Bilingualism: Language and Cognition

cambridge.org/bil

Exploring bimodal lexical access in unimodal bilinguals through videos with captions

Beatrice Giustolisi[1], Micol De Crescenzo[1] and Caterina Donati[2]

[1]University of Milan-Bicocca, Italy and [2]Université Paris Cité/LLF/Inidex EFL, France

Research Article

Cite this article: Giustolisi, B., De Crescenzo, M. and Donati, C. (2026). Exploring bimodal lexical access in unimodal bilinguals through videos with captions. *Bilingualism: Language and Cognition* 1–10. https://doi.org/10.1017/S1366728926101266

Received: 10 September 2025
Revised: 29 January 2026
Accepted: 1 March 2026

Keywords:
bimodal lexical process; semantic judgment task; audiovisual stimuli; captions; bilingual lexical access

Corresponding author:
Beatrice Giustolisi;
Email: beatrice.giustolisi@unimib.it

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Abstract

Studies on bimodal bilingualism showed that bimodal lexical access is not costly compared to unimodal lexical access, but that it can be even advantageous. We asked whether the same can be observed in unimodal bilinguals as long as some bimodal conditions are provided. We exploited the ecological bimodal setting of subtitled videos and designed four versions of a semantic categorization task, with unimodal (spoken or written language) and bimodal (speech and captions) stimuli. Regardless of the status of the language (L1/L2) and regardless of the bimodal stimuli being also bilingual, answers to bimodal stimuli were systematically faster than speech-only stimuli and slower than written-only stimuli. These results indicate that: i) bimodal stimuli were processed differently from unimodal stimuli, ii) both modalities were taken into account simultaneously, iii) the integration between modalities occurred automatically, even if focusing on one modality only would have been advantageous.

Highlights

- Unimodal bilinguals process bilingual audio (speech)–visual (written) simultaneously
- Integration of bilingual audio (speech)–visual (written) stimuli is automatic
- Effects in processing are driven by modality and not by language dominance

1. Introduction

Code-switching is a behaviour common to all bilinguals, and the only form of linguistic code-mixing that unimodal bilinguals can use when interacting face to face. In fact, unimodal bilinguals master (at least) two languages in the same modality, which means that the main articulators are shared, and simultaneous oral production of both languages is physically impossible. Bimodal bilinguals, on the other hand, are those bilinguals who know (at least) one spoken and one sign language and can use both simultaneously due to the independence of the main articulators. This behaviour is called code-blending, and several studies highlight that it is the preferred form of code-mixing in this population. For example, it has been shown that bimodal bilinguals produce more code-blending compared to code-switching (Emmorey et al., 2008).

A few studies on code-blending processing by bimodal bilinguals showed that, at the lexical level, perceiving two languages simultaneously is not costly as far as reaction times are concerned. On the contrary, bimodal lexical access can be even faster compared to unimodal lexical access (Donati et al., 2025; Emmorey et al., 2012; Giustolisi et al., 2024). All three studies employed a semantic categorization task in which participants had to decide whether different lexical items corresponded to an edible or a not-edible entity. Items were presented in three different conditions: two unimodal and monolingual conditions (sign language only, spoken language only) and one bimodal and bilingual condition (code-blending of sign and spoken language). None of these studies found a cost for bimodal lexical access; rather, in most cases, bimodal lexical access was rather advantageous compared to unimodal lexical access.

The exact mechanism causing this bimodal bilingual facilitation effect is still under-investigated. A possible explanation considers the redundant signals effect (Emmorey et al., 2012), that is, the fact that people are faster in responding to two (non-linguistic) stimuli sharing the same meaning with respect to one single (non-linguistic) stimulus (Miller, 1986). On the other hand, specific linguistic mechanisms might be at play, for example, the double phonological information activated by the spoken language and by the sign language might rapidly integrate and narrow the possible candidates in each language's lexical access (Emmorey et al., 2012).

Another aspect that is unknown is whether this possible advantage of double lexical access is specific to bimodal bilingual individuals, as it might be intrinsically related to the interplay of the spoken and the sign modality, or whether it is a hallmark of general bilingualism, only made visible in bimodal bilinguals due to the exceptional availability of code-blending.

To shed some light on these important questions, we asked if the same effect can be observed in unimodal bilinguals when given access to bimodality, focusing on the spoken and the written

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (http://creativecommons.org/licenses/by-nc/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

modalities. As being exposed to bimodal stimuli when watching movies with subtitles is a very common experience, we decided to design a new version of the task previously used with bimodal bilinguals with spoken–written language bimodal stimuli. We hypothesize that if double lexical activation is cost-free regardless of the modality of the languages involved, then also unimodal bilinguals should show no cost in responding to items presented simultaneously in the two languages compared to unimodal monolingual items, and possibly even an advantage.

Previous studies focusing on monolingual lexical processing have reported word recognition benefits for written and bimodal (simultaneous presentation of orthographic representation and phonological forms) modalities compared to auditory modality alone using lexical decision tasks (e.g., Lopez Zunini et al., 2020). Responses have been found to be faster and more accurate for words presented in the visual and bimodal modalities compared to the auditory modality. The present study builds on these pieces of evidence, yet introduces the important research question of what happens when a bilingual dimension is added to the simple bimodal setting, as we aim at investigating the impact of bimodality on unimodal bilinguals.

For this first study aiming at addressing the question of the possible existence of a bimodal lexical advantage in unimodal bilinguals, we opted for an easily accessible population, that of Italian/English consecutive bilinguals. The two languages share some vocabulary, being both heavily tributary of Latin, but are sufficiently distant in their global lexicon and their phonology, while they share the same writing system. The four experiments that we designed all exploit the same methodology, as will be described in detail, but differ on a number of parameters:

- Experiment 1 involved L1 Italian speakers with good competence in English L2. Spoken stimuli were given in Italian (L1), while subtitles were given in English (L2);
- Experiment 2 involved the same materials as Experiment 1, but targeted L1 English speakers with good competence in Italian L2. Spoken stimuli were given in Italian (L2), while subtitles were given in English (L1);
- Experiment 3 targeted the same population as Experiment 1, but both the spoken stimuli and the written ones were in Italian (L1).

In these three experiments, the mouth of the speaker shown in the video clip was blurred to avoid having visual information associated with the spoken stimulus.

- Experiment 4 was a replication of Experiment 1 but without blurring the speaker's mouth.

In what follows, we describe the four experiments in separate sections (2–5). A general discussion and some conclusions close the article.

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The experimental protocol was positively evaluated by the local commission for minimal-risk studies of the Department of Psychology of the University of Milano-Bicocca.

2. Experiment 1: Italian (L1) stimuli with English (L2) subtitles and blurred mouth

In our first experiment, participants were Italian (L1)–English (L2) bilinguals and were presented with short video clips showing a person uttering a word. The spoken language was the L1 (Italian) and the written language was the L2 (English). The two unimodal conditions were as follows: i) Spoken L1 and ii) Written L2. The bimodal condition was Spoken L1 + Written L2.

In the video clips, the mouth of the speaker was blurred. We took this decision as we did not want any L1 linguistic interference in the muted stimuli showing the L2 written word, and it is well known that watching lip movements contributes to speech perception (e.g., Grant & Walden, 1996; McGurk & MacDonald, 1976; Peelle & Sommers, 2015).

Our research questions were as follows: Is bimodal bilingual simultaneous lexical access costly for unimodal bilinguals? If it is costly due to the difficulty in processing two languages at the same time, we would expect longer reaction times in the bimodal condition compared to the unimodal conditions. On the contrary, if double lexical access is an advantage, we should expect faster reaction times in the bimodal condition compared to both unimodal conditions. Moreover, what is the impact of bilingualism on the processing advantage of written versus auditory stimuli? Being the written language participants' L2, it might be that participants rely more on their L1, thus lowering any advantage of the written modality over the spoken one. In this case, any bimodality effects could be overridden. This scenario is plausible, also considering that participants in the present study are unbalanced bilinguals.

2.1. Methods

2.1.1. Participants

Participants were recruited through *Prolific* using the following screening: Bilingual (native language + one other language, native language + two other languages, native language + three or more other languages); First language = Italian; Fluent languages = English; Language-related disorders = none; Hearing difficulties = no; Dyslexia = no; Primary language = Italian.

Among those who completed the task (N = 50), we had to remove one participant since no data were recorded during the experimental session.

To further check for L2 competence, all participants performed the LexTALE (Lemhöfer & Broersma, 2012), a test of lexical knowledge for English. We selected participants with a LexTALE score corresponding to a level of at least B2 (CEFR), that is, with a score (averaged % correct, % $correct_{av}$) of at least 60%. Two participants were removed from the database due to a low LexTALE score. Mean LexTALE score of the remaining participants was 82.9% (SD = 10.1%). Individual LexTALE scores are plotted in Appendix 1.A, stored on OSF.

The final sample of participants was composed of 47 bilinguals (L1 = Italian, L2 = English; 21 females, 26 males) with a mean age of 32 years (SD = 9).

2.1.2. Materials

To select the material, we started with a list of nouns in Italian corresponding either to an edible or to a not-edible object. Each noun in Italian was paired with the corresponding English noun. For each pair, we used the cross-linguistic overlap scale for phonology (Kohnert et al. 2004) to exclude cognates. We then selected 120 Italian–English pairs, 60 corresponding to an edible object and 60 to a non-edible object. An Italian L1 female speaker was videorecorded while uttering all 120 nouns in Italian.

Videos were cut so that speech began approximately six frames (i.e., 360 ms) after the video start. Videos were then modified using OpenCV in Python. First, in each video, the mouth was blurred. Then, the written English word corresponding to the audio was

https://doi.org/10.1017/S1366728926101266 Published online by Cambridge University Press

added, written in white on a black rectangle, and placed centred at the bottom of the video (Figure 1, top). The written word was synchronized with the speech onset. With these two manipulations, we obtained the bimodal bilingual condition (spoken Italian + written English). Then, to obtain the only speech condition, we used the same script, selecting black as the colour of the English word, so that the written word was not visible. This way, a black rectangle also appeared in the speech-only condition, and not only in the conditions with subtitles (Figure 1, bottom). To obtain the written-only condition, we muted the bimodal condition videos.

The 120 target items were divided into three lists of 40 items each: 20 edible and 20 non-edible. The three lists were balanced for frequency, based on film subtitles corpora (for English: Subtlex-UK, van Heuven et al. 2014; for Italian: Subtlex-IT, Crepaldi et al., 2015). For each condition, each participant received one list, with lists counterbalanced across participants, so that all items appeared in all conditions and no participant received the same item twice.

2.1.3. Procedure

The experimental task was implemented on Labvanced (Finger et al. 2017) and distributed online through Prolific. When entering the task, participants had to agree on data treatment. Then, they completed the LexTALE first and then the experimental tasks. For each item, they were asked to indicate whether it was edible or not edible by pressing the A and L keys.

2.1.4. Data analysis

We analysed accuracy on all items and RTs on correct answers using R version 4.3.1 (R Core team, 2021).

Before data analysis, we checked the mean accuracy by participants to exclude any participant with an overall mean accuracy lower than 80%. No participant was removed following this criterion.



Figure 1. Top: Frame of an experimental item in the written-only or in the bilingual condition (depending on the presence of the audio) showing the English text corresponding to the Italian audio "aceto." Bottom: Frame of an experimental item in the audio-only condition. Videos are stored on the OSF repository linked to the present study.

Accuracy data were analysed with generalized mixed-effects regression (family: binomial, link function: logit; Jaeger, 2008) with condition as fixed factor (treatment coding with bilingual bimodal [Spoken L1 + Written L2] as the reference level, to compare accuracy in the Spoken L1 condition with accuracy in the bimodal bilingual condition and accuracy in the Written L2 condition with accuracy in the bimodal bilingual condition).

RTs that were higher than 3000 ms and lower than 200 ms were excluded from the analysis. RT data were analysed with generalized mixed-effects regression (family: inverse Gaussian, link function: identity, Lo & Andrews, 2015). Condition was entered in the models as a fixed factor, employing treatment coding as in the accuracy analysis.

Both in the accuracy and in the RTs data analyses, we used the maximal random effect structure justified for the design, that is, random intercepts for participants and items and by participants and by items random slopes for the effect of condition ([condition|participant] + [condition|item]). If the model with this random structure did not converge, we simplified the random structure, omitting random slopes until we reached convergence, and we always avoided random structures resulting in a singular fit.

2.2. Results

Results of Experiment 1 are reported in Figure 2.

The analysis of accuracy showed a significant effect of condition ($\chi2(2) = 28.22$, $p < .0001$). The model's parameters are reported in Table 1. Accuracy was lower in the Written L2 condition compared to the Spoken L1 + Written L2 condition, whereas the difference between Spoken L1 and Spoken L1 + Written L2 conditions was not significant. The analysis of RTs also showed a significant effect of condition ($\chi2(2) = 1905.5$, $p < .0001$). RTs in the Spoken L1 condition were slower than in the Spoken L1 + Written L2 condition, whereas RTs in the Written L2 condition were faster than RTs in the Spoken L1 + Written L2 condition.

2.3. Interim discussion

The analysis of accuracy revealed lower accuracy in the Written L2 condition compared to the Spoken L1 + Written L2 condition, as expected, given the lower competence in the L2 compared to the L1 in our group of bilinguals.

As for RTs, responses in the bimodal bilingual Spoken L1 + Written L2 condition were faster than those in the unimodal monolingual Spoken L1 condition. On the contrary, responses in the bimodal bilingual Spoken L1+ Written L2 condition were slower than in the unimodal monolingual Written L2 condition.

Interestingly, regardless of the fact that the written stimulus was in the L2, participants were faster in the bimodal condition compared to the spoken language (L1)-only condition, whereas the bimodal condition was processed slower than the unimodal written condition, thus paralleling monolingual results focusing on modality effects (Lopez Zunini et al., 2020).

In order to better interpret these results, we designed three new versions of this experiment, each controlling for factors possibly at play.

Is there any effect of language dominance in these results? We might hypothesize that since speech was in the dominant language, participants might be biased to process also the L1, hence being slower in the Spoken L1 + Written L2 modality compared to the Written L2 modality. To assess this possibility, we designed Experiments 2 and 3.

https://doi.org/10.1017/S1366728926101266 Published online by Cambridge University Press

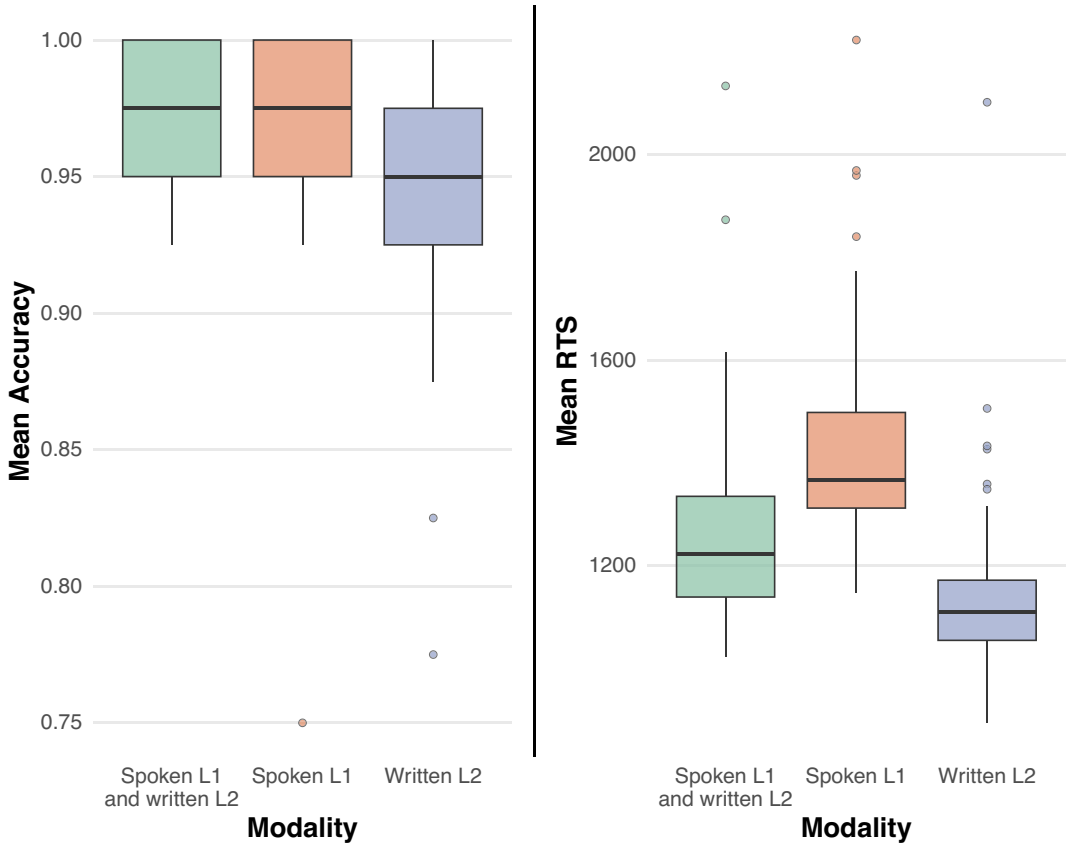


Figure 2. Boxplots showing participants' mean accuracy and RTS by modality in Experiment 1.

Table 1. Parameters of the models for Experiment 1

Dependent variable				
Accuracy	Random factors	Variance	SD	
	Item	1.69	1.30	
	Participant	0.34	0.58	
	Fixed factors	Estimate	SE	z-value
	Condition Written L2 vs. Spoken L1 + Written L2	−0.88	0.19	−4.57
	Condition Spoken L1 vs. Spoken L1 + Written L2	−0.08	0.21	−0.40
RTs	Random factors	Variance	SD	
	Item	3032	55.06	
	Participant	7202	84.86	
	Fixed factors	Estimate	SE	t-value
	Condition Written L2 vs. Spoken L1 + Written L2	−112.13	5.04	−22.24
	Condition Spoken L1 vs. Spoken L1 + Written L2	168.25	5.91	28.47

In Experiment 2, we used the same materials as in Experiment 1, but participants were bilinguals with English as L1 and Italian as L2. If slower RTs in the bimodal condition compared to the written-only condition in Experiment 1 were due to speech being in the L1, then we should not expect such an effect in Experiment 2, where speech was in L2.

In Experiment 3, participants were bilinguals with Italian as L1 and English as L2, as in Experiment 1, but we modified the videos so that subtitles were in Italian (as well as speech). This is a second control of the effect of language dominance. Being both speech and writing in the same language (the L1), no difference across conditions could be attributed to bilingual-related factors, such as language dominance. This experiment also allows us to determine whether the facts that we observe in the first experiments are properties related to bilingualism or simply to bimodality.

Another possibility that we wanted to explore was the effect of the blurred mouth. Since spoken language processing is multimodal and mouth movements have an impact on language comprehension (e.g., Grant & Walden, 1996; McGurk & MacDonald, 1976; Peelle & Sommers, 2015), it might be that in our experiment, spoken language processing was hindered by the blurred mouth. To explore this possibility, in Experiment 4 we tested again Italian (L1)–English (L2) bilinguals using the same materials as in Experiment 1, but with videos modified so that the mouth was not blurred.

3. Experiment 2: Italian (L2) stimuli with English (L1) subtitles and blurred mouth

This section presents our second experiment, which has the same materials as Experiment 1, but targets a different population.

3.1. Methods

3.1.1. Participants

Participants were recruited through *Prolific* using the same screening as in Experiment 1, with the following modifications: First language = English (instead of Italian); Fluent languages = Italian

https://doi.org/10.1017/S1366728926101266 Published online by Cambridge University Press

(instead of English); Primary language = English (instead of Italian).

To further check for their L2 competence, all participants performed the LexITA (Amenta, Badan, & Brysbaert, 2021), a test of lexical knowledge for Italian based on the LexTALE (Lemhöfer & Broersma, 2012), the test we used to assess English knowledge in L2 English speakers.

We selected participants with a LexITA score of at least 26, which corresponds to one standard deviation between the mean score of B2 participants in Amenta et al. 2021. Twelve participants were removed from the database due to a low LexITA score. Mean LexITA score of the remaining participants was 43.8 (SD = 11.2). Individual LexITA scores are plotted in Appendix 1.B, stored on OSF.

The final sample of participants was composed of 42 bilinguals (L1 = English, L2 = Italian, 23 females, 18 males, 1 not declared) with a mean age of 42 years (SD = 14).

3.1.2. Materials

Same as in Experiment 1.

3.1.3. Data analysis

Same as in Experiment 1. One participant was removed because of an overall accuracy of less than 80% correct.

3.2. Results

Results of Experiment 2 are depicted in Figure 3.

The effect of condition was significant both in the analysis of accuracy ($\chi2(2) = 51.40$, $p < .0001$) and in the RT analysis ($\chi2(2) = 154.1$, $p < .0001$). The models' parameters are reported in Table 2: accuracy was lower in the Spoken L2 condition compared to the bimodal bilingual condition (Written L1 + Spoken L2), whereas there was no significant difference between accuracy in the Written L1 condition and the bimodal condition. As for RTs, responses to the bimodal bilingual condition were faster compared to the Spoken L2 condition and slower compared to the Written L1 condition.

4. Experiment 3: Italian (L1) stimuli with Italian (L1) subtitles and blurred mouth

This section presents our third experiment, which has the same population as Experiment 1, presented with the same items as in Experiment 1, but with captions in Italian instead of English.

4.1. Methods

4.1.1. Participants

Participants were recruited through *Prolific* using the same screening as in Experiment 1. Participants who took part in Experiment 1 could not participate in Experiment 3. Forty-six participants completed the task; however, one participant did the task twice and their data were removed from the database.

English competence was again tested using the LexTALE (Lemhöfer & Broersma, 2012), and the same threshold of at least 60% $correct_{av}$ was adopted. Two participants were removed following this criterion. Mean LexTALE score of the remaining participants was 81.7% (SD = 10.2%). Individual LexTALE scores are plotted in

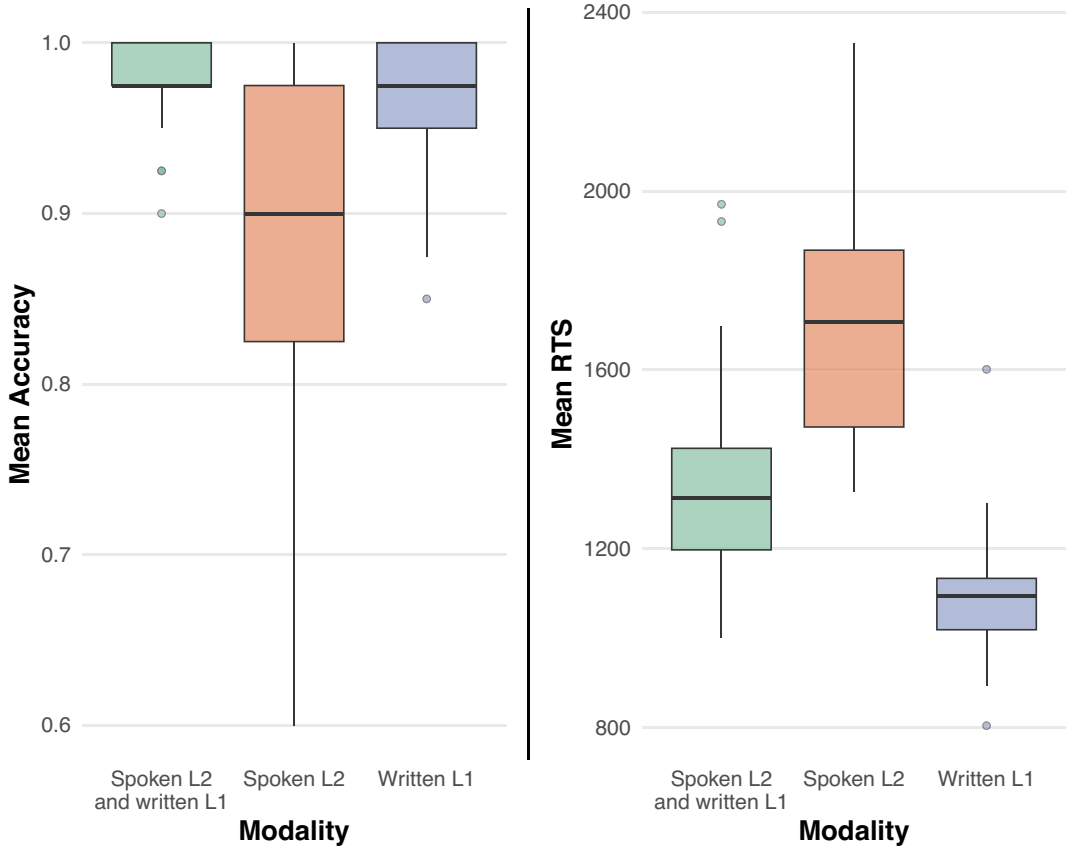


Figure 3. Boxplots showing participants' mean accuracy and RTS by modality in Experiment 2.

https://doi.org/10.1017/S1366728926101266 Published online by Cambridge University Press

Table 2. Parameters of the models for Experiment 2

Dependent variable				
Accuracy	Random factors	Variance	SD	
	Item	2.17	1.47	
	Condition (Written L1) | Item	0.10	0.31	
	Condition (Spoken L2) | Item	2.39	1.55	
	Participant	0.62	0.79	
	Fixed factors	Estimate	SE	z-value
	Condition Written L1 vs. Spoken L2 + Written L1	−0.22	0.46	−0.47
	Condition Spoken L2 vs. Spoken L2 + Written L1	−2.41	0.42	−5.82
RTs	Random factors	Variance	SD	
	Item	3860	62.1	
	Condition (Written L1) | Item	2830	53.2	
	Condition (Spoken L2) | Item	12470	11.2	
	Participant	4651	68.2	
	Fixed factors	Estimate	SE	z-value
	Condition Written L2 vs. Spoken L1 + Written L2	−218.84	8.96	−24.42
	Condition Spoken L1 vs. Spoken L1 + Written L2	334.417	15.77	21.21

Appendix 1.C, stored on OSF. We asked participants to perform LexTALE even if English was not required to complete the task because we wanted to introduce only one difference from Experiment 1, that is, the language of subtitles. Moreover, we wanted to be sure, also in this case, to have participants with a certain level of L2, and to have them in a similar language mode (Grosjean, 2010) as in Experiment 1.

The final sample of participants was composed of 43 bilinguals (L1 = Italian, L2 = English; 20 females, 21 males, 2 not specified) with a mean age of 34 years (SD = 11).

4.1.2. Materials

Same as in Experiment 1, but with subtitles in Italian.

4.1.3. Data analysis

Same as in Experiments 1 and 2. Two participants were removed because of an overall accuracy of less than 80% correct.

4.2. Results

Results of Experiment 3 are shown in Figure 4.

The analysis of accuracy showed no effect of condition ($\chi2(2) = 0.51$, $p < .78$). The analysis of RTs, on the contrary, showed a significant effect of condition ($\chi2(2) = 3084.8$, $p < .0001$). The models' parameters are reported in Table 3: considering accuracy, no significant differences among conditions emerged. As for RTs, the bimodal condition (Spoken L1 and Written L1) was processed faster than the Spoken L1 condition, and slower than the Written L1 condition.

5. Experiment 4: Italian (L1) stimuli with English (L2) subtitles without blurred mouth

This section presents Experiment 4, which replicates Experiment 1, but with stimuli without a blurred mouth.

5.1. Methods

5.1.1. Participants

Participants were recruited through *Prolific* using the same screening as in Experiments 1 and 3. Participants who took part in Experiment 1 or 3 could not participate in Experiment 4. Fifty participants completed the task; however, one participant did the task twice and their data were removed from the database. English competence was tested using the LexTALE, and no participant was removed due to a low performance (threshold of at least 60% $correct_{av}$). Individual LexTALE scores are plotted in Appendix 1. D, stored on OSF.

The final sample of participants was composed of 49 bilinguals (L1 = Italian, L2 = English; 18 females, 31 males) with a mean age of 30 years (SD = 8.7).

5.1.2. Data analysis

Same as in the previous experiments.

5.2. Results

Results of Experiment 4 are reported in Figure 5.

The analyses of accuracy and RTs revealed that the effect of condition was significant ($\chi2(2) = 15.99$, $p = .0003$ and $\chi2(2) = 1300$, $p < .0001$, respectively). The models' parameters are reported in Table 4: accuracy was lower in the Written L2 condition compared to the Spoken L1 + Written L2 condition. On the contrary, the difference between Spoken L1 and Spoken L1+ Written L2 conditions was not significant. Considering RTs, responses in the Spoken L1 condition were slower than in the Spoken L1 + Written L2 condition, whereas those in the Written L2 condition were faster than those in the Spoken L1 + Written L2 condition.

6. General discussion and conclusions

This study presents a series of four experiments we conducted to investigate whether processing bimodal lexical items (spoken and written stimuli) results in a cost in lexical access for unimodal bilinguals, or not. The starting point for this investigation is some recent work investigating bimodal bilinguals, which indicates that for this population, bimodal bilingual lexical access can even be faster than monolingual lexical access. Our general purpose was to shed light on the mechanisms underlying lexical access in general, beyond the bimodal bilingual population.

As Table 5 summarizes, regardless of the status of the language (whether the spoken language is L1 or L2, and the written language L1 or L2) and regardless of the bilingual status of the bimodal condition (i.e., we found the same results when the bimodal stimulus was also bilingual, as in Experiments 1, 2 and 4 and when it was not, as in Experiment 3), answers to bimodal stimuli were faster than speech-only stimuli, and slower than written-only stimuli. This stable result[1] suggests that: i) bimodal stimuli are processed

[1]Considering that spoken language processing is multimodal and mouth movements do have an impact on language comprehension (e.g., Grant & Walden, 1996; McGurk & MacDonald, 1976; Peelle & Sommers, 2015),

https://doi.org/10.1017/S1366728926101266 Published online by Cambridge University Press

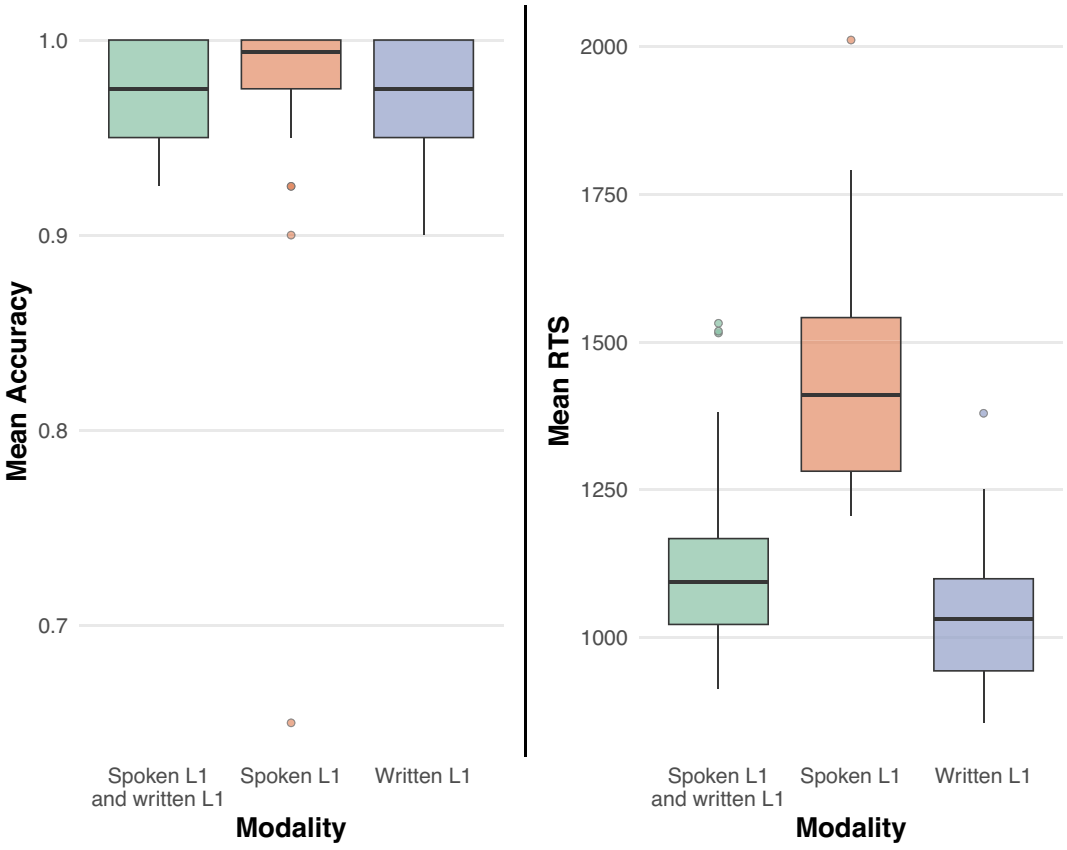


Figure 4. Boxplots showing participants' mean accuracy and RTS by modality in Experiment 3.

Table 3. Parameters of the models for Experiment 3

Dependent variable				
Accuracy	Random factors	Variance	SD	
	Item	1.11	1.05	
	Participant	0.44	0.67	
	Fixed factors	Estimate	SE	z-value
	Condition Written L1 vs. Spoken L1 + Written L1	−0.11	0.21	−0.53
	Condition Spoken L1 vs. Spoken L1 + Written L1	0.03	0.22	0.15
RTs	Random factors	Variance	SD	
	Item	2360	48.58	
	Participant	3646	60.38	
	Fixed factors	Estimate	SE	t-value
	Condition Written L1 vs. Spoken L1 + Written L1	−75.63	4.89	−15.46
	Condition Spoken L1 vs. Spoken L1 + Written L1	311.85	6.50	47.95

Experiment 1 was replicated in Experiment 4, employing the same stimuli, but without the blurred mouth. Even in Experiment 4, we found the same pattern of RTs as in the previous experiments, clearly indicating that the blurred mouth in this context did not cause a behaviourally measurable processing difficulty for the spoken language stimuli.

differently from unimodal stimuli, ii) both modalities are taken into account simultaneously and iii) the integration between modalities occurs automatically, even if focusing on one modality only would have been in principle advantageous.

As already mentioned, the rationale for this investigation was that, in bimodal bilinguals, double lexical access (spoken language + sign language) is never costly and can rather be advantageous compared to single lexical access. At first sight, we cannot say the same about double lexical access (spoken language + written language) in unimodal bilinguals, as the advantage we observe only holds with respect to the speech-only condition. A close comparison between the two sets of results deserves more discussion.

The present results, as well as the ones reported in the bimodal bilingualism literature (Donati et al., 2025; Emmorey et al., 2012; Giustolisi et al., 2024), show that bimodal stimuli are bimodally processed, that is, both modalities are taken into account, and that this bimodal process is automatic, that is, it does not depend on language dominance: it is not the case that it is activated only when it is *a priori* advantageous. However, the bimodality advantage is not observed in the spoken + written stimuli with respect to the written modality. What the present set of studies and those with bimodal bilinguals have in common is the presence of the spoken modality. The visual modality was, however, different, and we believe the reason why results do not converge has to do with this difference in the visual stimuli employed. Crucial points are sequentiality versus simultaneity, timing and strict or partial bimodality.

The first important factor is sequentiality versus simultaneity. Written stimuli are shown in their entirety in a single moment,

https://doi.org/10.1017/S1366728926101266 Published online by Cambridge University Press

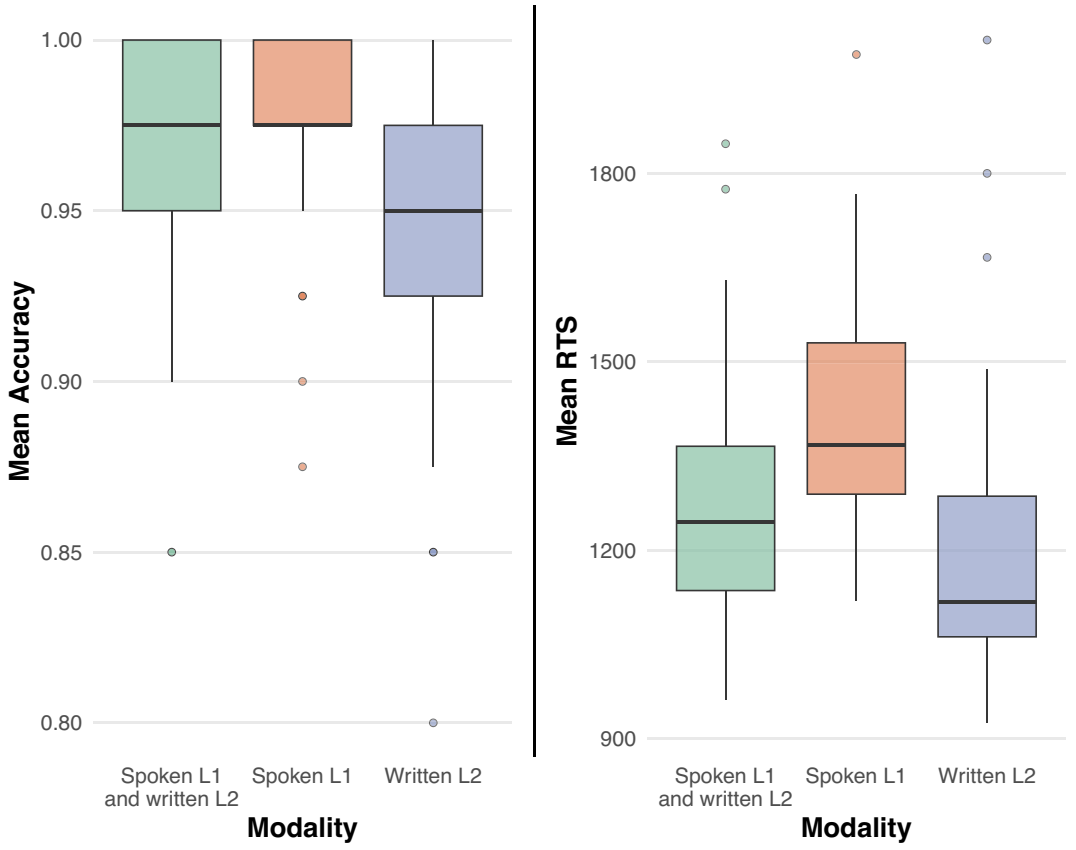


Figure 5. Boxplots showing participants' mean accuracy and RTS by modality in Experiment 4.

Table 4. Parameters of the models for Experiment 4

Dependent variable				
Accuracy	Random factors	Variance	SD	
	Item	1.47	1.21	
	Participant	0.55	0.74	
	Condition Written L2 | Participant	0.86	0.93	
	Condition Spoken L1 | Participant	0.56	0.75	
	Fixed factors	Estimate	SE	z-value
	Condition Written L2 vs. Spoken L1 + Written L2	−0.67	0.26	−2.62
	Condition Spoken L2 vs. Spoken L1 + Written L2	0.45	0.31	1.45
RTs	Random factors	Variance	SD	
	Item	2910	53.95	
	Participant	6243	79.01	
	Fixed factors	Estimate	SE	t-value
	Condition Written L2 vs. Spoken L1 + Written L2	−87.31	5.43	−16.08
	Condition Spoken L1 vs. Spoken L1 + Written L2	149.24	6.25	23.88

whereas signed stimuli unfold over time. More generally, processing speech and processing signs is incremental due to the nature of the input, which, although with a different pace and involving different perceptual systems, is sequential. As for reading, on the other hand, visual words up to a length of 10 letters are known to be processed simultaneously (e.g., Adelman et al., 2010; Radeau et al., 1992). It is thus possible that when it comes to lexical access, the reading modality introduces an advantage that can only be *slowed* by any simultaneous incremental processing. Some results comparing auditory, visual and audio–visual inputs in lexical decision tasks confirm this scenario (Lopez Zunini et al., 2020)[2].

A second important difference between our experiments and the ones assessing code-blending in bimodal bilinguals concerns timing. In the experiments we are discussing, written stimuli were locked with the beginning of speech, that is, items in both modalities shared the same starting point. As for sign language stimuli as well, the beginning of the sign was, in principle, synchronized with the beginning of speech (this was done spontaneously by the native bimodal bilinguals who produced the experimental stimuli). However, it is difficult to define exactly what counts as the beginning of a sign. Conventionally and from a phonological point of view, a sign begins when the hand(s) reach the place of articulation and begin(s) to move. However, some phonetic features of the sign are visible

[2]Experiment 3 results parallel previous work on audiovisual lexical access (Lopez Zunini et al., 2020), even if, importantly, the methodology we employed was different, considering both the task (lexical decision task vs. semantic judgment task) and the complexity of the stimuli (instead of written words and audio files, we had [captioned] video displaying a person speaking).

https://doi.org/10.1017/S1366728926101266 Published online by Cambridge University Press

Table 5. Summary of the experiments and their results

Experiment	1	2	3	4
Participants	Italian L1 – English L2	English L1 – Italian L2	Italian L1 – English L2	Italian L1 – English L2
Materials	Videos with blurred mouth, spoken Italian and written English	Videos with blurred mouth, spoken Italian and written English	Videos with blurred mouth, spoken Italian and written Italian	Videos without blurred mouth, spoken Italian and written English
Reason	To replicate, with unimodal bilinguals, previous work with bimodal bilinguals	To test the effect of language dominance	To test the effect of bilingualism	To test the effect of the blurred mouth
Results	Spoken L1 > Spoken L1 + Written L2 > Written L2	Spoken L2 > Spoken L2 + Written L1 > Written L1	Spoken L1 > Spoken L1 + Written L1 > Written L1	Spoken L1 > Spoken L1 + Written L2 > Written L2

before this moment (for a discussion, see Jantunen, 2015). This is well reflected by the fact that when the transition movement of the hand(s) from a relaxed position to the place of articulation associated with a sign is considered, sign identification often occurs very early and actually well before the moment corresponding to its conventional beginning (see Cecchetto et al., 2016; Emmorey & Corina, 1990). If we consider that bilinguals co-activate their languages (e.g., Thierry & Wu, 2007) and that this occurs also across modality (e.g., Morford et al., 2011; Villameriel et al., 2022), the bimodal bilingual facilitation effect that is observed in lexical access with sign language–spoken language code-blendings could reflect a boost in the integration of the two modalities due to spoken language pre-activation triggered by sign language. Such an effect has no counterpart when spoken language with a written caption is considered, since no written information appears before the beginning of the audio stimulus.

A third important difference between our experiments and the ones assessing code-blending concerns the extent of bimodality[3]. While with bimodal bilinguals the two simultaneous inputs are completely independent (one being acoustic, the other visual), hence strictly bimodal, the visual items in our experiments, being words written in an alphabetic system, activate a phonological representation potentially competing with the one given with speech. The consequence is that in this setting, the bimodality is only partial. There is indeed ample evidence that healthy, skilled silent readers are unable to avoid phonological activation, even in cases in which the use of phonological information complicates the meaning resolution process (e.g., Folk, 1999; Folk & Morris, 1995; Van Orden, 1987).

This effect of the potentially competing phonological activation prompted by written words in the bimodal condition is clearly visible if we compare the strength of our results in the bilingual Experiment 1 (see Figure 2) and in the monolingual Experiment 3 (see Figure 3)[4]: the facilitating effect of bimodality over speech only is bigger in Experiment 3 than Experiment 1, whereas the inhibiting effect of bimodality over written only is reduced in Experiment 3 compared to Experiment 1. Remember that all the items in the bilingual experiments were controlled for their phonological resemblance, and cognates were excluded. The fact that we did find a bimodal advantage over speech, notwithstanding the "interference" of a diverging phonological activation prompted by the written stimulus, makes our result all the more significant. Not surprisingly, the advantage is amplified in the monolingual experiment (Experiment 3), where indeed the phonological activation provides a possible boosting redundancy of the stimulus.

All in all, we showed that bilinguals presented with a semantically redundant (partially) bimodal lexical item are sensitive to both modalities and process them both. This results in an advantage over speech only, which appears to be independent from language dominance, but whose size varies according to whether the setting is bilingual or monolingual.

This result encourages us to pursue exploiting the bimodal setting provided by subtitled videos to run further experiments in a number of directions. One direction is to investigate other languages, including in particular languages that do not have (fully) alphabetical writing systems, and where thus written words might prompt a reduced phonological activation, such as Korean or Chinese.

Some results discussed in Donati et al. (2025), which did not find a consistent bimodal advantage in L2 learners of a sign language (LIS), might suggest that the factor of age of exposure might play a role in determining it. In Donati et al. (2025), the advantage of code-blending with respect to spoken language that was identified as the crucial hallmark of native bimodal bilinguals in previous studies did not emerge in L2 learners of LIS. This difference between native and L2 hearing signers is prima facie interesting and might suggest that the advantage of lexical code-blending is out of reach for some L2 users. However, some limitations in the methodology of the study (shallow measures of frequency of LIS use and no measure of its competence) make it difficult to interpret these negative results beyond speculation.

The results presented here, which suggest that bimodal stimuli are bimodally processed in L2 learners independently from language dominance, appear to point towards a different direction and encourage us to better explore the role of age of acquisition and/or language dominance in the processing of bimodal stimuli in bimodal bilinguals.

A final direction is to replicate our experiments with stimuli that go beyond single words, manipulating the temporal alignment between words and speech (including languages sharing the same word order, as Italian and French, or different word orders, as, say, French and Korean), similarly to what has been done in Giustolisi et al. (2024) with bimodal bilinguals.

More generally, we think subtitled videos might be an important methodology to shed light on simultaneous words and sentences processing in bilinguals' grammars.

Data availability statement. All materials, data and scripts used for the analyses are available on OSF (https://osf.io/rk5py/overview).

[3]We thank an anonymous reviewer for making us think about this important aspect of our study.

[4]In order to do this, we merged together the databases used for analysis in Experiment 1 and Experiment 3 and employed the same analysis as for the single experiments. Details are provided in Appendix 2, stored on OSF.

https://doi.org/10.1017/S1366728926101266 Published online by Cambridge University Press

Funding statement. This work was funded by the French Government and administered by the National Research Agency (ANR) as part of the "Investments for the Future" Programme (grant no. ANR-21-EXES-0002).

Competing interests. The authors declare none.

References

Adelman, J. S., **Marquis, S. J.**, & **Sabatos-DeVito, M. G.** (2010). Letters in words are read simultaneously, not in left-to-right sequence. *Psychological Science*, **21**(12), 1799–1801. https://doi.org/10.1177/0956797610387442

Amenta, S., **Badan, L.**, & **Brysbaert, M.** (2021). LexITA: A quick and reliable assessment tool for Italian L2 receptive vocabulary size. *Applied Linguistics*, **42**(2), 292–314. https://doi.org/10.1093/applin/amaa020

Cecchetto, C., **Giustolisi, B.**, & **Mantovan, L.** (2016). Short-term memory and sign languages: Sign span and its linguistic implications. *Revista de Estudos Linguísticos da Universidade do Porto*, **11**, 59–89.

Crepaldi, D., **Amenta, S.**, **Mandera, P.**, **Keuleers, E.**, & **Brysbaert, M.** (2015). *SUBTLEX-IT. Subtitle-based word frequency estimates for Italian. Poster presentation*. Annual meeting of the italian association for experimental psychology. Rovereto, Italy.

Donati, C., **Geraci, C.**, **Giustolisi, B.**, **Multineddu, G.**, & **Branchini, C.** (2025). How do L2 learners of LIS process code-blending? *RGG*, **47**(15), 1–19. Lingbuzz Press.

Emmorey, K., **Borinstein, H.**, **Thompson, R.**, & **Gollan, T.** (2008). Bimodal bilingualism. *Bilingualism: Language and Cognition*, **11**(1), 43–61. https://doi.org/10.1017/S1366728907003203

Emmorey, K., & **Corina, D.** (1990). Lexical recognition in sign language: Effects of phonetic structure and morphology. *Perceptual and Motor Skills*, **71**(3), 1227–1252. https://doi.org/10.2466/pms.1990.71.3f.1227

Emmorey, K., **Petrich, J. A.**, & **Gollan, T. H.** (2012). Bilingual processing of ASL–English code-blends: The consequences of accessing two lexical representations simultaneously. *Journal of Memory and Language*, **67**(1), 199–210. https://doi.org/10.1016/j.jml.2012.04.005

Finger H., **Goeke, C.**, **Diekamp D.**, **Standvoss, K.**, & **Konig P.** (2017). *LabVanced: A unified JavaScript framework for online studies*. International Conference on Computational Social Science. Cologne, DE.

Folk, J. R. (1999). Phonological codes are used to access the lexicon during silent reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **25**(4), 892–906. https://doi.org/10.1037/0278-7393.25.4.892

Folk, J. R., & **Morris, R. K.** (1995). Multiple lexical codes in reading: Evidence from eye movements, naming time, and oral reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**(6), 1412–1429. https://doi.org/10.1037/0278-7393.21.6.1412

Giustolisi, B., **Jaber, A.**, **Branchini, C.**, **Geraci, C.**, & **Donati, C.** (2024). Processing code-blending beyond the lexical level: Evidence for a double syntactic derivation? *Bilingualism: Language and Cognition*, **27**(5).927–938. https://doi.org/10.1017/S1366728923000901

Grant, K. W., & **Walden, B. E.** (1996). Evaluating the articulation index for auditory–visual consonant recognition. *Journal of the Acoustical Society of America*, **100**(4), 2415–2424. https://doi.org/10.1121/1.417950

Grosjean, F. (2010). *Bilingual: Life and reality*. Harvard University Press. https://doi.org/10.4159/9780674056459

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, **59**(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jantunen, T. (2015). How long is the sign? *Linguistics*, **53**(1), 93–124. https://doi.org/10.1515/ling-2014-0032

Kohnert, K., **Windsor, J.**, & **Miller, R.** (2004). Crossing borders: Recognition of Spanish words by English-speaking children with and without language impairment. *Applied Psycholinguistics* **25**(4), 543–564. doi:10.1017/S0142716404001262

Lemhöfer, K., & **Broersma, M.** (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, **44**(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lo, S. and Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology* **6**, 1171. doi: 10.3389/fpsyg.2015.01171

Lopez Zunini, R. A., **Baart, M.**, **Samuel, A. G.**, & **Armstrong, B. C.** (2020). Lexical access versus lexical decision processes for auditory, visual, and audiovisual items: Insights from behavioral and neural measures. *Neuropsychologia*, **137**, 107305. https://doi.org/10.1016/j.neuropsychologia.2019.107305

McGurk, H., & **MacDonald, J.** (1976). Hearing lips and seeing voices. *Nature*, **264**(5588), 746–748. https://doi.org/10.1038/264746a0

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics*, **40**, 331.

Morford, J. P., **Wilkinson, E.**, **Villwock, A.**, **Piñar, P.**, & **Kroll, J. F.** (2011). When deaf signers read English: Do written words activate their sign translations? *Cognition*, **118**(2), 286–292. https://doi.org/10.1016/j.cognition.2010.11.006

Peelle, J. E., & **Sommers, M. S.** (2015). Prediction and constraint in audiovisual speech perception. *Cortex*, **68**, 169–181. https://doi.org/10.1016/j.cortex.2015.03.006

Radeau, M., **Morais, J.**, **Mousty, P.**, & **Bertelson, P.** (1992). A listener's investigation of printed word processing. *Journal of Experimental Psychology: Human Perception and Performance*, **18**(3), 861–871. https://doi.org/10.1037/0096-1523.18.3.861

Thierry, G., & **Wu, Y. J.** (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, **15**(3), 181–198. https://doi.org/10.3758/BF03197716

van Heuven, W. J., **Mandera, P.**, **Keuleers, E.**, **& Brysbaert, M.** (2014). SUBTLEXUK: a new and improved word frequency database for British English. *Q J Exp Psychol (Hove)*. **67**(6), 1176–90. doi: 10.1080/17470218.2013.850521

Villameriel, S., **Costello, B.**, **Giezen, M.**, & **Carreiras, M.** (2022). Cross-modal and cross-language activation in bilinguals reveals lexical competition even when words or signs are unheard or unseen. *Proceedings of the National Academy of Sciences of the United States of America*, **119**(36), e2203906119. https://doi.org/10.1073/pnas.2203906119