The International Journal for
Translation & Interpreting
Research
trans-int.org

Speaking your translation: students' first encounter with speech recognition technology

Barbara Dragsted
Copenhagen Business School
bd.isv@cbs.dk

Inger M. Mees
Copenhagen Business School
im.isv@cbs.dk

Inge Gorm Hansen
Copenhagen Business School
igh.isv@cbs.dk

Abstract: In this article we discuss the translation processes and products of 14 MA students who produced translations from Danish (L1) into English (L2) under different working conditions: (1) written translation, (2) sight translation, and (3) sight translation with a speech recognition (SR) tool. Audio output and keystrokes were recorded. Oral and written translation data were examined in order to investigate if task times and translation quality differed in the three modalities. Although task times were found to be highest in written translation, the quality was not consistently better. In addition, since students were dictating in their L2, we looked into the number and types of error that occurred when using the SR software. Items that were misrecognised by the program could be divided into three categories: homophones, hesitations, and incorrectly pronounced words. Well over fifty per cent of the errors were caused by students' mispronunciations.

Keywords: translation processes; oral and written translation; sight translation; speech recognition software; translation quality; pronunciation.

1. Introduction

In this paper we examine the translation processes and performance of 14 Danish MA translation and interpreting (T&I) students at Copenhagen Business School (CBS), who produced translations into English (their L2) under different working conditions: written translation, sight translation and sight translation using a speech recognition program, i.e. software which automatically converts spoken output into written text (see Jurafsky and Martin, 2000, pp. 235–284 for an introduction to SR technology). On the basis of analyses of task times, translation quality and pronunciation challenges, we discuss the benefits and drawbacks of using SR and provide suggestions for improved interaction with the system.

The research questions which will be addressed are the following:

1) What are the task times in the three modalities? Specifically, are there any time savings in sight translation with SR (henceforth SR translation) compared with written translation? Normally, one would expect spoken translation (including SR translation) to be a good deal faster than written translation, but both the fact that students were unfamiliar with the SR software and the fact that they were dictating in their L2 might result in a larger number of errors having to be corrected, and therefore make this modality more time-consuming.
2) Is there any difference in the quality of translation in the three

modalities? A small-scale previous study (see 2.1 below) showed significant time savings in oral compared with written translation without output quality being noticeably affected (Dragsted and Hansen, 2009). Since translators using SR – like translators working in the written modality – have a written representation on the screen, this might lead to better quality in the final SR output than in traditional sight translation output.

3) What type of misrecognitions occur when students sight translate with SR? How many are caused by students' erroneous pronunciations? Are there any other factors which may result in misidentifications? It should be remembered that the students were working in their L2, and even though Danes find it easier to pronounce English correctly than do students from many other countries, it is nevertheless likely that problems will occur.

2. Background

2.1 A three-stage project

The present study reports on the initial experiments of the third stage of a larger project which investigates the coordination of comprehension and text production processes in translation, interpreting, and T&I hybrids, and the potential for convergence between the written and oral modalities of translation. The scheme was originally motivated by a desire to discover if there are advantages to be gained from encouraging students to draw on oral strategies when they produce written translations.

As teachers we have often had the experience that students produce better translations if they trust their first intuition to a greater extent, and think in terms of processing meaning rather than individual words. When writing translations, many learners appear to fall into the trap of endlessly seeking to optimise the text, and rephrasing sentences over and over again, the result of which is all too often a not particularly coherent or natural text. We therefore decided to introduce SR as a means of simulating an interpreting situation where both the source and target texts were visible. A further motivation for employing SR was that language technology increasingly dominates professional translators' lives, thereby making it ever more essential that students are familiarised with the various tools of the trade. Apart from an introduction to translation memory systems and terminological data bases, the CBS T&I curriculum at present does not include any language technology tools.

The project was planned so as to consist of the following three steps:

1) A pilot study (reported in Dragsted and Hansen, 2007) and a more detailed comparative study of written and sight translation (Dragsted and Hansen, 2009), both drawing on experimental data combining keystroke logging, eye-tracking and quality ratings of the spoken and written output. The studies were based on experiments with professional translators and interpreters. Both pointed to the relevance of translators speaking their translation as an alternative to typing, and prompted exploratory studies of the use of SR software in translation.
2) A small-scale in-depth experimental study (reported in Dragsted, Hansen and Sørensen, 2009) involving three professional translators using speech recognition software. The translators were selected on the basis of their general expertise and their varying degrees of previous experience with SR. Only the experienced SR user had a substantial time saving with the

SR tool.

3) These two rounds of experiments were followed by a longitudinal study involving recordings and analyses of process data from a group of 14 T&I students. The study comprised:
   - an initial experiment in which students used SR technology for the first time (the present paper)
   - a period of eight months in which half of the group were given a copy of the SR tool and asked to translate experimental texts and submit them to the researchers at regular intervals. In addition, the students were encouraged to use the program for other assignments produced in the course of their studies and to keep an activity log recording when and how they worked with the SR software, what problems they encountered, etc.
   - a final round of experiments with the same 14 students, followed by retrospective interviews with each participant.

The findings on task times, quality, and pronunciation challenges reported here will serve as a basis for pedagogical studies and in-depth translator behaviour analyses.

2.2 Translating into L2

As stated above, translations were from Danish into English, which was the students' L2. Thus the participants spoke English when they dictated their text. At first sight it may appear an odd choice to ask the students to translate into their L2 rather than their L1. But here it should be pointed out that Danish is a language of limited diffusion, and as such the demand for translation into and from Danish cannot be compared to that of countries whose languages have sizeable numbers of speakers (e.g. English, German, French, Spanish, not to mention Chinese and Arabic). To sustain a professional career, most Danish translators earn their livelihood by working bi-directionally. Consequently, translator training in Denmark focuses equally on translation in both directions.

Although dictating texts in an L2 may be thought to be a major challenge (at least, if the speaker's pronunciation has to be adequate for recognition by the SR software), it was our feeling that it would be less of a daunting task for Danes than for speakers of many other languages (such as Chinese, Japanese or Spanish). In the first place, Danish and English are both Germanic languages, and there are many correspondences in their sound systems (and, in addition, many close relationships in grammar and vocabulary). This holds true both for segmental features (vowels and consonants) and supra-segmental features (e.g. stress and rhythm). Secondly, in Denmark, English is taught from Grades 3 or 4 (age: 9/10), so at the time of the experiment, students had been working with English for at least 13 years. Finally, Danes are constantly exposed to English in the media; films are subtitled rather than dubbed; universities employ it as a medium of instruction; and many companies use English as a lingua franca.

Despite these advantages, it was by no means a foregone conclusion that using SR would be successful for Danish students, and consequently, one major aim of this study was to discover how well the software was able to deal with their audio input. Note that at the time of the recordings the participants had no experience using the software.

3. Features of spoken and written translation

Working methods and strategies adopted by interpreters producing spoken language are generally assumed to be fundamentally different from those employed by translators producing written language (Gile, 1995, pp. 111–114; Agrifoglio, 2004). Sight translation is a hybrid of the written and oral modality, and can be defined as "a specific type of written translation as well as a variant of oral interpretation" (Lambert, 2004, p. 298), where the source text (ST) is written and the target text (TT) is spoken. Sight translation using SR technology adds a further dimension to the spoken-written complex in that the ST is written and the TT is produced orally, but subsequently converted into a written text. While the sight translation process has been investigated by several scholars under varying conditions (e.g. Agrifoglio, 2004; Lambert, 2004; Setton and Motta, 2007), the same does not hold true for sight translation using SR technology.

Although sight translation has been said to have much in common with simultaneous interpretation (e.g. time pressure, anticipation and the oral nature of the task (Lambert, 2004, p. 298; Pöchhacker, 2004, p. 19), it differs from both consecutive and simultaneous interpreting in a number of ways. Firstly, the ST segment continues to be visually accessible to the interpreter/translator (Gile, 1997, p. 204; Agrifoglio, 2004, p. 44), which implies that there is no memory effort of the kind involved in traditional simultaneous and consecutive interpreting (Gile, 1997, p. 203; Shreve, Lacruz and Angelone, 2010, p. 66). Secondly, since sight translation is not paced by the source language speaker, the interpreter/translator has more flexibility in terms of speed of delivery. Nevertheless, it seems that the interpreter/translator will, under normal circumstances, be intent on producing a smooth delivery (Gile, 1995, p. 166; Mead, 2002, pp. 74, 82; Agrifoglio, 2004, p. 45), and the time constraints characterising interpreting are also to some extent present in sight translation.

In the first two steps of the T&I hybrid project (see 2.1 above) we found, in step 1, that interpreters sight translated up to 12 times faster than translators producing written translations, but only the most experienced SR user (step 2) achieved a substantial time saving under the SR condition compared with written translation. In the present study, all participants were subjected to all three modes of translation (written translation, sight translation and SR translation), one of the aims being to examine the different task times (see research question 1 above).

In addition to process differences between written and spoken translation, the products also vary (Chafe and Danielewicz, 1987; Chafe and Tannen, 1987). Spoken and written language are generally characterised by dissimilarities in the variety of vocabulary and "how speakers and writers choose words and phrases appropriate to what they want to say", because "speakers must make such choices very quickly whereas writers have time to deliberate, and even to revise their choices when they are not satisfied. As a result, written language, no matter what its purpose or subject matter, tends to have a more varied vocabulary than spoken" (Chafe and Danielewicz, 1987, p. 86). In other words, writers can, in principle, take as long as they want to find the perfect word or phrase, whereas speakers "may typically settle on the first words that occur to them" (Chafe and Danielewicz, 1987, p. 88).

Since interpreting and translation are subcategories of spoken and written language, the same type of features can be expected to characterise these modalities (Schäffner, 2004, p. 1), notably that the additional time available in the production of written translations can possibly improve the quality. In the study mentioned above (Dragsted and Hansen, 2009), comparisons of interpreters' and translators' sight translation and written output of identical texts showed significant time savings in the oral modality without seriously compromising the output quality. Translators using SR – like translators working in the written modality – have the possibility of revising their choices as they appear on the screen, which might lead to better quality in the final SR output than in traditional sight translation output. Our second research question examines the overall quality of translations produced under the three different conditions.

Using SR as a means of recording output to speed up the process is, of course, only of value if at the same time it leads to an equally good result as compared with other solutions. SR software for English has been developed for various varieties (American English, Australian English, South East Asian English, Indian English and UK English), but not for speakers of English as a foreign language. Therefore we were interested in discovering what sort of misrecognitions would occur when our Danish participants employed the program. Would these be caused mainly by the users' mispronunciations? (See Mees and Collins, 2000, pp. 171–178, for an error analysis of Danish speakers' problems.) Our third research question addresses this issue.

4. Research design and methods

4.1 Participants

The experiments involved 14 Danish T&I students, all in their fourth year of language and translation studies. The students were volunteers recruited from a class of 20 students. All had Danish as their L1 and English as L2. None had previously used speech recognition technology.

4.2 Procedure and data

Data were collected from three different experimental tasks: a written translation task, a traditional sight translation task (without SR) and a sight translation task with SR. All translations were from Danish into English. The three source texts were excerpts, all from the same report, namely the chairman's statement at the 2009 annual general meeting of a major Danish financial institution (*Danske Bank*). Every effort was made to select passages which were as similar as possible with respect to number of words and level of difficulty.

Each passage consisted of approximately 110 words (Text A: 111 words, Text B: 109 words and Text C: 113 words) and dealt with the same subject (the financial crisis). On the basis of an examination of the excerpts with respect to comprehensibility, style and general vocabulary, three professional translators rated them as being approximately equally difficult. Nevertheless, we cannot be certain that this was the case as it is almost impossible to predict what aspects will cause translation problems, and this varies from one individual to another. Therefore it was decided to rotate the order of the tasks to ensure that differences identified between the written and oral modalities were indeed owing to the specific translation mode and not, for instance, to varying levels of difficulty.

Four of the translators produced a written translation of Text A, a sight translation of Text B and an SR translation of Text C; five produced a written translation of Text B, a sight translation of Text C and an SR translation of Text A; and five produced a written translation of Text C, a sight translation of Text A and an SR translation of Text B. The participants were not allowed to use dictionaries or other resources (see 4.3 below).

In the written modality, the ST was displayed in the top window of the screen, and the TT was produced in the bottom window in the standard version of the keystroke logging program Translog (Jakobsen and Schou, 1999).[1] For the sight translation task (without SR), the ST was also displayed on the screen, and the oral translations were recorded in Translog Audio, a special version of Translog which creates an mp3 file of the speech produced during the translation process. In the SR translation task, the participants again produced an oral translation of the text, this time using speech recognition software.[2]

Before embarking on the translation, the participants received a brief introduction to the SR program (including basic oral commands for text revision); in particular, they were advised to speak fluently. After this they performed the program's basic user training in order for the SR tool to be able to recognise and become familiar with their voices and idiosyncratic pronunciation features. During the experiment, the participants were instructed to refrain from using the keyboard for online revision of the transcribed target text, and employ oral commands only. This of course imposed a serious restriction on the participants' ability to work with the program (see 4.3 below). As in the written modality, the ST was displayed in the top window of the screen, and the TT appeared in the bottom window as the participants' oral output was converted into text by the SR system. As in the case of the sight translation, the spoken output was recorded in Translog Audio.

For all three tasks, we tracked the participants' eye movements using a Tobii 1750[3] remote eye-tracker. For an introduction to eye-tracking during reading, see Rayner (1998); Radach, Kennedy and Rayner (2004); Clifton, Staub and Rayner (2007). For studies using eye-tracking in translation research, see for instance Göpferich, Jakobsen and Mees (2008). The eye-tracking data will not be reported here, but the eye-tracking recordings were used as a means of replaying the translation process in the Tobii eye-tracking analysis program ClearView.[4]

The oral translations without SR were transcribed. The transcribers were instructed to write what they heard without altering the text. They were told to add punctuation, but not to indicate hesitation markers, off-the-cuff remarks and transient versions; thus only the final version of the translation was to be written out. The transcriptions, together with the written translations and the written representations of the SR translations, were assessed by three independent evaluators, who were all experienced translators/teachers/examiners. Quality scores were given on a scale from 1 to 5, where 5 indicated highest and 1 lowest quality. The evaluators were requested to give a global score as they would normally do when grading student translation assignments. Apart from this, no specific criteria were provided.

4.3 Limitations of the experimental set-up

As in many experimental translation process studies, the ecological validity of the experiment can be said to have shortcomings, since the

---

[1] http://www.translog.dk
[2] *Dragon Naturally Speaking 10 Preferred* (Nuance Communications, Inc.).
[3] http://www.tobii.se
[4] http://tobii-clearview.software.informer.com

participants found themselves in an unusual situation in a lab facing the challenge of dealing with SR technology for the first time. On the other hand, the excerpts the participants were asked to translate did resemble the texts they are expected to produce as part of their translator training at CBS, and the feeling of being monitored and evaluated may not be all that different from what students regularly experience in the course of T&I training.

The participants did not have Internet access and were not allowed to use dictionaries or other similar support, which rendered the situation very different from the conditions under which students normally translate. However, allowing the participants to access external resources would have created unequal conditions in the three modalities – thus seriously distorting the time differences between the written and the oral translation modes – since students would probably have spent more time on information retrieval under the written condition (cf. Immonen, 2006, p. 319). It would also have been problematic to filter out time spent on the actual translation task as opposed to time spent on the Internet.

Another limitation of the experimental set-up was that under the SR condition the students were only allowed to use oral commands for text revision and not the keyboard. Normally, when working with an SR system, the user can supplement the oral interaction with the program with keystrokes, for instance to correct words which have not been recognised by the program (something which can be expected to happen regularly, especially when working in one's L2 - see section 6) or to edit the text either online or at the end. There were two reasons for this restriction with respect to editing. One was the technical constraint caused by the complex experimental set-up. Three recording programs were running simultaneously during the SR task (keylogging, eye-tracking and SR), and pilot experiments had shown that keyboard activity in the SR task caused Translog to crash resulting in loss of data. However, we did not want to reduce the complexity of the recording and monitoring procedure, because data from eye-tracking and keylogging provide the fine-grained measurements we need for more elaborate analyses of translator behaviour (these findings will be reported in subsequent articles).

A second reason for not allowing keyboard activity under the SR condition was our suspicion that if allowed to type during the SR task, some students would be tempted to fall back on deep-rooted habits of producing translations with the keyboard whenever they experienced a problem with the SR system. In this experiment we were interested in how successful the students' oral interaction was when using SR technology for the first time. These results will subsequently be compared with results from the second round of experiments in the longitudinal study (see 2.1 above).

5. Translation process results

5.1 Task times

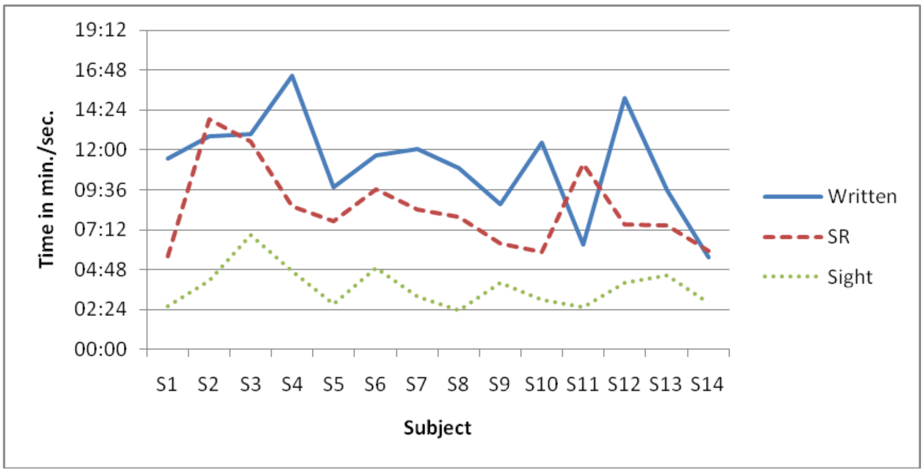
As Table 1 makes clear, average task times for the 14 students were generally longest under the written translation condition, and shortest under the sight translation condition, with SR translation placed in between. There did not seem to be an effect of individual text properties on task times. For example, under the written condition, text A was produced faster than the other texts, whereas under the sight translation condition, Text A took longest. Means for each of the texts are provided in Appendix A.

Table 1: Mean task times in sight translation, SR translation and written translation

	Sight	*SR*	*Written*
Task time (min./sec.)	03:44	08:28	11:07

Let us now look at individual task times to see how well these are reflected in the means.

Figure 1: Individual task times in sight translation, SR translation and written translation



As expected, all 14 students sight translated fastest. All except two (S2 and S11) produced written translations most slowly – which again was not surprising. We were curious to see whether SR translation task times were closer to written or to sight translation. One might have expected that task times in the two oral modes would be very similar, but it turned out that the largest time differences were found between sight translation and SR translation (Figure 1). This can partly be explained by the pronunciation-related challenges (see section 6 for discussion). In addition, monitoring one's own output once it is physically represented on the screen may add time and effort to the spoken translation process (Dragsted et al., 2009), though this assumption will need to be investigated further; it is one of the issues which will be considered in the last phase of the longitudinal study.

5.2 Quality ratings

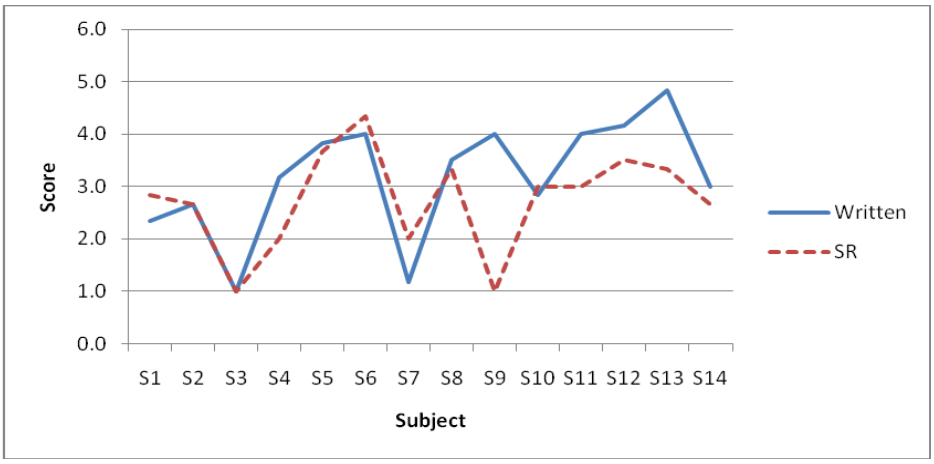
Overall, in terms of time savings, there seems to be a case for using SR technology as an alternative to typing. However, if it turned out that higher productivity was achieved at the cost of output quality, and that any time savings would be cancelled out by the time needed to remedy inaccuracies caused by working in the oral modality, there would be no rationale – at least from a productivity perspective – for using SR in translation. To investigate the output quality aspect, all translations were assessed by three evaluators. The average quality scores for the three tasks can be seen in Table 2.

Inter-rater agreement was high although rater 3 generally gave somewhat higher scores in Sight and SR translation. On average, the translators scored higher under the written translation condition than under the oral conditions. Individual scores for SR translation versus written translation (Figure 2) also revealed a tendency for the quality of the written output to be superior to that of the SR output.

Table 2: Mean of raters' quality scores for 14 students in sight translation, SR translation and written translation (1 = lowest quality, 5 = highest quality)

Quality score	*Sight*	*SR*	*Written*
Rater 1 (mean)	2.6	2.6	3.0
Rater 2 (mean)	2.6	2.6	3.3
Rater 3 (mean)	2.9	3.1	3.2
Mean	2.7	2.8	3.2

Figure 2: Individual quality scores in written translation and SR translation[5]



However, with the exception of S9, who might be considered an outlier, the quality differences are not substantial, and the written output is not consistently better than the SR output. In several cases, the spoken output is as good as (S2, S3, S10) or even better (S1, S6, S7) than the written output.

6. Errors affecting the output in the SR system

In order to discover to what extent mispronunciations or other factors resulted in the SR system displaying unintended text requiring correction, and thus taking up additional time, the audio recordings and the SR-produced written output of all 14 participants were analysed in depth. Although these data could not be accessed using a single device, it was possible to map and trace the entire process by playing the Translog audio file simultaneously with a ClearView reproduction of the output originally produced in the SR system.

When creating a new user in the SR software one can select the type of English accent preferred. All participants opted for British English, even though this may not necessarily have been the best choice – for some, American English might have been a better option (see 6.1.3 below for an example).

[5] Only Written and SR quality scores are included in this figure since the scores for the two oral conditions were very similar (see Table 2); in addition, for this study we were mainly interested in seeing if the time savings in the SR modality were accompanied by poorer quality.

In order for the SR software to recognise that text had to be removed, the participant had to use the oral command "Scratch that", causing the system to delete the word or phrase and enabling the user to make a new attempt. For instance, one participant wanted to say *contributors and creditors*, but because she pronounced the unstressed syllables *con-* and *-tors* in *contributors* with a full rather than a reduced vowel and rendered the /t/ in *creditors* as [ts], the program heard her articulations as the ways indicated in Table 3. (Note that here and below, mispronounced segments and resulting errors are shown in bold-faced type.)

Table 3: Example of the way the SR program interpreted one participant's incorrect pronunciation of the phrase *contributors and creditors*.

Words intended by student	*SR guess*
contribu**tors**	contribute **to this**
contributors	**coun**try riches
contribu**tors**	contribu**ted us**
that **con**tribu**tors**	that **coun**try be**cause**
and other credi**tors**	and other credi**ts his**
contributors and credi**tors**	contributors and credit**s his**
con**tri**butors and credi**tors**	Con**ser**vatives and credi**ts as**

Misrecognitions of this type are obviously time-consuming and frustrating, so in order to help students produce translations with SR more efficiently, it was decided to examine in more detail how many errors occurred and what caused them.

6.1 Types of error

Altogether 173 misrecognitions were identified. Not all participants rectified every error, either because they did not notice them, or because their attempts at correction were unsuccessful. The misrecognitions were divided into a number of different categories which will be explained and exemplified below (see Appendix B for a complete record of errors by student and by category). It can be seen from Table 4 that well over 50 per cent of the incorrectly identified items were caused by the participants' own incorrect pronunciations, which means that it would certainly be worthwhile investing more effort in providing students with a stronger awareness of potential pronunciation pitfalls.

Table 4: Typology of errors

Source of confusion	*Homophones*	*Word boundary problems and hesitations*	*Students' mis-pronunciations*	*Inexplicable*	*Total*
Number of errors	20	33	96	24	173
Percentage of errors	11.6%	19.1%	55.5%	13.9%	100%

As can be seen from Table 4, our analyses showed that the incorrect transcriptions could be divided into three types of error. Firstly, there were errors caused by words that were homophonous. Secondly, we found incorrect transcriptions resulting from hesitations and the software's difficulties locating word boundaries. Both these types (see 6.1.1 and 6.1.2 for examples) are also likely to occur when native speakers use the program, but since our data did not contain a control group with native English speakers, this assumption still has to be confirmed. Thirdly, there were misrepresentations which could be attributed to students' incorrect pronunciations, this being by far the largest group. The remaining errors did not appear to have been caused by the way the participant pronounced the words, but seemed either to result from the inadequacy of the software, or were quite simply inexplicable. Examples are given below from each category.

6.1.1 Homophones

An instance of a homophone which the program perceived incorrectly occurred when a participant wanted to say *the econom**ies*** but where the SR system recorded this as *the econom**y is***. This is presumably because *is* (/ɪz/)[6] is frequently reduced to /z/ in connected speech (shown orthographically as *'s*), e.g. *the econom**y's** improving*, and consequently the system has been programmed to identify such sequences, but appears not always to be able to guess which of the alternatives is intended. In fact, in this particular case there is yet a further complication. The possessive **'s** (as in *the econom**y's** impact*) is also pronounced in the same way. Thus it is impossible to hear the difference between *econom**ies***, *econom**y's*** (gen.) and *econom**y is/'s***. In cases of such homophones, or homophonic sequences, a sophisticated program will be able to draw on contextual clues and statistical information, and it is evident from other occurrences that the SR program indeed does operate in such a manner (see 6.1.2). Another example of a problem arising as a result of words being homophonous is where a participant intended *sub-prime **led** losses*, but which the program registered as *sub-prime **lead** losses*.

6.1.2 Hesitations and word boundary problems

The second type of error is formed by a wide-ranging category comprising hesitations and word boundary problems. Included in this group are errors caused by participants hesitating, prolonging sounds or stopping in mid-word. While it would appear that the speech recognition program was able to an admirable degree largely to ignore not only the most usual manifestations of hesitation (such as *uh*, *uhm*, *mmm*) but also sighs and laughter, all these types of phenomena were nevertheless a not infrequent source of error. Table 5 shows examples of the types of phenomena covered by this category.

Let us now have a closer look at the different types. When a participant said *At the same time* followed by ***uh***, the program wrote *At the same time **as***. This particular example shows that the software uses "contextual clues and statistical information to guess what to transcribe",[7] and types something which is statistically likely; …*same…as* is a frequent collocation, and thus the program comes up with a suggestion that would have worked on many other occasions. But in this particular case it was not what the participant had in mind. Another example occurred when a participant said …*in the GDP **uh***, which was transcribed as *in the GDP **per***…. Conversely, the program

---

[6] Phonetic symbolisation is as in the *Longman Pronunciation Dictionary* (Wells, 2008).

[7] Dragon NaturallySpeaking® Version 10, *End-User Workbook*, p. 26, retrieved 10 November 2010 from http://www.accessamericaat.com/nuance/Dragon%2010%20User%20Workbook%20Watermark.pdf

occasionally interpreted something as a hesitation marker which was not actually such. One student wanted to say *was **a** very difficult*, where the indefinite article *a* was presumably interpreted as ***uh*** by the SR program, and was therefore ignored, the utterance being represented as *was very difficult*.

Table 5: Examples of errors caused by hesitations and word boundary problems

Source of confusion	*Words intended*	*Pronounced as*	*Transcribed as*
Hesitation (uh, uhm) misinterpreted	at the same time	at the same time **uh** ([ə])	at the same time **as**
Word incorrectly interpreted as hesitation and deleted by SR program	was a very difficult	was [ə] very difficult	was very difficult
Prolonging of sound	spread	**ss**pread (the initial consonant /s/ was prolonged)	**this** spread
Pausing before completion of word	subsequently	subsequent---**ly**	subsequent **leak**
Word boundary problems	the so-called	th**e s**o-called	th**is s**o-called

An example of a hesitation error caused by abrupt pausing in a word before it was completed occurred when a participant wanted to say *subsequently* but hesitated after *subsequent-* before uttering ***-ly***, which the SR program displayed as *subsequent **leak***. There were also some instances of a participant lengthening a sound while considering what to say next. One student prolonged the /s/ in ***s**pread*, which was subsequently interpreted as *thi**s** **s**pread*. Another prolonged the /s/ in *di**s**trust*, which the program transcribed as *di**sc** trust*. Finally, the program sometimes found it difficult to determine word boundaries, i.e. to establish where one word ended and the next began. For instance, *th**e** **s**o-called* was interpreted as *thi**s** **s**o-called*, and *ne**w** **s**ub-prime* was registered as *new**s** **th**at prime*.

6.1.3 Students' mispronunciations

The most interesting misrecognitions are perhaps those that can be prevented through instructing students on how to remedy repeatedly occurring erroneous pronunciations of vowels and consonants which are in consequence incorrectly transcribed. Although the SR program can be trained to identify an individual's utterances with an increasing degree of accuracy, this is often a cumbersome way of dealing with the problem. A better approach is, of course, to train the *user* in the correct pronunciation of a particular speech sound, which will result in many other instances of words containing that sound being perceived correctly by the program.

In our sample of 14 Danish students, six pronunciation problems were responsible for the majority of the misrepresentations.[8] It is clear from our analyses that one of the most obvious sources of interference is the mispronunciation of function words belonging to closed grammatical classes, e.g. articles, pronouns, auxiliary verbs, prepositions and conjunctions. Notably, items such as *the, a, their, were, was, to, into, of, that, or* turned out to be stumbling blocks (see Table 6 for examples). There are two chief reasons for this: firstly, the students tended to pronounce these grammatical

[8] For more detail on pronunciation errors of Danish learners, see Davidsen-Nielsen (1994), Livbjerg and Mees (1997) and Mees and Collins (2000).

items as strong forms with a full vowel, rather than as weak forms (Wells, 2008, p. 891) with a reduced vowel; secondly, these words are all very short, and words consisting of one or two syllables are far more difficult to recognise than longer words, since the SR program has more items to choose from, but fewer clues to help it identify what is intended.

Table 6: Examples of incorrectly pronounced function words misinterpreted by the SR program

Words intended by student	*SR guess*
was	worse/wires
were	where
the	these
To	two

Another error related to the mispronunciation of the weak forms of grammatical words is the failure to reduce vowels in unstressed syllables (Wells 2008, p. 892). Unaccented syllables of words are sometimes misunderstood by the program, being heard as separate words (see Table 7 below).

Table 7: Examples of incorrectly pronounced unstressed syllables misinterpreted by the SR program

Words intended by student	*SR guess*
a discom**fort**	and this come **for**
a discom**fort**	there is come **for**ward
contributors	**coun**try riches
contribu**tors**	contribu**ted us**

A widespread Danish error (also true of speakers of many other languages e.g. Dutch, German, Russian, Polish, Turkish, Cantonese, Mandarin, Malay and Japanese (Collins and Mees, 2008, p. 211)) is failure to distinguish voiceless and voiced consonants, especially in syllable-final position. Speakers confuse, for example, /p - b/, /t - d/, /k -g/ and /s -z/. In our sample, *le**d** to* was heard as *le**t** to*, and the item *rate**d*** was interpreted as *rate **it***. On another occasion the /t/ in *ra**t**ed* was registered as *rai**d**ed*. (This particular error would presumably not have occurred if the student had selected American English rather than British English, but this assumption was not tested.) Several participants had trouble with *bi**g***, which was deciphered as *ba**k**e* (or even in one case as *ma**k**e*).

What exacerbates the problem is that English vowels are shortened before voiceless consonants (technically termed "pre-fortis clipping" (Wells, 2008, p. 155)) but retain full length before voiced consonants, so that the vowel in *feet* is shorter than that in *feed*. The SR system has been made sensitive to such length differences since it is one of the clues native speakers use to identify final consonants. To give a possible example, if a speaker accidentally prolongs a vowel that ought to be shortened, the program will interpret the item (say, *wick)* either as a word with the vowel occurring before a voiced consonant (*wig*), or suggest a word with a longer vowel (e.g. *wake*). In our sample, a combination of neutralising the contrast between voiceless and voiced consonants and pronouncing incorrect vowel length resulted in *sin**ce*** being transcribed as *send**s***, *grea**t*** as *gra**d**e*, *wor**s**e* as *word**s***.

Another consonant error heard from many non-native speakers is the replacement of voiceless **th** by /s/ or /t/. In our sample, the items *wor**th*** and *four**th*** were said with final /s/, so that the program registered these words as *wor**se*** and *for**ce***. The risk of the SR system guessing an incorrect word is presumably higher if the pronunciation error results in an existing word. Thus *worth* pronounced with /s/ for **th** is more likely to result in an erroneous rendering (e.g. being mistaken for *worse*), as compared with a mispronunciation of *month*, where the error cannot easily be confused with any existing word.

An error characteristic of the students taking part in this sample, and also typical of Danish speakers in general, is that caused by affrication of /t/, namely releasing the consonant with an [s]-like off-glide, thus [t^s] (Mees and Collins, 2000, p. 28). This results in some words and verb endings being heard incorrectly, e.g. *preven**t*** as *preven**ts***, *credi**t**ors* as *credi**ts** his/credi**ts** as*. Combined with the loss of contrast between voiced and voiceless consonants mentioned above, this has the effect of the software interpreting *sprea**d*** as *sprea**ds*** and *happen**ed*** as *happen**s***.

Finally, we need to consider the difficulty Danish speakers have with the contrast between the vowels in *st**u**ck* /ʌ/ and *st**o**ck* /ɒ/ (Mees and Collins, 2000, pp. 108–111, 176). Danish has a vowel (as in Danish *st**o**k* 'stick') located between these two English sounds, so that when one of the participants said ***lo**sses*, the SR software represented it as *cl**a**sses*, whilst ***sub**-prime* was heard as *s**o**ft crime*. Before certain consonants, Danish attempts at the *stock* vowel also sometimes resulted in confusion with the vowel in *thought*, /ɔː/ (Mees and Collins, 2000, pp. 110–111); *great l**o**sses* was heard as *great l**a**ws*.

In addition to the above-mentioned mistakes, which were found with more than one participant, there were also a number of idiosyncratic errors. See Appendix B for a full overview.

6.1.4 Errors for which there is no obvious explanation

Finally, there were a number of errors which could not be accounted for by any of the above. These could most probably be attributed to the inadequacy of the program, and can be illustrated by means of the examples shown in Table 8.

Table 8: Errors for which there is no obvious explanation

Words intended by student	*SR guess*
50 years	50 units
statistics	acoustics
show a major drop	and nature dropped
since the post-war period	since the post war careered
loans	gnomes
rated	rages

6.2 Numbers and percentages of incorrect guesses

In Table 4, we stated the total number of misrepresentations, but it is also interesting to investigate to what extent there was inter-individual variation. Table 9 presents the scores for each participant for each of the categories of error.

Table 9: Number of errors by category and participant

Students	*Homophones*	*Word boundary problems and hesitations*	*Students' mis-pronunciations*	*Inexplicable*	*Total*
S1	0	8	5	5	18
S2	13	1	15	3	32
S3	0	5	5	1	11
S4	0	2	1	1	4
S5	0	3	5	1	9
S6	2	1	4	0	7
S7	0	0	12	2	14
S8	0	1	6	4	11
S9	1	5	5	4	15
S10	1	2	1	0	4
S11	2	0	14	1	17
S12	0	1	7	1	9
S13	0	4	4	2	10
S14	0	0	12	0	12
Total	19	33	96	25	173

It can be seen that there is a certain amount of individual variation. For instance, most students have few problems (or none) with homophones, but a single participant accounted for 14 of the 20 incorrect transcriptions that were noticed in this area, potentially skewing the overall result. To remedy this, we calculated the overall percentage in two steps: first we determined the percentage of errors for each participant in each of the categories. Then, we took the mean of these by-participant by-category percentages to reach the figures reported in Table 10.

Table 10: Percentage of errors by category and by participant

	Homophones	*Hesitations*	*Students' mis-pronunciations*	*Inexplicable*
S1	0%	44%	28%	28%
S2	41%	3%	47%	9%
S3	0%	45.5%	45.5%	9%
S4	0%	50%	25%	25%
S5	0%	33%	56%	11%
S6	29%	14%	57%	0%
S7	0%	0%	86%	14%
S8	0%	9%	55%	36%
S9	7%	33%	33%	27%
S10	25%	50%	25%	0%
S11	12%	0%	82%	6%
S12	0%	11%	78%	11%
S13	0%	40%	40%	20%
S14	0%	0%	100%	0%
Total	8%	24%	54%	14%

It can be seen that this method of calculating the figures alters the results only slightly. The percentage of homophones is somewhat reduced, while the percentage of hesitations increases. Crucially, the percentage of errors caused by the participants' own mispronunciations remains well over 50%.

6.3 How can the SR transcription quality be improved?

There are various ways of reducing the error rate. One method, adopted very occasionally by our participants, is to use a synonym for the word that has been identified incorrectly. When the word *big* was displayed as *bake*, one participant replaced it by *large*. When another participant attempted to say *USA*, the program represented it as *USE* and *USC*, after which the student replaced it by *America*. This strategy, however, is not always unproblematic because it may change the meaning, style or register in the translation.

Another technique that can be employed is to train the SR program to identify idiosyncratic pronunciations. This is a good approach if a speaker consistently mispronounces a particular word or a restricted number of words. But if a large number of words are transcribed incorrectly, and false transcriptions occur owing to the same vowels and consonants consistently being pronounced erroneously, perhaps a better strategy is to focus on improving the speaker's pronunciation.

As mentioned above, the SR program most frequently represents words incorrectly if they are short words, notably if they belong to the category of grammatical items. Our sample indicates that it is well worth investing some effort in teaching students to pronounce the weak forms of these function words. One participant repeated the definite article *the* many times, pronouncing it as [ðiː] rather than [ðə] on every occasion. The program initially guessed *E*, and subsequently *their/there/their*, but was unable to arrive at the correct item. In addition, students should be advised to concentrate on rendering unstressed syllables of longer words correctly. As illustrated in Table 4, one student was unable to get the program to identify *contributors* because she failed to weaken the first and last syllables of the word. Since it appears that most languages do not have weakening of syllables to the same extent as English, making students aware of this rule will greatly reduce the number of errors.

As stated above (4.2), the recommendation is for the user to produce continuous speech streams. In the same vein, it is important to point out that a good correction strategy is to repeat longer sequences and not merely the word that has been rendered incorrectly. The program finds it easier to identify longer units and can draw on more statistical information and in-built syntactic rules if a stretch of speech is repeated instead of a single word or syllable. Finally, students' attention should be drawn to the significant problem posed by homophones. There is simply no audible difference between *lead* ("type of metal") and *led*, and the program can consequently employ only statistical frequency when guessing what the speaker is aiming at.

7. Conclusions and future perspectives

The findings on our three research questions are summarised below:

1. Written translation was the slowest modality and sight translation the fastest. Surprisingly, SR translation was closer to written than to sight translation.
2. On a five-point quality scale, the average written translation scores are 3.2 while the means of sight and SR translation are 2.7 and 2.8 respectively.

3. The majority of SR recognition problems are caused by students' mispronunciations.

As described in section 2.1, this ongoing study on SR in translation is motivated to a large extent by a desire to integrate spoken and written translation strategies in translator training. We believe that encouraging students to produce translations more spontaneously and fluently whilst drawing on oral translation strategies may not only have certain pedagogical advantages but can also result in better translations. The purpose of the study reported here has been to test the practical consequences (task time, quality and pronunciation challenges) of using an SR system compared with typing.

Our findings indicate that there is a case, in terms of productivity, for using SR, thus accentuating the viability of modernising T&I curricula. Future research into SR in translation will explore the pedagogical implications of integrating SR into translator training, in addition to investigating more generally the effect of SR on the translation process, for instance drawing on eye movement and keylogging data.

It emerged from retrospective interviews carried out in connection with phase 2 of the longitudinal study that virtually all the students were enthusiastic about working with SR, and envisaged SR as part of their tool kit in a future career as professional translators. This means that we have here a useful inexpensive translation tool (EUR 149)[9] that seems to have a motivating influence.

The decision to use SR may be a matter of individual preference: some students (and professional translators) may experience considerable time savings and in general prefer speaking their translation to writing it, whereas others (e.g. experienced touch typists) might be more comfortable with typing their translations. However, with more training and familiarity with the SR system (something that had been achieved when phase 2 of the longitudinal study was carried out), greater time savings and higher quality are likely to be achieved as technical obstacles are either reduced or overcome. We hypothesise that with more practice and training, SR time consumption will approach that of sight translation, and SR quality will approach that of written translation.

Acknowledgments

We wish to thank Beverley Collins (Leiden University Centre for Linguistics) and the two anonymous reviewers for their many valuable and constructive comments and suggestions.

[9] Retrieved 11 Feb 2011 from http://shop.nuance.com/store/nuanceeu/DisplayProductDetailsPage/ProductID.202232300/Currency.EUR.

References

Agrifoglio, M. (2004). Sight translation and interpreting: A comparative analysis of constraints and failures. *Interpreting*, *6*(1), 43–67.

Chafe, W., & Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz, & S. J. Samuels (Eds.), *Comprehending oral and written language* (pp. 83–113). San Diego: Academic Press.

Chafe, W., & Tannen, D. (1987). The relation between written and spoken language. *Annual Review of Anthropology*, *16*, 383–407.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). Amsterdam: Elsevier.

Collins, B., & Mees, I. M. (2008). *Practical phonetics and phonology* (2nd ed.). London: Routledge.

Davidsen-Nielsen, N. (1994). *An outline of English pronunciation.* (Transl. and rev. by Fritz Larsen and Hans Frede Nielsen). Copenhagen: Gyldendal.

Dragsted, B., and Hansen, I. G. (2007). Speaking your translation: Exploiting synergies between translation and interpreting. In F. Pöchhacker, A. L. Jakobsen, & I. M. Mees (Eds.), *Interpreting studies and beyond: A tribute to Miriam Shlesinger*. (Copenhagen Studies in Language 35) (pp. 251–274). Copenhagen: Samfundslitteratur.

Dragsted, B., & Hansen, I. G. (2009). Exploring translation and interpreting hybrids: The case of sight translation. *Meta*, *54*(3), 588–604.

Dragsted, B., Hansen, I. G., and Sørensen, H. S. (2009). Experts exposed. In I. M. Mees, F. Alves, & S. Göpferich (Eds), *Methodology, technology and innovation in translation process research*. (Copenhagen Studies in Language 38) (pp. 293–317). Copenhagen: Samfundslitteratur.

Gile, D. (1995). *Basic concepts and models for interpreter and translator training*. Amsterdam/Philadelphia: John Benjamins.

Gile, D. (1997). Conference interpreting as a cognitive management problem. In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 196–214). Thousand Oaks: Sage.

Göpferich, S., Jakobsen, A. L., & Mees, I. M. (Eds.) (2008). *Looking at eyes. Eye-tracking studies of reading and translation processing*. (Copenhagen Studies in Language 36). Copenhagen: Samfundslitteratur.

Immonen, S. (2006). Translation as a writing process: Pauses in translation versus monolingual text production. *Target*, *18*(2), 313–335.

Jakobsen, A. L., & Schou, L. (1999). Translog documentation. In G. Hansen (Ed.), *Probing the process in translation: Methods and results.* (Copenhagen Studies in Language 24) (pp. 151–186) Copenhagen: Samfundslitteratur.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. New Jersey: Prentice Hall.

Lambert, S. (2004). Shared attention during sight translation, sight interpretation and simultaneous interpretation. *Meta*, *49*(2), 294–306.

Livbjerg, I., & Mees, I. M. (1997). *Practical English phonetics* (2nd ed.). Copenhagen. Schønberg.

Mead, P. (2002). Exploring hesitation in consecutive interpreting. In G. Garzone, & M. Viezzi (Eds.), *Interpreting in the 21st century. Challenges and opportunities* (pp. 73–82). Amsterdam/Philadelphia: John Benjamins.

Mees, I. M., & Collins, B. (2000). *Sound English: A practical pronunciation guide for speakers of Danish* (3rd ed.). Copenhagen: Copenhagen Business School Press.

Pöchhacker, F. (2004). *Introducing interpreting studies*. London/New York: Routledge.

Radach, R., Kennedy, A., & Rayner, K. (2004). *Eye movements and information processing during reading*. Hove/New York: Psychology Press.

Rayner, K. (1998). Eye movement in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Schäffner, C. (2004). Researching translation and interpreting. In C. Schäffner,

(Ed.), *Translation research and interpreting research: Traditions, gaps and synergies*. (pp. 1–9). Clevedon: Multilingual Matters.
Setton, R., & Motta, M. (2007). Syntacrobatics: Quality and reformulation in simultaneous-with-text. *Interpreting*, *9*(2), 199–230.
Shreve, G. M, Lacruz, I., & Angelone, E. (2010). Cognitive effort, syntactic disruption, and visual interference in a sight translation task. In G. M. Shreve and E. Angelone (Eds), *Translation and cognition* (pp. 63–84). Amsterdam/Philadelphia: John Benjamins.
Wells, J. C. (2008). *Longman pronunciation dictionary* (3rd ed.). Harlow: Pearson Education.

Appendix A: Mean task times by text and modality

Written A		Written B		Written C	
S9	525	S2	768	S1	688
S11	379	S3	777	S6	700
S13	575	S4	989	S7	725
S14	332	S5	585	S8	653
		S10	747	S12	908
Mean	**452,75**		**773,2**		**734,8**

SR A		SR B		SR C	
S1	337	S9	382	S2	831
S6	577	S11	668	S3	749
S7	504	S13	446	S4	518
S8	478	S14	355	S5	464
S12	452			S10	352
Mean	**469,6**		**462,75**		**582,8**

Sight A		Sight B		Sight C	
S2	247	S1	158	S9	242
S3	414	S6	296	S11	153
S4	283	S7	189	S13	268
S5	165	S8	141	S14	168
S10	178	S12	242		
Mean	**257,4**		**205,2**		**207,75**

Appendix B: Misrecognitions by participant and by category

Name	**S1**	
	Intended	**Misrecognised as**
Incorrect pronunciation	…in [ən] the (countries)	…and the (countries)
	…our concern is doing [esdüɪŋ] business	…our concern is still in business
	…local losses [lɑːsəs]	…local classes
	…rated [reɪːdəd̥] (issues)	…raided (issues)
	…into [ɪn tu]	…in to
Homophones		
Dragon error	…since the post-war period	…since the post-war careered
	…and became a	…and became the
	…and thus	…and bus
	…and thus	…and in the ass
	…distrust	…this trust
Hesitation error or word boundary problem	…in the countrie-s [kʌntriː:#s] where	…and the country is where
	…what has been go(ing) [bɪn g̊oːɪ] (taking place)	…what has been caught (taking place)
	…spread [s:: spred]	…this spread
	…was that the [ðəː]	…was that they
	…segmented into [ˈɪn:: # tuː]	…segmented in two
	…and thus distrust spread [dɪstrʌst spreːd]	…and thus distrusts spread
	…distrust [dɪs##ˈtrʌst]	…this trust

Name	S2	
	Intended	**Misrecognised as**
Incorrect pronunciation	…for the [diː]	…for E
	…for the [ðë̞ː]	…their
	…the [ðë̞ː]	…there
	…the [ðë̞ː]	…their
	…the [ðë̞ː]	…there
	…Danske [ˈdɛnsgə] Bank	…Dental Bank
	…by the [ðë̞ː] financial crisis	…by their financial crisis
	…(stagnation) of [ɔːːf]	…(stagnation) or off
	…quarter, we [β]	…quarter, be
	…a [ɐː] large	…our large
	…of the [ðë̞ː]	…of their
	…(show) a sudden [ˈsə̞d̥n]	…(show) a certain
	…in the [ðë̞ː]	…in their
	…to the [ðë̞ː] fourth [fɔːs] quarter [kɔːdɔ]	…to their force called
	…the [ðɛː]	…their
Homophones	…have to [tuː] go back	…have two go back
	…to [tuː]	…two
	…to [tuː]	…2
	…to [tuː]	…two
	…to [tuː] go back	…two go back
	…to [tuː]	…two
	…to [tʊː]	…2
	…to [tuː]	…two
	…to [tuː]	…2
	…to [tuː]	…two
	…have to to [tuː tuː] go back	…2 to go back
	…to [tuː]	…two
	…to [tuː] go back to 1955	…two go back to 1955
Dragon error	…to the [ðiː]	…to see
	…we actually have to go back	…we actually had to go back
	…to [tëu]	…cheered on
Hesitation error or word boundary problem	…(in the GDP) /ɜː/	…(in the GDP) per

Name	S3	
	Intended	**Misrecognised as**
Incorrect pronunciation	…Danske [ˈdɛnsgə] Bank	…desk at
	…a great [[eːi greɪːd]	…any grade
	…2% [prəˈsent]	…two present
	…in [en] more than	…and more than
Homophones		
Dragon error	…50 years	…50 units
Hesitation error or word boundary problem	2008 became a [ə] very	2008 became very
	…the [ðiː əː]	…the EU
	…there was [wəs eːːə] in total	…there were severe and told
	…(compared to) this year uh [ɜːː]	…(compared to) this year who
	…(we) have [hɛːʊ] to	…(we) had that

Name	S4	
	Intended	**Misrecognised as**
Incorrect pronunciation	…the [ðiː] statistics	…these statistics
Homophones		
Dragon error	…statistics	…acoustics
Hesitation error or word boundary problem	…was a [əː] very (difficult) ["a" possibly heard as hesitation "uh"]	…was very (difficult)
	The concern [kənˈsɜːnh]	The concerns

Name	S5	
	Intended	**Misrecognised as**
Incorrect pronunciation	…Danish [dẹːnisj] Bank	…then each bank
	…Danske [ˈdɛnsgə] Bank	…then skip back
	...acceleration [ɛksəˈreɪʃn̩]	…expiration
	...figures [ˈfɪgərᵊs]	…vigorous
	…worse [wɜːːrsː]	…words
Homophones		
Dragon error	…show a major drop	…and nature dropped
Hesitation error or word boundary problem	..uh [ɜːː] the group was	…that the group was
	…uh [ɜːː] we are witnessing	…but we are witnessing
	…one point thee three [θi θriː]	…one point eight three

Name	S6	
	Intended	**Misrecognised as**
Incorrect pronunciation	…since [se̝nːs]	…sends
	…collected and and [ɛnd end]	…collected and in
	…rated [reɪdət^{h}]	…rate it
	...rated [reɪt^{s}ət^{h}]	…rates at
Homophones	…economies [iːðs]	…economy’s
	…economies [iːðs]	…economy’s
Dragon error		
Hesitation error or word boundary problem	…subsequent##ly [long pause between “subsequent” and “-ly”]	…subsequent leak

Name	**S7**	
	Intended	**Misrecognised as**
Incorrect pronunciation	…occurrence [ɐˈkjʊərəns]	…appearance
	…where [wə::]	…were
	…great losses [ˈlɔːsᵊs] occurred [ɐˈkjʊəd]	…great laws is cured
	…in USA [juː e seː]	…in USC
	…USA [juː es eː]	…USE
	...local losses [ˈlɔːses]	…lossless
	...losses [ˈlɔːses]	…lossless
	…loss [lɔːs]	…laws
	…loans [lɔ̆uːns]	…looms
	…rated [ˈreːⁱdəd] bonds	…raided bonds
	…all [ɔ̆ːl] over [ɔ̆uwə] the world	…or all-weather world
Homophones		
Dragon error	…mentioned loans [lɔ̆ːuns]	…mentioned gnomes
	…mentioned loans [lɔ̆ːᵘns]	…mentioned illness
Hesitation error or word boundary problem		

Name	S8	
	Intended	**Misrecognised as**
Incorrect pronunciation	…as well as [æs weðːs]	…as worlds
	…as well as [æs weðæðs]	…as worse
	…2007, where [wɛːːə]	…2007, where a
	(crisis) was [wɔ̈ːz]	(crisis) worse
	…in connection with the [ðœ]	…in connection with their
	…were [wøːːə] joined up	…where joined up
Homophones		
Dragon error	…nationally [ˈnæʃənli]	…and the
	...it is [eːs]	…it moves
	…happened. The [ðiː]	…happened with the
	…distrust	…this trust
Hesitation error or word boundary problem	…occurred in the [ðəsːəʊ] so-called	…occurred in this so-called

Name	S9	
	Intended	**Misrecognised as**
Incorrect pronunciation	…big [bḙg̊]	…make
	…big [bë:g̊]	…bake
	…if there occurred [ɔˈkjʊəd] losses [ˈlɑsɪs]	…if there are cured larcenous
	…were [wø̞ː] (protected against)	…wear (protective against)
Homophones	…the rescues	…the rescue is
Dragon error	…if [ɪːf] there	…is there
	…Lehman [ˈliːmən] brothers	…I and others
	…Lehman	…leave and
	…protected [prəˈtektɪt]	…protective
Hesitation error or word boundary problem	…new [njuːːsʌʔb] sub-prime derived	…news that prime
	…financial [lːː] corporations	…financials corporations
	…a distrust [dɪsː trʌst]	…a disc trust
	…uhm [mː]	…Mom
	…uh happened [ɜːː]	…are happened

Name	**S10**	
	Intended	**Misrecognised as**
Incorrect pronunciation	…worse [wɜːθ]	…worth
Homophones	…from Q3 to Q4	…from Q32 Q4
Dragon error		
Hesitation error or word boundary problem	…a h(uge) [ə çː]	…as huge
	…to [s:::] find a year	…to as fine a year

Name	S11	
	Intended	**Misrecognised as**
Incorrect pronunciation	…a discomfort [æ dɪsˈkɔmfɔ̈ːʔt]	…at this com fort
	…a [æ]	…and
	…a [æ]	…there
	…a [æ]	…and
	…a [ɛː]	…air
	…a [æ]… discomfort [dɪsˈkɔmfɔːd]	…and this come for
	…discomfort [dɪsˈkɔmfɔːd]	…and these come for
	[ɛː] discomfort [dɪsˈkɔmfɔːd]	…there is come forward
	…a [æ] discomfort [dɪsˈkɔmfɔːd]	…and this come for
	…led [leʔd] to (difficulties)	…let to (difficulties)
	…big [beːʔk] (financial)	…bake (financial)
	…big [be̹ʔk] (difficulties)	…bake
	...[bɪk]	…it
	...[bɪk]	…pick
	...[bɪk^h]	…a key
	…happened [hæpənt^s]	…happens
Homophones	…sub-prime led	…sub-prime lead
	…recoveries	…recovery is (’s)
Dragon error	…dis-	…this
Hesitation error or word boundary problem		

Name	**S12**	
	Intended	**Misrecognised as**
Incorrect pronunciation	…in the economies [ɪˈkɐnəmiːs]	…in the Academy is
	…summary of what has [wɐʔd hæðs] happened	…summary of White House
	…(crisis) started [stɑːtɪtˢ]	…(crisis) started is
	…was that the [wɔðs ðæʔtˢ]	….(crisis) wires that is
	…had been gathered [ˈgæðɜːtʰ]	…had been gather at
	…into rated [reɪtʰɪtˢ]	…into rate it
	…and mistrust [mɪːs ˈtrʌstˢ]	…and Ms Trusts
Homophones		
Dragon error	…into rated [reɪdɪd]	…into rages
Hesitation error or word boundary problem	…turned in##to [ɪn:tu]	…turned in to

Name	S13	
	Intended	**Misrecognised as**
Incorrect pronunciation	…a distrust [ət˺ dɪstrʌst]	…at this trust
	…(entailed) that [ðɛːd]	…then
	…spread [spreːt^{s}] in	…spreads in
	…and prevent [prəˈvenʔd^{s}] (that)	…and prevents (that)
Homophones		
Dragon error	…got [gɔːd]	…go into
Hesitation error or word boundary problem	…at the same time uh [ɜː]	… at the same time as
	…globally sub- [səb] experienced	…globally is experienced
	…experienced [ɜː]	…experienced by the
	…creditors uhm [ɜːm] (suffered)	…creditors are (suffered)

Name	S14	
	Intended	**Misrecognised as**
Incorrect pronunciation	…these [ði] (sub-prime)	The (soft prime)
	…sub-prime [sɔbːpraɪm]	soft prime
	…a [ä] mistrust	I mistrust
	…companies had [hætsː səʊ] so	…companies that's showed
	…(and make sure) that contributors [kənˈtrɪbjutɪðs]	(and make sure) that contribute to this
	…(prevent that) contributors [kʌnˈtrɪ(bjutəs)]	(prevent that) country riches
	…contributors [təðs]	contributed us
	…contributors [kʌnˈtrɪbjutəːrs]	country because
	…and other creditors [ˈkredɪthɪs]	and other credits his
	…(that) contributors and creditors [ˈkredɪtˢəðz]	(that) contributors and credits his
	…contributors [kəntˢrə] and creditors [kredɪtˢəz]	Conservatives and credits as
Homophones		
Dragon error		
Hesitation error or word boundary problem		