# Developing Live Subtitling in Finland
## Moving from manual subtitling towards respeaking

Kaisa Vitikainen
University of Helsinki

## Abstract

Respeaking with speech recognition software is the preferred way of producing live subtitles worldwide, but it is not used at all in Finland. All live subtitling (the intralingual subtitling of live programs) in Finland is done manually. This paper presents a pilot study on respeaking in Finnish, comparing authentic manual subtitles to simulated respoken subtitles and assessing their accuracy through the NER model. While the accuracy rate of respeaking does not reach the ideal 98%, it is found to be higher than that of manual subtitles, which suffer from significant loss of content. With trained respeakers and some improvements to the speech recognition software, respeaking in Finnish could reach 98% accuracy, and is worth pursuing further.

**Keywords:**    Live subtitling, subtitling for the deaf and hard of hearing, SDH, respeaking, speech recognition

## 1 Introduction

This paper presents the first results of the author's research on the use of speech recognition through respeaking in subtitling for the deaf and hard of hearing (SDH) in Finland, specifically in a live subtitling context. The purpose of this paper is to determine whether respoken subtitles, using Aalto ASR (automatic speech recognition), could surpass current methods of live subtitling in terms of accuracy, and to map areas in which Aalto ASR needs improvement in order to make it a functional tool for live subtitling.

The Finnish Broadcasting Company Yle is required by government decree to provide intralingual subtitles for all of its programming in Finnish and Swedish, save for sports and music programs.[1] The greatest challenge for Yle in trying to fulfill these requirements is the amount of live programming it offers. At the time of writing this paper, Yle is the only Finnish broadcaster to provide subtitles for any live programming, but most of its live programs are not subtitled. The afternoon and evening news are subtitled, as well as some special programs (such as election nights and the President's reception on Finland's Independence Day) (Vitikainen 2015: 20). For the news, the

---

[1] Detailed in the government decree on audio and text services to be added to television programs (A 292/2011) and the government decree on television and radio operation (A 1245/2014), available in Finnish and Swedish.

subtitlers get the scripts and teleprompter texts in advance; although news broadcasts are always cued live (pre-prepared subtitles timed manually during the broadcast), only a small portion of them are subtitled live (Vitikainen 2015: 5). The morning news, Aamu-tv (a daily morning show that lasts for several hours), talk shows and current affairs programs are not subtitled live at all, in part because current methods of live subtitling in Finland are not effective enough to provide quality subtitles for such programs. All live subtitling in Finland is done manually (ie. the subtitler types the subtitles on a regular QWERTY keyboard during the broadcast). Automatic subtitles (speech recognition) and respeaking are not used in Finland at all. Respeaking is defined by Romero-Fresco (2011: 1) as follows:

> A technique in which a respeaker listens to the original sound of a live programme or event and respeaks it, including punctuation marks and some specific features for the deaf and hard of hearing audience, to a speech recognition software, which turns the recognized utterances into subtitles displayed on the screen with the shortest possible delay.

The research done for this paper is a pilot study, and part of a project by Aalto University, Helsinki University and Humak University of Applied Sciences to improve intralingual subtitling in Finland, funded by the Kone Foundation. The material for this paper includes the authentic manual subtitles for one live segment in a news broadcast, respoken subtitles for the same segment, as well as a transcript of the respeaking audio. The manual and respoken subtitles are compared in terms of accuracy using the NER model.

## 2 Research Material

This is a pilot study, as respeaking is not yet used in Finland. There is no research on respeaking in Finnish, and this research faced many challenges in terms of research design and acquiring materials. This section presents the material used in this study, and the challenges faced and compromises made in order to acquire it.

### 2.1 The news segment and manual subtitles

The research material consists of an unscripted segment from a news broadcast and the subtitles as they were broadcast. The subtitles were produced live with the manual method by a professional subtitler during the live broadcast.

The live subtitling was done in block format, i.e. in full text blocks, rather than appearing on screen one word at a time (word scroll). At the time when the material for this research was gathered, this was the standard way of producing live subtitling within a news broadcast. The delay is much greater than in the word scroll method, but the segments within news broadcasts are relatively short (approximately 2 minutes), and it was in line with the scripted part of the broadcast being in block format as well.

In the segment analyzed in this paper, a correspondent covers Finnish prime minister Juha Sipilä's visit to Poland in January 2017. Around the same time, there was public

discourse in Finland about Sipilä's possible conflict of interest regarding a company partly owned by his relatives winning a major bid with the government-owned Terrafame (Yle News 2017). Both the visit and the conflict of interest are referred to in the segment, and the segment was preceded by a news story about the conflict of interest. The news anchor's lines are included for context, but will not be a part of the analysis, as they were scripted and pre-prepared for broadcast.

## 2.2 The respoken subtitles

Since respeaking is not used in Finland, there are no professional respeakers who could produce the respoken subtitles needed for the study. For the same reason, there is no Finnish speech recognition software that would work with the live subtitling software used at Yle. These challenges made authentic respoken subtitles impossible to acquire. Respeaking results were needed to continue the research, and as a compromise, the respoken subtitles are simulated.

   With no professional respeaker available, the author did the respeaking, after some practice with similar materials. While the author had no prior respeaking experience, they were experienced in live subtitling and the editing involved. This means that the following factors need to be taken into account: the respeaker was not trained or experienced in respeaking (which is likely to lower the overall quality of the respeaking), the respeaker was already familiar with the material (had seen the segment before, which is likely to increase the accuracy of the respeaking),[2] and the author is evaluating their own work.

   Because there was no Finnish speech recognition software that would work with the live subtitling program, the respeaking audio was recorded separately (in one take, as close to authentic as possible) and run through the Aalto ASR speech recognition program (Kurimo et al. 2017).[3] The results are treated as though they were subtitles appearing on screen one word at a time, as in the word scroll subtitling method, as opposed to appearing one full subtitle block at a time. The main challenge with this method is that the respeaker cannot edit the subtitles as they might do in an authentic situation, whether to correct errors or add punctuation. At the time the materials were gathered, Aalto ASR did not yet recognize punctuation commands.

## 3 The NER Model

This study is focused on the accuracy of the subtitles. The model best suited for the analysis is the NER model (**n**umber, **e**dition error, **r**ecognition error), developed by Pablo Romero-Fresco (Romero-Fresco 2011; Romero-Fresco & Martínez Pérez 2015).

---

[2] Although the respeaker had seen the material before, during the broadcast, there were some months between the broadcast and the respeaking. This may have mitigated the effect of the respeaker's familiarity with the material.
[3] Aalto ASR was used because it was available and familiar to the author, and because it was part of the larger research project.

Older methods such as the Word Error Rate (WER) model (e.g. Morris et al. 2004) and the CRIM model (Dumouchel et al. 2011) are not suitable for evaluating subtitles produced fully (manually) or in part (respeaking) by the subtitler, because the WER model sees every deviation from the broadcast audio as a mistake and does not account for the material being edited in the process, and the CRIM model only accounts for some deletions. The NER model is best suited for this paper because it allows for 'correct edits', deliberate differences between the original audio and the subtitles, acknowledging that not all differences are mistakes. This is important because subtitling practices vary between countries. In Finland, a good deal of editing is customary both in translated subtitles and in SDH (Peltola & Pöntys 2016; Pöntys 2016; Vertanen 2007). This is in part because the Finnish language has long words in comparison to many other languages, making it more challenging to fit full sentences in the space allowed, and because spoken Finnish and written Finnish are quite different from each other (Tiittula & Voutilainen 2016). If one were to subtitle everything exactly as it is said, the subtitles could be hard to follow; subtitlers at Yle are taught to create the illusion of spoken language rather than repeating everything verbatim (Peltola & Pöntys 2016). The NER model was created to be used with respoken subtitles, though Romero-Fresco and Martínez Pérez posit that it can be used to evaluate automatic subtitles as well (Romero-Fresco & Martínez Pérez 2015: 47).

The NER model calculates accuracy using the following formula:

$$Accuracy \frac{N-E-R}{N} * 100$$

CE (correct editings):
Assessment:

N is the number of words in the (respoken) subtitles. E stands for editing errors, which can be serious, standard or minor, scoring 1, 0.5 or 0.25 points respectively. R stands for recognition errors, with the same severity levels and scoring as editing errors. Serious errors change the meaning of the original text, creating a new, false meaning that could make sense in the context. Normal errors do not create a new meaning but might omit information or change the text into something that makes no sense in the context. Minor errors allow the viewer to follow the meaning of the original text and perhaps even guess the original word. Correct editings are instances in which the respeaker's/subtitler's editing has not led to a loss of information. Finally there is an assessment and analysis of the subtitles, taking priority over the accuracy rate figure. The ideal is to reach 98% accuracy.[4] (Romero-Fresco & Martínez Pérez 2015.)

In the NER model, error scores are given based on idea units (pieces of information) rather than individual words. According to Chafe (1985: 106), idea units are typically

---

[4] There is some variation in what is considered an acceptable accuracy rate. 98% is what Romero-Fresco and Martínez Pérez expect live subtitles to reach with the NER model (2015: 31). Should Finnish broadcasters adopt respeaking, they may set their own requirements for accuracy; such standards do not yet exist for respeaking in Finland.

formed out of one verb phrase along with whatever noun, prepositional or adverb phrase is appropriate, and usually consist of seven words. In Finnish, which lacks prepositions and articles, an idea unit would be shorter in terms of word count, but Chafe's definition is useful when judging the length of an idea unit to score.

The downside of the NER model is that it cannot be automated, but requires a human to do the analysis, judging each deviation as either a correct editing or an error, judging the severity of the error as minor, standard or serious, and producing the verbal assessment. While the model has clear descriptions on what constitutes a minor, standard or serious error, making such judgment calls can be challenging, and NER analyses made by two different people may well result in different accuracy scores.

In this paper, I will be using the NER model to analyze both the authentic manual subtitles and the simulated respoken subtitles, in order to compare their accuracy rates. The NER model is not designed to be used with manually produced live subtitles, but it has been used to analyze subtitles produced by stenography (Romero-Fresco & Martínez Pérez 2015: 33). The main concern with using NER analysis on manual subtitles is that the accuracy rating could be skewed due to the much lower word count of the subtitles compared to the original audio. This is something to keep in mind in the assessment.

There are several concerns with using NER analysis on the subtitles produced by the simulated respeaking: the respeaker was not a professional and had seen the source material before, and the respeaker could not correct mistakes or add punctuation. These concerns are to be taken into consideration in the assessment of the subtitles.


## 4 Analysis

In this section I will provide a detailed analysis of the authentic manual subtitles and the simulated respoken subtitles for one of the news segments in the material. Translations by the author are provided in brackets alongside the full texts in Table 1 and in examples selected for closer analysis and are not repeated in the full listing of errors and correct editings. Errors are sorted first by type (editing or recognition error) and then by severity. Scores and numbers are given as (first half of text) + (second half of text), the dividing line being the news anchor's second question. The news anchor's lines are included for context (in italics), but are not a part of the analysis, as they were scripted. The author's views of idea units are separated in the original text with a vertical line.

**Table 1:** The full texts

| Original text | Manual subtitles | Respoken subtitles |
|---|---|---|
| *Sipilän vierailua Puolassa on seurannut toimittajamme Pirjo Auvinen. Pirjo, pääministeri ei siis näe mitään ongelmia omistajaohjauksen kanssa. Onko tämä asia nyt loppuun käsitelty?* | *Sipilän vierailua Puolassa on seurannut Pirjo Auvinen. [Pirjo Auvinen has been following Sipilä's visit to Poland.]* <br><br> *Pääministeri ei siis näe ongel-* | |

| | | |
|---|---|---|
| *[Our reporter Pirjo Auvinen has been following Sipilä's visit to Poland. Pirjo, so the prime minister does not see any problems with ownership steering. Is this matter now closed?]* | *mia omistajaohjauksen kanssa.* *[So the prime minister does not see any problems with ownership steering.]*<br><br>*Onko asia loppuunkäsitelty?* *[Is the matter closed?]* | |

| | | |
|---|---|---|
| No asia ei varmasti ole loppuun käsitelty, \| mutta se tuli selväksi, \| että pääministeri on oman kantansa muodostanut \| ja siitä hän ei ole valmis millimetriäkään joustamaan.<br>[Well the matter is certainly not closed, but it was made clear that the prime minister has formed his own stance and is not willing to yield one millimeter from it.] | Ei se ole loppuunkäsitelty. [It is not closed.]<br><br>Sipilä on kantansa muodostanut. [Sipilä has formed his stance.] | asia ei varmasti ole loppuunkäsitelty jo se tuli selväksi että pääministeri on kantaansa muodostanut eikä hän siitä jousta [the matter is certainly not closed already it was made clear that the prime minister has formed of his stance and he will not yield from it] |
| Sipilä sanoi, \| että hän ei osallistu tähän keskusteluun sen takia, \| että oikeusasiamies käsittelee parhaillaan noita ensimmäisiä kanteluja \| ja hän on pääministerille antanut antanut sellaisen neuvon, \| että nyt ei saisi keskustella, \| mutta tuskin tämä keskustelu loppuu.<br>[Sipilä said that he is not taking part in this discourse because the ombudsman is currently processing those first complaints and he has advised advised the prime minister that now is not the time to take part in (public) discourse, but it is unlikely that this discourse will end.] | Hän ei osallistu keskusteluun, - [He is not taking part in the discourse -]<br><br>koska oikeusasiamies käsittelee asiaa. [because the ombudsman is processing the matter.] | sipilä sanoi että hän ei osallistu keskusteluun koska oikeus asiamies käsittelee kanteluja hän on antanut neuvon että nyt ei saisi keskustella herrat tuskin keskustelu loppuu [sipilä said that he is not taking part in the discourse because the ombudsman is processing the complaints he has advised that now is not the time to take part in (public) discourse gentlemen it is unlikely that the discourse will end] |
| Siitä pitää varmasti oppositio huoli-huolen \| ja kunhan tuo ensimmäinen ensimmäinen ratkaisu tulee \| niin varmaan sitten myös Sipilä lähtee osallistumaan tähän keskusteluun.<br>[The opposition will surely take car-care of that, and once that first first decision is made, likely then Sipilä will also take part in this discourse.] | Kunhan ratkaisu tulee, keskustelu jatkuu. [Once the decision is made, the discourse will continue.] | oppositio pitää siitä kuolen kunhan ensimmäinen ratkaisu tulee varmasti sipiläkin osallistuu keskusteluun [the opposition will take I die of that once the first decision is made likely sipilä will also take part in the discourse] |

| | | |
|---|---|---|
| *Entä itse tämä tapaaminen Varsovassa? Mikä oli pääministerien tapaamisen tärkein anti?* *[What about this meeting in* | *Mikä oli pääministerien tapaamisen tärkein anti?* *[What was the most important issue in the prime ministers'* | |

| *Warsaw? What was the most important issue in the prime ministers' meeting?]* | *meeting?]* | |
|---|---|---|
| No tärkeimmäksi anniksi nousi \| ei suinkaan mikään Suomen ja Puolan kahdenvälinen asia \| vaan tämä Puolan nykyisen hallituksen ja EU-komission välinen riita, \| joka on todella syvä.<br>[Well the most important issue became not any matter between Finland and Poland but this conflict between Poland's current government and the EU Commission, which is very deep.] | Tärkeimmäksi anniksi nousi Puolan hallituksen ja komission riita.<br>[The most important issue became the conflict between Poland's government and the commission.] | tärkeimmäksi anniksi nousi ei suinkaan suomen ja puolan asiat jo puolan hallituksen ja ee kuulu komission välinen riita se on todella syytä<br>[the most important issue became not any matters of finland and poland already conflict between poland's government and the ee wear commission which is very reason] |
| Komissio syyttää Puolan nykyistä hallitusta oikeusvaltioperiaatteiden rikkomisesta \| ja Puolan nykyinen tämänhetkinen pääministeri Beata Szydło pamautti tänään tuossa pääministerien tiedotustilaisuudessa, \| että että hänhän ei tuollaista arvostelua niele.<br>[The Commission accuses Poland's current government of breaking the principles of a state of justice and Poland's current present prime minister Beata Szydło said today in the prime ministers' press conference that that she will not accept such criticism.] | Hallitusta syytetään oikeusvaltioperiaatteen rikkomisesta.<br>[The government is accused of breaking the principle of a state of justice.]<br><br>Pääministeri ei sitä niele.<br>[The prime minister will not accept it.] | comissio tää puolan hallitusta ja oikeusvaltioperiaatteen rikkomisesta mutta puolan pääministeri pamautti tänään tv että hän ei tällaista arvostelua niele<br>[comission this poland's government and breaking the principle of a state of justice but poland's prime minister said today tv that she will not accept such criticism] |
| Komission tiedot ovat vanhat, ne ovat puutteelliset, \| ja sitä paitsi komissio puuttuu asiaan, joka ei sille kuulu, \| tällainen yhden yksittäisen jäsenmaan tilanne.<br>[The commission has old information, flawed information, and besides the commission is getting involved in a matter that is none of its business, this kind of situation in a single separate member country.] | Hänestä komissio puuttuu asiaan, joka ei sille kuulu.<br>[She thinks the commission is getting involved in a matter that is none of its business.] | komission tiedot ovat puutteellise ja se puuttuu asiaan joka ei sille kuulu jäsenmaan tilanteeseen<br>[the commission has flawe information and it is getting involved in a matter that is none of its business the situation of a member country] |
| Eli Puolan pääministeri on tässä erittäin jyrkkä \| ja tästä on syntymässä todella hankala tilanne.<br>[So the prime minister of Poland is very strict in this, and this is | Tästä tulee todella hankala tilanne.<br>[This will become a very difficult situation.] | puolan pääministeri on erittäin jyry b ja tästä syntyi hankala tilanne<br>[the prime minister of poland is very (nonsense) and this be- |

| | | |
|---|---|---|
| becoming a very difficult situation.] | | came a difficult situation] |
| Suomen, Suomen pääministeri Juha Sipilä sanoi, \| että tämä asia on sellainen, \| johon muut jäsenmaat ei suin surminkaan haluaisi ottaa kantaa.<br>[The prime minister of Finland, Finland Juha Sipilä said that this is the kind of matter that other member countries absolutely do not want to take a stance on.] | Sipilä sanoo, etteivät muut jäsenmaat halua ottaa kantaa.<br>[Sipilä says that the other member countries do not want to take a stance on (this).] | suomen pääministeri juha sipilä sanoi että asia on sellainen ja hår jäsenmaat eivät halua ottaa kantaa<br>[the prime minister of finland juha sipilä said that this is the kind of matter and (nonsense) member countries do not want to take a stance on] |
| Eli nyt odotetaan, \| että komissio saisi tämän ratkaistua, \| mutta tod… missään tapauksessa helpolta ei näytä.<br>[So now it is expected that the commission will resolve this, but it abs… definitely does not look easy.] | | komission toivotaan ratkaisevan tämän<br>[it is hoped that the commission will resolve this] |

## 4.1 Manual subtitles

The manual subtitles are heavily edited, due to the constraints of the method. The word loss compared to the source text is approximately 71% and character loss 70% in the sample analyzed here.

Most edits in the manual subtitles are omissions, and many of them are in line with standard subtitling practices in Finland. Repetition and some inserts, adjectives and adverbs were omitted (e.g. omission of *varmasti* 'certainly', and shortening *oman kantansa* to *kantansa*, 'his own stance' to 'his stance'), and some words and phrases were replaced by shorter synonyms (such as *sen takia, että* to *koska*, both meaning 'because'), and these were classified as correct editings. Many of the omissions classified as errors were minor, as the information lost was not significant (such as omitting *Sipilä sanoi, että* 'Sipilä said that' prior to Sipilä's reasoning for not taking part in the public discourse). There were some standard editing errors as well, where important pieces of information were lost, such as the omission of the ombudsman's advice to Sipilä and the opposition's role in keeping up the discourse. One omission that could cause confusion for the viewer was the shortening of *Puolan nykyinen tämänhetkinen pääministeri Beata Szydło* to *pääministeri*, 'Poland's current present prime minister Beata Szydło' to 'the prime minister'. While omitting 'current' and 'present' would cause no loss of key information in this context, omitting both the name and the country makes it possible to confuse Szydło with Sipilä. Using either 'Prime Minister Szydło' or 'Poland's prime minister' would have been sufficient here. The error was still classified as standard, as it did not convey false information to the viewer, though the potential for misunderstanding was there. The errors were spread fairly evenly between the first and

second half of the text, as separated by the news anchor's second question, with a slight emphasis on the second half (as is to be expected, since it is longer).

N: 21 + 38     59 words, not counting punctuation

E: 3.25 + 3.5    **minor**: 'mutta se tuli selväksi, että', 'ja siitä hän ei ole valmis millimetriäkään joustamaan', 'Sipilä sanoi, että', 'noita ensimmäisiä kanteluja' → 'asiaa', 'mutta tuskin tämä keskustelu loppuu', 'niin varmaan sitten myös', 'Sipilä lähtee osallistumaan tähän keskusteluun' → 'keskustelu jatkuu' // 'ei suinkaan mikään Suomen ja Puolan kahdenvälinen asia', 'vaan tämä', 'EU', 'Puolan', 'tänään tuossa pääministerien tiedotustilaisuudessa', 'komission tiedot ovat vanhat, ne ovat puutteelliset', 'tällainen yhden yksittäisen jäsenmaan tilanne', 'eli Puolan pääministeri on tässä erittäin jyrkkä', 'eli nyt odotetaan', 'mutta missään tapauksessa helpolta ei näytä'

              **standard**: 'ja hän on pääministerille antanut antanut sellaisen neuvon', 'että nyt ei saisi keskustella', 'Siitä pitää varmasti oppositio huoli huolen' // 'Puolan nykyinen tämänhetkinen pääministeri Beata Szydło' → 'pääministeri', 'että komissio saisi tämän ratkaistua'

              **serious**: -

R: -             -

CE:         'no', 'asia' → 'se', 'varmasti', 'pääministeri' → 'Sipilä', 'oman kantansa' → 'kantansa', 'tähän keskusteluun' → 'keskusteluun', 'sen takia, että' → 'koska', 'parhaillaan', 'tuo ensimmäinen ensimmäinen' // 'no', 'nykyisen', 'välinen', 'joka on todella syvä', 'komissio syyttää' → passive tense, 'nykyistä', 'tuollaista arvostelua' → 'sitä', 'ja sitä paitsi', 'hänestä', 'on syntymässä' → 'tulee', 'Suomen, Suomen pääministeri Juha Sipilä' → 'Sipilä', 'että tämä asia on sellainen, johon', 'ei suin surminkaan haluaisi' → 'eivät halua'

*Accuracy*

$$\frac{59 - 6.75 - 0}{59} * 100 = 88.6\%$$

Assessment:      Overall accuracy is fairly low, though higher than expected

               Recognition is good: there are no recognition errors[5]

               As is to be expected with manual live subtitling, the text is edited heavily. While the word loss is severe, the information loss is much less than that. Many of the edits are perfectly in line with Finnish subtitling standards and could be considered correct edits even in normal (non-live) subtitles. Most of the edition mistakes are minor. Some of the edition mistakes leave out pertinent information, but only a few might confuse or mislead the viewer. The core information is still relayed to the viewer, though in heavily edited form. The final sentence is completely omitted, typical of this method of live subtitling due to high delay (Vitikainen 2015: 46). The high delay also makes the subtitles harder to follow. The radically different word count compared to the source text might also skew the results.

---

[5] This is to be expected for manual subtitles, though the subtitler mishearing something could result in a recognition error even in the manual method.

## 4.2 The respeaking test

The word and character loss for the respoken subtitles were much smaller than for the manual subtitles, but still significant: ca. 35% and 34%, respectively. When examining the word loss and accuracy, one should bear in mind both that the respeaker had seen the material beforehand (likely to improve accuracy and reduce word loss) and that the respeaker was inexperienced (likely to decrease accuracy and increase word loss). While these factors may balance each other out to some degree, further research is needed where these factors are not in play.

While there are many omissions in the respoken subtitles, most of the errors are recognition errors by the speech recognition software. With few exceptions, the omissions can be considered correct editings. All of the editing errors are minor, such as using a singular form instead of plural (*oikeusvaltioperiaatteiden* to *oikeusvaltioperiaatteen*, 'the principles of a state of justice' to 'the principle of a state of justice') or using past tense instead of present (*on syntymässä* to *syntyi*, 'is becoming' to 'became'). Only one of the editing errors is an omission that leaves out information, the final *missään tapauksessa helpolta ei näytä*, 'it […] definitely does not look easy'. This is somewhat repetitive, with the earlier comments about the situation being difficult, so the information lost is very minor.

There are quite a few recognition errors, especially in the second half of the text. In the first half, the errors are mostly minor and do not impede comprehension of the subtitles, such as misspellings of compound words (*loppuun käsitelty* to *loppuunkäsitelty* ('closed') and *oikeusasiamies* to *oikeus asiamies* ('ombudsman')), or the wrong form of the right word (*kantansa* to *kantaansa*, 'his stance' to 'of his stance'), which are still easy for the reader to parse. In the first half there were two instances of a word being changed to a word with a different meaning: *mutta* to *jo* ('but' to 'even') and *huolen* to *kuolen* ('care' to 'I die'). These were still classed as minor errors, since in the first case the core message is still conveyed (the prime minister has formed his stance) and in the second case there is only a one-letter difference, and *huolen* is part of a fixed phrase, *pitää siitä huolen* ('will take care of that'), and it is easy for the reader to see what the word is supposed to be, especially since *kuolen* would make no sense in that context. The only standard error in the first half, confusing to the viewer but not giving false information, is the transformation of *mutta* to *herrat*, 'but' to 'gentlemen'.

The second half of the respoken subtitles has more errors than the first half, especially standard errors. The minor errors are misspellings (*komissio* to *comissio*, and *puutteelliset* to *puutteellise*) and minor omissions, as well as missing capital letters in proper nouns, none of which impede understanding of the subtitles. There are several standard errors, though, that make the text difficult to understand. Notably, there are some recognition errors where the words have no meaning in Finnish, such as *jyrkkä* ('strict') becoming *jyry b* and *johon muut* ('that other') becoming *ja hår*.[6] Such errors do not give the reader false information, but make the subtitles confusing to read in places. There were no serious errors.

---

[6] 'Ja' is Finnish for 'and', but 'hår' has no meaning in Finnish. The letter 'å' is not used in modern Finnish at all, though it is part of the Finnish alphabet for historical reasons.

Due to the restrictions of the software used for the testing, there is no punctuation in the respoken subtitles. In an authentic situation, the respeaker could either dictate the punctuation or add it manually, which is why no negative scores were given for missing commas or periods, though the impact of doing so is noted in the assessment below.

N: 55 + 80    135 words, no punctuation marks

E: 0.25 + 1    **minor**: 'ja' // 'oikeusvaltioperiaatteiden' → 'oikeusvaltioperiaatteen', 'on syntymässä' → 'syntyi', 'eli nyt odotetaan, että komissio saisi tämän ratkaistua' → 'komission toivotaan ratkaisevan tämän', 'missään tapauksessa helpolta ei näytä'
**standard**: -
**serious**: -

R: 2 + 6.25    **minor**: 'loppuun käsitelty' → 'loppuunkäsitelty', 'mutta' → 'jo', 'kantansa' → 'kantaansa', 'Sipilä' → 'sipilä' [missing capital letter], 'oikeusasiamies' → 'oikeus asiamies', 'pitää siitä huolen' → 'pitää siitä kuolen' // 'komissio' → 'comissio', 'ja', 'mutta', 'puutteelliset' → 'puutteellise', **missing capital letters**: 'suomen' x 2, 'puolan' x 5, 'juha, sipilä' [9 x 0.25]
**standard**: 'mutta' → 'herrat' // 'vaan' → 'jo', 'EU' → 'ee kuluu', 'syvä' → 'syytä', 'syyttää' → 'tää', 'tv', 'jyrkkä' → 'jyry b', 'johon muut' → 'ja hår'
**serious**: -

CE:    'no', 'oman kantansa' → 'kantansa', 'ja siitä hän ei ole valmis millimetriäkään joustamaan' → 'eikä hän siitä jousta', 'tähän keskusteluun' → 'keskusteluun', 'sen takia, että' → 'koska', 'parhaillaan', 'noita ensimmäisiä kanteluja' → 'kanteluja', 'hän on pääministerille antanut antanut sellaisen neuvon' → 'hän on antanut neuvon', 'tämä keskustelu' → 'keskustelu', 'siitä pitää varmasti oppositio' → 'oppositio pitää siitä', 'tuo ensimmäinen ensimmäinen ratkaisu' → 'ensimmäinen ratkaisu', 'niin varmaan sitten myös Sipilä lähtee osallistumaan tähän keskusteluun' → 'varmasti sipiläkin osallistuu keskusteluun' // 'no', 'mikään Suomen ja Puolan kahdenvälinen asia' → 'suomen ja puolan asiat', 'tämä', 'nykyisen', 'joka on…' → 'se on…', 'nykyistä', 'nykyinen tämänhetkinen', 'Beata Szydło', 'tuossa pääministerien tiedotustilaisuudessa', 'että että hänhän' → 'että hän', 'tuollaista' → 'tällaista', 'tiedot ovat vanhat, ne ovat puutteelliset' → 'tiedot ovat puutteelliset', 'sitä paitsi', 'komissio' → 'se', 'tällainen yhden yksittäisen jäsenmaan tilanne' → 'jäsenmaan tilanteeseen', 'eli', 'tässä', 'todella', 'Suomen, Suomen pääministeri' → 'suomen pääministeri', 'tämä', 'ei suin surminkaan haluaisi' → 'eivät halua'

*Accuracy*

$$\frac{135 - 1.25 - 8.25}{135} * 100 = 93\%$$

Assessment:    Overall accuracy does not reach 98%
The text is edited heavily, but most of the edits are correct edits and in line with standard subtitling practices in Finland. Almost all the errors are recognition errors by the speech recognition software, and many of those are minor, such as omitting the capital letter from a proper noun, or not forming compound words correctly. Especially in the second half of the analyzed segment there are some recognition errors that muddle the meaning of the subtitles and confuse the viewer. Notably, there are a few instances where the speech recognition results

are nonsense and not Finnish words at all.

A significant flaw is the lack of punctuation. If a minor error was assigned for each missing period and comma (23 x 0.25), the overall accuracy would be 88.7%, very close to that of the manual subtitles. On the other hand, if no errors were assigned for missing capital letters or punctuation, the overall accuracy would be 94.8%.

With the simulated respeaking, it was impossible to evaluate delay. Assuming a word scroll method, however, the delay would certainly be much lower than that of the manual subtitles, a mark in favor of respeaking. When comparing results, one should keep in mind the compromises made in the simulated respeaking. On the one hand, the accuracy may be improved because the respeaker had seen the material before. On the other hand, the accuracy may be decreased because respeaker was not experienced, nor trained in respeaking. Even so, the results give an idea of what the accuracy of authentic respoken subtitles could be.

## 5 Conclusions

While the accuracy score for manual subtitles was higher than expected, the respoken subtitles scored higher, even when giving the latter additional negative scores for missing punctuation. Despite the errors in the results, respeaking is much faster than typing manually, and the overall loss of words is significantly smaller than with the manual method. Considering that a more experienced respeaker would produce better results, that in an authentic respeaking situation punctuation could be added to the subtitles, and that the delay would be considerably shorter with respeaking, it is clear that respeaking could produce much higher quality results than manual subtitling when dealing with live broadcasts.

For more reliable results, further research is needed, with a more experienced respeaker who is not familiar with the material used. Based on these results, there are several ways in which the Aalto ASR can be improved: automatic capitalization of proper nouns, punctuation commands, phrase recognition,[7] and reducing the possibility of producing words and phrases that do not exist in Finnish (*jyry b*, *ja hår*).[8] With training, respeakers can produce higher quality subtitles, and in an authentic subtitling context, they can edit and correct subtitles before they are broadcast. With working punctuation and capitalization, and with trained respeakers, the ideal accuracy rate of 98% is not out of reach. What is of utmost importance for future research is to get Aalto ASR – or another Finnish speech recognition program – to function with subtitling software. That would benefit the broadcasters as well.

The data from this study should be considered more approximate than exact. Even so, and despite the challenges and compromises made in this study, the indications are clear: respeaking should be pursued in order to provide a higher quality (and likely

---

[7] This could help eliminate or at least reduce errors such as 'pitää huolen' (a fixed phrase) becoming 'pitää kuolen', as seen in this paper.
[8] Some of these changes have already been implemented since the time this paper was written, namely punctuation commands and automatic capitalization after a period and in some proper nouns.

quantity) for live subtitles in Finland. Further research is also needed, with more authentic material.

## References

Asetus televisio-ohjelmiin liitettävästä ääni- ja tekstipalvelusta 292/2011. [Government decree on audio and text services to be added to television programs.] Available at: http://www.finlex.fi/fi/laki/alkup/2011/20110292 [referenced 20.8.2017].

Asetus televisio- ja radiotoiminnasta 1245/2014. [Government decree on television and radio operation.] Available at: http://www.finlex.fi/fi/laki/alkup/2014/20141245 [referenced 20.8.2017].

Chafe, Wallace 1985. Linguistic differences produced by differences between speaking and writing. In: David Olson, Nancy Torrance & Angela Hildyard (eds) *Literacy, Language, and Learning: The Nature and Consequences of Reading and Writing.* Cambridge: Cambridge University Press, 105 – 23.

Dumouchel, Pierre, Gilles Boulianne & Julie Brousseau 2011. Measures for quality of closed captioning. In: Adriana Şerban, Anna Matamala & Jean-Marc Lavaur (eds) *Audiovisual Translation in Close-up: Practical and Theoretical Approaches*. Bern: Peter Lang, 161 – 172.

Kurimo, Mikko, Seppo Enarvi, Ottokar Tilk, Matti Varjokallio, Andre Mansikkaniemi & Tanel Alumäe 2017. Modeling under-resourced languages for speech recognition. *LANGUAGE RESOURCES AND EVALUATION,* vol 51, 961 – 987.

Morris, Andrew C., Viktoria Maier & Phil Green 2004. From WER and RIL to MER and WIL: Improved evaluation measures for connected speech recognition. Paper presented at *8th International Conference on Spoken Language Processing,* Jeju Island, Korea, 4–8 October 2004.

Peltola, Maire & Minna Pöntys 2016. Ylen ohjelmatekstityksen laatuvaatimukset. [Quality requirements for Yle's SDH.] Unpublished. Yle 22.1.2016.

Pöntys, Minna 2016. Ohjelmatekstitys – silmin nähtävää puhetta. In: Liisa Tiittula & Pirkko Nuolijärvi (eds) *Puheesta tekstiksi. Puheen kirjallisen esittämisen alueita, keinoja ja rajoja.* Helsinki: Suomalaisen Kirjallisuuden Seura, 135 – 161.

Romero-Fresco, Pablo 2011. *Subtitling through Speech Recognition: Respeaking.* Manchester: St. Jerome.

Romero-Fresco, Pablo & Juan Martínez Pérez 2015. Accuracy Rate in Live Subtitling: The NER Model. In: Jorge Díaz Cintas & Rocío Baños Piñero (eds) *Audiovisual Translation in a Global Context: Mapping an Ever-changing Landscape.* Houndmills, Basingstoke, Hampshire: Palgrave Macmillan, 28 – 50.

Tiittula, Liisa & Eero Voutilainen 2016. Puhe, kirjoitus ja puheen muuttaminen kirjoitukseksi. In: Liisa Tiittula & Pirkko Nuolijärvi (eds) *Puheesta tekstiksi. Puheen kirjallisen esittämisen alueita, keinoja ja rajoja.* Helsinki: Suomalaisen Kirjallisuuden Seura, 29 – 62.

Vertanen, Esko 2007. Ruututeksti tiedon ja tunteiden tulkkina. In: Riitta Oittinen & Tiina Tuominen (eds) *Olennaisen äärellä. Johdatus audiovisuaaliseen kääntämiseen.* Tampere University Press, 149 – 170.

Vitikainen, Kaisa 2015. Hyvää iltaa Yle Uutisista. Digiajan uutistekstitys Suomessa. [Pro gradu thesis.] Helsinki: University of Helsinki. Available at: http://urn.fi/URN:NBN:fi-fe2015091611654

Yle News 2017. Chancellor of Justice confirms complaints over PM Sipilä's biotech firm ties. Available at: https://yle.fi/uutiset/osasto/news/chancellor_of_justice_confirms_complaints_over_pm_sipilas_biotech_firm_ties/9398702

**About the author**

Kaisa Vitikainen is a PhD student at the Department of Modern Languages, University of Helsinki. She also works as a subtitler for the Finnish Broadcasting Company. Her research interests include SDH, live subtitling and respeaking.
Email: kaisa.vitikainen (at) helsinki.fi