Interpreter corpora and interpreter training

Guy Aston, University of Bologna, Italy

One of the areas offered by modern technology which has as yet been little exploited in the training of interpreters is the use of electronic corpora of interpreting data. A growing number of such corpora now exist, such as EPIC (Bendazzoli & Sandrelli 2005) and EJSPC (Shimuzu et al 2014), but these have primarily been used to focus on translation equivalences between input and output with a view to analyzing the interpreting process and/or developing automatic speech translation systems. Little has been done to investigate the potential use of such corpora in interpreter training.

Henriksen (2007) argues that professional simultaneous interpreters, like the Homeric poet(s) (Parry 1930), rely to a large extent upon extended phraseological units, or speech formulae (Wray 2005). In this paper I propose that the acquisition of such formulae, both in the source and the target language, constitutes an important requirement of training as a means of partly reducing the "effort" of simultaneous interpreters (Gile 1999), and thereby of increasing their potential fluency. I will illustrate how corpora of interpreting data, such as those derivable from the videos of European Parliament proceedings, provide a means of identifying useful formulae which can be placed in the hands of trainees, both as reference tools and as a means to develop their sensitivity to such formulae (Aston 2015), using interrogation software such as Wordsmith Tools (Scott 2012).

While few interpreting corpora align transcriptions with the original audio, it is also possible to use audio-aligned corpora of other data types (e.g. Hasebe 2015) to associate particular formulae with their prosodic patterns, thereby providing acoustic models for trainees to imitate. My paper will illustrate applications of these procedures in the interpreter training context and outline their potential for autonomous learning.