

Chapter 11

Audiovisual speech decreases the number of cognate translations in simultaneous interpreting

Anne Catherine Gieshoff

FTSK Germersheim, Johannes-Gutenberg-Universität Mainz

A large body of research suggests that audiovisual speech facilitates listening comprehension, especially in adverse conditions like noisy environments or hearing impairment, but previous studies on simultaneous interpreting focusing on the interpreting performance failed to demonstrate the benefit of visual input. One explanation might be that conference interpreters increase their cognitive effort to maintain the quality of their rendering. Hence, the impact of visual input might not directly be visible in the interpretation. In order to elucidate this question, I concentrated on self-monitoring in simultaneous interpreting and analyzed the number of cognate translations in a 2×2 factorial design with presence/absence of lip movements and presence/absence of white noise as levels. The results showed an increase of cognate translations when the interpreters worked without visible lip movements, indicating a less effective monitoring in this condition. The findings of this study point out the importance of visual input in simultaneous interpreting and its integration in models of simultaneous interpreting.

1 Simultaneous interpreting as interaction of auditory and visual information processing

Conference interpreters engage in a highly complex task, the oral translation of live speech, where several processes take place more or less simultaneously: comprehension of the source text, storage of the message, retrieval and integration of general knowledge or other previously stored information, rendering of the message in the target language, and (speech or error) monitoring (see for example Seeber & Kerzel 2012; Gile 2009; Setton 1999; Gerver 1975. Beside the



Anne Catherine Gieshoff. Audiovisual speech decreases the number of cognate translations in simultaneous interpreting. In Silvia Hansen-Schirra, Oliver Czulo & Sascha Hofmann (eds.), *Empirical modelling of translation and interpreting*, 313–330. Berlin: Language Science Press. DOI:10.5281/zenodo.1090978

auditory input of the speaker, the interpreter processes a wide range of different visual information: lip and head movements of the speaker, facial expressions and gestures of the speaker and the audience, presentations or projected images, graphs and text elements, written information on a screen or on printed documents like a handout or glossaries and more.¹ While interpreters in general claim to rely on visual input and especially, on the visual contact with the speaker,² process models of simultaneous interpretation failed so far to fully integrate this aspect (see for example Seeber & Kerzel 2012; Gile 2009; Gerver 1975. So the question arises how visual information affects the interpreting process and the interpreting performance and whether interpreters benefit from visual input or not.

A small number of studies have been conducted to elucidate the impact of visual input in simultaneous interpreting. Rennert (2008) opted for a rather direct approach: she asked student interpreters to translate two live speeches and deprived them from any visual information during one speech each. On the whole, she observed no difference in their interpreting performance, except of very few moments where visual input provided necessary complementary information. The author admitted:

“In many instances, visual information was quite redundant, since the information was contained in the verbal message as well. Here it was often difficult to judge the influence of visual input, as the information was conveyed by subjects from both groups. There are several cases where the group with visual contact and the blind booth conveyed information present in both the verbal and the nonverbal material, but it cannot be determined conclusively whether the visual nonverbal information was helpful.” (Rennert 2008: 214)

Despite the fact that participants delivered comparable renderings in terms of quality, they expressed a considerable unease when they had to interpret without visual input and rated the speech as being more difficult than when they had visual contact (Rennert 2008).

¹ISO 2603 states that “booth shall be located at the back and/or the sides of the hall, making sure there is good visual contact between all booths and with the control booth. They shall be raised no further above the floor of the hall than is necessary for a clear view [...] of all proceedings in the hall, i.e. all participants, lecturers, the chairman, etc. as well as visual aids (projection screen, etc.)” (International Organization for Standardization, 2603:1998: 3) Further, booths should be equipped with work-lighting that cover the whole working table and that is “positioned as to avoid shadows being cast by the working interpreter, on the working surface: on documents, equipment, fixtures, etc.” (International Organization for Standardization, 2603:1998: 7)

²The largest professional association of conference interpreters, the AIIC, recommends for example to position the booth so that interpreters have visual contact with the speaker, the audience and the screen (AIIC 2016).

Her results are in line with an earlier experiment conducted by Anderson (1994) who found that a video of the conference setting (no further information about what exactly the video showed are provided) did not improve intelligibility nor “informativeness” (Anderson 1994: 106) of the translations (Anderson 1994). According to a review of several remote interpreting studies by Moser-Mercer (2005), interpreters suffered from concentration difficulties and fatigue when visual input was limited due to the video recordings of the conference setting (Moser-Mercer 2005). A candidate for visual input that improves performance might be the written speech manuscript, provided that the speaker does not deviate from his manuscript (Lambert 2004; but see De Laet & Plas (2005) for the influence of preparation time on performance during simultaneous interpreting with text). To sum up: contrary to what interpreters might expect, these studies did not reveal any differences between interpreting with or without visible input. However, they highlight the unease conference interpreters experience when they have to work without or with limited visual input.

These counter-intuitive results might be due to multiple reasons. First, in most studies the sample is very small with a large variability between subjects that could have covered the effects of the independent variables (Anderson 1994: 108). Appropriate statistical techniques that account for this variability may provide a solution. Second, simultaneous interpreting is a very complex process and visual input covers a range of different information of varying complexity. That is, while some kinds of visual information might facilitate source language comprehension or interpreting in general, for example lip movements of the speaker, others might require additional resources or processing capacities, even if they provide useful information, like presentation charts or additional written information. Researchers studying simultaneous interpreting need to be very careful in their experimental set-up and control for possible confounds in order to tear apart the effects of the various factors. Third, experiments with interpreters usually use the interpreter’s performance, the target text, as dependent variable. This is problematic because common standards of how performances are to evaluate, are lacking. Consequently, target text evaluations might consider different aspects (intelligibility, information content, use of terminology, intonation, etc.) or use different methods (source text analysis, expert judgements, and subjective ratings) and therefore the studies might not be comparable. Moreover, effects of visual input might be absent on a semantic or syntactic level because interpreters increase their cognitive effort to maintain interpreting quality even in adverse conditions. If this is the case, effects would then either be visible at a more fine-grained level, for example in the richness of their vocabulary or in

effective speech monitoring, or under higher cognitive load, for example when working in noisy conditions.

In order to deal with these shortcomings and eliminate as far as possible potentially confounding variables, I opted in the present study for a more systematic and controlled approach and chose to concentrate on lip movements only. A large body of psychological research demonstrates that listening comprehension clearly benefits from visible lip movements, especially in adverse listening conditions, like noise or hearing impairments (Calvert & Thesen 2004; Giraud & Truy 2002; von Kriegstein et al. 2008; Lewandowski 1993; McGettigan et al. 2012; Matys 2011; Rosenblum 2008; McGurk & MacDonald 1976). According to connectionist models of bilingual speech processing, correspondent visual information at a phonological level (lip movements) enhance phoneme recognition by reducing phonological ambiguity and limiting the number of possible candidates (Shook & Marian 2013). If interpreting with visible lip movements facilitates listening comprehension, more resources should be available for other internal processes in simultaneous interpreting, like self-monitoring.

2 Cognates and speech monitoring in bilinguals

One way to assess how visual input might affect self-monitoring in simultaneous interpreting is to check for cognate translations. Cognates are words that share the same etymological roots in two languages and whose orthographic and phonetic representations overlap considerably. According to Paradis (2004), they are immediately understood, even in a foreign language (Paradis 2004). For instance, an English native speaker will immediately understand the German word *Haus* without any knowledge of German because *Haus* resembles very much its English equivalent *house*. A somewhat special case are false friends that are orthographically and phonetically very similar but do not have the same meaning, for example *bekommen* ('to get') and *become*.

In the last years, researchers have made extensive use of cognates to understand bilingual language processing. Their studies have revealed a cognate facilitation effect in a large variety of paradigms, such as lexical decision tasks, priming or picture naming. That is, participants respond faster and more reliably to cognates than to non-cognates or false friends (Peeters et al. 2013; Christoffels et al. 2007; Costa et al. 2005; Christoffels et al. 2003; Dijkstra et al. 1999; van Hell & de Groot 1998; de Groot & Nas 1991), especially in the L2 (Starreveld et al. 2015). This is even the case in sentence processing (Van Assche et al. 2011; 2009; Schwartz 2006). For simultaneous interpreting, a marginal cognate facilitation effect has been shown in the L2 (Dong & Lin 2013).

Cross-lingual interference may also have the reverse effect. In a lexical decision task, Dijkstra and colleagues found significantly lower reaction times for false friends than for non-cognates or cognates. The authors attributed this effect to the inhibition of false friends (Dijkstra et al. 1999). A similar inhibitory effect for cognates was especially observed in mixed language paradigms and in language decision tasks where participants needed to distinguish between both languages (Dijkstra et al. 2015; Acheson et al. 2012; Dijkstra et al. 2010; 1998) see also Christoffels et al. (2007). This cross-lingual interference has not only been shown at an orthographical, but also at a phonological level (Costa et al. 2003; Jared & Kroll 2001). These findings are taken as evidence that competition between cognates is particularly strong and that inhibition mechanisms are necessary to block inappropriate lexical candidates.

Taken together, these findings may indicate that cognates are more strongly activated and therefore, more easily accessed, but also more difficult to inhibit. In order to offer an interpretation of high quality and intelligibility, interpreters need to avoid false friends or cognates that are not very common in the target language (low frequency cognates). This requires them to closely monitor their output so as to detect these inappropriate cognates and to block them. In this view, inhibition can be seen as the result of successful monitoring. However, research so far suggests that monitoring depends on the amount of available resources (Postma 2000; Kessel et al. 2014). For instance, participants made more filled pauses in a story telling task while concurrently exploring figures, compared to a single task condition (Oomen & Postma 2001). In a recent study, Oster (2017 [this volume]) demonstrated that cognate translations reflect the degree of self-monitoring in translation tasks. She observed that translators realize fewer cognates in a written translation than in a spontaneous oral translation where self-monitoring is lowered due to time constraints (see her publication in this volume, 23).

In the present study, I made use of cognate translations in a simultaneous interpreting task to assess the impact of visible lip movements of the speaker and of background noise (white noise) on self-monitoring in simultaneous interpreting. Based on the aforementioned studies demonstrating the benefit of visible lip movements for listening comprehension especially in adverse listening conditions, I hypothesized that

1. interpreters might be able to spend more of their resources on speech production and monitoring when working with visible lip movements, and therefore, might more effectively inhibit low frequency cognates or false friends,

2. masking the source text by adding white noise on the audio track of the source text (signal to noise ratio approximately -10 decibel) should further impact listening comprehension and lead to an increase of cognate translations,
3. the benefit from visible lip movements should be larger in adverse (noise added to source text) than in normal listening conditions (no noise added).

The experiment described below was a pilot study and part of a larger research project on the impact of visual input on cognitive load in simultaneous interpreting. I used different methodologies to assess cognitive load. In this paper, I will focus on cognate translations.

3 Experiment

3.1 Experimental material

The experimental material consisted of four speeches chosen from the basic level of the EU speech repository that makes test speeches available for candidates who prepare to be admitted as freelance interpreter at the European Union. The chosen speeches covered four different topics (air travel, the Greek economic crisis, work conditions and the demographic change). They were in great parts rewritten and edited in order to reduce text complexity as far as possible and obtain a higher comparability between the texts. Words that did not belong to the 5000 most frequent words of American English (Davies 2009) were substituted (mean word length 4.63, SD=0.2). Passive sentence constructions were omitted (with one exception: “born” in “Many children were born.” was accepted, as it is the most frequent form of this verb). Long sentences were split up in order to obtain sentences with maximally one subordinate clause (mean number of words per sentences: 12.5, SD=2.2) The number of functional words (articles, prepositions and other words with a purely grammatical function) and type token relation served as indicator for information density. In every text, functional words made up approximately 40 % of all words (ratio functional words mean=0.4, SD=0.03; mean type token relation: 0.48, SD=0.05). Finally, every text was shortened to approximately 590 words (mean= 588, SD=5.23).

The speeches were read out by an American native speaker and recorded on video. A training session with a metronome helped to ensure a constant speech rate of 140 words per minute within and between texts. When necessary, smaller

adjustments of the speech rate were made by slowing the video down or accelerating it. In the end, all videos were 3'30 to 3'50 minutes long, much less than the speeches of twenty minutes length the participants were used to.

These videos were used to create a 2×2 factorial design: lip movements/ no lip movements \times noise/no noise (see Table 1). The video (condition with visible lip movements) showed the whole face of the speaker. In the audio condition (no visible lip movements), the video stream was replaced by a freeze image of the speaker's face (audio condition). This method allowed to keep screen brightness in all four conditions constant and to reduce light adaptations of the pupil. In the noise condition, white noise was added to the audio stream in the same volume. In order to reduce potential effects of the speeches, I created two groups and inverted the speech in the audio/video-condition. Moreover, I randomized the order of presentation of the conditions for each participant.

Table 1: Experimental conditions

	Group 1		Group 2	
	Audio	Video	Audio	Video
No noise	No lip movements no noise S1	Lip movements no noise S2	No lip movements no noise S2	Lip movements no noise S1
Noise	No lip movements noise S3	Lip movements noise S4	No lip movements noise S4	Lip movements noise S3

3.2 Participants

8 interpreting students in their final year at the Johannes-Gutenberg-Universität Mainz agreed to participate.³ Participants were assigned randomly to one of both groups and interpreted each of the four experimental texts. Due to missing voice recordings, two participants were excluded from the analysis. Participants received 10 euro for participation (and a bar of Swiss chocolate, which was for most of them the main incentive for participation).

³The low number of participants is due to the preliminary character of the study. A follow-up study will be conducted to confirm the results reported in this paper.

3.3 Procedure

Participants received explanations about the procedure beforehand. The experiment encompassed four blocks. Participants pressed a key to start the video and orally translated the speech they heard, while their pupil sizes were measured. Participant's translations were recorded on a second computer with the program audacity (Audacity 2015). After having translated the speech, participants were asked to rate the sound and video quality, the text difficulty, the speech rate and the clarity of the speaker's articulation on a scale from 1 to 4.

3.4 Analysis of cognate translations

For analysis purposes, I extracted all words from the English source texts that showed considerable phonological or orthographical similarity with an existing German word. 77 % of all cognate pairs shared more than 66% of their orthographic forms. The remaining 23 % were phonologically very similar, even if their orthographic form differed (for example: *techniques* – *Technik*) For each English cognate, all possible translations were checked on two online dictionaries (linguee; dict.cc) and their frequency class according to the corpus of the University of Leipzig (Quasthoff et al. 2013) was noted. A cognate translation was considered as “high frequency cognate” if the German cognate was indeed the most frequent translation or only one frequency class below the most frequent translation (example: *international* – *international*). If another translation was considerably more frequent, e.g. at least two frequency classes higher, the cognate translation was judged to be a “low frequency cognate” (example: *to implement* – *implementieren*, the more frequent German translation in this context is *einführen* or *umsetzen*). If the meaning of the cognate translation did not correspond to the contextual meaning of the English source word, the cognate translation was categorized as “false friend” (example: *company* – *Kompanie*: the German word to designate a ballet group or a military unit). The raw figures are displayed in Table 2.

Using R (R Core Team 2014) and the lme4 package (Bates et al. 2015), I constructed a generalized linear mixed model. The dependent variable was the number of cognate translations. It describes if an English cognate has been translated as a German cognate or not. As recommended by Barr et al. (2013), I specified a maximal random effect structure covering intercepts for word category, speech, and participant. Fixed effects included visibility of lip movements, presence of noise, cognate category, the interaction of cognate category and presence of noise, and the interaction of visibility of lip movements and presence of noise.

11 Audiovisual speech decreases the number of cognate translations

Table 2: Total number of items in each cognate category and in each experimental condition. 428 cognates were not translated (missing data).

	audio			video			NA
	high frequency	low frequency	false friend	high frequency	low frequency	false friend	
no noise	220	110	34	182	84	32	428
noise	164	110	34	160	94	38	

P-values for the fixed effects were obtained by comparing the full model against the reduced model without the effect in question with Laplace approximation.

Variability between participants ($SD < .001$), speeches ($SD = <.001$) and word category ($SD < .001$) was low.

The model revealed main effects for the video condition ($z = 2.47$, $p < 0.05$, standard error: 0.247, log odd estimate: 0.5167), low frequency cognates ($z = -8.99$, $p < 0.05$, standard error: 0.256, log odd estimate: -2.295), false cognates ($z = -6.72$, $p < 0.05$, standard error: 0.459, log odd estimate: -3.0866) as well as an interaction between presence of noise and false friends ($z = -2.038$, $p < 0.05$, standard error: 1.126, log odd estimate: -2.2952). The addition of background noise ($z = -2.23$, $p < 0.238$, standard error: 0.24, log odd estimate: 0.283) failed to be significant.

The transformed estimates for the fixed effects show that the probability for cognate translation decreases in the video condition by approximately 8 %. The strongest effect is observed for the cognate category: the probability of cognate translation decreases by 52 % for low frequency cognates and by 64 % for false cognates compared to frequent cognates in the video condition without added noise.

Contrary to our expectation, participants did not make significantly more cognate translations in adverse interpreting conditions where noise was added to the source speech than in normal interpreting conditions. This is surprising in the sense that noise was expected to hamper listening comprehension and therefore to drain more cognitive resources to listening comprehension which should have affected the monitoring of cognate translations. In fact, the interaction between cognate category and addition of background noise estimated by the statistical model indicates that background noise had the opposite effect for high frequency cognates on one hand and false friends on the other hand (see Figure 1), as can also be seen in the observed data (Figure 2). This paradoxical pattern is certainly due to the imbalanced distribution of cognate translations in the three categories:

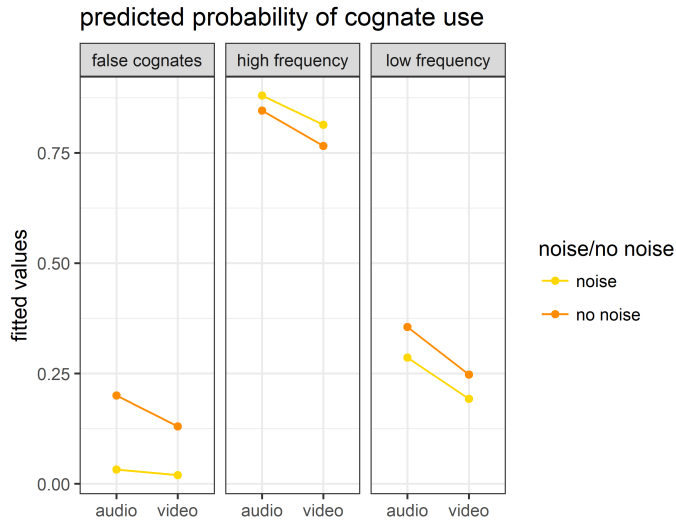


Figure 1: Estimates for cognate translation in the different conditions as predicted by the generalized linear mixed model. The estimate for the noise condition was not significant. For reasons of readability, the log odd estimates are transformed to probability estimates.

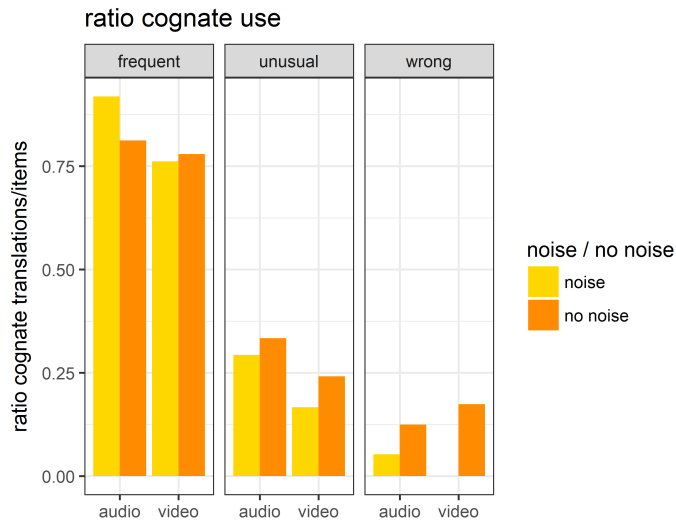


Figure 2: Observed ratio of the number realized cognate translations and the total number of cognates in each cognate category and for each condition.⁴

the number of low frequency and false cognates was substantially lower than the number of high frequency cognates. Each text counted only five to six false friends, but 25 to 35 high frequency cognates. The probability for translating a false friend was thus much lower than for translating a high frequency cognate. A more balanced design with an equal number of cognates in each category and a higher number of participants should help to counter this problem.

4 Discussion

The purpose of the study was to investigate the impact of audiovisual speech, e.g. visible lip movements, and white noise on self-monitoring in simultaneous interpreting by analyzing the number of cognate translations in each condition. On the whole, participants produced very few false friends, regardless the experimental condition. This is in line with research by Van Assche et al. (2011) and by Schwartz (2006) who noted that semantic information influences lexical competition and contributes to suppress lexical candidates that do not fit the semantic context.

Conforming to our hypothesis, participants translated more English cognates as German cognates in the audio condition without visible lip movements than in the video condition with visible lip movements. Compared to the effect of the cognate category (a decrease of cognate translations by 52 % for low frequency cognates and 64 % for false friends), the effect of audiovisual speech (a decrease by 8 % for visible lip movements) seems rather small. Nevertheless, the effect is reliable and is not covered by the larger effect of the cognate category which underlines the importance of audiovisual speech. Participants seemed to be less able to detect and to inhibit cognate translations when they interpreted the source text without seeing the lip movements of the speaker.

One possible explanation is that visible lip movements facilitate listening comprehension in simultaneous interpreting and allows freeing resources for self-monitoring. Researchers observed that listening comprehension benefits from lip movements, especially in adverse listening conditions (cognitive load, background noise, hearing impairment, Mattys 2011; von Kriegstein et al. 2008; Brancazio et al. 2006; Bernstein et al. 2004; Massaro & Light 2004). To account for this observation, Massaro developed a Fuzzy Logical Model of Perception 1999. He assumes that neither auditory speech nor visual speech inputs are unambiguous, or to put it in other words: no signal, whether it comes from the eye or from

⁴Plot created with ggplot (Wickham 2009) in R (Venables et al. 2017)

the ear, is perfect. According to his model, a fuzzy value expresses the extent to which the new sensory information (auditory, visual, haptic or else) corresponds to a certain *prototype*. A prototype describes the features of a perceptual unit of language. Auditory features for language could for example include voicing or formant information; visual features could describe the articulatory movements you see when someone pronounces a sound. The value 1, for instance, corresponds to a complete match, while the value 0 corresponds to a complete mismatch. For example, if the auditory information is ambiguous and corresponds with a value of 0.6 to prototype A and with a value of 0.5 to prototype B, but the visual input is clearly assignable to prototype A (match of value 1), the decision is taken in favor of the prototype A and against prototype B. In this example, the visual input provided complementary information and contributed to disambiguate the auditory input (Massaro & Cohen 1999).

If the auditory and visual input provided complement each other and thereby provide a clearer signal, speech perception processes may need fewer cognitive resources and leave more resources for other processes in simultaneous interpreting. In his effort model for simultaneous interpreting, Gile (2009) describes four efforts which, summed up, indicate the overall resource requirements during simultaneous interpreting: listening and comprehension, speech production, memory and coordination. He presents several examples that illustrate how increased demands of one effort affect the other ones. For instance, a foreign accent or bad pronunciation constrains the interpreter to allocate more resources on the listening and comprehension effort. As a consequence, speech production suffers (clumsy formulations, errors) or memory gets overloaded (information loss, Gile 2009: 173). A similar account could hold for cognate monitoring. Detecting inappropriate cognates needs cognitive resources. If the signal is “noisy” or blurred, the interpreter might devote too much of his resources to the listening and comprehension processes, which leaves insufficient resources for speech production and monitoring. Inversely, if the signal is clearer or less ambiguous, as it is the case audiovisual speech, interpreters may need fewer resources for listening and comprehension and can monitor more closely their output in order to detect and inhibit uncommon cognate translations or false friends.

In addition to its implications for audiovisual speech in simultaneous interpreting, the findings reported in this paper, even though preliminary due to the low number of participants, have also methodological implications. Previous translational studies using a global evaluation of the interpreting performance, like the informational content of the interpretation or other aspects of interpretation quality, failed to demonstrate a benefit for audiovisual speech or visual input in

SI. The present study proposes a new method that succeeded in demonstrating a benefit for audiovisual speech in SI. If these results can be confirmed with a larger sample, the cognate translations analysis could prove itself a suitable method to analyze the influence of other types of visual input or even more global problem triggers in SI, such as high source speech delivery, foreign accent, and concurrent use of other media.

An important limitation of the study concerns the assumptions underlying our hypotheses. For instance, I assumed that interpreters can better monitor low frequency or false cognate translations in audiovisual speech because they benefit from visible lip movements and have more resources available for monitoring in the video condition. However, the experimental set-up does not allow distinguishing between self-monitoring and other processes that could explain a decrease of cognate translations, like for example a larger activation of the semantic networks or a deeper understanding of the source text. In this respect, our study hints towards a benefit from lip movements in simultaneous interpreting but is non-conclusive when it comes to the nature of this effect.

Furthermore, I would like to stress that the study reported in this paper was a pilot study and that the participants were student interpreters. During the years of their professional activity, interpreters acquire a certain expertise that may have an influence on how they process visual input. For instance, their knowledge of their working languages and the ability to discriminate the sounds of these languages improves over the years. Consequently, the benefit of visible lip movements could diminish. Further research is necessary to extend the findings to professional interpreters and to confirm them with a larger sample.

To summarize, the results showed an increase of cognate translations when the interpreters worked without visible lip movements. One explanation might be that self-monitoring is less effective in this condition because conference interpreters need to allocate more of their resources to the comprehension of the source text. The findings of this study point out the importance of visual input in simultaneous interpreting and its integration in models of simultaneous interpreting.

Acknowledgments

I am very grateful to Joachim Kildau and Katharina Oster for their helpful comments on earlier drafts of this paper.

References

- Acheson, Daniel J., Lesya Y. Ganushchak, Ingrid K. Christoffels & Peter Hagoort. 2012. Conflict monitoring in speech production: Physiological evidence from bilingual picture naming. *Brain & Language* 123(2). 131–136.
- AIIC. 2016. *Hiring SI equipments: Tips for conference organisers*. <http://aiic.net/page/151/hiring-si-equipment-tips-for-conference-organisers/lang/1>, accessed 2016-03-28.
- Anderson, Linda. 1994. Simultaneous interpretation: Contextual and translation aspects. In Sylvie Lambert & Barbara Moser-Mercer (eds.), *Bridging the gap: Empirical research in simultaneous interpretation*, 101–120. Amsterdam: Benjamins.
- Audacity. 2015. <http://audacityteam.org/copyright>.
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.
- Bernstein, Lynne E., Edward T. Auer Jr. & Sumiko Takayanagi. 2004. Auditory speech detection in noise enhanced by lipreading. *Speech Communication* 44(1–4). 5–18.
- Brancazio, Lawrence, Catherine T. Best & Carol A. Fowler. 2006. Visual influences on perception of speech and nonspeech vocal-tract events. *Language & Speech* 49(1). 21–53.
- Calvert, Gemma A. & Thomas Thesen. 2004. Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris* 98(1–3). 191–205.
- Christoffels, Ingrid K., Annette M. B. de Groot & Lourens J. Waldorp. 2003. Basic skills in a complex task: A graphical model relating memory a lexical retrieval to simultaneous interpreting. *Bilingualism: Language and Cognition* 6(3). 201–211.
- Christoffels, Ingrid K., Christine Firk & Niels O. Schiller. 2007. Bilingual language control: An event-related brain potential study. *Brain Research* 1147. 192–208.
- Costa, Albert, Angel Colomé, Olga Gómez & Nuria Sebastián-Gallés. 2003. Another look at crosslanguage competition in bilingual speech production: Lexical and phonological factors. *Bilingualism: Language and Cognition* 6. 167–179.

- Costa, Albert, Mikel Santesteban & Agnès Caño. 2005. On the facilitatory effects of cognate words in bilingual speech production. *Brain & Language* 94(1). 94–103.
- Davies, Mark. 2009. *Word frequency data: Corpus of Contemporary American English*. <http://www.wordfrequency.info/>, accessed 2014-03-15.
- de Groot, Annette M. B. & Gerard L. J. Nas. 1991. Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language* 30(1). 90–123.
- De Laet, Frans & Raymond Vanden Plas. 2005. La traduction à vue en interprétation simultanée: Quelle opérationnalité ambitionner? *Meta* 50(4).
- Dijkstra, Ton, Jonathan Grainger & Walter J. B. van Heuven. 1999. Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language* 41(4). 496–518.
- Dijkstra, Ton, Henk van Jaarsveld & Sjoerd ten Brinke. 1998. Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition* 1(1). 55–61.
- Dijkstra, Ton, Koji Miwa, Bianca Brummelhuis, Maya Sappelli & Harald Baayen. 2010. How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language* 62(3). 284–301.
- Dijkstra, Ton, Janet G. van Hell & Pascal Brenders. 2015. Sentence context effects in bilingual word recognition: Cognate status, sentence language, and semantic constraint. *Bilingualism: Language and Cognition* 18(4). 597–613.
- Dong, Yanping & Jiexuan Lin. 2013. Parallel processing of the target language during source language comprehension in interpreting. *Bilingualism: Language and Cognition* 16(3). 682–692.
- Gerver, David. 1975. A psychological approach to simultaneous interpreting. *Meta* 20(2). 199–128.
- Gile, Daniel. 2009. *Basic concepts and models for interpreter and translator training*. Amsterdam: Benjamins.
- Giraud, Anne Lise & Eric Truy. 2002. The contribution of visual areas to speech comprehension: A PET study in cochlear implants patients and normal-hearing subjects. *Neuropsychologia* 40(9). 1562–1569.
- Hermetsberger, Paul (ed.). 2002. *dict.cc. Deutsch–Englisch Wörterbuch*. (dict.cc GmbH). <http://www.dict.cc/>, accessed 2014-12-10.
- International Organization for Standardization. 1998. *Booths for simultaneous interpretation: General characteristics and equipment*. Vernier, Geneva.

- Jared, Debra & Judith F. Kroll. 2001. Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language* 44(1). 2–31.
- Kessel, Ramona, Judith Gecht, Thomas Forkmann, Barbara Druke, Siegfried Gaugel & Verena Mainz. 2014. Metacognitive monitoring of attention performance and its influencing factors. *Psychological Research* 78(4). 597–607.
- Lambert, Sylvie. 2004. Shared attention during sight translation, sight interpretation and simultaneous interpretation. *Meta* 49(2). 294–306.
- Lewandowski, Lawrence J. 1993. The effects of redundancy of bimodal word processing. *Human Performance* 6(3). 229–239.
- Linguee GmbH. 2015. *Linguee: Deutsch–Englisch Wörterbuch*. <http://www.linguee.de/>, accessed 2014-12-10.
- Massaro, Dominic W. & Michael M. Cohen. 1999. Speech perception in perceivers with hearing loss: Synergy of multiple modalities. *Journal of Speech, Language and Hearing Research* 42(1). 21–41.
- Massaro, Dominic W. & Joanna Light. 2004. Using visible speech to train perception and production of individuals with hearing loss. *Journal of Speech, Language and Hearing Research* 47(2). 304–320.
- Mattys, Sven L. 2011. Effects of cognitive load on speech recognition. *Journal of Memory and Language* 65(2). 145–160.
- McGettigan, Carolyn, Andrew Faulkner, Irene Altarelli, Jonas Obleser, Harriet Baverstock & Sophie K. Scott. 2012. Speech comprehension aided by multiple modalities: Behavioral and neural interactions. *Neuropsychologia* 50(5). 762–776.
- McGurk, Harry & John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264. 746–748.
- Moser-Mercer, Barbara. 2005. Remote-interpreting: Issues of multi-sensory intergration in a multilingual task. *Meta* 50(2). 727–738.
- Oomen, Claudy C. E. & Albert Postma. 2001. Effects of divided attention on filled pauses and repetitions. *Journal of Speech, Language and Hearing Research* 44(5). 997–1004.
- Oster, Katharina. 2017. The influence of self-monitoring on the translation of cognates. In Silvia Hansen-Schirra, Oliver Czulo & Sascha Hofmann (eds.), *Empirical modelling of translation and interpreting* (Translation and Multilingual Natural Language Processing 7), 23–39. Berlin: Language Science Press. DOI:10.5281/zenodo.1090948
- Paradis, Michael. 2004. *A neurolinguistic theory of bilingualism*. Amsterdam: Benjamins.

- Peeters, David, Ton Dijkstra & Jonathan Grainger. 2013. The representation of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language* 68(4). 315–332.
- Postma, Albert. 2000. Detection of errors during speech production: A review of speech monitoring models. *Cognition* 77(2). 97–131.
- Quasthoff, Uwe, Dirk Goldhahn & Gerhard Heyer. 2013. *Technical Report Series on Corpus Building: Deutscher Wortschatz 2012*.
- R Core Team. 2014. *A language and environment for statistical computing*. Vienna. <http://R-project.org/>.
- Rennert, Sylvi. 2008. Visual input in simultaneous interpreting. *Meta* 31(1). 204–217.
- Rosenblum, Lawrence D. 2008. Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science* 17(6). 405–409.
- Schwartz, Ana I. 2006. Bilingual lexical activation in sentence context. *Journal of Memory and Language* 55(2). 197–212.
- Seeber, Kilian G. & Dirk Kerzel. 2012. Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism* 16(2). 228–242.
- Setton, Robin. 1999. *Simultaneous interpretation: A cognitive-pragmatic analysis*. Amsterdam: Benjamins.
- Shook, Anthony & Viorica Marian. 2013. The bilingual language interaction network for comprehension of speech. *Bilingualism: Language and Cognition* 16(2). 304–324.
- Starreveld, Peter A., Annette M. B. de Groot, Bart M. M. Rossmark & Janet G. van Hell. 2015. Parallel language activation during word processing in bilinguals: Evidence from word production in sentence context. *Bilingualism: Language and Cognition* 17(2). 258–276.
- Van Assche, Eva, Denis Drieghe, Wouter Duyck, Marijke Welvaert & Robert J. Hartsuiker. 2011. The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language* 64(1). 88–107.
- Van Assche, Eva, Wouter Duyck, Robert J. Hartsuiker & Kevin Diependaele. 2009. Does bilingualism change native-language-reading? *Psychological Science* 20(8). 923–927.
- van Hell, Janet & Annette M. B. de Groot. 1998. Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition* 1(3). 193–211.
- Venables, William N., David M. Smith & R core Team. 2017. *An introduction to R*. <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>.

- von Kriegstein, Katharina, Özgür Dogan, Martina Grüter, Anne-Lise Giraud, Christian A. Kell, Thomas Grüter, Andreas Kleinschmidt & Stefan J. Kiebel. 2008. Simulation of talking faces in the human brain improves auditory speech recognition. *Proceedings of the National Academy of Sciences of the United States of America* 105(18). 6747–6752.
- Wickham, Hadley. 2009. *Ggplot2: Elegant graphics for data analysis*. New York: Springer.