

# 3

## Accuracy Rate in Live Subtitling: The NER Model

*Pablo Romero-Fresco and Juan Martínez Pérez*

### 3.1 Introduction

Over the past few years, the focus of audiovisual translation (AVT) seems to have shifted from quantity to quality. As is demonstrated by international conferences, such as Media for All 3 in 2009 ([www.mediaforall.eu/all3](http://www.mediaforall.eu/all3)) and Media for All 4 in 2011 ([www.imperial.ac.uk/humanities/translationgroup/mediaforall4](http://www.imperial.ac.uk/humanities/translationgroup/mediaforall4)), this shift applies to industry as well as to academia. In the case of live subtitling, and more specifically respeaking, the most common method used to evaluate the quality of subtitles produced in real time consists of assessing their accuracy. Needless to say, where quality is concerned there are also a number of other features to be considered, such as delay, positioning, character identification and speed, as well as factors relating to their reception by viewers (opinion, comprehension, perception). These issues have all been discussed by Romero-Fresco (2011) with particular reference to the UK market. Yet, what concerns broadcasters, regulators such as Ofcom and subtitling companies is the accuracy of live subtitles and it is this that constitutes the main focus of this chapter.

Up until now, subtitling companies have tackled the issue of quality in very different ways. In some companies, trainers are in charge of error calculation, whereas in others this is done by the subtitlers themselves. The calculation methods vary greatly, some being much more 'generous' than others. In addition, the approach to live subtitling is different depending on the country in question. Whereas, in the UK, live subtitles are nearly always verbatim, in other countries, such as Germany or Switzerland, they are variously edited. This heterogeneous picture raises several questions, namely, are the accuracy rates obtained by different companies at all comparable? Do the methods used take

into account differences between languages? Do they only provide a final score or do they also give an indication of the improvements necessary to obtain better results?

The aim of this chapter is to present the NER model, a new model for assessing the accuracy of live subtitles in different countries and in different languages by analysing the extent to which errors affect the coherence of the subtitled text or modify its content. An emphasis will be placed on respoken subtitles, the type in most common use nowadays. The model is also applicable to automatic subtitles, which, given the rapid development of speech recognition technology, are likely to become more widespread in the near future.<sup>1</sup> Following an introduction outlining the basic requirements that such a model might be expected to fulfil, an overview of the traditional methods used in what is known as word error rate (WER) is given. This is then followed by examples of the different types of error assessed by the NER model in English, Spanish, Italian and German with, finally, an explanation of the application of the NER model to real-life subtitles.

## **3.2 Basic requirements and traditional methods**

Before presenting the NER model, it seemed important to outline the basic requirements of such a model in order to ensure its success in academia as well as in the industry as a whole. It also seemed essential to assess traditional methods so as to illustrate their deficiencies and, consequently, the areas that ought to be targeted by the new model.

### **3.2.1 Basic requirements**

Models to assess the quality of live subtitling should meet the following basic requirements. They should:

1. Be functional and easy to apply. Although the use of multiple variables might conceivably be helpful for the researcher, respeakers and trainers should also be able to apply the model on a daily basis.
2. Include the tried and tested principles of WER calculations from speech recognition theory.
3. Take into account the different degrees of editing entailed/required by different programmes. For example, sports commentating is often heavily edited, whereas subtitles for news programmes, especially in the UK, are reproduced almost verbatim (Eugeni 2009).
4. Take into account the fact that the approach to live subtitling may differ from country to country, thus allowing for the possibility