

Assessment of consecutive interpretation by different user groups Jana Rejšková, Libor Nekula

The paper is divided into two parts:

- 1) description of an MA thesis I have recently supervised
- 2) comments and reflections on the thesis and on the discussion we had in Graz

- 1) The experiment conducted as part of the thesis involved three recent graduates (two female, one male) of our Institute of Translation Studies, Charles University, Prague. Their language combinations were English and German (2) and English and Spanish(1). The interpreters interpreted consecutively an 8 minute speech from English into Czech - their mother tongue. The original speech was delivered by a native speaker of English. The topic of the speech was Intellectual Property Rights and their Protection as provided for in the ECJ case law. Two “clients” were present to make the setting more authentic. The three performances were recorded on a camera and posted on the Internet to make the assessment as user-friendly as possible. Three groups of users, comprising ten people each, were invited to assess the performance, using a questionnaire designed for this purpose, but drawing on Ingrid Kurz’s questionnaire.

The first group comprised professional conference interpreters, all members of the Czech National Association or accredited with SCIC. In the second group there were ten lawyers, and in the third ten regular users of consecutive interpretation, all employees of supranational companies based in Prague - most of them with background in engineering. The thirty users were sent detailed instructions on how to proceed, with special emphasis on how important it is *not* to assess the three performances in the sequence in which they had been posted on the Internet. Unfortunately very few followed this instruction. The questionnaire included a 10 point scale for each criterion assessed (fluency of delivery, logical cohesion, sense consistency, completeness of interpretation, correct grammatical usage, use of correct terminology, reception by the user – eye contact, suprasegmental features working with notes, confidence in the interpreter and faithfulness to the original). Drawing on papers by Bühler, Reiss, Gile, Kurz, Peter Moser, Miriam Shlesinger, F. Pöchhacker, and others, the author of the thesis wanted to test the hypothesis that different user groups would attribute different weight to different criteria, or more specifically that e.g. the lawyer group would consider correct terminology most important, the interpreters would stress language, fluency, consecutive skills, and the regular users would be inclined to compare these performances with what they are used to in their everyday practice. This hypothesis has not been confirmed.

Findings:

- a) With the exception of two criteria - reception and intonation – where surprisingly interpreters were significantly *more lenient* than the other two groups, all other criteria yielded no significant differences.
- b) No significant differences were found between the regular user group and the lawyer group.
- c) “Individuality” of the assessor seems to be of much greater significance than his/her professional background – in the conference interpreter group one interpreter was assessed as the best by one user and as the worst by another user (!).
- d) Smooth delivery can disguise flaws and the performance may be assessed as very good with correspondingly high scores. On the contrary, hesitation,

“noises” and other factors, i.e. body language making delivery less smooth, will mean less confidence on the part of the respondent and consequently lower score.

- e) The assessor does not have to be a language expert to be able to identify incorrect language use, linguistic style etc.
- f) In a number of instances there was no immediate correlation between comments to open questions - often rather strict – and the actual score.
- g) The sequence in which the three performances were assessed was key. E.g. the C interpreter (interpreters were referred to as A,B,C, and had been posted on the Internet in this order) produced a swift, “likeable” interpretation with a lot of small mistakes, repetitions, shifts and yet was assessed by many as the best. Our explanation is that the assessors had grasped the main messages while listening to A and B, and paid less attention to the content. Another explanation would be *fatigue*. Assessors suffer from fatigue and sometimes from too little positive stress, i.e. they may get bored after the first two renditions. This point is, in our opinion, immediately relevant to consistency or lack of consistency in the exam situation.

Summary of findings:

- the results have shown that assessors do not necessarily have to be language experts to be able to detect incorrect language use;
- nor do they need to be regular users of interpretation services to be able to identify what is good and what is less good
- nor do they have to be experts to identify correct/ incorrect terminology, assessors in the regular user group and the interpreter group pointed out the same mistakes – this may be topic dependent, a more technical, e.g. a medical text would have probably yielded different results. Nevertheless, the finding is relevant, as very technical topics are hardly ever used in exam situations;
- Peter Moser in his research concludes that experience with conference interpreting influences the user’s priorities. This was not borne out in our experiment - perhaps due to the difference between consecutive and simultaneous interpreting;
- Kopczyński concludes that professional background will have only a small impact on assessment, our findings confirmed that conclusion;
- Kahane claims that interpreters tend to be more demanding assessors – this was not confirmed, on the contrary non-interpreters can be very sensitive and their “detection”abilities should not be underrated. Shouldn’t more non-interpreters be invited to sit on exam boards?
- Miriam Shlesinger is right in saying that a “flawless” delivery may disguise a lot of flaws.

2) Comments and reflections on our Graz discussion:

- a) terminology is not just individual terms. Our experiment showed that the intellectual property rights terminology was assessed in much broader sense, which may explain why there was no significant difference between the lawyer group and the other groups. It seems that understanding the concept rather than individual terms is key;

- b) fatigue and lack of “positive stress” is truly a problem and may influence the exam board’s assessment and final verdict;
- c) exam boards must be well balanced and should include users. It is equally important to have “users” present in class, whenever possible.
- d) To conclude I would like to borrow Marianna Sunnari’s borrowed quote from Lansing:

He seemed almost to take over the character of the individual whose words he translated, and to reproduce his emotions as well as his thoughts.

I believe this is what makes consecutive interpreting so exciting. However, in the class/exam setting it is this kind of excitement or positive stress that is usually lacking . On the other hand there is usually too much negative stress and I am convinced it is our task to make the “environment” closer to life, i.e. full of positive stress and excitement.