of right or wrong; they are a matter of what feels most natural" (Barr and Barr 1989: 3). Levesque (2001) uses the MBTI as a basis for helping people to identify and develop their creative talents.

The MBTI is the assessment tool employed in the current study and is discussed in greater detail in Section IV. In sum, "one of the most natural things in the world is the mind trying to make sense out of the data of everyday life ... It does that by codifying and putting things into categories" (Hogan, cited by Oldenburg 1988: C5).

2.   Rationale for the current study

Interpreter trainers have long been involved in the development and refinement of screening devices which attempt to best identify those individuals who have the greatest chance of success in an interpreter training program (Herbert 1952; Keiser 1978, 1964; Kurz 1996; Longley 1968; Moser 1978; Nilski 1967; Pfloeschner 1965; Schweda Nicholson 1986b, Sofr 1976; Szuki 1988). The search for a "perfect" screening examination, i.e., one which would consistently select potentially successful interpreters and weed out those who are unsuitable, goes on. Screening devices can include a variety of components. First and foremost, though, it is critical to ascertain that candidates have a high level of competence in their working languages. Exceptional facility in their "A" language(s) is of paramount importance. An oral interview as well as written tests of synonyms, antonyms and reading comprehension may be employed. Some trainers include additional assessment components like shadowing and sight translation, as well as consecutive and simultaneous interpretation (Schweda Nicholson 1986b). The present study grew out of the author's interest

in adding new elements with the goal of improving the current screening exam. Inasmuch as it has been demonstrated that particular careers tend to attract certain personality types, the author thought it might be useful to gather personality data from interpreter trainees. An assessment device such as the MBTI is not viewed as a replacement for a traditional screening test; rather, it is proposed that such a personality inventory could be one component of a broader exam. As a result, it might provide interpreter trainers with an additional perspective on their potential trainees.

3.   Personality and interpretation

A. What makes a good interpreter?

   Practicing interpreters and interpreter trainers have wondered and spoken about the "ideal" personality traits for the successful interpreter for many years. Within the field of interpretation, the classic approach to the identification of personality characteristics has been an introspective one. To be more specific, interpreters have often examined their own personalities and attempted to generalize based on their personal assessments. For example, an individual may express the following ideas: "I am a good interpreter. I am outgoing, intellectually curious, good at analysis and synthesis, and have an eye for organization and detail. Therefore, all good interpreters are/should be like me and possess these same qualities." In this connection, if one asks an interpreter what he or she believes to be the perfect temperament and personality for a new trainee, the interpreter will, almost without exception, describe his or her own personality. The requirements of the interpretation task such as speed with accuracy, grace and calm under pressure, intense powers of concentration, the ability to internalize large amounts of unfamiliar material quickly, and analytical talent (just to name a few) have been projected into the arena of personality. Hence, one finds a compendium of numerous characteristics from which to choose. It will be interesting to identify which of these hypothesized traits actually materialize in the personality inventories of those surveyed. While introspective data can be useful for research purposes, its value should be viewed in perspective. Such information may be included as one component of a study in which more objective measures are also employed.

B. Review of the interpretation literature

   Before proceeding to a more detailed description of the research method and analysis of the data, it is useful to include a brief review of the interpretation literature regarding personality. What has been written to date is based on both observation and introspection, primarily within the field of conference interpretation. The observation data come from people both within and outside

the profession. In an early article, Paneth (1962) speculates on the procedures for identifying those candidates who have the greatest chance for success in an interpreter training program. She stresses the "qualities of split-mindedness" and "concentration" and suggests that there are certain "right personality traits"; however, Paneth does not elaborate on what these might be (1962: 109). Longley (1968, 1978), Keiser (1978), and Seleskovitch (1978) also emphasize the importance of concentration. Gerver *et al* (1984), Henderson (1980), and Longley (1968) discuss the interpreter's ability to work as a member of a team. In simultaneous interpretation, interpreters work in glass-fronted enclosures with a partner, generally two colleagues per booth. Longley goes on to state: "Some of us have sometimes wondered if it is the need to work constantly and faithfully in a team that has made so many interpreters impossible individualists outside the cabin" (1968: 52). In fact, interpreters are sometimes characterized as "arrogant" (Henderson 1980: 222).

Interpretation can be a frustrating occupation for some. Those who make it a career as well as those who move on to other professions often discuss the need for interpreters to subjugate their own personality to that of the speaker, as it is the lecturer's thoughts which are being expressed and not the interpreter's. Over the years, many have remarked that interpretation requires one to suppress personal ego and ideas. The interpreter is not the originator of what is said; rather, he/she is the human conduit through which ideas expressed in one language are transferred via/to the structure of another. This is not to say that interpreters do not have personal feelings and/or knowledge about the subjects they interpret; their opinions, however, are not permitted to surface in the context of the interpretation. This "suppression of ego" (constantly expressing another's thoughts and not one's own) may become difficult and frustrating for some interpreters. An article by Henderson (1980) includes the observation that the role of the interpreter is a "subordinate" one (225). Longley (1978: 55) discusses the fact that interpreters provide a service to others and are "constantly under control of an outside will (the speaker)". To wit, they (interpreters) facilitate communication between individuals who, without their assistance, would be unable to establish meaningful verbal contact.

Henderson (1980) conducted a study designed to examine personality characteristics of professional interpreters and translators[1]. He asked these two groups to indicate those personality traits which they (1) ascribe to themselves as well as those they (2) believe best describe their colleagues. More

---

1   Although Henderson's respondents completed two separate questionnaires, one which covered "biographical data, education, experience, career goals and attitudes" (217), and a second which consisted of the 16 PF Test (Form C 1969 Edition), the 1980 article discussed solely responses to Questions 6 and 7 from the **first** questionnaire. None of the data gathered from the 16 PF Test was included.

specifically, interpreters not only suggested characteristics for their own group but also were asked to describe translators as well. Henderson primarily discusses the answers to two open-ended questions: (1) "In terms of personality, how would you describe a 'typical' translator?" and (2) "Similarly, how would you describe a 'typical' interpreter?" (217). For purposes of this study, the responses to question #2 are of greatest interest. Some respondents offered only one "terse" response while others provided as many as ten or more characteristics. Of particular interest is that "... generally each group's view of the other tended to corroborate that group's own self-image, e.g. the views of interpreters on translators largely confirm those of translators on translators" (218). In this connection, David C. Funder, a psychologist at the University of Illinois at Urbana, is attempting to demonstrate that people's instincts are generally on target when they are asked to evaluate another's personality (Oldenburg 1988).

Much has always been made of the tremendous stress of the job (Longley 1968; Gerver *et al* 1984; Cooper *et al* 1982). In this connection, many agree that it is particularly important for interpreters to be cool under pressure, to have strong self-control, and "nerves of steel" (Henderson 1980, Keiser 1978, Seleskovitch 1978). Related to the stressful nature of the occupation, the interpreter is also expected to be quick-witted and provide interpreted material in a split second (Gerver *et al* 1984, Henderson 1980, Seleskovitch 1978). However, many have observed that interpreters are, as a result, "high-strung", "temperamental", "touchy" and "prima donna" types (Henderson 1980: 222). Under Henderson's category, "Empathy", the interpreter is also characterized as "sensitive" by some respondents (221).

Additional traits which are generally agreed upon include "inquisitiveness" and "curiosity" (Henderson 1980, Keiser 1978, Seleskovitch 1978). Interpreters are thought to prefer variety, to be tolerant, versatile, adaptable, and open-minded. As might be surmised, they are expected to be articulate and have a "knack for communicating" (Seleskovitch 1978: 78). Moreover, proficiency in analyzing and synthesizing material (Keiser 1978, Seleskovitch 1978) as well as attention to detail (Longley 1968) are often discussed. Interpreters are also thought to be self-confident, possessing the ability to take control of difficult situations (Henderson 1980, 1987). In this connection, Seleskovitch writes about the requirement that interpreters exhibit "great self-control" as well (1978: vi).

"Extraversion"[2] is perhaps the characteristic employed most often when talking about interpreters (Carroll 1978; Cattell 1971; Henderson 1980, 1987; Seleskovitch 1978; Szuki 1988). People frequently say that interpreters can be

---

2   The most frequent dictionary spelling of this word is "extr**o**version." However, Jung wrote it as "extr**a**version", and those who work in the field today have adopted this spelling (Keirsey and Bates 1978).

compared to actors, who enjoy appearing in public and have a flair for public speaking (Henderson 1980; Keiser 1978; Longley 1978). Although rarely mentioned in the literature, discussion has centered recently on the possibility that introverts may actually make better interpreters because they are more focused on the "inner world" (Myers 1987: 5) and are unlikely to be susceptible to internal or external distractions. In fact, one respondent in Henderson's survey characterized an interpreter as "not a good mixer/often a loner" (1980: 221). One can draw a potential connection between the requirement for lengthy concentration and the inner focus of the introvert.

Longley (1968) believes the interpretation profession is not for the "happy-go-lucky" type. She writes: "The need to concentrate for long hours on end, to prepare carefully for meetings, to pay attention to detail, all of which are part of an interpreter's job, do not usually go hand in hand with a bohemian nature" (68). On the other hand, Henderson (1987) does include "happy-go-lucky" in his interpreter profile.

Kurz (1996) employed the 'communication value orientation model' (Casse 1981), which is targeted for use in intercultural communication training. The four major categories are: (1) action-oriented; (2) process-oriented; (3) people-oriented; and (4) idea-oriented. Although Casse believes that everyone possesses all four characteristics to a certain extent, each individual has one orientation that dominates, one whose comfort level is clearly higher than the other three. His instrument consists of first-person statements arranged in forty pairs that deal with personality traits as well as attitudes. Respondents are asked to select the one in each pair that they believe is most reflective of their own personalities. Kurz, however, modified the approach. Instead of asking her sample (which consisted of 31 beginning and 39 advanced students[3] who were taking both translation and interpretation courses) to respond for themselves, she asked students to go through the sentences twice. During one round, they were asked to answer as they thought a translator would and, during the other, as an interpreter would. Kurz's analysis showed that the results were generally in line with the literature cited in her review. Translators were considered to be both "process-" and "people-oriented" whereas interpreters were judged to be "people-" and "action-oriented." Both beginning and advanced students associated "process" more with translators and "people" more with interpreters. Although Kurz's study is an interesting one that uses a personality inventory which differs from the MBTI, the reader is cautioned when interpreting her results. Inasmuch as trainees were not responding with their *own* preferences in mind, the author believes that it is possible that the answers reflected and/or

---

3   Kurz actually began with a set of 57 questionnaires from the beginners and 42 from the advanced students. However, not all were usable, primarily because a significant number were incomplete.

reinforced existing stereotypes within the fields of translation and interpretation. Kurz herself reflects on this particular limitation in her conclusions (1996).

In sum, although the literature does include some studies as well as much introspective data regarding interpreters' personalities, a detailed and multi-faceted investigation employing the Myers-Briggs Type Indicator (MBTI) (a well-known, standardized personality inventory) has yet to appear. As a result, the current research seeks to fill that void by examining personality characteristics of interpreter trainees using the MBTI.

Figure 1. Four scales of the MBTI

A. Extravert (E)/Introvert (I): gather energy

E	I
--interaction	--inner focus
--talkative	--quiet, shy
--active, outgoing	--prefer to work alone

B. Sensing (S)/Intuitive (N): collect information

S	N
--five senses	--abstract
--live in present	--live in future
--clarity, simplicity	--complexity ("big picture")
--just right word	--flexible

C. Thinking (T)/Feeling (F): make decisions

T	F
--objective	--subjective
--head	--heart
--like problem-solving	--thrive on harmony

(NOTE: Among T/F types, there is a clear gender distinction in the general population: Women = 2/3 F and 1/3 T.)

D. Judging (J)/Perceiving (P): stance toward external world

J	P
--decisive	-- go with flow", "play by ear"
--plan/organize	-- pontaneous
--control life	-- nderstand life

(Sources: Keirsey and Bates 1978; Kroeger and Thuesen 1992, 1988; Myers 1987; Silver and Hanson 1980)

---

4.   The Myers-Briggs Type Indicator (MBTI)

The Myers-Briggs Type Indicator (MBTI) is used to examine personality characteristics. This assessment tool has become a standard in business, education, career counseling, and government agencies. It is especially useful in team-building and evaluation of learning styles (Pauley 2002; Scherdin 1994; Sullivan 1994). In addition, research on personalities and careers has demonstrated that certain types of people gravitate toward particular professions because they allow individuals to exercise their favorite ways of doing things (Myers and McCaulley 1985). A description of the four bipolar scales (or "preferences") measured by the MBTI follows along with a hypothetical suggestion regarding the traits of an "ideal" interpreter at the end of each section.

A. The Extraversion (E)/Introversion (I) Scale
    The first scale defines one's preferences in gathering energy: Extraversion (E) vs. Introversion (I). Extraverts gain energy from direct interaction with people and things. Talkative and gregarious, they tend to have a wide scope of interests and prefer to live through experiences and talk about them later. Extraverts like to act rather than take a passive role, and they often make decisions spontaneously. Moreover, Extraverts are sociable and tend to like to meet new people. They enjoy seeking out novel experiences. In contrast, Introverts gather energy from within themselves. Quiet and sometimes even shy, they favor depth over breadth and often devote considerable time to thinking things through before acting. Many Introverts are overwhelmed by the outside world and prefer to work alone. Based on the information provided in Section III, one could hypothesize that the "ideal" interpreter would be an Extravert.

B. The Sensing (S)/Intuition (N) Scale
    The second dimension of the MBTI, Sensing (S) versus Intuition (N), deals with how people prefer to collect information. Sensing types pay particular attention to their five senses: what they can see, feel, hear, touch, and taste. Living very much in the present, they prefer to take things one step at a time and have a knack for keen observation and an impressive memory for concrete details. Sensing individuals prefer tasks which require them to be careful and extremely thorough. Conversely, they generally dislike activities which demand intuition and imagination. Clarity and simplicity have great appeal for the Sensing type. On the other end of the scale, Intuitive (N) types tend to skip over the sensory data in order to focus on abstract ideas, possibilities, and concepts. They tend to live in the future and enjoy bouncing around various ideas in no fixed order. Intuitive people easily see how things are related; they are most

interested in the "big picture". They are intellectually curious and adaptable to the exploration of numerous relationships and connections among data. They are good at anticipation and prediction inasmuch as they are future-oriented. Moreover, Intuitive types are good guessers. Whereas the Sensing type has a tendency to want to find the "right" word to express an idea, the Intuitive person is flexible and can usually come up with various appropriate word choices easily. Barr and Barr (1989) offer yet another comparison between the Sensor and the Intuitor: "Sensors focus on what someone said. Intuitors focus on what they meant" (3). Complexity is particularly enticing to the Intuitive individual, who probably has a variety of intellectual interests. It appears that the "ideal" interpreter would be an Intuitive type.

C. The Thinking (T)/Feeling (F) Scale

The third bipolar scale of the MBTI focuses on how people prefer to make decisions: Thinking (T) vs. Feeling (F). Before proceeding to a discussion of the Thinking and Feeling types' preferences, it is important to mention that "[t]he T-F dimension is the only pair of preferences which shows a sex trend" (Keirsey and Bates 1978: 20). More specifically, approximately two-thirds of women are Feeling types, while only one-third of women are Thinking types (Kroeger and Thuesen 1988: 20). Conversely, then, the great majority of men are Thinking types as opposed to Feeling ones. The wide disparity between males and females within this particular preference will be discussed in greater detail in Section VI.C.

Thinking types favor an objective, logical approach. Problem solving has great appeal, for it encourages their impersonal analytical skills. Thinking individuals may be perceived by others as cold, even arrogant. They tend to be critical and skeptical. "The Thinker appears to be head-dominated, while the Feeler appears to be heart-dominated" (Barr and Barr 1989: 4). In this connection, Thinkers like to focus on content and ideas rather than the individual who generates the ideas. They dislike redundancy. In contrast, Feeling (F) types take a subjective view and assess personal values, their own and those of others. They focus much more on social relationships and social climate. They thrive on friendship and harmony and are likely to be socially aware and active in humanitarian causes. One can hypothesize that interpreters would be Thinking types.

D. The Judgement (J)/Perception (P) Scale

The last scale deals with control: Judging (J) vs. Perceiving (P). Judging types prefer to control their environment. They are decisive, and constantly move toward closure, toward the completion of tasks. They like to plan and organize; they have a strong sense of duty and prefer to be on time. Making

decisions comes easily and quickly to the Judging type. Conversely, Perceiving (P) types prefer to control their participation in the environment. They like to remain spontaneous, and are always open to new possibilities. Perceptive types are curious and flexible, preferring to "go with the flow" and play things by ear. Once again, based on data referred to earlier, one would surmise that many interpreters would be Judging types, although characteristics of Ps are certainly relevant as well. See Section VI.D. for additional discussion.

To sum up, it is hypothesized that many interpreters will be Extravert (E), Intuitive (N), Thinking (T), and Judging (J) or, in the vernacular of the MBTI, "ENTJ".

However, it must be remembered that most personality tests demonstrate that "... no one is a pure amalgam ... no pure introvert, no pure extrovert, no pure type" (Hogan, cited by Oldenburg 1988: C5).

5.   Subjects

A. Groupings

The current study examines the MBTI personality data of several groups of interpreter trainees. First, the group classified as "Regular" (R) (N = 28) is composed of those students who enrolled in and finished a one-year conference interpretation program at either the University of Delaware or the University of Hawaii. The "Vancouver" (V) category (N = 12) includes those trainees who completed a seven-day intensive seminar in primarily simultaneous interpretation at Vancouver Community College. "Government" (G) trainees (N = 19) are those who participated in a five-day intensive course in consecutive interpretation. Unlike most in the "Regular" and "Vancouver" groups, these individuals are currently employed as "Language Specialists" in an agency which utilizes language-skilled individuals for numerous purposes. Those codified as "Not Finish" (NF) (N = 9) are trainees who began the one-year program either at the University of Hawaii or Delaware but, for any number of reasons, did not complete the course of study. The "Hawaii Applicants" (HA) group (N = 56) comprises those individuals who took the Screening Examination (which includes the MBTI) at the University of Hawaii but did not enroll in the program.

In another attempt to garner relevant information from the sample, subjects were also divided by language groups. All trainees had English as a working language, so there is no separate English sample table. However, tables for Spanish, Chinese, and Japanese are included. (See Section VI for detailed discussion of all groupings.)

Of interest is that many in the current sample had no prior interpretation experience whatsoever. Others possess various lengths and types of experience

in language-related fields. Experience or lack of it, however, is not of concern in this study. The fact that these individuals chose to become involved in interpretation and judged it to be a career for which they were suited is of principal interest.

B. Gender

    With respect to the subjects' gender, the current primary sample (N = 68) is represented by 51 females and 17 males. In percentages, one finds that a full 75% are women, while only 25% are men. These statistics closely parallel those of Zeller (cited by Kurz 1989) who examined enrollment figures at the Institute of Translation and Interpretation at the University of Vienna during the period 1983-84. During this particular academic year, "84.2% of the students were girls and only 15.8% were boys" (Kurz 1989: 73). In Kurz's 1996 study, she surveyed two different groups: beginning and advanced students. In terms of complete, usable surveys, the beginning students' gender breakdown was 27 females (90%) and 3 males (10%), whereas that of the advanced students was 32 females (84%) and 6 males (16%). The reader notes the striking consistency among all of these groups. In fact, the 1989 statistics re: male/female make-up are virtually identical to those of the advanced students sampled in 1996.

    In the author's experience of over 20 years in the interpreter training field, groups of interpretation students are generally overwhelmingly female. Most recently, participants in two, two-week consecutive interpretation courses at the University of Delaware in 2000 and 2001 numbered 10 females and 4 males (2000) and 6 females and 0 males (2001). People who self-select into orientation classes for prospective court interpreters in Delaware are predominantly female as well, usually between 75 and 80%.

    The International Association of Conference Interpreters (AIIC) also maintains long-range statistics regarding membership. During the period 1978-84, women constituted a full two-thirds of AIIC's members. Moreover, during the same period, "2.5 times as many women as men joined AIIC" (Kurz 1989: 73). As of June 18, 2003, AIIC membership is at 2667, and 75% are women (www.aiic.net 2003). Zeller's thesis posits a number of sociological and linguistic reasons for the increased presence of women in the interpretation profession, and concludes by saying that men may not be interested in the field because they view it as a "serving" profession (Kurz 1989). Although not a subject of the current study, the reasons for the feminization of the profession are certainly worthy of further investigation.

6.   Results and discussion

TABLE 1.		Interpreter trainees (R, G, V, NF)	N = 68
ISTJ N = 12 % =17.65	ISFJ N = 4 % = 5.88	INFJ N = 2 % = 2.94	INTJ N = 2 % = 2.94
ISTP N = 2 % = 2.94	ISFP N = 1 % = 1.47	INFP N = 4 % = 5.88	INTP N = 6 % = 8.82
ESTP N = 5 % = 7.35	ESFP N = 2 % = 2.94	ENFP N = 5 % = 7.35	ENTP N = 5 % = 7.35
ESTJ N = 6 % = 8.82	ESFJ N = 3 % = 4.41	ENFJ N = 3 % = 4.41	ENTJ N = 6 % = 8.82

I=Introvert, E=Extravert, S=Sensing, N=Intuitive, F=Feeling, T=Thinking, J=Judging, P=Perceiving.

Distribution by the four preferences for TABLE 1:

	N	%
E	35	51.47
I	33	48.53
S	35	51.47
N	33	48.53
T	44	64.71
F	24	35.29
J	38	55.88
P	30	44.12

In order to provide an overview analysis, several groups are joined together in TABLE 1. It represents the type distribution of a primary sample of interpreter trainees (R, V, G and NF categories: N = 68). First, it is important to note that the profession attracts all sixteen types. There is at least one in each category. The greatest number, however, appears in the top left corner: ISTJs constitute approximately eighteen percent of the total. Of interest, however, is that a preponderance of the "Government" group falls into this category. These, as noted, are already language professionals and cannot be considered "typical"

trainees. On the other hand, the fact that they have self-selected into the profession is significant and cannot be overlooked.

The columns underneath TABLE 1 show the distribution across the eight preferences. The sample is about evenly divided between Extraverts and Introverts as well as between Sensors and Intuitors. However, there is a preponderance of Thinkers over Feelers (65% versus 35%). Finally, Judgers outnumbered Perceivers, but only slightly. A detailed discussion of the four scales follows.

A. Extraversion vs. Introversion

The hypothesis that most interpreter trainees and, subsequently, interpreters are outgoing and gregarious Extraverts is not supported. The sample contains about the same number of Extraverts as Introverts. The common impression held by those both within and outside the field is not verified by the current data.

A discussion of working languages may serve to elucidate at least partially the reason for the belief that interpreters are Extraverted. The most common language combination for conference interpreters is English-French-Spanish (Schweda Nicholson 1986a; 1989). Although the United Nations' working languages also include Russian, Arabic, and Chinese in addition to English, French, and Spanish, interpretation from and into Arabic and Chinese is a relatively recent phenomenon in international organizations (Schweda Nicholson 1986a). English, Spanish and French have dominated over the years. An examination of AIIC statistics shows that, of its 2667 current members, 100% have English as a working language while approximately 2300 have French and about 1150 have Spanish[4] (www.aiic.net 2003).

It is, of course, imprudent to make gross generalizations about groups of people and cultures, but many people comment on the friendliness, openness, and charm of Hispanics. Among those who count Spanish as a working language, Extraverts dominate almost two to one over Introverts (total N = 28; E = 18 and I = 10. See TABLE 2.) Moreover, without seeming too simplistic, one can also cite the "*joie de vivre*" mentality of many French speakers. Although the French sample is very small (N = 6), it is worthy of note that approximately 85% are Extraverts (E = 5 and I = 1).

As a result, the data from the current study do support the hypothesis that members of both the Spanish and French groups are highly Extraverted.

---

4   The numbers for AIIC's Spanish and French interpreters are approximations. On the website, the statistics are represented with bar graphs, which only have general reference point numbers on the sides. As a result, the author had to make a good faith estimate as to the approximate size of these two groups. An e-mail request sent to the AIIC Secretariat for exact information went unanswered.

TABLE 2.		Spanish (SP) trainees	N = 28
ISTJ N = 4 % = 14.29	ISFJ N = 0 % = 0.00	INFJ N = 0 % = 0.00	INTJ N = 0 % = 0.00
ISTP N = 0 % = 0.00	ISFP N = 0 % = 0.00	INFP N = 4 % = 14.29	INTP N = 2 % = 7.14
ESTP N = 2 % = 7.14	ESFP N = 2 % = 7.14	ENFP N = 3 % = 10.71	ENTP N = 4 % = 14.29
ESTJ N = 1 % = 3.57	ESFJ N = 2 % = 7.14	ENFJ N = 1 % = 3.57	ENTJ N = 3 % = 10.71

I=Introvert, E=Extravert, S=Sensing, N=Intuitive, F=Feeling, T=Thinking, J=Judging, P=Perceiving.

Distribution by the four preferences for TABLE 2:

	N	%
E	18	64.29
I	10	35.71
S	11	39.29
N	17	60.71
T	16	57.14
F	12	42.86
J	11	39.29
P	17	60.71

The English component is more complicated. Inasmuch as one finds native English-speakers in numerous countries (which are often characterized by widely differing cultural norms), it would be extremely hard to generalize. For example, the following contrast clearly illustrates the point: Americans are known throughout the world for their gregarious, friendly, and easygoing nature. On the other hand, the British (also native English-speakers) generally have a reputation for being more reserved and formal. Unfortunately, it is not possible to compare the English-speakers in the same fashion as the Spanish and French groups because all subjects have English as a working language.

Whether the individuals would be equally as outgoing when speaking either language is another question[5]. Important to this study is the fact that the interpreter trainees control these particular languages and, as a result, also are knowledgeable about and/or members of the cultures in which they are spoken. With all of this said, one could return to the original premise and state with a fair amount of confidence that perhaps the "Extravert" trait among interpreters has been perpetuated over the years simply because there are more Spanish, French, and English interpreters than any others. In other words, people are more likely, just because of sheer numbers, to come into contact with interpreters of these languages.

Also of interest is the fact that Introverts react to stressful situations "primarily by decreasing activity" (Barr and Barr 1989: 42). Inasmuch as stress is a major part of the interpreter's everyday life, it is a bit surprising to find so many Introverts because it is not possible for interpreters to "retreat" when things become difficult; they must persevere under all circumstances/conditions, which are often difficult at best. Kroeger and Thuesen (1992) include a section on stress management in their book. Of particular interest is their characterization of the strategies employed by Introverts to deal with stress:

> [...] because the workplace usually rewards Extraversion over Introversion, there is a tendency for Introverts to 'sell out', giving up their natural preference in favor of living and working on Extraverted terms. So, they act Extraverted during the workday ... Co-workers are shocked to learn that these chatty souls are Introverts in Es' clothing. For the Is it is simply a survival technique, but it can carry a high price in the form of stress and related health issues. Indeed, Introverts tend to be plagued with a range of stress- related illnesses" (234).

Once again, the above analysis (coupled with the previous one regarding language combinations) may serve as a partial explanation for the impression that all interpreters are Extraverted. The 50% in this sample who are Introverts may act Extraverted in the workplace because of the reward attached to the outgoing behavior. As such, the general perception of all interpreters being Extraverted has perhaps been reinforced by the fact that many Introverts behave like Extraverts. This notion is also treated cross-culturally by Allik and McCrae (2002). For example, Asians generally respond like Introverts and are part of collectivistic cultures. However,

---

5   A study of Spanish/English coordinate bilinguals suggests, however, that people may exhibit different personalities when speaking different languages (Simon 1987).

> Asians living in close social groups may attribute sociability not to themselves, but to their collectivistic circumstances. They may act like extraverts, but believe it is their duty rather than their disposition (318).

A related point is that Extraverts like to talk with others as a means of sorting out their experiences. On the other hand, Introverts prefer to think quietly by themselves before acting on anything. Analysis is easier for the Introvert than for the Extravert (Myers 1980). The large number of Introverts in the current sample is once again unanticipated, as interpreters earn their living by talking and interacting with others. Although it was predicted that most interpreters would be Extraverts, the data clearly demonstrate that the profession attracts quiet and retiring Introverts as well. Along with additional evidence cited to this point, this result may partially derive from the fact that interpreters dwell in the mind when working.

B. Intuitive vs. Sensing

Secondly, the hypothesis that most interpreters are Intuitive types is not sustained. Rather, the profession attracts both Intuitors and Sensors in about equal numbers. Examining the American population in general, it is interesting to note that approximately 75% are Sensing types and only 25% are Intuitive types (Myers and McCaulley 1985). Comparatively speaking, the current sample includes a higher percentage of Intuitive types than would be found in the general population.

Interpretation seems to offer opportunities for those who are highly proficient at managing concrete details (S) as well as those who favor broad abstractions (N). By way of explanation, upon examination of the Sensing category in a more in-depth fashion, one notes that these types have a good memory for facts and details and are talented at dealing with specifics. Sensing types are performance-oriented. They tend to focus on the here and now. "Sensors at their best are clear and accurate readers of the facts in the immediate situation" (Barr and Barr 1989: 56). Moreover, they prefer tasks that require soundness of understanding. These are all traits which can easily be related to the task of interpretation.

C. Thinking vs. Feeling

With respect to the third bipolar division, the hypothesis that interpreters tend to be logical, analytical Thinkers (T) is buttressed by the data. Thinking types outnumber Feeling types almost two to one. By way of further elaboration, Thinkers prefer precise work and tend to speak and write straight to the point. They are not only good at organizing information but at synthesizing it as well. Setting high standards of achievement for both themselves and others

is another characteristic of Thinkers (Silver and Hanson 1980). It is not surprising, then, that a majority of the sample is represented by Thinking types.

A comparison of Thinkers' and Feelers' reactions to stress may shed some light on the fact that the interpretation profession attracts more Thinkers than Feelers: "One big difference between Thinkers and Feelers is that Ts want to confront a stressful situation head-on, get it out of the way, and get back on track. Fs want to avoid it at all costs, hoping that it will simply go away" (Kroeger and Thuesen 1992: 235). Moreover, Thinkers "are able to stay cool, calm and objective in situations when everyone else is upset" (Kroeger and Thuesen 1988: 18). As previously stated, interpreters must constantly manage the stress of not knowing a word, interpreting for a fast speaker, and so on. They simply cannot avoid tense situations, which is what Feeling types prefer to do.

D. Judging vs. Perceiving

Finally, the prediction that interpreters would be mostly Judging (J) types is not supported. Although the distribution is not as close as it is for the E/I and S/N scales, approximately 56% of our sample are "J" and 44% are "P". This is also quite unexpected, for it was thought that interpreters would be extremely concerned with organization and closure, not easygoing as is the P's characterization.

On the other hand, one can offer an explanation for the high percentage of Ps. Perceiving types have a tendency to be curious, open-minded and often "fly by the seat of their pants". Of course, interpreters are required to do just this quite regularly inasmuch as they are under extreme pressure to convey the source language message on the spot. The perfect word may not always come to them in a split second, and so they often have to choose a less attractive alternative. Similarly, if interpreters are forced to omit a word because they don't know it and cannot glean the meaning from context, they simply have to accept the fact that they missed it and go on. In these cases, good interpreters cannot and will not allow themselves to become bogged down by focusing on what should have been said, but rather must continue/persevere and interpret subsequent material to the best of their ability. Some interpreters pride themselves on "winging it" and often discuss how they handled a difficult concept or vocabulary problem (or, conversely, did not). Judgers become stressed when they lose control of a situation (Kroeger and Thuesen 1992). This brings to mind the previous discussion of interpreters being required to play a subordinate role to the speaker. Although there are more Js than Ps in our sample, the 44% which are Perceivers may be better at dealing with some of the stressful situations which typify the interpretation profession.

E. The ISTJ profile

Inasmuch as the largest group in the current sample is ISTJ (approximately 18%), it is useful to examine this personality type in greater detail. According to Kroeger and Thuesen (1988), ISTJs are the "most private of the sixteen types" (215). ISTJs can be best characterized as "trustees" (Keirsey and Bates 1984: 189). If one single adjective had to be selected for the ISTJs, it would be "dependable". Of interest in that ISTJs represent only about 6% of the population in general (Keirsey and Bates 1984). They are quite sedate and serious, and prefer to perform their assignments without fanfare or flourish. ISTJs are interested in being thorough, and pay great attention to detail. In this connection, they "absorb and enjoy using an immense number of facts" (Myers 1980: 104). Kroeger and Thuesen write that ISTJs are "contemplative, quiet, grounded, objective, accountable, and conservative" (1992: 240). Keirsey and Bates (1984) continue:

> [ISTJs] … communicate a message of reliability and stability. [They] ... make excellent bank examiners, auditors, accountants. ... ISTJs will see that resources are delivered when and where they are supposed to be; material will be in the right place at the right time. And ISTJs would prefer that this be the case with people, too. (190)

Moreover, one of the ISTJs' strengths is the ability to act quickly, and they are "rock solid" in emergencies (Kroeger and Thuesen 1992: 303). On the other hand, the unknown is considered to be a stress producer for the ISTJ (Kroeger and Thuesen 1992).

Levesque (2001), in her book on creativity and personality characteristics, names the ISTJ the "Navigator" (55). She writes:

> Knowledge of facts and events and a sense of history are important in making sense out of new situations and bringing invaluable experience to bear on problems (72).

Scherdin (1994) reports on an MBTI study of 1600 librarians sponsored by the Association for College and Research Libraries (ACRL). Of the 16 possible type configurations, ISTJ ranked first, a full 16.5% of the sample. When the general population is examined, however, one finds that only 7% were ISTPs based on 1985 data (Myers and McCaulley 1985) and a mere 5.4% fell into this category in a 1998 sample (Quenk 1998). As a result, ISTPs are more than twice as numerous (1985 data) and over three times as numerous (1998 data) among librarians as they are within the general population. It is interesting to think about the general traits of ISTPs and ponder how/why these individuals would be drawn to professions as seemingly diverse as interpretation and library science.

Inasmuch as 75% of the current sample is female, it is also useful to examine particular characteristics of ISTJ women. Kroeger and Thuesen (1988) offer the following observations:

> While all Thinking females swim upstream in our society, this is particularly true for female ISTJs. The responsible, driven nature of this type, while admirable, flies in the face of traditionally 'feminine' traits ...
>   ... ISTJ is often dubbed 'the macho type' - a label with which few women would feel comfortable (but which doesn't necessarily bother those ISTJ women. (216)

Of interest is a parallel which can be drawn between comments from Henderson's 1980 study and the ISTJ profile offered by Keirsey and Bates (1984). Cited in Henderson's survey is the description "outwardly cool but emotionally unstable" (222). Keirsey and Bates (1984) write: "Often this type seem [sic] to have ice in their veins, for people fail to see an ISTJs [sic] vulnerability to criticism" (190). However, stability is considered to be one of the ISTJ's strongest characteristics. On the other hand, Keirsey and Bates (1984) write: "ISTJs have a distaste for and distrust of fanciness in speech, dress, or home" (191). Henderson's data include a comment made about interpreters by translators: "if female, dresses elegantly and presents herself well" (1980: 223). Whereas the first one appears to be quite accurate, the second is in sharp contrast.

F. National (US) type distribution statistics

If one examines a "national representative sample" (Myers *et al* 1998)[6] of types in the United States, the group which is most represented across the board (excluding gender differences) is ISFJ at 13.8%. The least common type is INFJ (1.5%) followed closely by ENTJ (1.8%). The most prevalent type among females is also ISFJ (19.4%); among males, it is ISTJ (16.4%). Least common among females are ENTJ and INTJ (tied at 0.9%), and the rarest type among males is INFJ (1.2%) The reader will remember that ISTJ is the most common type found in our data (and the one which occurs most frequently among males in the national sample), yet the majority of our subjects are women.

G. Additional group - Characteristics analyses

1. Actors

Inasmuch as interpreters are often compared to actors because of their exuberance, flair for public speaking, and desire to be in the public eye, it was

---

6   The "National Representative Sample" consists of 1,478 males and 1,531 females, totaling 3,009.

decided to compare personality data gathered by the MBTI on a sample of 52 actors[7].

The results are quite interesting, as the comparison showed only one major difference: While actors tend to favor Intuition strongly (81% vs. 19% Sensing), slightly more than 51 per cent of the interpreter sample prefers Sensing ($\underline{I}$ = 2.23, p <.01). As a result, the current group showed higher representations of the ST ($\underline{I}$ = 2.39, p <.01), SJ ($\underline{I}$ = 2.73, p < .01), and IS ($\underline{I}$ = 3.63, Fisher's exact p = .01) combinations. While some interpreters may well display considerable acting skill, they prefer Sensing more than twice as often as might be expected if acting talent were that helpful. The reader will remember that the current interpreter data contain an approximately equal number of Sensors and Intuitive types. When compared with actors, however, the current study includes a much larger group of Sensors than the actor sample. This result also holds true for the "ST" type combination. Moreover, there are fewer "NT"s among actors than in our population. Surprisingly, the data show an almost equal distribution between the Extraverts and Introverts. Finally, the interpreter group is represented by "IS" types much more frequently than the actor group.

2. Level of education

Interpreters generally tend to be well-educated. Schweda Nicholson's survey of United Nations and free-lance interpreters demonstrates that virtually all interpreters have a Bachelor's degree, and many have a Master's (1986a, 1989). It is for this reason (and for the recurring emphasis on "intellectual curiosity" as a trait ascribed to interpreters) that a MBTI comparison between the current sample and college graduates is included. With only one major exception, interpreter trainees are very much like college graduates[8]. The EI, SN, and TF scales showed no differences. On the JP scale, however, interpreters include a higher than expected proportion of Perceptive types. While about 32 per cent of college graduates are Perceivers, 44 per cent of the interpreters appear in this category ($\underline{I}$ = 1.39, p < .05). As a result, the interpreters had higher proportions of EPs ($\underline{I}$ = 1.56, p < .01), and almost twice as many TPs as would be expected ($\underline{I}$ = 1.96, p < .01). Thus, interpretation appears to attract a greater percentage of Perceptive types who are college graduates.

3. Smaller sample analyses

Dividing our sample of 68 cases into smaller groups and running Chi-square analyses of various proportions loses statistical power. Thus, the comparisons

---

7   The personality profile data for the "actor" and "level of education" comparative analyses were taken from the *CAPT-MBTI Atlas* (Macdaid *et* al 1986).

8   For this analysis, *CAPT-MBTI Atlas* tables for males and females were combined (N = 14,769) (Macdaid *et al* 1986).

discussed below can only be considered suggestive at best. More research is required to gather sufficient data to make such comparisons reliable.

TABLE 3.	Regular (R) trainees		N = 28
ISTJ N = 5 % = 17.86	ISFJ N = 2 % = 7.14	INFJ N = 1 % = 3.57	INTJ N = 1 % = 3.57
ISTP N = 0 % = 0.00	ISFP N = 0 % = 0.00	INFP N = 1 % = 3.57	INTP N = 2 % = 7.14
ESTP N = 3 % = 10.71	ESFP N = 1 % = 3.57	ENFP N = 3 % = 10.71	ENTP N = 1 % = 3.57
ESTJ N = 2 % = 7.14	ESFJ N = 1 % = 3.57	ENFJ N = 2 % = 7.14	ENTJ N = 3 % = 10.71

I=Introvert, E=Extravert, S=Sensing, N=Intuitive, F=Feeling, T=Thinking, J=Judging, P=Perceiving.

Distribution of the four preferences for TABLE 3:

	N	%
E	16	57.14
I	12	42.86
S	14	50.00
N	14	50.00
T	17	60.71
F	11	39.29
J	17	60.71
P	11	39.29

TABLE 3 shows the type configurations of the Regular(R) trainees (N = 28). Dominant in this group is ISTJ, a result which is consistent with our overall analysis in TABLE 1. Extraverts dominate, but not strongly. Sensors and Intuitors are evenly distributed, while the T/F and J/P scales show a relatively strong preference for the TJ combination.

TABLE 4.		Vancouver (V) trainees	N = 12
ISTJ N = 1 % = 8.33	ISFJ N = 0 % = 0.00	INFJ N = 0 % = 0.00	INTJ N = 0 % = 0.00
ISTP N = 1 % = 8.33	ISFP N = 0 % = 0.00	INFP N = 1 % = 8.33	INTP N = 2 % = 16.67
ESTP N = 1 % = 8.33	ESFP N = 1 % = 8.33	ENFP N = 1 % = 8.33	ENTP N = 2 % = 16.67
ESTJ N = 0 % = 0.00	ESFJ N = 0 % = 0.00	ENFJ N = 1 % = 8.33	ENTJ N = 1 % = 8.33

I=Introvert, E=Extravert, S=Sensing, N=Intuitive, F=Feeling, T=Thinking, J=Judging, P=Perceiving.

Distribution of the four preferences for TABLE 4:

	N	%
E	7	58.33
I	5	41.67
S	4	33.33
N	8	66.67
T	8	66.67
F	4	33.33
J	3	25.00
P	9	75.00

The 12 Vancouver subjects (TABLE 4) show a high percentage of Perceptive types ($\underline{I}$ = 1.70, Fisher's exact p = .03), and especially TP ($\underline{I}$ = 1.89, p < .05). As the reader shall see below, this preference may be influenced by those specializing in Spanish.

TABLE 5.	Government (G) trainees		N = 19
ISTJ N = 6 % = 31.58	ISFJ N = 1 % = 5.26	INFJ N = 1 % = 5.26	INTJ N = 0 % = 0.00
ISTP N = 0 % = 0.00	ISFP N = 1 % = 5.26	INFP N = 2 % = 10.53	INTP N = 1 % = 5.26
ESTP N = 1 % = 5.26	ESFP N = 0 % = 0.00	ENFP N = 0 % = 0.00	ENTP N = 0 % = 0.00
ESTJ N = 3 % = 15.79	ESFJ N = 2 % = 10.53	ENFJ N = 0 % = 0.00	ENTJ N = 1 % = 5.26

I=Introvert, E=Extravert, S=Sensing, N=Intuitive, F=Feeling, T=Thinking, J=Judging, P=Perceiving.

Distribution of the four preferences for TABLE 5:

	N	%
E	7	36.84
I	12	63.16
S	14	73.68
N	5	26.32
T	12	63.16
F	7	36.84
J	14	73.68
P	5	26.32

The 19 Government language specialists (TABLE 5) prefer Sensing almost one and one-half times more than the rest of the sample (I = 1.43, Fisher's exact p = .03). In fact, there are nearly twice as many SJs (I = 1.72, p < .01) than in the full group. A full one-third of the Government subset falls into the ISTJ category.

TABLE 6.		Not Finish (NF) trainees	N = 9
ISTJ N = 0 % = 0.00	ISFJ N = 1 % = 11.11	INFJ N = 0 % = 0.00	INTJ N = 1 % = 11.11
ISTP N = 1 % = 11.11	ISFP N = 0 % = 0.00	INFP N = 0 % = 0.00	INTP N = 1 % = 11.11
ESTP N = 0 % = 0.00	ESFP N = 0 % = 0.00	ENFP N = 1 % = 11.11	ENTP N = 2 % = 22.22
ESTJ N = 1 % = 11.11	ESFJ N = 0 % = 0.00	ENFJ N = 0 % = 0.00	ENTJ N = 1 % = 11.11

I=Introvert, E=Extravert, S=Sensing, N=Intuitive, F=Feeling, T=Thinking, J=Judging, P=Perceiving.

Distribution of the four preferences for TABLE 6:

	N	%
E	5	55.55
I	4	44.44
S	3	33.33
N	6	66.66
T	7	77.77
F	2	22.22
J	4	44.44
P	5	55.55

There were 9 students who enrolled in but did not finish the one-year, four-course program (TABLE 6: NF). Although the NF category is very small, the data show that Intuitors outnumber Sensing types two to one and that Thinking types are more than three times as prevalent as Feeling types.

TABLE 7.	Hawaii Applicants (HA)		N = 56
ISTJ N = 7 % = 12.50	ISFJ N = 4 % = 7.14	INFJ N = 1 % = 1.79	INTJ N = 8 % = 14.29
ISTP N = 1 % = 1.79	ISFP N = 0 % = 0.00	INFP N = 4 % = 7.14	INTP N = 4 % = 7.14
ESTP N = 0 % = 0.00	ESFP N = 1 % = 1.79	ENFP N = 6 % = 10.71	ENTP N = 0 % = 0.00
ESTJ N = 8 % = 14.29	ESFJ N = 1 % = 1.79	ENFJ N = 2 % = 3.57	ENTJ N = 9 % = 16.07

I=Introvert, E=Extravert, S=Sensing, N=Intuitive, F=Feeling, T=Thinking, J=Judging, P=Perceiving.

Distribution by the four preferences for TABLE 7:

	N	%
E	27	48.21
I	29	51.79
S	22	39.29
N	34	60.71
T	37	66.07
F	19	33.93
J	40	71.43
P	16	28.57

A comparison of 40 trainees from Delaware, Hawaii (R) and Vancouver (V) with an additional 56 individuals who applied to the Hawaii Program ((HA) TABLE 7) who did not enroll is also included. The most significant difference appeared among Judging types who predominated in the non-enrollers 71% vs. 29% ($\underline{I}$ = 1.46, Fisher's exact p = .05). Over one-half of these applicants were TJs (58%; $\underline{I}$ = 7.80, p = .05). The reader will remember that this profile agrees with the preliminary hypothesis regarding the "typical" interpreter. Under-represented groups include EPs, SPs, and TPs (Fisher's ps < .05). It is difficult to know why these people did not enroll. Some reports indicated personal obstacles (e.g., inability to find a babysitter, conflict between a work schedule and the hours at which the courses were offered) and other circumstances

beyond their control. There is no clear indication that those who did not register refrained from doing so because of personality preferences.

    4. Specific language groups in combination with English
      a. Spanish

    With respect to possible differences based on working languages, it is important to note that TABLE 2 shows that almost half of the subjects are Spanish speakers. This subset differed from the whole sample in its higher percentage for Perceptive ($\underline{I}$ = 1.38, p < .05). Nearly one half of these, 13, were NPs ($\underline{I}$ = 1.58, p < .01). The interest in flexibility and spontaneity may be related to the native language or the culture from which the interpreters come (Simon 1987).

TABLE 8.	Chinese (C) trainees		N = 17
ISTJ N = 3 % = 17.65	ISFJ N = 3 % = 17.65	INFJ N = 1 % = 5.88	INTJ N = 1 % = 5.88
ISTP N = 2 % = 11.76	ISFP N = 0 % = 0.00	INFP N = 0 % = 0.00	INTP N = 3 % = 17.65
ESTP N = 2 % = 11.76	ESFP N = 0 % = 0.00	ENFP N = 0 % = 0.00	ENTP N = 0 % = 0.00
ESTJ N = 1 % = 5.88	ESFJ N = 1 % = 5.88	ENFJ N = 0 % = 0.00	ENTJ N = 0 % = 0.00

I=Introvert, E=Extravert, S=Sensing, N=Intuitive, F=Feeling, T=Thinking, J=Judging, P=Perceiving.

Distribution by the four preferences for TABLE 8:

	N	%
E	4	23.53
I	13	76.47
S	12	70.59
N	5	29.41
T	12	70.59
F	5	29.41
J	10	58.82
P	7	41.18

b. Chinese

Seventeen of the current sample specialize in Mandarin Chinese (TABLE 8). A high proportion of these were Introverts ($\underline{I}$ = 1.58, Fisher's exact p < .05), especially IS ($\underline{I}$ = 1.68, p < .05), about one half of this subset. Conversely, none of the Chinese fall into the EN category ($\underline{I}$ = 0.00, Fisher's exact p < .05). The idea of "inscrutable Asians" finds tentative support in this particular group. Introverts are more reserved and less communicative than Extraverts, the favorite American preference (Myers and McCaulley 1985).

TABLE 9.	Japanese (JA) trainees		N = 11
ISTJ N = 2 % = 18.18	ISFJ N = 1 % = 9.09	INFJ N = 0 % = 0.00	INTJ N = 1 % = 9.09
ISTP N = 0 % = 0.00	ISFP N = 0 % = 0.00	INFP N = 0 % = 0.00	INTP N = 1 % = 9.09
ESTP N = 0 % = 0.00	ESFP N = 0 % = 0.00	ENFP N = 0 % = 0.00	ENTP N = 0 % = 0.00
ESTJ N = 2 % = 18.18	ESFJ N = 0 % = 0.00	ENFJ N = 1 % = 9.09	ENTJ N = 3 % = 27.27

I=Introvert, E=Extravert, S=Sensing, N=Intuitive, F=Feeling, T=Thinking, J=Judging, P=Perceiving.

Distribution by the four preferences for TABLE 9:

	N	%
E	6	54.55
I	5	45.45
S	5	45.45
N	6	54.55
T	9	81.82
F	2	18.18
J	10	90.91
P	1	9.09

c. Japanese

Out of 11 trainees specializing in Japanese (TABLE 9), 10 are Judging types ($\underline{I}$ = 1.63, Fisher's exact p = .02). In comparison, a sample of 47 students of

Elementary Japanese at the University of Hawaii included 30 (about 64%) Judging types (Moody: Personal communication 1991). Although personality is basically genetic (Bouchard *et al* 1990; Bouchard and McGue 1990; Myers 1980), it may be true that a particular culture encourages development which favors qualities of organization and decisiveness.

   d. Arabic and French

   Only six trainees fall into each of the Arabic and French groups. Unfortunately, these numbers are simply too small to permit speculation.

7.   Conclusions

Henderson (1980) offers a summary profile after analyzing all of his data:

> What then is the 'typical' interpreter like? A self-reliant, articulate extrovert, quick and intelligent, a jack of all trades and something of an actor, superficial, arrogant, liking variety and at times anxious and frustrated - such are only the major features of a complex picture which ... is of course a caricature, but the picture is composed from informed observations (223).

   Interpretation attracts people of all personality types. At least one subject appears in each of the 16 categories. Looking back to some of the personality characteristics listed by other authors in Section III.B. (Review of the Interpretation Literature), the variety of traits represented there also figures in the current sample.

   One immediately sees the qualities of the Extraverts in their preference for variety, their versatility and their knack for communicating. On the other side of the EI scale, one notes the analytical skills and a tendency to be a loner among the Introverts. Among the SN group, attention to detail clearly characterizes the Sensing individual. Curiosity, versatility, and open-mindedness are traits of the Intuitive person. Proceeding to the TF scale, Thinkers are represented by their concentration, arrogance, analysis skills, and the ability to remain cool under pressure. Feelers are sensitive, seek harmony, and work well as members of a team. Finally, on the JP preference scale, Judging types are decisive, self-confident, strong in their convictions, and self-controlled. Perceivers, on the other hand, are versatile, tolerant, open-minded, spontaneous and happy to "go with the flow". As such, it appears that the profession may offer opportunities for all personality types to exercise their preferred ways of interacting, deciding and being.

   However, there are some favorites. While the trainees and language specialists in the sample were about evenly divided between E-I, S-N, and J-P,

the T-F scale showed a meaningful difference: Thinking types outnumber Feeling types two to one. This finding is extremely significant. In this connection, in an examination of the general population, approximately 60% of males are Ts while about 65% of females are Fs (Myers and McCaulley 1985). In this sample, females outnumber males about four to one, yet Thinking types predominate. To conclude, then, most of the predominantly female participants in the current study display a preference for impersonal, logical analysis as well as content and ideas ("head") as opposed to focusing on traditional feminine, subjective values and the promotion of group harmony ("heart"). Based on the limited scope of this study, it is interesting to note that the current sample includes a great number of "Thinking" females. This result is not a surprising one, given the demands of the interpretation profession. The fact that there are just about equal numbers of Extraverts and Introverts goes against conventional wisdom in terms of peoples' off-the-cuff impressions of interpreters' personalities.

In terms of the value and potential use of these data, having the personality profiles of interpreter trainees is of great interest, in principle. However, in terms of looking at other components of a screening exam, for example, the author is confident that all interpreter trainers would agree that skills such as L1 and L2 abilities are far more important than personality type.

However, one's personality may definitely have an effect on that person's comfort level in different situations as well as on processing and organizational behavior. Of course, some areas of work life can be controlled by the individual worker but others cannot. Clearly, many factors play a role in one's professional and personal development over time. This study shows that the personality profiles of interpreters can be as varied as the topics with which they work.

As far as suggestions for further research, other types of interpreters could be surveyed. The emphasis in the existing literature has been on conference interpreters and interpreter trainees. It would be worthwhile to investigate other groups of interpreters, such as those who work in the courts and community service settings. An expanded sample of various ethnic groups might inform us about cross-cultural differences and similarities, perhaps breaking down some of the stereotypical images. Additionally, it would be desirable to include a larger number of subjects.

In summary, the MBTI is an interesting, broadly-used and widely-accepted personality inventory. It is clear that people seem to enjoy learning about their preferences for interacting, working, socializing, thinking and organizing. Isabel Briggs Myers writes:

> The MBTI is primarily concerned with the valuable differences in people
> that result from where they to focus their attention, the way they like to

take in information, the way they like to decide, and the kind of lifestyle
they adopt (1987:4).

Understanding how these differences appear in the interpreter population
provides insight into the personalities of those choosing this profession. This
perspective will assist both those involved in training and individuals interested
in learning whether they may be suited to interpretation careers.

Author's Note: The author wishes to thank Prof. Ray Moody of the University of
    Hawaii for his invaluable assistance with the statistical analysis and insights
    regarding the study's findings.

References

Allik J. and McCrae R.R. (2002) "A five-factor theory perspective", in *The
            Five-Factor Model of Personality Across Cultures*. Ed. by R.R
            McCrae and J. Allik, New York, Kluwer Academic / Plenum
            Publishers, pp. 303-322.

Barr L. and Barr N. (1989) *The Leadership Equation: Leadership, Management
            and the Myers-Briggs*, Austin, Texas, Eakin Press.

Bayne R. (1995) *The Myers-Briggs Type Indicator: A Critical Review and
            Practical Guide*, London, Chapman and Hall.

Block J. (2002) *Personality as an Affect-Processing System*, Mahwah, NJ,
            Erlbaum.

Bouchard Th.J., Lykken D.T., McGue M., Segal N.L. and Tellegen A. (1990)
            "Sources of human psychological differences: The Minnesota
            study of twins reared apart", *Science* 250 (12 October), p. 223-
            250.

Bouchard Th.J. and McGue M. (1990) "Genetic and rearing environmental
            influences on adult personality: An analysis of adopted twins
            reared apart", *Journal of Personality* 58/1, pp. 263-292.

Carroll J.B. (1978) "Linguistic abilities in translators and interpreters", in
            Language Interpretation and Communication. Ed. by D. Gerver
            and H.W. Sinaiko, New York, Plenum Press, p. 119-129.

Casse P. (1981) *Training for the Cross-cultural Mind: A Handbook for Cross-
            Cultural Trainers and Consultants*, Washington, Sietar.

Cattell R.B. (1971) *Abilities: Their Structure, Growth, and Action*, Boston,
            Houghton Mifflin.

Cooper G.L., Davies R. and Tung R.L. (1982) "Interpreting stress: Sources of
            job stress among conference interpreters", *Multilingua* ½, pp. 97-
            107.

Digman J.M. (1990) "Personality structure: Emergence of the Five-Factor Model", *Annual Review of Psychology* 41, pp. 417-440.

Gerver D. and Sinaiko H.W. (1978) *Language Interpretation and Communication*, New York, Plenum Press.

Gerver D., Longley P., Long J. and Lambert S. (1984) "Selecting trainee conference interpreters: A preliminary study", *Journal of Occupational Psychology* 57/1, pp. 17-31.

Hampson S.E. (ed.). (2000) *Advances in Personality Psychology* (Volume One), Philadelphia, Taylor and Francis Inc.

Henderson J. (1980) "Siblings observed", *Babel* XXVI/4, pp. 217-225.

Henderson J. (1987) *Personality and the Linguist*, Bradford, UK, Bradford University Press.

Herbert J. (1952) *Manuel de l'interprète*, Genève, Georg et Cie.

Jung C. (1923) *Psychological Types*, New York, Harcourt Brace.

Jung C. (1971) *Psychological Types*, Bollingen Series XX, The Collected Works of C.G. Jung, vol. 6., Princeton, Princeton University Press.

Keirsey D. (1998) *Please Understand Me II*, Del Mar, CA, Prometheus Nemesis Books.

Keirsey D. and Bates M. (1978) *Please Understand Me*, (3rd Edition), Del Mar, CA, Prometheus Nemesis Books.

Keirsey D. and Bates M. (1984) *Please Understand Me: Character and Temperament Types*, (4th Edition), Del Mar, CA, Prometheus Nemesis Books.

Keiser W. (1964) « Tendances récentes dans la formation d'interprètes de conférences. », *L'Interprète* No. 2, p. 5-10.

Keiser W. (1978) "Selection and training of conference interpreters", in *Language Interpretation and Communication*. Ed. by D. Gerver and H.W. Sinaiko, New York, Plenum Press, p. 11-24.

Kroeger O. and Thuesen J.M. (1988) *Type Talk*, New York, Delacorte Press.

Kroeger O. and Thuesen J.M. (1992) *Type Talk at Work*, New York, Delacorte Press.

Kurz I. (1989) "Book Review: Causes and Effects of the Feminization of the Profession of Translating and Interpreting", (Thesis of Christa Maria Zeller), *The Interpreters' Newsletter* No. 2, pp. 73-74.

Kurz I. (1996) "Scribe or actor? A survey paper of personality profiles of translators and interpreters", *The Interpreters' Newsletter* 7, pp. 3-18.

Levesque L.C. (2001) *Breakthrough Creativity*, Palo Alto, CA, Davies-Black Publishing.

Longley P. (1968) *Conference Interpreting*, London, Sir Isaac Pitman and Sons Ltd.

Longley P. (1978) "An integrated programme for training interpreters", in *Language Interpretation and Communication*. Ed. by D. Gerver and H.W. Sinaiko, New York, Plenum Press, p. 45-56.

Macdaid G.P., McCaulley M.H. and Kainz R.I. (1986) *Myers-Briggs Type Indicator: Atlas of Type Tables*, Gainesville, FL, Center for Applications of Psychological Type, Inc.

McCrae R.R. and John O.P. (1992) "An introduction to the Five Factor Model and its applications", *Journal of Personality* 60, pp. 175-215.

McCrae R.R. and Allik J. (eds.) (2002) *The Five-Factor Model of Personality Across Cultures*, New York, Kluwer Academic/ Plenum.

Mohan J. (ed.) (2000) *Personality Across Cultures*, Oxford, Oxford University Press.

Moser B. (1978) "Simultaneous interpretation: A hypothetical model and its practical application", in *Language Interpretation and Communication*. Ed. by D. Gerver and H.W. Sinaiko, New York, Plenum Press, p. 353-368.

Myers I.B. (1962) *Manual: The Myers-Briggs Type Indicator*, Palo Alto, CA, Consulting Psychologists Press.

Myers I.B. (with P.B. Myers) (1980) *Gifts Differing*, Palo Alto, CA, Consulting Psychologists Press, Inc.

Myers I.B. (1987) *Introduction to Type*, Palo Alto, CA, Consulting Psychologists Press.

Myers I.B. and McCaulley M.H. (1985) *Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator*, Palo Alto, CA, Consulting Psychologists Press.

Myers I.B., McCaulley M.H., Quenk N.L. and Hammer A.L. (1998) *Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator*, (3rd Edition), Palo Alto, CA, Consulting Psychologists Press.

Nilski T. (1967) "Translators and interpreters - Siblings or a breed apart?", *Meta* 12/2, pp. 45-49.

Nomi Toshitaka and Besher A. (1983) *You Are Your Blood Type*, New York, Pocket Books.

Oldenburg D. (1988) "Typing a cast of character", *The Washington Post*, p. C5.

Paneth E. (1962) "The interpreter's task and training", *The Incorporated Linguist* ¼, pp. 102-109.

Pauley J.A. (2002) *Here's How to Reach Me: Matching Instruction to Personality Types in Your Classroom*, Baltimore, MD, P.H. Brookes Publishing Co.

Pfloeschner F. (1965) « Des tests d'aptitude à l'interprétation », *L'Interprète* 1, pp. 5-7.

PRO (Performance Requirement Options) (1989) Wellesley Hills, MA, Praendex Incorporated.

Quenk N.L. (1993) *Beside Ourselves: Our Hidden Personality in Everyday Life*, Palo Alto, CA, Consulting Psychologists Press Books.

Quenk N.L. (2000) *Essentials of Myers-Briggs Type Indicator Assessment*, New York, John Wiley and Sons.

Saucier G., Hampson S.E. and Goldberg L.R. (2000) "Cross-language studies of lexical personality factors", in *Advances in Personality Psychology*. Ed. by S. Hampson, Philadelphia, Taylor and Francis Inc., pp. 1-36.

Scherdin M.J. (1994) "Vive la différence, Exploring librarian personality types using the MBTI", in *Discovering Librarians: Profiles of a Profession*. Ed. by M.J. Scherdin, Chicago, IL, Association of College and Research Libraries, pp. 125-156.

Schweda Nicholson N. (1986a) "A United Nations interpreter survey: The specialist/generalist controversy I", *Multilingua* 5/2, pp. 67-80.

Schweda Nicholson N. (1986b) "Screening interpretation candidates at Delaware: A comparative study", in *Proceedings of the First FIT North American Translators Congress*. Ed. by P.E. Newman, A. Zenker and J.-F. Joly, Medford, NJ, Learned Information, Inc., pp. 213-229.

Schweda Nicholson N. (1989) "An interpreter survey: The specialist/generalist controversy II", *The Interpreters' Newsletter* 2, pp. 38-45.

Seleskovitch D. (1978) *Interpreting for International Conferences*, Washington, D.C., Pen and Booth.

Silver H. and Hanson J.R. (1980) *User's Manual: The TLC Learning Preference Inventory*, Moorestown, NJ, Hanson Silver and Associates.

Simon Ch. (1987) "Dr. Jekyll, Señor Hyde", *Psychology Today*, December issue, p. 16.

Sofr O. (1976) "Test of aptitude for interpreting", *L'Interprète*, No. 3/4, p. 3-8.

Sullivan M. (1994) "Using the MBTI for team building", in *Discovering Librarians: Profiles of a Profession*. Ed. by M.J. Scherdin, Chicago, IL, Association of College and Resource Libraries, pp. 173-180.

Szuki Atsuko (1988) "Aptitudes of translators and interpreters", *Meta* XXXIII/1, pp. 108-114.

Tieger P.D. (1995) *Do What You Are: Discover the Perfect Career for You Through the Secrets of Personality Type*, Boston, MA, Little, Brown.

Williams J.E., Satterwhite R.C. and Saiz J.L. (1998) *The Importance of Psychological Traits: A Cross-Cultural Study*, New York, Plenum Press.
www.aiic.net (International Association of Conference Interpreters website)

QUALITY RESEARCH REVISITED

Franz Pöchhacker
Center for Translation Studies, University of Vienna

1.   Introduction

Considering the broad range of topics and the great diversity of research approaches in the field of interpreting studies, research on quality in interpreting stands out as an impressively rich and cohesive area of study. One line of investigation in particular – survey research on interpreters' and users' quality expectations and preferences – has been around for about twenty years and could be said to form a distinct research model, or 'paradigm' (in the narrower sense often used in various sciences). As such it is productive in various ways: it embodies a set of underlying theoretical assumptions and thus supplies the necessary conceptual framework for empirical research. Crucially perhaps, it also consolidates a set of methodological choices, thereby facilitating repeated application (replication). This in turn helps extend the base of empirical data from which conclusions may be drawn. As an accepted standard of sorts, the research model offers a working method that can readily be adopted also by less experienced investigators.

At the same time, and on a different level, a research model's prominence may also expose it to closer scrutiny within the scientific community. Careful (re)examination of its conceptual and methodological choices will put the research model to the test and either confirm or question its validity. Either way, such methodological criticism serves to consolidate and refine research practices and results. It is this hopeful assumption that lies at the heart of the present paper, which revisits and critiques some studies on interpreters' and users' quality expectations and preferences. Most of the revisiting will be done in rather practical methodological terms, with an emphasis on statistical procedures for the analysis of survey data. Aside from this re-analysis component, the paper also doubles as a review of some recent research, with special emphasis on methodological issues and on the gatekeeping function of the editorial process leading to quality publications. In either dimension, my discussion will pivot on a recent paper by Delia Chiaro and Giuseppe Nocella, of the University of Bologna, which both raises important methodological doubts about previous studies and prompts some concerns about research published in our field.

2.   A reliable springboard

Like any piece of serious research, the present contribution should begin by reviewing the state of the art. Given the breadth of the topic, however, the scope of such a review must be strictly limited. It would be impossible here to summarize the expansive literature on quality in interpreting, as reflected, for instance, in the bibliography by Shlesinger (2000) and in the two proceedings volumes of the international conference on the topic convened in 2001 by Ángela Collados Aís of the University of Granada (Collados Aís *et al.* 2003a, 2003b). The same is true of survey research on interpreting quality, which has been the subject of several review papers (e.g. Kurz 2001a, 2003; Pöchhacker 2001). Indeed, I will (have to) narrow my focus to one particular line of investigation, namely questionnaire-based surveys on the quality criteria and expectations of conference interpreters and users of simultaneous interpreting (SI) – QE research, for short.

QE research was pioneered in the 1980s by colleagues at the University of Vienna, Hildegund Bühler (1986) and Ingrid Kurz (1989). Their work proved seminal to most subsequent efforts, including the user expectation study commissioned by AIIC (Moser 1996) and the 'matched-guise' experiments by Collados Aís (1998, 2002) and Garzone (2003). Most recently, an innovative survey using the World Wide Web (Chiaro and Nocella 2004) has again shone the spotlight on these 'classic' studies, albeit in a rather exposing way. Before reporting their empirical study, Chiaro and Nocella (2004) offer a review of methodological issues in quality-oriented research, including a rather harsh critique of Bühler (1986) and Kurz (1989, 1993). Their paper can therefore serve as a convenient peg both for a more detailed account of the studies in question and for addressing some basic methodological problems.

2.1.   The interpreters' perspective

Chiaro and Nocella (2004) depart from the observation that "there appears to be little harmony concerning which perspective to take when undertaking research" (279). Framing their choice as one between the perspectives of the interpreter and the user, they opt for the former to provide "a helpful starting point" and hope for their findings to serve as "a reliable springboard for further research" (279).

Though Chiaro and Nocella supply no further rationale for adopting the interpreters' perspective, it is obvious from their research design that it was actually Bühler (1986) who provided the springboard for their survey: "The criteria used in this investigation are the same as those used by Bühler (...)" (Chiaro and Nocella 2004: 283).

As described very briefly by Chiaro and Nocella (2004: 282), "the well-known study conducted by Bühler (1986)" was based on a list of sixteen "linguistic" (performance-related) and "extra-linguistic" (interpreter-related) criteria which Bühler suggested AIIC members might consider more or less important when sponsoring candidates for membership. Bühler's all too sparse description, in an endnote, of her sample of 41 interpreters who received and returned the questionnaire "at the Council Meeting and the International Symposium [...] convened by AIIC in Brussels in January 1984" (1986: 233-234) does not draw any critical remarks; rather, it is her results that lead Chiaro and Nocella to conclude that "something was faulty in the research design of the study" (2004: 283). According to Chiaro and Nocella (2004: 282), "interpreters valued most of the items as important or highly important, thus highlighting their difficulty in assigning an order of importance". This assessment, according to which "the interpreters were incapable of discriminating and were giving equal importance to all the criteria" (283), invites a look at Bühler's actual findings. Figure 1 was drawn up on the basis of the percentage values published as an annex to Bühler's paper (1986: 235).
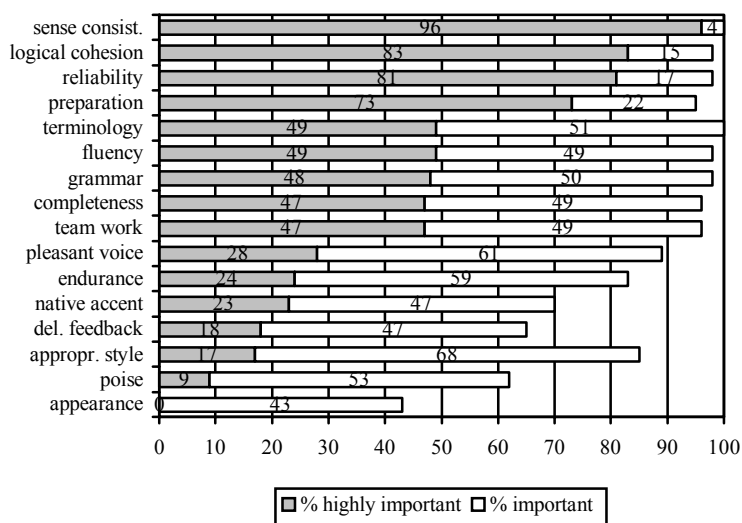
Figure 1. Quality criteria rated as "(highly) important" by 47 AIIC members (Bühler 1986)

Ordered according to the percentage of respondents who gave a rating of "highly important", the sixteen criteria displayed in Figure 1 reflect a rather clear-cut differentiation, from the top-rated demand for "sense consistency with original message" to the least important criterion, the interpreter's "pleasant

appearance", which a majority of respondents considered "less important" (43%) or "irrelevant (13%). While it is true that all other criteria received a rating of at least "important" from a clear majority of respondents, exclusive use of the two highest ratings was made for only two criteria – "sense consistency with original message" and "use of correct terminology".

It may also be noted that among the nine top-ranking criteria in Figure 1 (at least 47% "highly important") there are three interpreter-related ("extra-linguistic") qualities: "reliability", "thorough preparation of conference documents" and "ability to work in a team". This is of interest here because subsequent QE surveys – up to the study by Chiaro and Nocella – largely neglected Bühler's extra-linguistic criteria, so that comparisons have been possible only for her output-related ("linguistic") criteria.

2.2.   Interpreters vs. users

The shift from conference interpreters' criteria for sponsoring AIIC candidates – and, presumably, for a "first class interpretation" (cf. Bühler 1986, note 2) – to the expectations of end-users was brought about by Ingrid Kurz, who questioned Bühler's (1986: 233) assumption that her criteria "reflect the requirements of the user as well as [the] fellow interpreter". Narrowing down the list of criteria to the first eight items in Bühler's questionnaire, Kurz (1989) introduced a comparative view on quality expectations, most famously presented in her 1993 paper on "expectations of different user groups" in *The Interpreters' Newsletter* (reprinted in *The Interpreting Studies Reader*).

While there is no need here to say more about Kurz' (1993) widely noted findings, the ostensible methodological weaknesses of her work, as pointed out by Chiaro and Nocella, require closer examination. Chiaro and Nocella (2004: 282) observe that "Kurz' samples were very small and uneven" and even speak of "discouragingly poor returns". Given the actual number of respondents (124), this critique is hardly justified. One might point out, for instance, that the sample size of the AIIC survey (Moser 1996), in which 94 interpreters conducted questionnaire-based interviews at 84 different meetings with a total of 201 conference participants, by no means dwarfs what was achieved single-handedly by Kurz in three conferences. Her sample, made up of participants in a medical conference (47), a meeting of engineers on quality control (29) and a Council of Europe meeting on equivalences (48), also compares well with the work of Vuorikoski (1993) and Mack and Cattaruzza (1995), who had 177 and 75 questionnaires, respectively, completed at five meetings with SI.

Elsewhere in their paper, Chiaro and Nocella (2004: 284) speak of "the 10-15% rate of questionnaire returns that is normal for traditional surveys". Assuming that not all participants would make use of the SI services offered,

Kurz' four dozen questionnaires each from two of her meetings could easily amount to a 15% response rate in a conference with some 400 participants. Admittedly, though, this conjecture may well err on either side, and it is indeed regrettable that no information on the number of questionnaires distributed is available. A laudable model in this regard is provided by Mack and Cattaruzza (1995: 40), whose return rate, incidentally, was three times higher (roughly 80% to 90%) in meetings where the survey had been announced to the participants than in meetings without such announcement (roughly 25%). Again, it is not known for Kurz' surveys how the questionnaires were brought to the attention of the conference participants.

While these methodological shortcomings go unmentioned, Chiaro and Nocella level a different, rather curious charge against Kurz' (1993) work, namely that her questionnaire was "administered in three very different moments in time and in different contexts, thus weakening the rigour of the experiment" (2004: 282). Though one may well ask for more detailed information on the meetings concerned, it is hard to see how the aim of studying different user groups could be achieved without surveying participants in different meetings, as was indeed done purposely also in the AIIC survey (Moser 1996).

In the abstract of their paper, Chiaro and Nocella (2004: 278) note that "research undertaken so far is surprisingly lacking in methodological rigour". In the text, at the outset of their review of methodological issues, they similarly state that "attempts at more scientific research in interpreting often appear to be based on rather uncertain methodological principles" (279). Aside from the shortcomings mentioned above, the most serious criticism brought against the studies by Bühler (1986) and Kurz (1993) would seem to concern their statistical analysis of the data. According to Chiaro and Nocella (2004: 283), "a substantial shortcoming of this particular study is that the mean was used as the descriptive statistic for analysing and discussing data and drawing conclusions when dealing with ordinal data". And here they have a point. Though Chiaro and Nocella voice this criticism, erroneously, with reference to Bühler's (1986) study (cf. Fig. 1) and are more benign toward Kurz' statistical analysis, the latter does indeed suffer from the infelicitous choice of using the arithmetic means to describe her ordinal data. Having asked her respondents, as Bühler did, to rate the individual quality criteria on a four-category scale ("highly important" – "important" – less important" – "irrelevant"), Kurz (1993) should have described her results, as Bühler did, in terms of the percentages for the various ratings. Essentially, the intervals between the four items making up the scale cannot be assumed to be the same, so metric conversion is, strictly speaking, not permissible. But even if Kurz had used a four-point metric scale, e.g. ranging from "least important" to "most important", with numbered values in-between,

statisticians would be wary of using the arithmetic mean to describe the data because too much of the variability and actual distribution of the data between "1" and "4" may be lost to an average value in the middle. Whereas some would accept such calculations for a five-point metric scale, many authors suggest that rating scales analyzed in terms of means should consist of at least seven points (cf. also Gile 1983: 241).
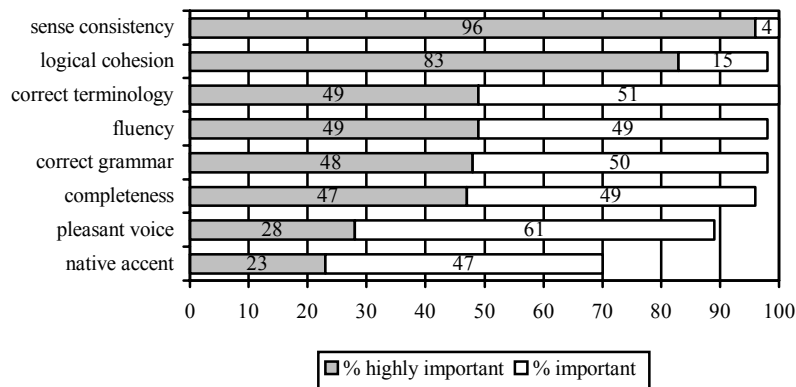
Figure 2a. Eight criteria as rated by 47 AIIC members
(based on Bühler 1986)
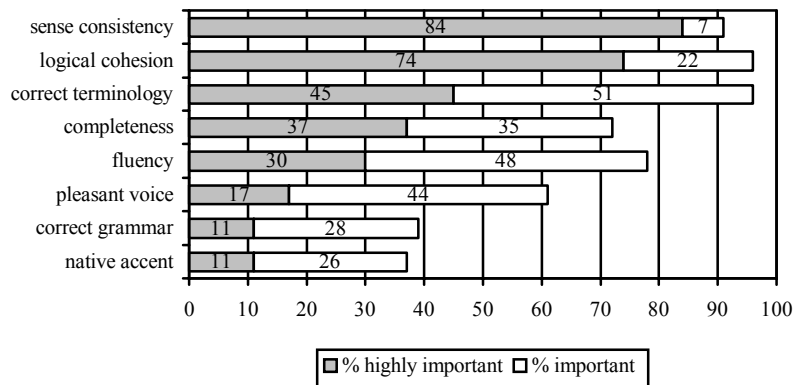
Figure 2b. Eight criteria as rated by 47 medical conference participants
(based on Kurz 1989)

It should be noted, however, that a description in percentages was in fact offered in Kurz (1989), where the values for ratings of "highly important" and

"important" by AIIC interpreters and medical conference participants (47 each) were juxtaposed in a table. Using 'valid percent' of responses, i.e. percentages adjusted for the 2 missing values in Bühler's and the 9 missing responses in Kurz's data,[1] the results can be visualized as shown in Figures 2a and 2b.

As discussed in detail by Kurz (1989), conference participants (MDs) generally tended to give lower ratings than the AIIC members in Bühler's study. A noteworthy exception is "use of correct terminology", which was rated "important" by 51% of interpreters and users alike and for which the interpreters' ratings of "highly important" were only slightly higher (49% vs. 45%). It is also evident that the two criteria given the least importance, "pleasant voice" and "native accent", have distinctly lower percentage ratings among the SI users at the medical conference.
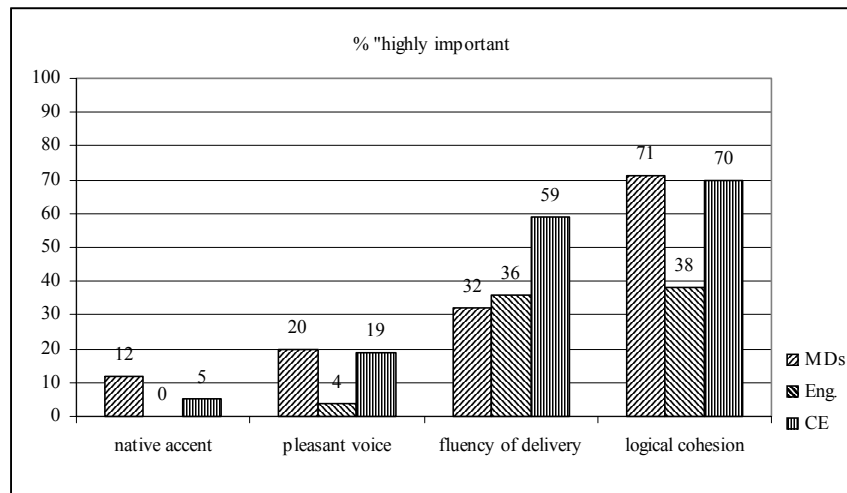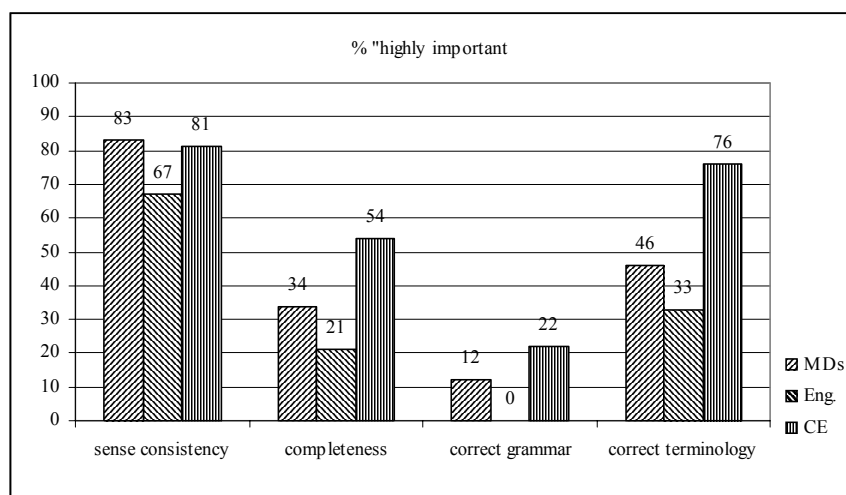

Figure 3a. Quality criteria ratings (% "highly important") by three user groups (cf. Kurz 1993)

While it is thus quite feasible to compare the findings of Bühler (1986) and Kurz (1989) in terms of percentages, the matter is more difficult in the case of Kurz (1993), which requires a comparative analysis of at least three sets of findings.[2] Tables or charts of percentages describing Kurz' user expectation

---

1 Special thanks are due to Ingrid Kurz, who kindly provided me with her original survey materials for statistical reanalysis.

2 It was for this purpose that Kurz (personal communication) enlisted the help of someone with training in statistics – and was supplied with mean values for comparative analysis of her data (see in particular Kurz 1993: 16-17).

dataset are likely to yield a rather complicated picture. Figures 3 is an attempt to describe the ratings of "highly important" for the three different user groups: medical conference participants (MDs), engineers (Eng.) and Council of Europe delegates (CE). To facilitate comparison with the charts published in Kurz (1993: 17), the eight criteria are shown in two charts (Figs. 3a and 3b).

Figure 3b. Quality criteria ratings (% "highly important") by three user groups (cf. Kurz 1993)

Such charts, which offer a rather detailed but cumbersome description, could be drawn up for all four response options. Ideally, however, our statistical analysis should not stop at mere description but should help us understand what the various differences in the data mean – if they mean anything at all. In other, statistical words, we would ask whether these differences are significant, that is, based on some principled relationship in the data, or whether they are equally likely to result from pure chance. Although I cannot claim any special statistical expertise, I will attempt such an analysis in the section below, using some widely available analytical tools.[3] While my main goal here is to illustrate a few basic methodological options in processing ordinal data, the analysis will also serve to test some of the longest-standing findings in QE research for their statistical significance.

---

3   The statistics software SPSS for Windows (version 12.0) was used to process the data and perform the various calculations and tests.

3. Significance

3.1. Crosstabulation

In examining Kurz's (1993) ordinal data for significant relationships between the three user groups, the most elementary option would be crosstabulation. This involves the cross-classification of two categorical variables – in our case, a given criterion's degree of importance (an ordinal variable) and the nominal variable of 'user group'. The four response options ("highly important", "important", "less important", "irrelevant") and the three user groups (MDs, Eng., CE) result in a three-by-four data matrix for each of the eight criteria. It is on the basis of such contingency tables that various measures of association can be calculated. Chief among them is the chi-square test, a nonparametric test that compares observed frequencies to their expected values.

Unfortunately, the sample of 124 respondents is not quite large enough to ensure an adequate number of expected values in all twelve cells of the three-by-four table. For each criterion the distribution yields at least two cells (20% of cells and more) for which the expected frequency in the chi-square test is smaller than five, which renders any interpretation of the test invalid.

For a chi-square test to be viable for the given data set, the values should have a more balanced distribution. This can be achieved by collapsing some categories containing low-frequency values. When this is done by recoding "less important" and "irrelevant" into a single value ("not important"), crosstabulation yields better results. Though there are still too many cells with low-frequency values in the tables for four of the criteria, Pearson's chi-square test indicates a significant relationship in two cases, namely "completeness of interpretation" (Table 1) and "correct grammatical usage" (Table 2).

		User group			Total
		MDs	Eng.	CE	
not important	Count	13	6	5	24
	% of group	28.3%	20.7%	10.4%	19.5%
important	Count	16	17	16	49
	% of group	34.8%	58.6%	33.3%	39.8%
highly important	Count	17	6	27	50
	% of group	37.0%	20.7%	56.3%	40.7%
Total	Count	46	29	48	123
	% of group	100%	100%	100%	100%

Pearson chi-square = 13.103; p = .011
(0 cells with expected frequency < 5; min. exp. = 5.66)

Table 1. Crosstabulation for "completeness of interpretation"

As can be seen from the percentages in Table 1, the participants in the Council of Europe meeting attributed significantly more importance to "completeness" (56% "highly important" vs. 10% "not important") than either medical doctors (37% vs. 28%) or engineers (21% vs. 21%). According to Pearson's chi-square test, this difference is clearly significant at the 95% confidence level ($p < .05$) and even approaches significance at a probability level of 99% ($p < .01$).

		User group			Total
		MDs	Eng.	CE	
not important	Count	28	22	19	69
	% of group	60.9%	78.6%	40.4%	57.0%
important	Count	13	6	19	38
	% of group	28.3%	21.4%	40.4%	31.4%
highly important	Count	5	0	9	14
	% of group	10.9%	0%	19.1%	11.6%
Total	Count	46	28	47	121
	% of group	100%	100%	100%	100%

Pearson chi-square = 12.512; $p = .014$
(1 cell (11.1%) with expected frequency < 5; min. exp. = 3.24)

Table 2. Crosstabulation for "correct grammatical usage"

As regards users' differential appreciation of "correct grammatical usage", the significant relationship confirmed by Pearson's chi-square test clearly holds between the ratings of engineers and Council of Europe delegates. Whereas the former assign particularly little importance to grammatical correctness (79% "not important"), a majority of CE delegates consider it "important" (40%) or even "highly important" (19%). Again, the difference is highly significant ($p = .014$).

3.2.   Other nonparametric tests

Aside from the chi-square test, there are other nonparametric tests for identifying significant relationships among different sets of rank-ordered data. The most appropriate procedure here is the Kruskal-Wallis *H*-test, applied to multiple independent samples for determining whether the values of a particular variable differ between two or more groups. The Kruskal-Wallis test, which involves comparisons of rank orders, can be viewed as the nonparametric

equivalent of the one-way analysis of variance (ANOVA) commonly used to determine whether the means of various groups are significantly different.[4]

	Chi-square	df	Asymptotic significance
1. native accent (n=123)	.595	2	.743
2. pleasant voice (n=121)	.987	2	.610
3. fluency of delivery (n=113)	12.468	2	.002
4. logical cohesion of utterance (n=118)	10.798	2	.005
5. sense consistency with original message (n=120)	1.843	2	.398
6. completeness of interpretation (n=123)	9.558	2	.008
7. correct grammatical usage (n=121)	11.766	2	.003
8. use of correct terminology (n=124)	19.122	2	.000

Table 3. Results of Kruskal-Wallis test for quality ratings by user group

The results of the Kruskal-Wallis test for the ordinal data under study (Table 3) indicate group-related differences significant at the 99% confidence level for five of the eight criteria (cf. note 4). For "native accent" and "pleasant voice" as well as "sense consistency with original message", quality expectations are not significantly different among the three user groups. For the remaining criteria, paired tests are required to identify the nature and location of the differences between groups. This can be done using the Mann-Whitney *U-*

---

4   If the means used in Kurz (1993) were accepted as a valid descriptive statistic, the test used to identify significant differences among the three user groups would be an analysis of variance. Its results, calculated for illustration, indicate significant relationships in four of the eight criteria: fluency, $F(2,110) = 7.037$, $p = .001$; logical cohesion, $F(2,115) = 3.79$, $p = .025$; completeness, $F(2,120) = 5.056$, $p = .008$; and correct terminology, $F(2,121) = 9.958$, $p = .000$. (The values for correct grammar fail the preliminary test for homogeneity of variances and must therefore be excluded from the interpretation.) Upon further examination in paired post-hoc tests (e.g. Bonferroni), particularly clear-cut differences are found for completeness and correct terminology, where the mean ratings of Council of Europe delegates differ significantly from each of the other groups (cf. Table 4).

test, which tests for significant differences between two independent samples. The Mann-Whitney test results for the three possible comparisons (MDs vs. Eng., MDs vs. CE, Eng. vs. CE) suggest that the medical doctors have the least to do with the overall between-group differences: Only one criterion in comparison with engineers shows a significant relationship (logical cohesion, $p = .002$), and four criteria are significantly different in relation to Council of Europe delegates (fluency, $p = .001$; completeness, $p = .022$; correct grammar, $p = .046$; correct terminology, $p = .002$). It is the comparison between the latter and the engineers that yields significant differences for all five of the criteria identified as significant by group in the Kruskal-Wallis test (Table 3). For illustration, detailed results are shown in Table 4.

	Group	N	Mean Rank	Sum of Ranks	Mann-Whitney U	Asymptotic Significance (2-tailed)
fluency of delivery	Eng.	26	29.29	761.5	410.5	.037
	CE	43	38.45	1653.5		
logical cohesion of utterance	Eng.	28	29.70	831.5	425.5	.013
	CE	44	40.83	1796.5		
completeness of interpretation	Eng.	29	30.10	873.0	438.0	.003
	CE	48	44.38	2130.0		
correct grammatical usage	Eng.	28	27.84	779.5	373.5	.001
	CE	47	44.05	2070.5		
use of correct terminology	Eng.	29	27.09	785.5	350.5	.000
	CE	48	46.20	2217.5		

Table 4. Results of Mann-Whitney *U*-test for differences between groups "Eng." and "CE"

3.3. Significance and meaning

This (re)analysis of Kurz' user surveys has focused on the statistical options and tools for describing the data and examining them for significant associations between them. It has highlighted in particular the importance of choosing the appropriate procedures in accordance with the nature of the data and the assumptions holding for various analytical tools. While a thorough understanding of statistics would be highly desirable for anyone carrying out such analyses, it is suggested here by way of demonstration that PC-based statistics software has become accessible enough to be used, with proper guidance, also by the 'semi-skilled' analyst.

However, as much as some statistical know-how can and should well be expected of interpreting reseachers today, the above exercise in significance testing should not obscure the fact that analyzing empirical data, whether from survey research, fieldwork or experiments, is not a question of mathematical skills but, essentially, a matter of meaningful interpretation, of making sense of the relationships indicated by the data. In other words, a statistical significance test does not explain anything but merely points reliably to what needs to be explained. Such (possible) explanations of their survey findings are amply discussed in the papers by Bühler (1986) and Kurz (1989, 1993), and there is neither need nor space in this methodology-oriented paper to revisit this – crucial – part of QE research. Two comments may be in order, though, since they relate to fundamental issues of research methodology (see also section 4.2 below).

One is prompted by the rather striking findings for the role of terminological correctness. "Use of correct terminology" ranked high in Bühler's (cf. Figs. 1 and 2a) as well as Kurz's (1989) findings (Fig. 2b), and was also given special attention by Mack and Cattaruzza (1995), who even found correct terminology to be the top-rated criterion (cf. also Kopczyński 1994). Bühler, herself an expert in the area of terminology, had argued that "[o]ne has to use correct terminology if one aspires to render the message faithfully" (1986: 232). Acknowledging this reasoning, Kurz (1989: 144) also suggested that "the strong emphasis on correct terminology observed here may well be a specific feature of medical (and other highly technical) conferences". When she put this assumption to the test in her subsequent surveys, the prominent role of correct terminology was undiminished but showed a clear peak among Council of Europe delegates (cf. Fig. 3b). Kurz (1993) sought to explain this finding with reference to the institution-specific terminology of international organizations. Judging from the program of the CE conference in question, however, one should also consider an alternative explanation, as suggested also by Mack and Cattaruzza (1995: 46-47). The conference, held in Vienna and Budapest under the auspices of the Council of Europe, was devoted to equivalences in education, that is, the comparability and recognition of certificates and degrees granted by institutions of secondary and higher education in Europe. On the face of it, interpreters at that meeting would have grappled with the rendition of concepts linked to different sociocultural traditions and institutions – a daunting translational task in any case, which was probably not made any easier by the organizers' request, in the preliminary conference program, that speakers limit their oral presentations to five minutes. In this light, it is quite conceivable that the thematic context of the meeting made terminology a prized asset to the proceedings, and that the CE delegates' high expectations for terminological correctness were a function of the conference topic, if not the actual interpreting

services received. In her conclusions, Kurz (1993: 20) makes explicit reference to "the importance of situationality and communicative context" for her comparative study as such; based on the information available, it appears that this awareness should extend also to the situational and thematic context in which her QE survey data were collected.

This methodological issue in data collection, which bears on the interpretation of the survey findings, is connected to another point that may deserve further consideration, namely the language used to collect responses. Kurz used a bilingual (English/German) questionnaire (see Kurz 1996: 57) in the first two of her surveys (MDs and engineers) and an English-only version in the CE meeting. One might therefore ask whether the language in which respondents (MDs and engineers) filled in the questionnaire could have influenced the results. Crosstabulation of the (three-category) ratings by language indeed reveals such an effect for the criterion of completeness, which received significantly higher ratings from the 39 respondents using the English version than from the 36 German-language users (Pearson's chi-square; p = .005). When analyzed by conference (MDs vs. Eng.), this effect appears to obtain irrespective of user group (Mann-Whitney U-test; MDs: p = .034, Eng.: p = .018). As for a possible explanation of this finding, it may again be of a methodological nature. Bühler's English term "completeness of interpretation" was rendered in German as "vollständige Wiedergabe des Originals" (complete rendition of the original). One might speculate whether the greater redundancy of the German version, which foregrounds "rendition" rather than completeness ("Vollständigkeit"), led German-language users to give lower ratings to this criterion, not least because it followed immediately upon "sense consistency with original message", another "a priori" feature of interpreting. Additional support for this hypothesis might be seen in the fact that the CE delegates, who received only the English version of the questionnaire, gave significantly higher ratings to completeness than the other two groups (see Table 1).

4.   The way forward

The re-examination of previous QE research findings undertaken in the previous section essentially suggests that progress in interpreting studies, especially with regard to research methodology, may come not only from the introduction of novel techniques but also from a more detailed, critical engagement with previous work. This applies in particular to the recent contribution by Chiaro and Nocella (2004), whose criticism of previous QE research prompted the discussion offered in the preceding sections, and whose own research will be reviewed and used as a starting point for additional methodological reflections in the sections to follow.

4.1.   Interpreters on the Web

With a keen awareness of methodological limitations in previous QE research, apparently inspired by Gile's (1994) critical view of research skills in interpreting studies, Chiaro and Nocella report an innovative study in which "great care was taken (...) not to fall into the traps that previous studies had failed to avoid." (2004: 283). With Bühler's (1986) criteria as their starting point, the authors drafted a questionnaire which included quality criteria as well as background variables (age, place of birth, qualifications, experience). Rather than a rating of individual criteria on a scale with several response options, the survey instrument designed by Chiaro and Nocella (2004) called for a ranking of the criteria in descending order of importance, i.e. from the most important to the least important item in the list. The questionnaire was administered through the World Wide Web by sending out 1,000 invitations by e-mail "to interpreters belonging to several professional associations" (284). A total of "286 conference interpreters across five continents" responded to the web-based survey (279).

     The sample was 29% male and 71% female, with a mean age of 45 years and an average of 16 years of experience. 44% of respondents had their birthplace in Western Europe and had a degree in interpreting. Chiaro and Nocella also report that the interpreters in the sample are mostly freelancers and that, rather strikingly, "most respondents do not interpret into their mother tongue" (285).

     To facilitate the ranking task, the list of quality criteria was offered to the respondents in two groups, "linguistic" and "extra-linguistic", the first of which comprised the first nine items in Bühler's questionnaire (i.e. the eight used by Kurz plus "appropriate style"). Displaying the percentages for the various ranks (first to ninth) for three sets of three criteria, Chiaro and Nocella (2004: 287) find the following pattern of relative importance: "consistency with original", "completeness of information" and "logical cohesion" as the three most important factors, followed by "fluency of delivery", "correct grammatical usage" and "correct terminology", with "appropriate style", "pleasant voice" and "native accent" ranking lowest. These findings are further explored by multidimensional scaling, a statistical technique for plotting the similarity structure found in the data in a two- or three-dimensional conceptual space. The three most important and the three least important criteria are found to cluster at opposite ends of a "discriminating quality" dimension, while grammar and terminology occupy a middle ground and "fluency of delivery" appears in a unique intermediate position.

     As regards the set of extra-linguistic criteria, the authors do not find a neat pattern, except for the two top-rated items, "concentration" and "preparation of

conference documents". Results are given as summary scores (from 1932 to 1024), the calculation of which is left unexplained in the paper.[5]

4.2.   Methodological issues

There is no doubt that Chiaro and Nocella have tread new ground by harnessing the Internet for QE research among interpreters, and their innovative study deserves praise and recognition. Their use of advanced statistical methods for data analysis is likewise apt to encourage the use of more sophisticated analytical techniques in future studies. And yet, in light of the authors' aspirations to methodological soundness and their somewhat heavy-handed criticism of previous studies, one cannot but question some aspects of research design and presentation that would have demanded more attention.

The first of these weaknesses concerns the authors' conceptual framework as reflected in their use of basic terms. Aside from their liberal use of the term 'experiment' in referring to Kurz' surveys, Chiaro and Nocella base their review section on a two-fold distinction between product analysis and "field work (based upon the results of questionnaire surveys)" (2004: 280). While there are indeed many ways of distinguishing various types of approach, it is not clear how the authors' categorization improves on earlier proposals, such as the four-fold distinction made by Vuorikoski (1993) specifically for the purpose of research on interpreting quality. More critically, though, Chiaro and Nocella use the term "perception" as the principal keyword in their work (and its title), obscuring the fundamental distinction between QE research on generic expectations (as pioneered by Bühler and Kurz) and the direct assessment, or judgment, of an actually perceived interpreting performance, as introduced by Gile (1990) and combined with QE research by Mack and Cattaruzza (1995). This distinction is crucial to the work of Collados Aís (1998, 2002) and Garzone (2003), which has taken user-oriented studies of interpreting quality to a new level. Confounding preferences and perception could therefore be said to fall short of the state of the art.

Another methodological uncertainty concerns the authors' survey instrument, with regard to both design and distribution. Though Chiaro and Nocella (2004: 283) state that their criteria "are the same as those used by Bühler", they actually use 17 rather than 16 criteria, several of which are not the same as those in Bühler's (1986) questionnaire. While a critical appraisal and, if

---

5   The scores become clear from the questionnaire which the authors kindly provided to me after receipt of a first draft of this paper: Respondents were instructed to give "8 to the most important and 1 to the least important"; the scores were thus calculated by multiplying the rank values by the number of respective responses.

necessary, appropriate modification of previous instruments would certainly be welcome, Chiaro and Nocella do not offer any discussion of this part of their work. There is mention of "several interviews" and "endless brainstorming sessions" with interpreters as the basis for devising the questionnaire (2004: 283), but no explanation why two of Bühler's linguistic criteria were apparently rephrased and five new ones substituted for items in the extralinguistic category.[6] At any rate, it would have been desirable to reproduce the relatively short (one-page) questionnaire in an annex to the paper.

Most consequentially perhaps for a paper boasting an innovative approach to QE research, Chiaro and Nocella (2004) give an all too sparse description of their sampling procedure (see section 4.1). It would be interesting to know *which* professional associations were targeted for the survey and, if AIIC was among them, how individual interpreters were selected from the membership list (which in the case of AIIC includes more than 2,600 entries. It is thus not even clear whether the survey was addressed to conference interpreters only: The indication of workload in terms of "hours per month" (with the minimum reported as 0 and the maximum as 200 hours = about 30 days per month), and the baffling finding that "most respondents do not work into their mother tongue" (285) raises some serious doubts which could easily have been dispelled by asking respondents to indicate their professional affiliation and domain of work.[7]

Another methodological issue in survey research of such a comprehensive scope is the language and cultural context of survey administration. With one third of respondents originating from (though not necessarily residing in) South and Central America and Eastern Europe, one cannot be sure that the questionnaire was equally accessible to all recipients (unless they were included in the sample for having English among their working languages). Moreover, there is some evidence in the literature that preferred interpreting styles may differ from one sociocultural context to another (e.g. Ločmele 2001); Chiaro and

---

6   A number of critical comments are on record regarding the criteria used in QE research, beginning with detailed reflections on possible misunderstandings by Bühler (1986) herself and the immediate "Comment" by Seleskovitch (1986). The fact that Chiaro and Nocella (2004: 290) use "intonation" as a synonym of "fluency of delivery" highlights the problem of definition and the need for terminological clarity.

7   As it happens, the clue can be found in the poorly worded questionnaire item (cf. note 5): "Do you interpret mostly esclusively [sic] towards your mother tongue? (Yes/No)". Nevertheless, further information on respondents' professional domain could also be expected from the last item in the questionnaire ("Is your interpreting: Mostly consecutive / Both consecutive and simultaneous / Mostly simultaneous"), the results for which are not reported.

Nocella do not examine their findings for such differences, or do not report any such attempts in their paper. Even if the interpreting profession in various parts of the world were homogeneous enough to render such linguistic and cultural effects negligible, translation scholars conducting surveys across cultural boundaries should probably be the first to demonstrate an awareness of this delicate methodological issue (see, e.g., Harkness *et al.* 2003).

Contextual effects ought to be considered also in a more concrete sense, as illustrated in connection with particular user expectations in Kurz' (1993) surveys (see section 3.3 above). At least since the comprehensive survey commissioned by AIIC (Moser 1996), QE researchers have been aware that users' (and possibly interpreters') quality criteria may differ depending on the type of conference (large vs. small, technical vs. general). Studies on quality requirements for interpreting in media settings (e.g. Elsagir 2001, Kurz 2001b) are another case in point. Asking interpreters to give an opinion regardless of meeting type (cf. Gile 1989, Pöchhacker 1995) therefore precludes a more differentiated view of quality among the respondents.

The way respondents were asked to give their opinion deserves special attention also in a more technical sense. Asking interpreters to rank rather than rate the individual criteria is of course perfectly valid, and represents an innovative aspect of the study. However, there is some evidence in the literature (e.g. Bradburn and Sudman 1979) that ranking more than five to six items may be an overly difficult task for reliable performance. (As explained by Chiaro and Nocella, ranking Bühler's first nine criteria requires 36 mental comparisons.) In light of the authors' interpretation of Bühler's findings, that respondents had "difficulty in assigning an order of importance" (Chiaro and Nocella 2004: 282), their forced-choice approach for a list of nine items therefore seems less than ideal for bringing out subtle distinctions. In future studies it may be preferable – and more user-friendly – to design the questionnaire as a combination of rating scales and rankings, e.g. with a list of criteria to be rated on a multi-point scale followed by a request to rank the three or five most-important ones in the list.

Another option is the paired-choice approach adopted by Gourevich and Mateeff (1989), who asked 50 experienced interpreters to state a preference for one of each pair of criteria offered to them on 28 test cards (which reflected all possible combinations of eight criteria, including completeness, correctness, usefulness, smoothness, calmness and pleasantness). Though the mathematics of their scaling analysis are daunting, the findings suggest that, despite disagreement among the experts concerning the importance of various characteristics of SI, "correctness" and "usefulness" outweigh prosodic characteristics on the scale of relative importance.

The study by Gourevich and Mateeff (1989) offers an interesting parallel to the work of Chiaro and Nocella. Admittedly, the latter could not easily have

been aware of that paper, published as it was in a rare journal and language.[8] Still, the comparative discussion, or lack thereof, of the survey findings is yet another broadly methodological issue to note. Since Bühler's (1986) pioneering survey constituted their basis and point of departure, Chiaro and Nocella (2004) could be expected to draw some explicit comparative conclusions. Instead, the authors vaguely state that "contrary to common belief, results highlight that interpreters do not consider all the criteria in question as being of more or less equal importance" (291). Leaving aside the rather crude interpretation of Bühler's findings (cf. Fig. 1), the conclusion drawn by Chiaro and Nocella is circular, since the design of the web-based questionnaire did not allow respondents to assign equal importance to any two or more items.

The various problems noted for the authors' handling of the literature and of their own findings, and the methodological issues raised by the design and presentation of their study, bear strongly on the broader theme of research standards in interpreting studies, as addressed most consistently by Daniel Gile (e.g. 1994, 1999). Research training, international and interdisciplinary cooperation, and joint supervision of theses have been suggested as measures to improve the quality of research done in interpreting studies. Such initiatives notwithstanding, a crucial aspect of quality assurance in our field, as in any other scholarly/scientific discipline, is a screening procedure prior to publication. With the article by Chiaro and Nocella (2004) as a case in point, this issue will be discussed in the following, final section of this paper.

4.3.   Into print?

For a research paper to be published in an edited volume or academic journal, it has to meet certain requirements with respect to both substance and presentation. One or more editors will usually be responsible for making sure that this is the case. For scientific journals in particular, the editorial process relies heavily on a peer review system, in which colleagues with appropriate expertise examine the manuscript for its contribution to the state of the art, making sure that the research reported is theoretically and methodologically sound. A highly informative description of this process is offered by Gile and Hansen (2004) with reference to the proceedings volume of the 2001 EST Congress in Copenhagen. The following remarks on the paper 'under review' will have to be more anecdotal, but should serve to highlight some of the issues nevertheless.

---

8   Knowledge of that study came to me through Ingrid Kurz, whose cooperation in this endeavor is again gratefully acknowledged.

The research reported by Delia Chiaro and Giuseppe Nocella in volume 47 (2004) of the Canada-based translators' journal *Meta* was conducted in the fall of 2000, prior to the International Conference on Interpreting at Forlì, where the survey and preliminary findings were presented by Giuseppe Nocella. Nocella subsequently submitted his paper for publication in the proceedings which were to be edited by the conference organizers, Giuliana Garzone and Maurizio Viezzi. Instead of the editors' original plan to publish two volumes with a leading international publisher in translation studies, only one book was eventually published in John Benjamins' Translation Library series (Garzone and Viezzi 2002). A second volume was published locally in the same year (Garzone *et al.* 2002). Neither volume contains the paper by Nocella, with whom the present author had exchanged manuscripts by e-mail at the time of submission for the proceedings. Instead, an extended version co-authored by Delia Chiaro appeared in *Meta* two years after the publication of the Forlì Conference proceedings volume(s).

It is difficult to establish to what extent and at what stage in this process the author(s) received feedback from any editorial screening or peer reviewing. A comparison between Nocella's original paper and the joint version, mainly enlarged by the critical review of previous studies, suggests that this was not the case for the shortcomings noted here.

Aside from the fact that peer reviewers might have suggested that Chiaro and Nocella include some key references in their discussion of methodological issues (e.g. Moser-Mercer 1996, Shlesinger *et al.* 1997), referee reports by colleagues with a background in QE research would most probably have pointed out the authors' imprecise use of key terms (e.g. perception); their erroneous criticism of Bühler's analysis; the ambiguity surrounding the criteria in their questionnaire; the missing information on the sampling procedure; and the highly unlikely finding that most interpreters would not work into their mother tongue. Assuming the necessary degree of motivation (cf. Gile and Hansen 2004: 301) and active editorial interest in the reviewer(s), the authors might also have received feedback and recommendations on making their text more focused, particularly in the introductory and concluding sections, and making their statistical analysis more accessible to a wider readership.

Moreover, formal defects of the paper, though not as consequential as issues of research design and interpretation, should not be ignored. A keen reviewer or editor might have noticed, for instance, that the three subheadings in section 2 are on different levels (2.1, 2.1.1, 2.1.2) and thus at odds with the authors' conception of three different methodological perspectives (product analysis, user surveys, interpreter surveys) to which the subheadings refer. (In Nocella's original manuscript, the headings were numbered 2.1, 2.1.2 and 2.1.3, indicating some, albeit unsuccessful, editorial intervention or revision.) A finer point,

which deserves comment only in the context of aspirations to maximum methodological rigor, is the use of unequal scales for the visualization of comparable percentages, as in the authors' Figure 2 (Chiaro and Nocella 2004: 287). More blatantly, in contrast, the consistent misspelling of 'Kopczyński' as "Kopezynski" (282, 293) and other infelicities in the bibliography (Bassnett misspelled; entry for Kopczyński truncated; Kurz 1989 listed as 1988; no data for Tommola's 1995 volume) suggests that the editorial process in this case proved less than fully effective in ensuring optimum standards for the quality of published research.

5.   Conclusion

As illustrated by the present review paper on methodological issues in QE research, the field of interpreting studies reflects an evolution toward higher scientific standards at the same time as leaving ample room for improvement with regard to both analytical rigor and editorial procedure. The aspiration to greater methodological sophistication underlying the paper by Chiaro and Nocella (2004) thus deserves special acknowledgment. The authors point to a number of issues in previous research which deserve more critical attention, and their paper is greatly appreciated as a starting point for this endeavor. Unwittingly, however, Chiaro and Nocella, in their commendably innovative study, also provide material for a critical discussion of methodological rigor in quality research. While offering a convincing demonstration of the power of the Internet and advanced statistical analyses in QE research, the authors give insufficient consideration to various aspects of design and presentation for the paper to meet their own stringent requirements for high-quality research. The fact that these weaknesses were not corrected in the course of the – rather extended – editorial process suggests that quality assurance in the academic publishing process in translation studies is not as systematic and reliable as it could and should be.

   Apart from constructive criticism sought from fellow researchers before submission, the peer review system for scholarly manuscripts is mostly anonymous, and its content and effect remain hidden to the research community at large. That a critique of published papers should be offered here is therefore rather delicate. In the case of Ingrid Kurz, a colleague at the University of Vienna as well as in professional interpreting practice, such published scrutiny and comment might be considered awkward, were it not for her active cooperation to allow a reassessment and elaboration of her data. As regards the work of Delia Chiaro and Giuseppe Nocella, this public feedback ex post facto is offered in support of their welcome ambition to raise the methodological standards of research in this field. Understandably, these colleagues would

rather not see their published work become an object of methodological criticism. However, while we certainly owe respect and appreciation to fellow members of our scientific community, we also owe it to the next generation of researchers, in search of guidance and inspiration for their work, to refine our research models and methodological standards as much as our skills and resources will permit. This paper, and the present issue of *The Interpreters' Newsletter*, will hopefully serve to further promote quality research in our field and help the discipline of interpreting studies earn the academic recognition it deserves.

References

Bradburn N.M. and Sudman S. (1979) *Improving Interview Method and Questionnaire Design*, San Francisco, Jossey-Bass.

Bühler H. (1986) "Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters", *Multilingua* 5 (4), pp. 231-235.

Chiaro D. and Nocella G. (2004) "Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the World Wide Web", *Meta* 49 (2), pp. 278-293.

Collados Aís Á. (1998) *La evaluación de la calidad en interpretación simultánea: La importancia de la comunicación no verbal*, Granada, Comares.

Collados Aís Á. (2002) "Quality assessment in simultaneous interpreting: The importance of nonverbal communication", in *The Interpreting Studies Reader*. Ed. by F. Pöchhacker and M. Shlesinger, London/New York, Routledge, pp. 327-336.

Collados Aís Á., Fernández Sánchez M.M. and Gile D. (ed.) (2003a) *La evaluación de la calidad en interpretación: Investigación*, Granada, Comares.

Collados Aís Á., Fernández Sánchez M.M., Pradas Macías E.M., Sánchez Adam C. and Stévaux E. (eds) (2003b) *La evaluación de la calidad en interpretación: Docencia y profesión*, Granada, Comares.

Elsagir I.M. (2001) "Anforderungen an Dolmetschleistungen im Fernsehen aus Zuschauersicht: Eine Fallstudie", in *Dolmetschen: Theorie, Praxis, Didaktik*. Ed. by S. Kalina, S. Buhl and H. Gerzymisch-Arbogast, St. Ingbert, Röhrig Universitätsverlag, pp. 107-123.

Garzone G. and Viezzi M. (eds) (2002) *Interpreting in the 21st Century: Challenges and Opportunities*, Amsterdam/Philadelphia, John Benjamins.

Garzone G., Mead P. and Viezzi M. (eds) (2002) *Perspectives on Interpreting*, Bologna, CLUEB.

Garzone G. (2003) "Reliability of quality criteria evaluation in survey research", in *La evaluación de la calidad en interpretación: Investigación*. Ed. by Á. Collados Aís, M.M. Fernández Sánchez and D. Gile, Granada, Comares, pp. 23-30.

Gile D. (1983) "Aspects méthodologiques de l'évaluation de la qualité du travail en interprétation simultanée", *Meta* 28 (3), pp. 236-243.

Gile D. (1989) "Le flux d'information dans les réunions interlinguistiques et l'interprétation de conférence: premières observations", *Meta* 34 (4), pp. 649-660.

Gile D. (1990) "L'évaluation de la qualité de l'interprétation par les délégués: une étude de cas", *The Interpreters' Newsletter* 3, pp. 66-71.

Gile D. (1994) "Methodological aspects of interpretation and translation research", in *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. Ed. by S. Lambert and B. Moser-Mercer, Amsterdam/Philadelphia, John Benjamins, pp. 39-56.

Gile D. (1999) "Use and misuse of the literature in interpreting research", *The Interpreters' Newsletter* 9, pp. 29-43.

Gile D. and Hansen G. (2004) "The editorial process through the looking glass", in *Claims, Changes and Challenges in Translation Studies*. Ed. by G. Hansen, K. Malmkjær and D. Gile, Amsterdam/Philadelphia, John Benjamins, pp. 297-306.

Gourevich A. and Mateeff S. (1989) "Study of characteristics of simultaneous interpretation by the method of paired comparisons" (in Bulgarian), *СЪПОСТАВИТЕЛНО ЕЗИКОЗНАНИЕ / СОПОСТАВИТЕЛЬНОЕ ЯЗЫКОЗНАНИЕ / Contrastive Linguistics* 14 (1), pp. 32-38.

Harkness J.A., van de Vijver F.J.R. and Mohler P.P. (eds.) (2003) *Cross-Cultural Survey Methods*, Hoboken NJ, John Wiley.

Kopczyński A. (1994) "Quality in conference interpreting: Some pragmatic problems", in *Translation Studies: An Interdiscipline*. Ed. by M. Snell-Hornby, F. Pöchhacker and K. Kaindl, Amsterdam/Philadelphia, John Benjamins, pp. 189-198.

Kurz I. (1989) "Conference interpreting – user expectations", in *Coming of Age: Proceedings of the 30th Annual Conference of the American Translators Association*, Medford NJ: Learned Information, pp. 143-148.

Kurz I. (1993) "Conference interpretation: Expectations of different user groups", *The Interpreters' Newsletter* 5, pp. 13-21.

Kurz I. (1996) *Simultandolmetschen als Gegenstand der interdisziplinären Forschung*, Wien, WUV-Universitätsverlag.

Kurz I. (2001a) "Conference interpreting: Quality in the ears of the user", *Meta* 46 (2), pp. 394-409.

Kurz I. (2001b) "Mediendolmetschen und Videokonferenzen", in *Dolmetschen: Theorie, Praxis, Didaktik*. Ed. by S. Kalina, S. Buhl and H. Gerzymisch-Arbogast, St. Ingbert, Röhrig Universitätsverlag, pp. 89-106.

Kurz I. (2003) "Quality from the user perspective", in *La evaluación de la calidad en interpretación: Investigación*. Ed. by Á. Collados Aís, M.M. Fernández Sánchez and D. Gile, Granada, Comares, pp. 3-22.

Ločmele G. (2001) "Interpreting norms in Latvia", in *Dolmetschen: Beiträge aus Forschung, Lehre und Praxis*. Ed. by A.F. Kelletat, Frankfurt, Peter Lang, pp. 179-184.

Mack G. and Cattaruzza L. (1995) "User surveys in SI: A means of learning about quality and/or raising some reasonable doubts", in *Topics in Interpreting Research*. Ed. by J. Tommola, Turku, University of Turku, Centre for Translation and Interpreting, pp. 37-49.

Moser P. (1996) "Expectations of users of conference interpretation", *Interpreting* 1 (2), pp. 145-178.

Moser-Mercer B. (1996) "Quality in interpreting: Some methodological issues", *The Interpreters' Newsletter* 7, pp. 43-55.

Pöchhacker F. (1995) "Simultaneous interpreting: A functionalist perspective", *Hermes. Journal of Linguistics* 14, pp. 31-53.

Pöchhacker F. (2001) "Quality assessment in conference and community interpreting", *Meta* 46 (2), pp. 410-425.

Seleskovitch D. (1986) "Comment: Who should assess an interpreter's performance?", *Multilingua* 5 (4), p. 236.

Shlesinger M. *et al*. (1997) "Quality in simultaneous interpreting", in *Conference Interpreting: Current Trends in Research*. Ed. by Y. Gambier, D. Gile and C. Taylor, Amsterdam/Philadelphia, John Benjamins, pp. 123-131.

Shlesinger M. (2000) "Evaluation issues in interpreting. A bibliography", *The Translator* 6 (2), pp. 363-366.

Vuorikoski A.-R. (1993) "Simultaneous interpretation – user experience and expectations", in *Translation – the Vital Link. Proceedings. XIIIth World Congress of FIT*, vol. 1. Ed. by C. Picken, London, Institute of Translation and Interpreting, pp. 317-327.

LOOKING QUALITY RESEARCH IN THE EYE:
ARE WE BEHOLDING THE SPLINTER AND IGNORING THE BEAM?

Delia Chiaro
University of Bologna at Forlì

Giuseppe Nocella
University of Bologna at Cesena

1.   Introduction

The aim of Chiaro and Nocella's article (2004) which has been much quoted by Franz Pöchhacker in this issue, was not to boast their competence in slick and sophisticated statistical techniques. Neither was it to be excessively harsh on researchers who were, after all, pioneers in bringing survey techniques to Interpreting Studies (IS) in the first place. Chiaro and Nocella's aim, nonetheless, was to underscore a certain lackadaisical attitude rampant in several attempts at questionnaire based quality research (QBQR) in this field. However, as Pöchhacker casts doubts on the reliability of their study, the authors cannot do otherwise but jump to their own defence[1]. In fact, much as Pöchhacker's re-elaboration of existing data does him honour (this issue: 150-154), as we intend to demonstrate in this essay, it cannot, and indeed does not disguise the existing general lack of methodological expertise and rigour present in many attempts at QBQR in IS. Furthermore, while grateful to Pöchhacker for having pointed out a series of shortcomings in their work, Chiaro and Nocella wish to accept total responsibility for each and every weakness, rather than take refuge behind the shield of poor refereeing. Presumably all attempts at research have their strengths and weaknesses. What is important is that the latter do not outnumber the former, otherwise our incessant quest for knowledge could well go awry.

Moreover, the present authors would like to highlight the fact that they are flattered to see that their infinitesimal contribution to the field has triggered off an animated response by such an eminent scholar. In fact, Chiaro and Nocella

---

1   The authors are grateful to the editorial board of *The Interpreters' Newsletter* for having given them the opportunity to respond and go into print in the same issue in which Franz Pöchhacker's article appears. The editorial decision to allow two lesser known researchers to respond so openly to such a renowned scholar is evidence of transparency and a true credit to the journal. Furthermore, they would also like to express their appreciation of Pöchhacker's sense of fair play and sportsmanship for having given them prior access to his critique and consequently the opportunity to elaborate the present reply.

believe that IS could benefit from some lively, albeit constructive discussion, a common practice in other scientific discourse communities but, until now, rather lacking in this one. In other words, with the present discussion, the authors welcome the opportunity to defend their work and intend, good heartedly, not only to stick to their guns, but also (hopefully) to trigger off a wider debate.

Taking Pöchhacker's re-visiting of QBQR in this issue as a starting point, we too will follow the same path and respond to his critique while simultaneously providing our own (over)view, where relevant, of other, similar existing research. In addition, similarly to Pöchhacker, most of our revisiting will also be carried out in practical methodological terms, with special emphasis on research hypotheses underlying previous studies, the nature of research design and finally statistical procedures for the analysis of survey data. We will also (re)visit Pöchhacker's detailed reanalyses of the work of his colleagues Bühler and Kurz and naturally bear out his critique of our own methods, results and conclusions. The above argumentation will be arranged in two major sections, the first regarding a detailed discussion and defence of what we consider to be a series of unjust criticisms of our work brought to light by Pöchhacker. In a separate section we will unearth a number of significant flaws in the works of others that Pöchhacker appears to have overlooked after which, we will attempt to demonstrate beyond reasonable doubt that our claim that "research undertaken so far (in QBQR) is surprisingly lacking in methodological rigour" (Chiaro and Nocella 2004: 278) is anything but inaccurate. However, we feel obliged to underscore the fact that we are not taking issue with the worth of research in IS *tout court.* We are not disputing the wealth of existing descriptive and experimental work in the field. Our criticism was originally, and still is, limited to QBQR alone.

Nonetheless, before embarking on this enterprise, the present authors would like to begin by seriously challenging the suggestion that survey based research offers "a working method that can readily be adopted also by less experienced investigators" (Pöchhacker this issue: 143).

2.   A working method for less experienced investigators?

It would appear that after the well known work of Bühler (1986) and Kurz (1989), survey work in IS has become trendy and *à la mode* as more and more researchers jump onto the questionnaire bandwagon (e.g. Meak 1990; Vuorikoski 1993; Mack and Cataruzza 1995; Moser 1996; etc.). Yet those who think that developing a questionnaire is simply a matter of sitting at a desk and thinking up a list of questions are mistaken. Questionnaire development is a demanding and challenging process which requires time and energy spent first and foremost in preliminary qualitative research methods. These consist of

preparatory processes such as setting up and conducting in depth interviews and/or focus groups or adopting projective techniques such as association, completion, construction or expressive techniques (Malhotra 1996) which provide essential input for setting up a survey. A glance at the extensiveness of the literature on interview techniques alone can provide us with a fair idea as to how far such pre-survey qualitative methods have developed, while closer examination reveals how complex such practices actually are (Malhotra 1996; Tull and Hawkins 1993). And even if the principal investigator wears two hats and is also an experienced practitioner, as often appears to be the case in IS, this should not exempt them from this preliminary phase. In a certain sense this stage is even more important when the investigator is a practitioner because a researcher-cum-practitioner by default may well be inclined to increase the "observer's paradox" (Labov 1972)[2] as such a researcher will be even more lacking in the psychic distance required for unbiased study.

Let us now turn to what we shall crudely define the second stage in questionnaire design. Once researchers have obtained sufficient input from an adequate number of external informants to enable them to outline a questionnaire, they will need to know exactly what, as well as how, information is to be collected from the population under examination. This may sound trite and obvious, yet poor judgment at this stage may lead to results that are not relevant to the purpose of the study, or else that are incomplete. Questions require choosing the appropriate measurement scales, formatting and careful wording, as well as proper sequencing and layout (Aaker *et al*. 1995: 291); tasks which are easier said than done. Less than careful framing of questions can lead to distorted results. The following anecdote should illustrate the point we are trying to make:

> Two priests, a Dominican and a Jesuit, are talking about whether it is a sin to smoke and pray at the same time. After failing to reach a conclusion, each goes off to consult his respective superior. The next week they meet again:
> "Well, what did your superior say?" asks the Dominican.
> "He said it was all right", the Jesuit responds.
> "That's funny", replies the Dominican, "my superior said it was a sin."
> "What did you ask him?" inquires the Jesuit. "I asked him if it was all right to smoke while praying"', says the Dominican.
> "Oh," says the Jesuit, "I asked my superior if it was all right to pray while smoking!" (Dillon *et al*. 1994)

---

2   We would like to point out that we are not using the term in its strictest Labovian meaning but in its broader sense to embrace all types of biases which can occur owing to the relationship between researcher and informant.

If a mistake occurs in a measurement scale, problems are bound to arise. Several drafts as well as extensive piloting are essential before arriving at a final version. Once satisfied with the instrument, aspects such as deciding upon a method of administration (e.g. face to face, telephone, mail etc.), selecting a random[3] sample, choosing *a priori* the statistical technique to test the research hypothesis, elaborating raw data and in the final stages, interpreting results are not aspects to be taken lightly. We believe that many of the shortcomings inherent to many such studies in IS have been due to this very underestimation of what designing a survey instrument actually entails. In fact, as far as we know, there are no Translation and Interpreting faculties which offer foundation courses in empirical research methods at either undergraduate or postgraduate level, so it is understandable that interpreters often lack in necessary know how. If the single investigator is unable to see beyond data collection they may well be walking up a blind alley. Thus investigations involving researchers with different types of expertise and the adoption of an interdisciplinary attitude to IS research can only be fruitful, as long as the single researchers do not work independently and are involved in every single stage of the study. In other words, a statistician brought in *a posteriori* is unhelpful. A statistician (or better, a researcher trained in methodology and statistical analysis) at this point will indeed be capable of elaborating existing data, but his or her cooperation would have been more productive at the stage of research design. What we are trying to say is that researchers should already have in mind the kind of statistical tests they want to carry out on resulting data in order to test the initial research objective <u>before</u> carrying out the survey. "Here's my data see what you can do with it" is out of order in serious empirical research.

Last but not least, one of Pöchhacker's many objections to our work is that "peer reviewers might have suggested that Chiaro and Nocella include some key references in their discussion of methodological issues (e.g. Moser-Mercer 1996, Shlesinger 1997)" (this issue: 162). With all due respect to the two studies which Pöchhacker suggests should have been included in our discussion and which apparently slipped the mind of the journal's referees, we would like to state that we preferred to refer the reader to authors specialized in qualitative and quantitative research methods (i.e. Aaker *et al*. 1995; Hair *et al*. 1995 and Schiffman *et al*. 1981) rather than scholars of interpreting. This was not to belittle the two renowned scholars in question but simply because surely it is IS which is drawing from well established methodologies of the Social Sciences rather than vice-versa. Now what we were suggesting from the start was that IS should look more closely at the rules of qualitative and quantitative research methods which were born and bred <u>outside</u> this discipline. Interesting as both

---

3   "A random sample allows a known probability that each elementary unit will be
      chosen." (Lapin 1990: 104)

articles may be, they appear to remain, however, within the somewhat self-referential boundaries of IS.

Before beginning our discussion proper, we would like to raise one more small issue. Over and over again we read that investigating quality in interpreting is not an easy task due to the huge number of variables involved not only in the process itself, but also in conditions which regard operators, users and even the contractors of the service (e.g. Shlesinger 1997; Garzone 2003 etc.). The general idea which comes across to the reader is that dealing with the enormous heterogeneity of circumstances in and around interpreting verges on the insurmountable. This may well be true and we certainly do not wish to claim that quality research in IS is unproblematic. But is not apparent insuperability typical of scientific enquiry? Was Watson and Crick's model easy to identify? And what of the excogitation of a formula that shows that distance and time are not absolute? And discovering penicillin? The list of seemingly intractable problems is endless. But is it not this very complexity that is what makes research fascinating and irresistible?

> A researcher is a detective or a spy who is out to discover or uncover something that is in some way, unnoticed, hidden, secret or problematic. Researchers, like detectives, find that their sources sometimes lie, sometimes offer conflicting stories, and sometimes behave in baffling ways. That is why research is so exciting … (Berger 1991: 7).

2.1.   A harsh critique or calling a spade a spade?

Pöchhacker accuses Chiaro and Nocella of offering a "rather harsh critique" of the work of his Viennese colleagues Bühler (1986) and Kurz (1989). We hereby express regret for our lack of tact and for having couched our criticism harshly. Our exact words were: "Unfortunately, a substantial shortcoming of this particular study (Bühler) is that the mean was used as the descriptive statistic for analyzing and discussing data[4]. Percentage, mode or median would have described the data more correctly." (Chiaro and Nocella 2004: 283) and, with regard to Kurz, our claim was that "percentage would have given a better comparison" (282). Admittedly, each turn of phrase could be seen as being rather heavy handed and inconsiderate. We could have perhaps been less direct and softened matters slightly by using words to the effect of: "Let us see what would have happened if the median/percentages had been used instead?" or possibly relegating the entire issue to a couple of footnotes. But would this have really changed anything if the analyses did not have a clear direction? More

---

4   For a detailed discussion on the concept of measurement see (Aaker *et al*. 1995: 56 and Tull 1993: 309).

seriously however, Pöchhacker criticises the present authors for "their erroneous criticism of Bühler's analysis", well, for the sake of diplomacy, much as we have tried to fault our analysis, mathematics is not an opinion and we will show that it is not in the least "erroneous" (4.1).

And if we were "rather harsh", Pöchhacker's critique is hardly tender. The use of subtle irony (or biting wit?) in the title of the essay[5], or indeed in the heading of the section entitled "Into print?" is not exactly gentle either. A question mark can be every bit as cutting, if not more so, than a word. And here we are talking in terms of our academic credibility. Are we certain that the words spent on Kurz and Bühler deserve such a scathing attack?

Our extensive, hands on experience in questionnaire based surveys (admittedly in other fields of research) led us both (foolhardily it would appear) to try our hand at applying our expertise to IS and also to feel (erroneously it now seems) that we had something to contribute to other less experienced researchers trying their hand at such surveys. Any impression of overconfidence surfacing from our study was quite unintentional, and by the same token, we wish to assert that, despite our experience, we are perfectly aware of how very little we do know and how much there is for us still to learn. However, we do feel that Pöchhacker is actually implying that as our study was less than perfect we should not have criticised others. And in a sense he is right. In an ideal world casting stones should be restricted to those without sin. Now we dared cast stones despite being less than immaculate ourselves. But the point is that our offences were venial rather than mortal and that most of the accusations for which we have been charged are fallacious. We sincerely believe that Pöchhacker's critique is disproportionate and that the faults in our work are in no way connected with methodological mishandling and will demonstrate that, in contrast, other studies quoted by Pöchhacker contain major inadequacies.

However, having said that, it would be a true pity if all the thought and energy which have gone into both Pöchhacker's critical assessment of Chiaro and Nocella's essay and this present retaliation were to degenerate into a lengthy scuffle of *quid pro quo*.[6] Rather it would be desirable that both

---

5   The present authors would like to point out that 'Revisiting and reanalyzing the work of Bühler and Kurz and replying to the work of Chiaro and Nocella on quality research' would have been a more fitting title to the essay to which they are responding. Nevertheless, as a scholar of Humour Studies, Delia Chiaro cannot help but relish in the clever and, admittedly, successful inherent paronomasia coined by Pöchhacker for his title.

6   In line with Pöchhacker's anecdotal style it is also perhaps worth mentioning that ample correspondence via e-mail as well as a lengthy and affable telephone conversation between the two parties involved in this discussion had occurred prior

Pöchhacker's critique and the present defence should serve to shed light upon what we still believe to be a shadowy area in IS, ie research design and implementation in QBQR and thus promote ample and, above all, <u>constructive</u> discussion.

3.   Beholding the splinters: interpreters on the Web

Having been accused of several deficiencies, we now intend to tackle each and every one throughout the course of this comeback. Although these faults appear in a somewhat jumbled order in Pöchhacker's critique, we have tried to disentangle them and present them, together with our rebuttal, in a logical order so as to facilitate both the reader and the force of our argument.

3.1.   Conceptual frameworks and operational definitions

3.1.1. Key terms

The liberal use of the word 'perception' is dangerous. Rather like cigarette smokers, consumers of the term should be made aware that its use may well present several hazards. In fact, Chiaro and Nocella dared to adopt the term liberally without defining it in operational terms and, as a result, have not only been accused of "imprecise use of key words" (this issue: 162) but, perhaps more significantly, also appear to have been severely misunderstood.

   In psychology and in the cognitive sciences the word 'perception' refers to the concept of acquiring, interpreting, selecting and organizing sensory information:

> The sense organs provide our brain with a steady flow of information about our environment and the brain's task is then to take this raw material and use it to help us make sense of that environment through the process of *perception*. And the brain does its job so smoothly and well that we're not even aware of what it does. (Statt 1997: 46)

   Now we have been, quite appropriately, criticized for our unclear use of the term in our study. And this is one criticism which we openly acknowledge. However, our use of the term 'perception' was quite deliberate. We were in no way confounding 'perception' with interpreters' 'generic expectations' as suggested by Pöchhacker (this issue: 158), such confusion would indeed have "fallen short of the art" (Pöchhacker this issue: 158). Besides, why should we

---

> to going into print. It would not be unfair to say that communication concluded in a reciprocal decision to remain united in our diversity.

confuse perception with expectations? Is the study of expectations in IS compulsory? Why cannot perception be taken as a starting point instead? While aware of the fact that there is a strong tradition of investigations into expectations in IS (e.g. Kurz 1993, Moser 1996 etc.), expectations were not what Chiaro and Nocella were investigating at all, yet Pöchhacker seems to imply that wanting to look at interpreting from a different angle is not viable. Or rather that what we were really studying were expectations. Well, let us put the records straight and underscore that we were not seeking to access respondents' awareness or judgment of performance. What we were trying to establish was interpreters' *consciousness of mental selections which they constantly make* (our emphasis). To put it another way, we were plainly asking respondents to consider and attempt to untangle a complex mental process and express their awareness in terms of how they weighted a set of essential criteria against each other in their effort to transform incoming sensory information into verbal output in a different language. If this was erroneously confused with the expectations of a final product we trust that we have now clarified our position and again are obliged to the Editors of *The Interpreters' Newsletter* for having given us the opportunity to make amends. Incidentally, is it not also the case that 'expectations' are more relevant when one is interviewing end-users, less so when the subjects are interpreters themselves? Surely, regardless of all, 'perception' seems a more appropriate term to refer to self-monitoring by an interpreter? Over and above this, our essay contains a perceptual map (290) which displays interpreters' mental image of the various criteria. Without wishing to be tautological, a perceptual map represents perception and not expectations. How can this have been construed as confusion on our behalf?

Furthermore, we are also accused of not having distinguished between research on "generic expectations […] direct assessment, or judgement […]" (Pöchhacker this issue: 158). Needless to say, this omission was not because we didn't know the difference or because we had deliberately decided to ignore the issue. Yet, operational definitions of these terms are nowhere to be found in any of the QBQR we have examined. Moser, for example, freely uses the term perception (1996: 148, 159) imprecisely when in effect what he was investigating were "judgements, needs and expectations" (145). We are criticised for using it. He gets away with it scot-free. Again, Mack and Cataruzza suddenly introduce the term with no further definition too (1995: 45) and more recently Garzone (2003: 23-24) also adopts it freely. Are we to be the first to be accused of a lack of operational definitions? Since Moser-Mercer introduced the concept of "optimum quality" (1996: 44), the issue of attempting to define the concept of quality itself any further appears to have slipped almost everyone's mind until quite recently (Kurz 2001: 395 and 2003: 17-18). Again, the concept of multi-dimensional models of quality begin to be mentioned

(Garzone 2003: 23) while the only serious attempt at modeling the multifaceted issue of quality in interpreting has been produced by Gile (2003: 110).

Now, let us turn to the term 'experiment', a word which, according to Pöchhacker, we have used improperly in reference to Kurz's survey (Chiaro and Nocella 2004: 282). For the sake of argument, let us accept that we did use the term 'experiment' inappropriately. By the same token, Pöchhacker sets off our work against the controlled laboratory studies of Gourevich and Mateeff (1989)[7]; Collados Aís (1998, 2002); and Garzone (2003) in his defence of sound QBQR. Is this because he ignores the difference between an experiment and a survey? Or does Pöchhacker wish to widen the present dispute to colleagues adopting different methods by paying them homage? We repeat, we were/are only criticizing QBQR. Collados Aís and Garzone have carried out laboratory style research with which we have no bones to pick. And yes, we are aware of the difference between an experiment and a survey; the former is:

> A controlled situation in which the experimenter systematically changes the values of one or more variables [the independent variable(s)] to measure the impact of these changes on one or more other variables [the dependent variable(s)]. (Tull 1993: G-6)

while the latter refers to the "systematic collection of information directly from respondents" (Tull 1993: 61). Furthermore,

> The important distinction between the survey and the experiment is that the survey takes the world as it comes, without trying to alter it, whereas the experiment systematically alters some aspects of the world in order to see what changes follow. (Simon 1969: 229)

Or would Pöchhacker prefer us to adopt Vuorikoski's vague definition of experimentation as something through which "… it is possible to arrive at clear causal inferences" (1993: 318)?

Moreover, lexical networks are created within texts by the writer and false or close synonymy are simply textual strategies of reiteration (for ample discussion see Halliday and Hasan 1976: 278-279 and Hoey 1983). For the purpose of textual cohesion special synonymy with words which are not 'normally' synonymous are often created – if such a thing as 'normal' or absolute synonymy exists. Of course, this not only applies to our specific use of terminology but also to the other researchers who we have quoted above as a counter-argument (see Moser, Mack and Cattaruzza and Garzone's use of the term 'perception' above). A similar argument can just as easily be constructed

---

7   We do not have access to this paper and are thus relying on Pöchhacker's description of the study (Pöchhacker this issue: 160).

for the criticism of our use of the term "intonation" as a synonym of "fluency of delivery" (Pöchhacker *this issue*: 159 note 6).

3.1.2. Conceptual frameworks

Although Pöchhacker has understood that "Chiaro and Nocella base their review section on a two-fold distinction between product analysis and 'field work (based upon the results of questionnaire surveys)' (this issue: 158), a more accurate reading reveals that our distinction was, in effect, threefold and that we had we split QBQR into "analyses of the product" (Approach number 1), field work on end-users (Approach number 2) and field work on interpreters (Approach number 3) (280)[8]. Furthermore, we are also accused of not having considered Vuorikoski's (1993) "fourfold distinction" created "specifically for the purpose of research on interpreting quality" (Pöchhacker *this issue*: 158). Pöchhacker is quite right, we do indeed ignore this "fourfold distinction". However, the reason we have done so is simply because we were unable to locate this distinction. The only mention of anything remotely "fourfold" in Vuorikoski's essay are the multi research methods she sets out to discuss. Therefore, the comparison Pöchhacker makes between our study and Vuorikoski's is quite vain. Vuorikoski's "fourfold distinction" regards the application of diverse research methods simultaneously. Our threefold distinction regards the ways in which quality research had been carried out so far in IS. In fact we state that

> …attempts at empirical research carried out so far on quality interpreting reflect these three perspectives [supplier of service, client and service itself] and have thus been based on a) analyses of product; b) field work based upon…end user perception and c) … interpreter perception…of interpretations in general. (Chiaro and Nocella 2004: 280).

We know full well of the existence of multimodal research methods but to the best of our knowledge, no attempts have as yet been made to adopt them in QBQR.

We are next charged with not providing sufficient rationale for choosing to examine interpreters rather than end-users. The pros and cons of one or the other have been argued at length in the field (Kurz 1993, Moser 1996 etc.) and we were (mistakenly it seems) convinced that we had argued our case adequately by explaining that interpreting is a service which is used by clients who presumably require assistance in understanding a language with which they are

---

8   Having argued about the meaning of perception (3.1.1.) it seems clear that
     Approaches 2 and 3 are diverse.

not familiar. This lack of knowledge of the source language renders end user quality judgement of the service of interpreting difficult as clients would be unable to judge a basic characteristic such as fidelity to the original (Chiaro and Nocella 2004: 281-282). Judging the quality of an interpretation is quite different from that of judging a regular marketable good. (We suggest that those convinced by our argument skip the rest of this section and move on to 3.2). A housewife asked to judge the quality of a pot of jam, for example, has a range of tangible and highly perceptible characteristics upon which to base her evaluation. The colour of the jam, how much it costs, it's shelf life, nutritional information on the label, packaging and, last but not least, it's flavour. In fact, we state that "Interpreting is a service and according to Economics a service is an intangible and non-transferable economic good and thus quite distinct from a physical commodity. Therefore the special nature of interpreting makes its evaluation difficult for people who consume the service but know very little about it" (281). Of course, a conference delegate can judge a variety of features connected to an interpreter's voice quality, he or she can judge clarity and coherence of speech as well as their command of the language. But a genuine delegate is likely to be hard put to be able to judge the fidelity of an interpreted speech with the original. Thus our choice of respondents naturally fell on interpreters, and with two seminal studies to rely on, using Bühler as a springboard seemed a natural choice.

3.2. The survey instrument

3.2.1. Design

Pöchhacker's first incursion regarding our survey instrument concerns the fact that we did not discuss the reasons for not adopting Bühler's criteria *tout court*. In fact, we adapted 7 criteria and we included a new one, namely "absence of stress" which twenty years ago may not have been an issue for Bühler's interpreters. And here Pöchhacker has a point so we shall immediately make amends. The input of the experts who helped us construct our instrument together with our own common sense led us to accept that Bühler's criteria "pleasant appearance" and "poise" could perhaps be cut as they possibly do not contribute to the quality of an interpretation. As for Bühler's inclusion of "positive feedback from delegates", this was considered to be a criterion which is not part of the interpreter's self-perception and therefore jars with the truly linguistic and extra-linguistic criteria which we had decided to examine. Again the concept of "reliability" was excluded for similar reasons. Our sample of interpreters were asked *verbatim* to "rank (a list of) factors contributing to the quality of interpreting". The concept of reliability was felt to be in a

hyperonymous relationship with the other factors. If an interpretation is of good quality it follows that it can be considered reliable precisely because it is made up of a positive relationship between the criteria listed. Finally, we changed Bühler's "completeness of interpretation" to "completeness of information" and "thorough preparation of conference documents" to "preparation of conference documents" upon the advice of our informants.

We also accept Pöchhacker's criticism of our somewhat cavalier description of how we constructed our instrument basing it on "several interviews" and "endless brainstorming sessions" with interpreters (283). With neither of us being a practitioner we had to look outwards and seek professionals and academics for help in devising our instrument. Few of the well known studies in QBQR appear to have bothered with any preliminary research or if they did, they certainly do not mention it in their work. Moser (1996) is the only scholar to describe a preparatory phase of his survey but we are sure that there can be no disagreement that he is as offhand as we are in his description of this stage.

Furthermore, we also acknowledge the fact that our questionnaire did not contain a request for information regarding respondents' specialized fields of expertise. Neither was information solicited regarding working language combinations. But why regard the lack of such information as methodological deficiencies? The exclusion of queries to elicit such data were choices which we intentionally made and not slips of the mind.[9] Firstly we had to keep the questionnaire as brief as possible as, in the days before the advent of widespread broadband connections, we did not know how much time people could spend on line, thus we opted for essentiality. More simply, interpreters may simply not want to waste their time filling out endless questions. Furthermore, the aim of the study was to provide a (reliable) springboard for further study. Thus, what we were searching for was broad-spectrum data. In other words, what we were interested in was obtaining a general idea of what the average conference interpreter perceived as being important and less important in his or her choices. In fact, our study was devised to be a starting point which might act as a spur for more particular, fine tuned studies. If, generally speaking, *n* conference interpreters reported that they perceive criteria *x* to be important when working to and from languages *a*, *b*, *c* or *d*, (to put it another way in and out from *any*

---

9   The exclusion of such data from our work was deliberate, however, let us imagine that we had simply skipped this variable for a number of reasons which could range from sloppiness and forgetfulness to sheer ignorance. Let us remind readers that the seminal works of Bühler and Kurz contain no socio-demographic variables at all, while one of Moser's is based on guesswork (for a discussion see section 4.5.). Also, one might wonder where one should stop when it comes to assembling socio-demographic information, surely you can always think up another variable that might be potentially relevant and that had not been taken into account!

non-specified language), it would then be interesting to see how the same test stands to trial when applied to specific language combinations. As Pöchhacker suggests "interpreting styles may differ from one sociocultural context to another"(this issue: 158) – well let's go out there and support this claim. Or else reject it. Who knows, perhaps interpreters' perception of choices they make may even prove to be universal. After all, surely scientific experimentation starts from the general to the particular rather than vice-versa?

As for challenging the language of the questionnaire's administration, the use of English was again a conscious choice. Unlike other surveys in which we have been engaged where the issue of language was indeed a concern, here we were looking for all-encompassing generalized data. Furthermore, we are quite certain that we were not erring in an excessive credence in the linguistic colonialism of the English language by assuming that the hypothetical average interpreter would be likely to have a working knowledge of English.

And in response to one of Pöchhacker's most critical charges, to wit, the fact that the majority of our respondents claimed that they did not work into their native tongue, again, "baffling" as this may sound, this is how the sample responded and, like it or not, the information needs to be taken at face value. Should we have excluded these findings just because he is not happy about them? Or should we have manipulated our data and claimed the contrary? We would also like to take issue with Pöchhacker's charge that the question which led us to the above claim was due to a "poorly worded questionnaire item" (Pöchhacker this issue: 159 note 7). In fact, Pöchhacker is basing this claim on an early draft of the instrument which we had sent him, and not to the final pluri-piloted version in which the wording had been improved and which we were unable to send him. However, if Pöchhacker is unhappy with our results and their subsequent interpretation, we suggest he rerun the test on a different sample.

3.2.2. Distribution

Next, we are accused of having given "an all too sparse description of their sampling procedure" (this issue: 159). Way back in 2000 when we conducted the survey, Web based questionnaires were indeed a novelty and today, the way we sampled at the time makes us both smile at our naïve techniques. What we did, which would be highly irregular today (as well as being almost impossible with the number of fire-walls and anti-spamming programs which have been widely installed in computers), was to spam an invitation to visit the site containing our questionnaire to a number of mailing lists of conference interpreters world wide. These lists were collected by networking and included a list of EU interpreters, and national associations across the world. At this point

Pöchhacker could easily argue that our sample is unreliable because it was restricted to Internet users and that we only invited about 1000 interpreters to participate. True, there surely are more than a 1000 interpreters in the world and of course we are aware that we did not contact every single one of them. But the point is that we were <u>sampling</u> and not contacting the entire population of interpreters. We are well aware that the 1000 interpreters we contacted had not been selected according to the table of random numbers. In other words, we cannot be sure that every interpreter with an e-mail address received our invitation to participate and that others did not receive the information twice. Nevertheless, we would like to call attention to our good faith by highlighting that neither of us are practitioners and between the pair of us we only knew about a score of interpreters at the time. This means that we were unable to use personal networking to create our sample, so at least Pöchhacker should give us our due and allow us to go down in IS history as being the first QBQR researchers who did not depend on a self selected, albeit a convenient, sample.

3.3.   The mathematics behind the scores

One comment of Pöchhacker's which the authors (partially) agree with is the lack of accessibility of the statistical analysis. Or rather, for an IS readership the statistics may well be inaccessible whereas in fields such as psychology, economics and marketing research there would no need to explain the mathematics behind well-known techniques unless data is being modelled introducing innovative elements. The "sum of the scores"[10] (Chiaro and Nocella 2004: 288) is a descriptive statistic, so if there is a need to explain it we should clarify every descriptive statistic from mean to mode and from standard deviation to range and so on.

Finally, Pöchhacker introduces "A finer point, which deserves comment only in the context of aspirations to maximum methodological rigor" and criticizes our "use of unequal scales for the visualization of comparable percentages, as in the authors' Figure 2 (Chiaro and Nocella 2004: 287)" (this issue: 163) We really do not understand this comment. How can the scales be unequal if all the data summarized in Figure 2, labelled "Distribution of the degree of importance given to each linguistic criterion",[11] was obtained from the same rank scale. Instead of presenting our data in one crowded graph which may have been confusing, we simply split the data into three different line graphs to allow

---

10   The sum or total of the values, across all the cases with non-missing values.
11   A typo which Pöchhacker did not spot is the plural form CRITERI**A** which appears instead of singular CRITER**ION** above figure 2.

readers to follow the distribution of the 9 criteria under investigation more easily.

4. Ignoring the beams

What follows is a brief overview of the QBQR quoted by Pöchhacker in this issue. We wish, however, to begin by reiterating that most of the contributions Pöchhacker mentions and sets off against our own work present a series of gross methodological deficiencies. Secondly, we would like to declare that we are somewhat uncomfortable with having to draw attention to these studies, but having had our own work publicly scrutinized for what are patently much lesser faults, we cannot but support our claims that "research undertaken so far is surprisingly lacking in methodological rigour" (Chiaro and Nocella 2004: 278) by pointing to these examples. In fact, if we had originally spoken of "uncertain methodological principles" (279), a sense of delicacy had led us to go no further and remain somewhat vague. Now, while we are aware that our lack of humility in criticizing others has led to the disparagement of our own work, what still remains a mystery is why Pöchhacker should consider the splinters in our eyes and yet demonstrably overlook the beams in those of others. Thus the necessity to safeguard our own faces internationally now leads us to bring these beams to light. This will be done following a chronological order and restricting the review to the field of surveys pertaining to quality alone.

Interestingly, most of the surveys which Pöchhacker plays off against Chiaro and Nocella's Web survey reveal a strikingly similar series of faults which principally regard three areas, namely the sampling frames, the measurement scales adopted and the choice of statistical test. We will briefly tackle all three with reference not only to the work of Bühler and Kurz, but also to that of Vuorikoski, Mack and Cataruzza, Moser and finally Pöchhacker.

4.1. Bühler

Pöchhacker appears to be puzzled by the fact that the present authors did not take issue with regard to the rather small sample of 41 interpreters who returned questionnaires in Bühler's well-known study. We really see no cause for bewilderment simply because this study as most of the others regarding QBQR is simply descriptive in nature i.e. there is no use of any technique of inferential statistics. As a result, in absence of any use of probability theory, there is no need to argue about sample size[12], Bühler was not inferring from the sample to

---

12 Most text books on general statistics and methods in social research tackle the issue of sample size (i.e. budget constraint, sampling error, interval estimation, etc.).

the population but simply commenting percentages on the criteria investigated. On the other hand, what Pöchhacker should really be asking is why Bühler decided to use a sample size of 41 as this cannot be deduced from her article. Why indeed 41? Budget constraints? Time constraints? Rule of thumb? Or was sample size determined according to statistical theory considering factors such as reliability, confidence, tolerable error and precision?

However, what we would like to highlight once more in Bühler's explorative work, the importance of which is still relevant in IS today, regards the way in which these criteria were assessed. In order to test the importance of these criteria Bühler adopted the following measurement scale:

Highly important, Important, Less important, Irrelevant

The fact that most of her respondents only chose the two highest points of the scale: "highly important" and "important" was what led us to question the validity of the instrument. This is also confirmed in Pöchhacker's graphic effort (bar chart this issue: 145, 148) to reanalyse Bühler's data. Instead of showing what he claims to be a "clear cut differentiation" (this issue: 145) it shows a clearly skewed sampling distribution with a tail on the right for almost all the criteria investigated. So in the light of this observation we asked ourselves, was the measurement scale adopted the most appropriate? Had the questionnaire been properly piloted? Furthermore, why use a scale which is unbalanced towards the importance of linguistic and extra-linguistic criteria with no mid-point of neutrality and no escape route for those who did not know what to answer?[13] How can we verify whether most of these items are really so important or highly important to interpreters?

In other words, we asked ourselves whether interpreters should evaluate each item independently or whether they should play off the items against each other. Therefore, the question of how to measure the importance of these criteria led us to consider sets of non-comparative scales (e.g. continuous rating scales, itemized rating scales such as Likert scales, semantic differential scales and staple scales) and comparative scales (e.g. paired comparisons, graded paired comparisons, constant sum scales and rank order scales) in order to decide whether to modify the scale used or to choose a new measurement scale.[14] With the intention of testing whether interpreters could discriminate in terms of importance, a rank order scale seemed to be the most appropriate because interpreters could compare these criteria in one fell swoop according to their

---

13   An example of such a scale could be: 'Highly important', 'Important', 'Neither important nor unimportant', 'Unimportant', 'Irrelevant', 'I don't know'.

14   For a detailed discussion on measurement scales see Aaker *et al.* (1995), Maholtra (1996), Tull and Hawkins (1993).

level of importance. What is more, as the cognitive exertion involved in choosing from 16 factors would have been extremely high, we split the criteria under investigation into two separate sets: linguistic criteria and extra-linguistic criteria. Hopefully, this should have somewhat eased respondents' efforts at selection. If there are any shortcomings in the chosen scale, they could be linked to the fact that the process of selection was controlled by an algorithm in *Java script* which did not allow interpreters to give the same level of importance to two or more factors. However, having noted in the initial piloting stages of the project that nobody took issue with this characteristic it remained unchanged throughout. At the end of the sampling a total of three interpreters complained about this restriction in choice.

For the sake of argument regarding our criticism of the mathematics employed by Bühler and challenged by Pöchhacker, we must remind readers the objects were measured on an ordinal scale which was also unbalanced, thus

> Because we don't know the amount of difference between objects, the permissible arithmetic operations are limited to statistics such as the median or mode (<u>but not the mean</u>). Our emphasis. (Aaker *et al*. 1995: 257)

Finally, we note in passing that, Bühler's survey contains no information about the socio-economic characteristics and professional experience of the interpreters who took part in the survey.

4.2.   Kurz

In order to defend the work of Kurz, Pöchhacker compares her sample with Moser's (1996) arguing in favour of the greater validity of the former sample of 124 end users which had been collected at only three conferences while Moser had to gather data at 84 different conferences in different parts of the world in order to collect a final sample of 201 end users. Clearly, being based on 84 extremely diverse conferences, Moser's sample could[15] surely have been more representative than Kurz's sample. But this is not the point. Let us examine the precise date in which data was gathered at Kurz's conference on general medicine. According to the original publication of this study, Kurz gathered her data in 1988 (Kurz 1993) while according to the reprinted version it was apparently collected in 1989 (Pöchhacker 2001). After the ruthless critique of the editorial process of the journal *Meta* with regard to Chiaro and Nocella, how could such a significant detail have escaped Pöchhacker's notice? Are we to

---

15   We are adopting a tentative conditional form because when we discuss the study by Moser we will illustrate why his sample is equally unrepresentative.

assume that the date was erroneous in the 1993 article? If so, why did not Pöchhacker add a footnote to clarify the point? Was 1989 the correct date of the study or was it a typo? Or is this to remain a mystery? However, as we referred to Kurz's original study (1993) we wondered how much time had lapsed between her three data sets. It is worth remembering that Kurz compares the data gathered from Bühler's 47 interpreters in 1986 with data from her own three samples collected between 1988 and 1989. Could it be that there were almost two years between Kurz's three sub-samples and almost four years between Bühler's study and the Council of Europe meeting? Now the point is whether sample size is so important with such a large time gap in sequential sampling? Sequential sampling is a technique adopted when taking decisions (usually in business and marketing) which depend upon laws of probability. If sampling had been deliberately sequential in nature we would need to know whether Bühler's interpreters and Kurz's end users inhabited an immutable world or a dynamic one. Surely four years must have brought a minimum of technological and scientific progress to the world of interpreting. Now if progress has zero impact on the world then a comparison between samples collected at different points in time may be plausible. However, as occurs in most human activity, technological and scientific progress travel at breakneck speed, thus a comparison is almost bound to present problems. Why? Progress (better working conditions in booths, use of PCs, more sophisticated technology, more competent interpreters, more fastidious end users etc.) could, on the one hand, have an impact on the average performance of interpreters and, on the other, on end users' power of assessment and thus the comparison of expectations of the different groups will be problematic, unless, of course, this is accounted for methodologically.

As for testing, Kurz appears to have adopted Bühler's measuring scales even though this is not clearly mentioned apart from a vague reference to evaluating "[…] the quality of interpretation on a four point scale" (15). It is clear that the issues we raised regarding Bühler's measurement scales apply here too. Furthermore, Kurz claims that she wishes to test the hypothesis that "different groups of end users have different expectations and needs" (15) and yet presents a set of descriptive data which remain untested. Furthermore, the same information in Table 1 (16) is repeated in Figs. 1, 2 and 3 (17) in the form of bar charts. Therefore, are we, like Pöchhacker to assume that peers were not consulted and/or that refereeing was slack?

Finally, Kurz does not attempt to compare the different groups (124 users and 47 interpreters). Instead of grouping her end users together as a single set and comparing them to the 47 professionals, she seems to lose her thread and goes on only to compare the three sub-sets to each other. Furthermore, the CACL group (6 experts) from Bühler's study are merged into the 47 interpreters

yet remain unmentioned. Again there is no mention of socio-demographic characteristics of her samples. Would this not have influenced opinions?

4.3.   Vuorikoski

Vuorikoski makes an attempt to import multi-method research to quality research in IS. Unfortunately, her efforts at innovation fall short as she brings neither methodological innovation nor any empirical contribution to the field. Although she mentions a variety of methods available, her own survey does not reflect the spirit of multi-method research which she so strongly advocates. In fact, it is quite unclear where exactly the "eclectic" (1993: 318) dimension in her study is. The author, in fact, claims that

> … the small size of typical fieldwork research was compensated for with survey techniques. By covering five different seminars, each having about 100 participants, there would be more ground for generalizing the results. The size of the seminars was closer to that of fieldwork, and consequently no statistical sampling method was necessary: the seminars were considered to be theoretically relevant populations as such, and large enough for statistical analysis when treated as one population. (Vuorikoski 1993: 318)

This is, of course, a clearly contradictory statement regarding sample size. Is Vuorikoski saying that both samples are large, or is she saying that they both are small, or is one large and the other small? It would appear that the author swings back and forth from population to sample making sweeping statements yet with no mention of theory when she should have quoted some law of probability theory in support of her argument.

As for the survey itself, Vuorikoski declares that respondents were asked to give a phone number so as to allow for follow up phone interviews. Here too we find a contradiction as the author claims that "Telephone interviewing was selected as an alternative to the more traditional face-to-face interview" (323). Now rather than an alternative which allows the comparison of two independent samples and would have thus given force to a multi-method approach, what we appear to have here is a paired sub-sample of the same respondents being interviewed before and after the conference. In terms of number of people interviewed telephonically and what they were asked, these elements remain unknown. Hardly multi-method. Finally, Vuorikoski entitles a chapter "The eleven statements in the questionnaire" (321). Overlooking the fact that some of these "statements" turn out to actually be "questions", in questionnaires statements are usually measured on 5 or 7 point Likert scales. And here Vuorikoski is finally innovative as she adopts a two point scale which includes

the two options "agree" and "disagree" with no mid-point. However, to do her justice, she did include an "I don't know" escape route.

Over and above all these problems, the study is purely descriptive, research hypotheses are vague and no attempt is made to test them or even to really argue in favour of the much quoted multi-research methods she so fervently upholds. In the light of this discussion, is Pöchhacker still certain that we should have quoted this study? And again, we are forced to ask ourselves, were the referees caught out sleeping on the job just for us?

4.4.   Mack and Cattaruzza

With regard to the descriptive survey carried out by Mack and Cattaruzza (1995) we do not wish to discuss sampling, measurement scales and statistical tests as we have with the others. The reason for this is to be found in the conclusions of their work in which they claim that

> Since this survey was conducted and elaborated using non-professional statistical means, no attempt was made to generalize its results nor achieve full comparability with previous studies, as this would require more sophisticated methods (47).

The awareness and unassuming nature of the two researchers admission of their lack in methodological know how makes criticism of their shortcomings absolutely unnecessary. Surprisingly too, of all the studies in QBQR, it is the study with fewest methodological weaknesses.

4.5.   Moser

Once more, taking Pöchhacker's comments as a starting point, we would like to specify that rather than defend the size of Kurz's sample in terms of having "by no means" being dwarfed by that of Moser (this issue: 146), we should instead ask ourselves how the giant (i.e. Moser's sample 1996) was produced. In fact, even if Moser's sample size is considerable, an impressive 201 respondents was pretty remarkable for the field of interpreting at that time, the way in which participants were interviewed at 84 different meetings clearly shows that the sample was self selected. How do we know this? First and foremost because an average of 2.4 interviews took place at each conference and of these 1.2 involved speakers as opposed to delegates. Surely speakers and end users cannot be considered as the same thing unless speakers are considered a particular

segment of end users?[16] Furthermore, out of scores of participants how were the respondents who were not speakers actually selected? Did each of the participants have an equal chance of being chosen? What we are saying is that sampling is not simply a question of size but, also importantly about <u>how</u> a sample is selected.

Furthermore, also in the case of Moser, the nature of the study is just descriptive and explorative. In fact, there is no application of any statistical test even if in this case the author in the central concerns of the survey states that he wants to investigate the "hypothesis … that different user groups would have different expectations of interpretation" (1996: 146). Yet how this hypothesis is going to be tested and the relative results are left to the imagination of the readers; similarly where the author sees the "positive correlation" (157) between increasing conference-going experience and the fact that users want the interpretation to match the original also remains statistically unexplained. Naturally, considering the self selected sample one could argue about the parameters of distribution involved in the statistical test chosen, but since this was not the case we can only leave the answer to the reader's imagination.

Now, in order to understand the way in which Moser measured his items we have been forced to draw on both the work published by AIIC (1995) and a different version of the same study published in *Interpreting* (1996). We have had to look at both articles because interestingly, the same study published in the journal omits a great deal of background information present in the initial study. Measurement scales, for example, are not stated in the 1996 article. So, let us pick on a couple of examples to examine how Moser in his survey measured some items relating to end users' needs and expectations. regarding "completeness of rendition", "clarity of expression" and "correct terminology" for which the following scale was used:

Very important	Fairly Important	Fairly unimportant	Unimportant	Ambiguous	I don't know

Firstly, we can see that the scale seems to be lacking in a central point (ie "neither important nor unimportant"), unless of course the reader is supposed to assume that the item "ambiguous" is filling the gap. Now, if our first assumption is true, that is that the scale is lacking in a central point, then it follows that the scale is incomplete. If, on the other hand, "ambiguous" is a deliberate choice in the scale for the central point, then there is obviously a problem of wording in communicating the points of the scale. Wording, as pointed out previously (see 2), is a very important aspect of setting up measurement scales. Moreover,

---

16   Of course we are well aware that at some conferences speakers are there also as delegates.

Moser only comments on criteria which respondents judged as being "very important" (162 and 163) according to conference type. In other words, he is giving readers a somewhat incomplete picture of user expectation by concealing other information.

Now let us take another example from the AIIC publication (1995: C1, C2). After having asked respondents to indicate the importance of three criteria ("completeness of rendition", "clarity of expression" and "correct terminology") according to the scale mentioned above, they were then asked, under the label "other", to identify criteria not specified in the preceding questions.[17] Now the issue of the word "other" followed by a list begs the following question: why after such extensive preliminary research were none of these criteria identified and inserted in the questionnaire to be measured on the same scale reported above? Could it be that the observer's paradox has reared its head? In other words, what we are trying to say is that if the target population were end users it should have clearly been tested on a group of end users before final administration. This is not clear from the final version of the questionnaire in German neither in terms of numbers, nor in terms of the people interviewed (interpreters again or end users?). However, over and above this, the list of assorted criteria detected by informants leads the reader, in any case, up a blind alley as their degree of importance is not measured with the same scale as the first three criteria. In other words, these criteria are incomparable with the first three. An example of a more gross error is finding the item "correct terminology" under "other" when the respondent had already given an opinion on that criterion in the previous question.

For the sake of argument, let us consider one more example of a scale adopted in the same study (1995: C4, C5):

Very irritating	Fairly irritating	Not really irritating	Unimportant	Don't know	Ambiguous

In this case the researcher is trying to measure end users' degree of irritation of particular behaviour of interpreters. The question which arises here is why include "importance" in a scale which is trying to measure irritation? And again we find the baffling item "ambiguous" occurring once more. Last but not least, we have yet another unbalanced scale, with no central point. How come?

---

17 The choice of criteria listed by Moser under a stark "Other" are "synchronicity, emotional congruence, pleasant voice, correct terminology, focus on essentials, technical knowledge, faithfulness to the original, faithfulness to the meaning of the original (*sic*.), clarity of expression, neutrality towards the speaker, lively, animated delivery, translation of jokes and asides, native sounding accent, stop when a mistake is made, other." (1995: C2)

So how did the editorial process work here?[18] Do high standards of refereeing apply only to the work and Chiaro and Nocella published in *Meta* or are these criteria universal? We will however return to this issue in a dedicated paragraph (5).

Curiously, at a certain point Moser's study introduces the concept of attitudes towards providers of the service in order "to shed additional light on the study" (p. 159) and sets out to ask end users what they consider to be particularly interesting and particularly difficult about the interpreting profession.[19] Well, how attitudes are used in this context is not clear despite the fact that already at the time in which this survey was conducted attitude models such as the Fishbein model (theory of reasoned action) and the Ajzen model (theory of planned behaviour) to try to measure attitudes were already well established (Ajzen 1991; Solomon 2004). However, the important conclusion at which Moser arrives using the term attitude (in the broadest possible sense?) is that it is linked to "the broad educational and cultural background for which they (interpreters) are envied" (160).

Finally, also in this survey, information about respondents' education and professions which could have played an important role in testing the unproven hypothesis were not solicited. And unusually, instead of directly asking respondents (end users?) how old they were, ages were supplied by interviewers (interpreters) who "were asked to estimate the age of persons interviewed" (1996: 151) thus bringing to mind vets who estimate the age of horses by examining their teeth. Why were interviewers not asked to guess other socio-demographic data too? Presumably because apart from evaluating a person's sex, the rest is quite difficult. Again, it would appear that *Meta* is not the only journal to suffer from lax refereeing.

4.6. Pöchhacker

According to Mark Twain there are three kinds of lies: "lies, damned lies and statistics." And it is undeniably a truism that statisticians can manoeuvre numbers at their will. And this is precisely what Pöchhacker attempts to do by offering readers fresh analyses of his colleagues' data.

---

18 We have chosen just a couple of the numerous faults in Moser's work simply for the sake of argument.

19 Moser's question "What do you find particularly interesting about the profession, and what particularly difficult?" (1996: 159) is actually two questions in one and thus would require rewording.

4.6.1. Kurz's calculations according to Pöchhacker

Pöchhacker occupies more or less a third of his essay re-elaborating the data of Ingrid Kurz. His re-elaboration of the percentages in Bühler and Kurz's and comments on figures 2a, 2b, 3a e 3b have already been amply discussed in paragraphs 4.1 and 4.2 above. However, before we offer our own interpretation of this recent amplification of the data, we wish to make a short premise. Without a shadow of a doubt the energy which Pöchhacker has exerted into the re-elaboration of his colleagues' data is, to say the least, admirable. Nevertheless, his efforts recall the period between the two World Wars when the first social scientists were lacking in a compass (a research hypothesis) to guide them through their studies. In fact, they would start off by gathering data willy-nilly and subsequently observing what emerged. The only guide they had at the time was their personal capacity to elaborate data with the means of sound techniques (Guidicini 1996). In other words, our predecessors possessed neither computers nor sophisticated, click-of-the-mouse software. Fortunately, their somewhat careless manner of conducting research was soon to be replaced by one which started off by forming a hypothesis, gathering and elaborating specific data and subsequently either confirming or rejecting the initial assumptions aided with new technology which was to come to their rescue.

In his discussion of the use of statistical tests on Kurz's data Pöchhacker seems to have lost his compass as he appears to oscillate between testing group differences on continuous variables, testing relationships among discrete variables and testing both together. Does he see the nature of the variable? Can the same variable be both discrete and continuous at the same time?[20] Pöchhacker is analyzing categorical data from an unbalanced ordinal scale.[21] First he applies chi squared testing to check whether there are any relationships between two or more categorical variables and then on the same data he explores differences amongst the groups treating the variables as though they were continuous.

Moreover, during the application of the chi squared test, Pöchhacker quite rightly observes that more cells have the expected frequency which is smaller than five and begins by admitting that " the expected frequency in the chi-square test is smaller than five, which renders any interpretation of the test invalid" (*this issue*) and that the sample is not big enough, and then that the data should have had a more balanced distribution. But, in order to resolve the problem he collapses the 2 categories of the scale adopted by Bühler and Kurz (i.e. "less

---

20   Obviously it is possible to transform a continuous variable into a discrete one but the reverse would be more complex.

21   For a detailed discussion on statistical tests on categorical data see Agresti 2002.

important" and "irrelevant") into a single category labelled "not important". Is it really plausible to collapse "irrelevant" and "less important" into a single category? In other words, surely "very important", "important" and "less important" have more in common semantically with each other than "less important" and "irrelevant"? Would it not have been more reasonable to collapse all the categories which measured importance so that the new dichotomic variable would have been acceptable? Pöchhacker would then at least have had all the dominions of "importance" in one category and "irrelevance" in the other. One last point, Pöchhacker includes an explanation of an elementary concept such as cross tabulation yet does not elucidate chi square distribution, the significance of probability p or acceptance or rejection of the null hypothesis. Surely if anything needed clarification it would be the latter concepts and not the former.

At the beginning of the sub paragraph "Other non parametric tests (*this issue*?)" Pöchhacker states that "Aside from the chi-square test, there are other nonparametric tests for identifying significant relationships among different sets of rank-ordered data" and he uses the Kruskal-Wallis H and the Mann-Witney U tests. Here again, it is unclear whether he is looking for relationships or differences. And in applying these tests, if respondents had originally been asked to express their opinion on a single unbalanced item scale for the various criteria, how did he obtain his rank ordered data? Is he still using mean scores? If so, once more is the variable continuous or discrete? Moreover he does not explain that the use of asymptotic significance for the exact test may not be a good measure of significance if the variables are poorly distributed, which seems the case with these data sets. It would appear that Pöchhacker is simply looking for anything significant in the dataset without clarifying how he is manipulating his data. Moreover to explain differences among the 3 groups on the criteria which resulted as being significant, he runs the Mann Witney U test taking into consideration the three combinations of the three independent samples. Well, in this case we would like to point out that by following this path Pöchhacker is falling into the so called "familywise" or "experimentwise" error rate (Field 2000)[22]. In other words, is he aware that the probability of making at least one type I error is increasing from 0.05 to 0.143? If the Bühler group was included, and we do not understand why in his re-analysis this group has been omitted, this probability would have jumped to 0.185.

But why, we wonder, twenty years on does Pöchhacker want to show significance at all costs? Undeniably, more than one statistical test can be

---

22 We apologize but space does not allow us to explain testing hypotheses. However, the topic is treated in almost all text books on general statistics. As regards the figure assuming the independence of the samples we apply the independent law of probability: in the case of 3 groups (0.95)3 ≈ 0.857 and (1-0.857) ≈ 0.143.

carried out on both Bühler and Kurz's data, on our data, or on anyone else's data come to that, but is there any point? If a researcher originally sets out to either accept or reject a particular hypothesis, why demonstrate that they could have done something different?

Our discussion of the beams seems to have highlighted the difficulty both of choosing and applying a statistical test in the studies examined so far. We would now like to dedicate a few words to this issue as succinctly as possible for obvious reasons of space.

4.6.2. The choice of a statistical test: an overview

In the light of our previous discussion, it now seems evident that the choice of a statistical test cannot be dictated simply by the significance or lack of significance deducted from p values produced by any "PC-based statistics software […] accessible enough to be used, with proper guidance, also by the 'semi-skilled' analyst" (Pöchhacker this issue: 154). On the contrary, the choice of a statistical test should be made in function of three general conditions at the same time: the research question, the nature of the data and the plan or design of the research.

The research question should veer in a clear direction. From the start, the researcher should know whether the aim is to find differences or correlations between or among the variables which are object of the study. Once the researcher has decided which direction the study will take, inferential statistics offer numerous tests which test the hypothesis underlying the research question: univariate, bivariate and multivariate techniques[23]. Naturally, the choice of a test also depends upon the nature of the data. Does the data consist of discrete or continuous variables? And what are the forces of the measurement scales upon which the variables were measured: nominal, ordinal, interval or ratio-scaled? As a result, descriptive statistics and statistical tests must be also gauged in function of the metric or non metric nature of the variable and of the force of the measurement scale. The research design used to generate the data also affects the choice of a statistical test. So, decisions regarding the independence of the samples, number of groups, number of variables and variable control must be taken *a priori*. Moreover, when a technique is used, the assumptions regarding that technique have to be satisfied before applying the technique. So if one wishes to apply ANOVA for example, to explore differences among groups, the assumptions of independent random samples, normality and equal variances of

---

23   There is really no room to explain even briefly the statistical techniques included in these three groups, however explanations can be found in basic and more higher level text books in general and advanced statistics.

all populations must be assessed. So, the analyst should explore the dataset in order to understand whether these conditions have been satisfied. But if the necessary assumptions are violated what should be done? Well, it depends on which assumptions have been violated. If normality or equal variances are involved then transformation to symmetry[24] could be applied (Ryan 1985) to approach a Gaussian distribution[25]. However, if one or more samples differ in a significant way from the population of interest then it could be very difficult to draw any conclusions from the dataset. An explorative analysis becomes more stringent and of paramount importance when more variables are involved, i.e. when multivariate techniques are used (Hair *et al.* 1995). In this case, the relationship among variables, the analysis of missing data, the detection of outliers through graphical output (e.g. stem and leaf diagrams or box plots) or statistics (e.g. Mahalanobis D^2) and the verifications of the assumptions such as normality, homoscedasticity[26] and linearity are something which cannot be solved just by a few clicks of a mouse.

Finally, while we are perfectly aware that the advent of user-friendly statistical software has facilitated the application of statistical tests and the mathematical calculations behind them, we firmly believe that these very packages require a sound knowledge of the field of statistics. We cannot possibly agree that

> […] analyzing empirical data […] is not a question of mathematical skills but, essentially, a matter of meaningful interpretation, of making sense of the relationships indicated by the data (Pöchhacker this issue: 155).

This does not do the field of research methodology justice. Without understanding what he or she is doing in terms of statistics, the researcher not only runs the risk of misapplication of tests but also of a poor interpretation of results. Not only, but whether a statistical significance test "does not explain anything but merely points reliably to what needs to be explained" (Pöchhacker this issue: 155) is highly debatable too.

---

24 Of all transformations made on data in practice, the three most popular are the square root (moderate), the logarithm (strong) and negative reciprocal (very strong).

25 When assumptions are violated one could also think of applying non parametric tests which are less stringent in matching assumptions; in the case of ANOVA one could use the Kruskal-Wallis test.

26 Homoscedasticity is an assumption related primarily to dependence relationships between variables. It refers to the assumption that dependent variable(s) exhibit equal levels of variance and homogeneity of variance across the range of predictor variable(s). (Hair *et al.* 1995: 67).

5.   To print or not to print

It is clear from his use of ironic punctuation ("Into print": 161) that Pöchhacker did not consider our study worthy of publication. For reasons of delicacy we would rather avoid the hearsay and the chitchat surrounding our article's journey from Italy to Canada.

Moreover, in defence of *Meta* and the referees of our article, we wish once more to take total responsibility of all shortcomings which are totally our own and not imputable to the journal[27]. From our point of view, our article was sent to Canada in mid-2001, refereed about six months later, corrected, accepted and finally published in the summer of 2004. However, perhaps it would have been more correct if Pöchhacker's article had appeared in *Meta* rather than *The Interpreters' Newsletter* seeing that it is the former journal which is under attack. Unfortunately, when Pöchhacker thoughtfully sent us his paper, he had already sent it to *The Newsletter* thus we too had no option but to respond in the same journal. But thinking more precisely on the matter, perhaps a Special Issue on quality is exactly where this discussion should take place. However

> Aristotle argues that there are three kinds of rhetorical proof; that is three ways in which a speaker can persuade an audience of his position – *ethos*, *pathos* and *logos*. *Ethos* is ethical proof, the convincing character of the speaker (…) *Pathos* is an appeal to the emotions of the audience (…) *Logos* is logical proof, or argument, the kind of proof that appeals to reason (Root 1987: 16-18).

And we have attempted to defend ourselves from Pöchhacker's accusations taking the philosopher's advice by blatantly appealing above all to the reader's understanding of our competence in research methodology, as well as to his or her emotions and reason. We hope to have clarified above all our use of the term perception (3.1.1.); that our sampling frame was accurate (3.2.2.) and demonstrated that our criticism of Bühler was all but erroneous (4.1.). In doing so, we have been forced, albeit unwillingly, to be harsh on others.

However, what emerges from the present discussion is that over the years, macroscopic faults in the refereeing process in this field have been common across the board. Admittedly our argument was perhaps circular. This is one criticism which Pöchhacker makes that we feel we must accept, but we still wish to claim that our study was methodologically sound in design,

---

27   Above all, we humbly apologize for our misspelling of Kopczyński, for depriving Susan Bassnet's surname of an 's' and the "infelicities in the bibliography." However, IS must be in an embarrassingly poor state if, in an attempt to punish two authors who have (apparently) stepped out of line, Pöchhacker feels he must include typos in a critique of methodology.

administration and data elaboration. Indeed, we could have extended our article with more detailed information, rationale and discussion. But could not the same be said of the other works mentioned? The methodological faults we have found in others are substantial and incomparable to the display of nitpicking displayed by our plaintiff. Circular? Maybe. But what have been the conclusions that other researchers adopting QBQR have reached so far? Have they been so insightful? But again, whatever their findings and conclusions at least they had a go, unlike our complainant who simply sits and looks and then comments from high with a critical eye. Indeed, one wonders whether such an eye is really critical. For us the word misguided seems more fitting. What is more, as two researchers looking in from the periphery, the argumentation put forward by Pöchhacker makes the field of QBQR in IS appear somewhat self referential to say the least.

And we certainly could have done without the author glibly offering the quasi-total demolition of our work "… in support of their welcome ambition to raise the methodological standards of research in this field" (Pöchhacker this issue: 163). Are we supposed to feel honoured by this insult to our intelligence? Yet still not satisfied, Pöchhacker turns the dagger in the wound by stating that "Understandably, these colleagues would rather not see their published work become an object of methodological criticism." Well, Pöchhacker certainly notches up full marks in insight and sensitivity on that score, yet at the end of the day, what we find most objectionable, is not so much the criticism itself, but the rather patronizing tone in which it is couched. Criticize us by all means, but superciliousness we can do without. As far as survey research is concerened, Pöchhacker is still in an early stage of infancy.

6.   Beams of light?

Last, but certainly not least, despite our criticisms of the studies mentioned, we would like to emphasize our respect for all those researchers who have tried their hand at field work. Our harshness towards these people has been dictated by the need to demonstrate that if we "Unwittingly … provide material for a case study of methodological rigor in quality research" (Pöchhacker this issue: 163) others provide just as much, if not more so and presumably, until this moment, just as inadvertently. Nevertheless, however faulty and elementary their instruments, only people who have actually rolled up their sleeves and personally tried to obtain answers from complete strangers will have experienced the blood, sweat and tears behind each single return. Which naturally makes the whys and wherefores of less than rigorous sampling understandable. Of course, it is simpler to announce a questionnaire to a roomful of delegates than to stop them one by one thus chancing a higher risk of refusal. But then we must be aware that the sample is no longer random.

Similarly, asking friends and colleagues to take part in studies is equally open to criticism. It was precisely this type of nonchalant way of surveying that made us want to contribute with our five (Euro)cents.

We would like nonetheless to express our discomfort with the present response. This time we are fully aware of our heavy handedness regarding the work of others. But if originally we had been vague, here we have had to argue our case as clearly as possible and hopefully readers and, above all, the researchers involved will understand that we had no option. From our point of view, Pöchhacker's critique of our work was short sighted and in places erroneous.

Re-reading the QBQR in IS in preparation for this reply, the lack of knowledge in the tools and methods of the social sciences is self evident. Measurement scales, sampling frames, statistics and statistical tests are constantly defective and studies are strikingly self-referential. If the field of IS aims to "earn the academic recognition it deserves" (Pöchhacker this issue: 164) at excellence in research design and applications, then it should be open to the views and criticism of outsiders who are free of the institutional shackles of unassailable individuals within that group. We hope to have shown that no one is exempt from developing clay feet. Having said that, if the field wishes to remain self-referential, then so be it.

However, we wish to conclude on a positive note. The present essay is a display of academic argument in which we have criticized rather old studies. Perhaps the time has come to let sleeping dogs lie. And perhaps it is also time that translation and interpreting faculties began introducing courses in research design and statistics so that students wishing to embark upon the fascinating field of research are well equipped to do so with a working knowledge of how to go about it. Perhaps now is the moment to learn from disciplines which have been working in social research for decades. In fact, in more recent publications, it is highly uplifting to find that IS scholars are beginning to look tentatively outside IS towards the social sciences for insights into quality research (e.g. Kurz 2003). Surely, if there is something to be learnt from the successful marketing of a good or a service it is the collaboration of experts with diverse expertise who together construct high quality products and facilities. If it is truly excellence which interpreters desire for themselves and their clients in the real world, then the path of interdisciplinary research of practitioners and objective outsiders is surely a good one. If, on the other hand, the issue of quality is to be restricted to the philosophical argument and mutual back patting of a few, then let it remain trapped and stagnant in its ivory towers.

References

Aaker D.A. *et al*. (1995) *Marketing Research*, New York, John Wiley.

Agresti A. (2002) *Categorical Data Analysis*, New Jersey, John Wiley.

Ajzen I. (1991) *Attitudes, Personality and Behaviour*, Milton Keynes, Open University Press.

Berger A. (1991) *Media Research Techniques*, Newbury Park, Sage.

Bühler H. (1986) "Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters", *Multilingua* 5 (4), pp. 231-235.

Berenson M. L. and Levine D.M. (1986) *Basic Business Statistics: Concept and Applications*, New Jersey, Prentice-Hall.

Chiaro D. and Nocella G. (2004) "Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the World Wide Web", *Meta* 49 (2), pp. 278-293.

Collados Aís Á. (1998) *La evaluación de la calidad en interpretación simultánea: La importancia de la comunicación no verbal*, Granada, Comares.

Collados Aís Á. (2002) "Quality assessment in simultaneous interpreting: The importance of nonverbal communication" in *The Interpreting Studies Reader*. Ed. by F. Pöchhacker and M. Shlesinger, London-New York, Routledge, pp. 327-336.

Dillon W.R, Madden T.J and Firtle N.H. (1994) *Marketing Research in a Marketing Environment*, Boston, MA, Irwin.

Field A. (2000) *Discovering Statistics using SPSS for windows*, London, Sage.

Garzone G. (2003) "Reliability of quality criteria evaluation in survey research", in *La evaluación de la calidad en interpretación: Investigación.* Ed. by Á. Collados Aís, M.M. Fernández Sánchez and D. Gile, Granada, Comares, pp. 23-30.

Gile D. (2003) "Quality assessment in conference interpreting: methodological issues" in *La evaluación de la calidad en interpretación: Investigación.* Ed. by Á. Collados Aís, M.M. Fernández Sánchez and D. Gile, Granada, Comares, pp 109-124.

Gourevich A. and Mateeff S. (1989) "Study of characteristics of simultaneous interpretation by the method of paired comparisons" (in Bulgarian), *СЪПОСТАВИТЕЛНО ЕЗИКОЗНАНИЕ / Сопоставительное языкознание / Contrastive Linguistics* 14 (1), pp. 32-38.

Guidicini P. (1996): "La ricerca sociologica nella sua evoluzione storica" in: *Nuovo manuale della ricerca sociologica* Ed. by P. Guidicini, Milano, Franco Angeli, pp 11-37.

Hair J.F., Anderson R.E., Tatham R.L., Black W.C. (1995) *Multivariata data analysis with readings*, Fourth edition, New Jersey, Prentice-Hall.

Halliday M.A.K. and Hasan R. (1976) *Cohesion in English*, London, Longman.

Hoey M. (1983) *On the surface of Discourse*, London, Allen and Unwin.

Kurz I. (1989) "Conference interpreting – user expectations", in *Coming of Age: Proceedings of the 30th Annual Conference of the American Translators Association*, Medford NJ, Learned Information, pp. 143-148.

Kurz I. (1993) "Conference interpretation: Expectations of different user groups. *The Interpreters' Newsletter*, 5, pp. 13-21.

Kurz I. (2001) "Conference interpreting: Quality in the ears of the user", *Meta* 46 (2), pp. 394-409.

Kurz I. (2003) "Quality from the user perspective", in *La evaluación de la calidad en interpretación: Investigación*. Ed. by Á. Collados Aís, M.M. Fernández Sánchez and D. Gile, Granada, Comares, pp. 3-22.

Labov W. (1972) *Language in the Inner City*, Philadelphia, Pennsylvania University Press.

Lapin L.L. (1990) *Statistics for Modern Business Decisions*, Fifth Edition, San Diego, Harcourt Brace Jovanovich.

Mack G. and Cattaruzza L. (1995) "User surveys in SI: A means of learning about quality and/or raising some reasonable doubts", in *Topics in Interpreting Research*. Ed. by J. Tommola, Turku, University of Turku, Centre for Translation and Interpreting, pp. 37-49.

Malhotra N.K. (1996) *Marketing Research: an applied orientation*, New Jersey, Prentice Hall.

Meak L. (1990) "Interprétation simultanée et congrès medical : attentes et commentaries", *The Interpreters' Newsletter* 3, pp. 8-13.

Moser P. (1995) *Survey on Expectations of users of conference interpretation*, Final report, Vienna, AIIC.

Moser P. (1996) "Expectations of users of conference interpretation", *Interpreting* 1 (2), pp. 145-178.

Moser-Mercer B. (1996) "Quality in interpreting: Some methodological issues", *The Interpreters' Newsletter* 7, pp. 43-55.

Pöchhacker F. (2001) "Quality assessment in conference and community interpreting", *Meta* 46 (2), pp. 410-425.

Root R.L. Jr. (1987) *The rhetorics of popular culture: Advertising, advocacy, and entertainment*, New York, Greenwood.

Ryan B.F., Joyner B.L., Ryan T.A. (1985) *Minitab handbook*, Boston, PWS Kent.

Schiffman S.S. *et al*. (1981) *Introduction to multidimensional scaling: Theory, methods, and applications*, Orlando, FL, Academic Press.

Shlesinger M. (1997) "Quality in Simultaneous Interpreting", in *Conference Interpreting: Current Trends in Research*. Ed. by Y. Gambier, D. Gile and C. Taylor, Amsterdam-Philadelphia, John Benjamins, pp. 123-132.

Simon J.L. (1969) *Basic Research Methods in Social Science: the art of empirical investigation*, New York, Random House.

Solomon M.R. (2004) *Consumer behaviour*, New Jersey, Prentice Hall.

Statt D.A. (1997) *Understanding the Consumer: a psychological approach*, Basingstoke, Palgrave.

Tull D.S and Hawkins D.I. (1993) *Marketing Research: measurement and method*, New York, Macmillan.

Vuorikoski A.R. (1993) "Simultaneous interpretation – user experience and expectations", in *Translation – the Vital Link. Proceedings. XIIIth World Congress of FIT*, vol. 1. Ed. by C. Picken, London, Institute of Translation and Interpreting, pp. 317-327.

BOOK REVIEW

Jesús Baigorri Jalón, *De Paris à Nuremberg : naissance de l'interprétation de conférence*, tr. de l'espagnol sous la dir. de Clara Foz, Les Presses de l'Université d'Ottawa, 2004.

De Paris à Nuremberg : naissance de l'interprétation de conférence, traduction en français de *La interpretación de conferencias*, constitue un ouvrage fondamental pour toute approche historique à l'interprétation de conférence.

La curiosité de l'historien se conjugue chez Jesús Baigorri Jalón avec le regard de l'interprète qui fait revivre, sous les yeux du lecteur, le travail passionné et passionnant de la première génération de professionnels de l'interprétation.

L'auteur a exploité sans réserves le privilège qu'il a eu d'accéder aux archives de la S.D.N., de l'ONU et à d'autres organisations internationales comme l'O.I.T. À ce travail de recherches s'ajoutent les entretiens qu'il a eus avec des témoins des débuts de l'interprétation de conférence. Le dépouillement minutieux des dossiers personnels des interprètes de l'époque a donné naissance à des portraits vivants et inoubliables, dans lesquels les noms célèbres des pionniers de l'interprétation se concrétisent en individus bien insérés dans le cadre historique de leur temps.

Dans le premier chapitre, consacré à la Conférence de la Paix de Paris de 1919, Jesús Baigorri Jalón aborde la question linguistique et nous montre l'intérêt et les enjeux qu'elle recelait pour les grandes personnalités du début du XXe siècle. La reconnaissance de l'anglais comme langue officielle de la Conférence à côté du français, montre clairement le rôle que les nations anglophones étaient en mesure de jouer grâce à la victoire remportée. De même, cela marque le déclin du français que l'auteur attribue à la décadence de la diplomatie internationale, désavouée par les insuccès qui avaient mené au premier conflit mondial et mise à l'écart par la participation directe des chefs d'État et de gouvernement aux travaux de la Conférence de Paris. C'est sur ce scénario que se distinguent des personnes bilingues ou polyglottes appelées à s'improviser interprètes et à le devenir de profession. C'est grâce à elles que la consécutive prendra son essor pour atteindre son apogée pendant la période de l'entre-deux-guerres (chapitre II).

Rien n'est laissé au hasard : Jesús Baigorri Jalón nous mène à l'intérieur de la Société des Nations et de l'Organisation Internationale du Travail pour nous montrer de près les conditions de travail de nos anciens collègues, leurs

difficultés, leur façon de travailler, les jugements dont ils furent l'objet ainsi que l'attitude des usagers de l'interprétation.

Le long chemin parsemé d'obstacles de l'interprétation simultanée est abordé dans le troisième chapitre. L'auteur y peint avec rigueur et minutie, les démarches entreprises par M. Edward Filene, "entrepreneur et philanthrope américain", désireux d'améliorer et d'accélérer la communication humaine multilingue. La consécutive, en effet, demandait au moins un redoublement des temps de parole.

Face aux avancées indéniables de la technique et à la volonté progressiste de quelques personnalités, vaine fut la ferme opposition, voire l'hostilité des consécutivistes. Les interprètes expérimentés furent ainsi remplacés par des jeunes interprètes formés sur le tas, qui, aux procès de Nuremberg, élevèrent la simultanée au rang de l'interprétation de conférence par excellence (chapitre V).

La narration du parcours qui avait mené à l'adoption de la simultanée à Nuremberg et ensuite à l'ONU, est interrompue – par souci d'ordre chronologique – par le IVe chapitre, dans lequel l'auteur décrit la personnalité et le travail des interprètes des dictateurs. Jouant un rôle de premier plan, ces interprètes furent les témoins des grands cataclysmes qui bouleversèrent le XXe siècle et en même temps le miroir de la redoutable association entre la fonction qu'ils exerçaient et la responsabilité qui découle de la connaissance des faits.

Les recherches rigoureuses menés par Jesús Baigorri Jalón débouchent sur un volume d'envergure, un "livre d'histoire" qui attire l'attention du lecteur grâce au goût que l'auteur a pour le détail et l'anecdote, et qui nous offre un récit toujours agréable et émouvant.

La traduction en français de l'ouvrage écrit en langue espagnole est le fruit d'un travail d'équipe, effectué à l'Ecole de traducteurs et d'interprètes de l'Université d'Ottawa, par un groupe d'étudiantes coordonnées par Clara Foz. Le résultat homogène et la lecture aisée en font une mise en abyme remarquable dans laquelle la traduction se penche sur le discours de l'interprétation.

Caterina Falbo