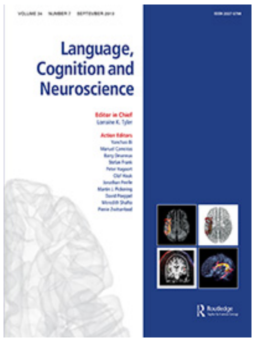

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Prediction and integration of semantics during L2 and L1 listening

Aster Dijkgraaf, Robert J. Hartsuiker & Wouter Duyck

LANGUAGE, COGNITION AND NEUROSCIENCE
2019, VOL. 34, NO. 7, 881–900
https://doi.org/10.1080/23273798.2019.1591469

REGULAR ARTICLE

Prediction and integration of semantics during L2 and L1 listening

Aster Dijkgraaf, Robert J. Hartsuiker and Wouter Duyck

Department of Experimental Psychology, Ghent University, Gent, Belgium

ABSTRACT
Using the visual world paradigm, we tested whether Dutch-English bilinguals predict upcoming semantic information in auditory sentence comprehension to the same extent in their native (L1) and second language (L2). Participants listened to sentences in L1 and L2 while their eye-movements were measured. A display containing a picture of either a target word or a semantic competitor, and three unrelated objects was shown before the onset of the auditory target word in the sentence. There were more fixations on the target and competitor pictures relative to the unrelated pictures in both languages, before hearing the target word could affect fixations. Also, semantically stronger related competitors attracted more fixations. This relatedness effect was stronger, and it started earlier in the L1 than in the L2. These results suggest that bilinguals predict semantics in the L2, but the spread of semantic activation during prediction is slower and weaker than in the L1.

ARTICLE HISTORY
Received 7 March 2018
Accepted 25 February 2019

KEYWORDS
Visual world paradigm; prediction; comprehension; bilingualism; semantic competitor

Smooth and efficient language comprehension involves prediction of upcoming information. Context information affects the language comprehension system before new bottom-up input is encountered, and this may involve pre-activation of linguistic information (see Kuperberg & Jaeger, 2016 for a recent review; but also see Nieuwland et al., 2018 for a multilab failure to replicate pre-activation of phonology). Linguistic predictions are made on the basis of cues from the linguistic (e.g. Altmann & Kamide, 1999; Otten, Nieuwland, & Van Berkum, 2007) and non-linguistic context information (Chambers, Tanenhaus, & Magnuson, 2004; Salverda, Brown, & Tanenhaus, 2011). The content of predictions also varies greatly. Predictions can consist of semantic properties of upcoming words (including object shape) (e.g. Altmann & Kamide, 1999, 2007; Rommers, Meyer, Praamstra, & Huettig, 2013), syntactic information (e.g. Arai & Keller, 2013), and possibly word form information (e.g. Dikker, Rabagliati, Farmer, & Pylkkänen, 2010; Ito, Pickering, & Corley, 2018). Predictive language processing is not an all-or-nothing phenomenon but rather something that occurs in a graded manner (Kuperberg & Jaeger, 2016). Several word candidates for prediction are activated in parallel, depending on how likely they are given the context. Here, we tested whether prediction of target word semantics by bilinguals, and spreading semantic activation to competitors with varying degrees of semantic associatedness, is equally strong in both of their languages.

How much or how strongly a person predicts seems to be affected by processing speed (Huettig & Janse, 2016), language experience (Foucart, 2015; Kaan, 2014; Kuperberg & Jaeger, 2016; Peters, Grüter, & Borovsky, 2015; Phillips & Ehrenhofer, 2015), and the availability of cognitive resources. Each of these factors is likely to differ between native language (L1) and second language (L2) processing, and can therefore potentially affect predictive language processing in each language differently. For example, increased lexical competition due to cross-lingual word coactivation affects speed of lexical access in bilinguals (Duyck, Assche, Drieghe, & Hartsuiker, 2007; Lagrou, Hartsuiker, & Duyck, 2013), particularly in the L2 (Weber & Broersma, 2012). Bilingual language users usually have much less experience using their L2 than their L1. This may result in weaker links between word forms and semantics (Gollan et al., 2011; Gollan, Montoya, Cera, & Sandoval, 2008) and this may in turn again result in slower or weaker retrieval of linguistic representations. Less use may also result in lower quality of linguistic representations and different frequency biases for prediction, because a particular continuation for a prior context may have been encountered less often (Kaan, 2014). Furthermore, prior knowledge and new input may be considered less reliable in a less familiar L2, and this may affect the degree of predictive processing (Kuperberg & Jaeger, 2016). Finally, L2 processing may tax working memory more than L1 processing (Francis & Gutiérrez, 2012; McDonald, 2006). Therefore,

CONTACT Aster Dijkgraaf aster.dijkgraaf@ugent.be
Supplemental data for this article can be accessed at https://doi.org/10.1080/23273798.2019.1591469

if working memory resources are required for predictive processing (e.g. Huettig & Janse, 2016), then prediction may be less efficient in L2 than in L1. In sum, less efficient retrieval of representations in L2 processing may hinder the construction of higher-level meaning (such as sentence meaning) used for generating a prediction. In addition, the L2 representation of the target for prediction itself may be retrieved less efficiently, leading to slower, weaker, and/or less accurate predictions.

In a recent theoretical account of predictive processing, Pickering and Gambi (2018) postulate two routes for prediction. The first one is based on spreading activation between associated representations. This "prediction-by-association" route is relatively automatic and not targeted. This entails that it should be mostly intact in populations with limited resources, such as L2 comprehenders. The second route to prediction uses covert imitation of the input, constructs a representation of speaker intention, and engages the production system to generate a targeted prediction (see Dell & Chang, 2014; Huettig, 2015; Pickering & Garrod, 2013, for other accounts assuming involvement of production). The authors hypothesise that this "prediction-by-production" route is optional and that its use depends on the availability of sufficient time and cognitive resources. Therefore, prediction-by-production is likely used less or fails more often in cognitively more demanding contexts, such as L2 comprehension.

Differences between prediction in L1 and L2 comprehension have been found when a language-specific morpho-syntactic or phonotactic rule needs to be applied quickly and accurately in order to pre-activate a target for prediction or when the target for prediction is a word form (Hopp, 2013, 2015; Ito, Pickering, & Corley, 2018; Martin et al., 2013; Mitsugi & Macwhinney, 2016). For example, in Martin et al.'s (2013) ERP study, native speakers of English and late Spanish-English bilinguals read English sentences with a predictable or unpredictable sentence ending (e.g. *Since it is raining, it is better to go out with an umbrella* [*EXPECTED*]/ *a raincoat* [*UNEXPECTED*]). The article preceding the sentence-final noun was always congruent with the final noun, but not always congruent with the expected noun. Martin et al. found an N400-effect on the processing of incongruent versus congruent articles for L1 readers, but not for L2 readers. The sentence-final noun elicited an N400-effect as well, in both groups, but the effect was larger for L1 than for L2 readers. Thus, the N400 elicited by the article showed that bilinguals reading in the L2 did not anticipate upcoming word forms like native readers did, but the noun-elicited N400 might indicate that target word integration was easier in both languages when the target word was predictable. Alternatively, the effect could be attributed to slower prediction in the L2. The two interpretations cannot be dissociated because the effect was not found before target word onset.

Ito, Pickering, and Corley (2018) studied prediction of word form using a visual world paradigm. Japanese-English bilinguals and native English controls listened to constraining sentences such as "*The tourists expected rain when the sun went behind the …* ". Visual displays contained a predictable target object (*cloud*; in Japanese: *Kumo*), a phonological competitor of the target object in English (*clown*), a phonological competitor of the target object in Japanese (bear; *kuma*), or an unrelated object (*globe*; *tikyuugi*). The bilinguals predictively looked at target objects, but slower than native listeners. They did not look more at English or Japanese phonological competitors than at unrelated objects. This finding suggests that the bilinguals predicted target word semantics when listening in their L2, but not word form. Native listeners fixated both target objects and English phonological competitors more than unrelated objects before hearing the target could affect fixations.

Hopp (2015) contrasted prediction based on morpho-syntactic cues and lexico-semantic cues. In a visual world paradigm study, Native German listeners and English-German bilinguals looked at picture displays including three possible actors and a control object while they listened to SVO (e.g. *The$_{NOM}$ wolf kills soon the$_{ACC}$ deer*) or OVS (e.g. *The$_{ACC}$ wolf kills soon the$_{NOM}$ hunter*) sentences in German. The native listeners looked at expected patients (*the deer)* before the onset of the second noun phrase in SVO sentences and at expected agents (*the hunter*) in OVS sentences. The bilinguals were more likely to look at patient objects before the onset of the second noun phrase, irrespective of first noun phrase case marking (nominative or accusative). Thus, even though Hopp found evidence for prediction based on lexical-semantic cues (verb information) in the L2, no prediction based on morpho-syntactic (case marking) information was found in the L2. Participants' knowledge of the German case marking system was not assessed separately, but German proficiency of the bilingual participants did not affect the pattern of results. Similarly, Mitsugi and Macwhinney (2016) found that English-Japanese bilinguals were unable to use case marking information as a cue for prediction in Japanese, even though the bilinguals had good offline knowledge of the Japanese case-marking system.

The findings of Ito, Pickering, and Corley (2018) and Hopp (2015) suggest that semantic prediction is relatively intact in L2 comprehension. Indeed, when no application of a language-specific (morpho-)syntactic rule is required for prediction (Dijkgraaf, Hartsuiker, & Duyck, 2017; Hopp, 2015; Ito, Corley, & Pickering (2018)), or when the same rule exists in the participants' L1 (Foucart, Martin, Moreno, & Costa, 2014; Foucart, Ruiz-Tada, & Costa, 2015; Van Bergen & Flecken, 2017), L2 listeners often do show prediction effects, like in L1. Dijkgraaf et al. (2017), for example, compared prediction between the L1 and the L2 of the same participants using an eye-tracking paradigm based on Altmann and Kamide (1999). Participants listened to simple SVO sentences with either a constraining (e.g. *Mary knits a scarf*) or a neutral verb (e.g. *Mary loses a scarf*). The visual display showed four objects that could all be lost, but only one that could be knitted (a scarf). Dutch-English participants listening to sentences in Dutch or English were more likely to fixate on the target object in the constraining condition than in the neutral condition, before exposure to the auditory target word could influence fixations. The bias in target fixations did not differ between the L1 and L2. Likewise, using a between-subject comparison, Ito, Corley, and Pickering (2018) found that bilinguals listening to constraining and neutral sentences in their L2 (English; various L1 languages) showed similar predictive looking behaviour as L1 listeners. Adding a cognitive load during the listening task (remembering 5 words) affected prediction, but in a similar way for L1 and L2 listeners. These findings indicate that at least under some circumstances, L2 listeners predict upcoming semantic information.

However, as Pickering and Gambi note, spreading activation in semantic prediction depends on the number and strength of links between representations (2018), which is in turn shaped by (linguistic) experience, and could therefore differ between L2 and L1. Different theories of bilingual lexicosemantic memory indeed assume that the mapping of words onto semantic memory is different in the L2 than in the L1. Specifically, L1 words may be semantically richer than L2 words (Finkbeiner, Forster, Nicol, & Nakamura, 2004; Schoonbaert, Duyck, Brysbaert, & Hartsuiker, 2009). Schoonbaert et al. based their model on the distributed feature model (Van Hell & De Groot, 1998) and suggested that L2 words have fewer semantic features than L1 words. Therefore, two words that share features in the L1 may have no, or fewer, shared features in the L2. Thus, even though bilinguals are able to make semantic predictions based on lexical-semantic information from the sentence context in the L2, perhaps they do not do so as strongly and quickly as monolinguals do. This should be the case especially when the semantic associations between the sentence content and the predicted information is weaker, or when remote spreading of activation to concepts semantically associated with the predicted concept is tested. Also, the strength of the links between word forms and semantics may be weaker in L2 than in L1 (Gollan et al., 2008, 2011), which may similarly affect strength and speed of semantic pre-activation.

In line with this hypothesis, Japanese-English bilinguals listening to predictable sentences anticipated a predictable target object later than English native speakers (e.g. *cloud*, when listening to *The tourists expected rain when the sun went behind the …*) (Ito, Pickering, & Corley, 2018). Also, using ERPs, Ito, Martin, and Nieuwland (2017) found no evidence of pre-activation of a semantic competitor of the predictable target word in non-native speakers, whereas such an effect was found in native speakers (Ito, Corley, Pickering, Martin, & Nieuwland, 2016). Similarly, Foucart, Moreno, Martin, and Costa (2015) found that value-inconsistent statements as compared to value-consistent statements (e.g. *Nowadays, paedophilia should be prohibited/tolerated across the world*) triggered an N400 response in native speakers but not in non-native speakers. One possible interpretation of this finding is that the valence of a concept is not retrieved from the word as efficiently in the L2 as in the L1, and that therefore, the L2 speakers did not generate predictions based on concept valence.

Peters et al. (2015) showed that highly proficient bilinguals pre-activated target word semantics faster than low proficient bilinguals. For instance, they fixated pictures of a *ship* faster when listening to the sentence *The pirate chases the ship*. In contrast, low-proficient bilinguals were more likely to fixate competitors that were locally related to the action verb, but not necessarily consistent with the sentence meaning (e.g. looking at a *cat* after hearing the verb *chases* in the above sentence). Finally, Kohlstedt and Mani (2018) presented discourse information in a visual world paradigm. When presenting two sentences in which the first contained a semantically associated or a neutral prime for a target in the second, predictive fixations were found in L1 listeners, but not in L2 listeners when analysing each group seperately. However, in an overall analysis the effect of context (biasing or neutral) on target fixations did not differ significantly between groups (bilinguals in L2 vs. native speakers).

In sum, bilinguals can predict upcoming information during L2 processing in some circumstances, but they do not always do so to a similar extent as native speakers when application of a language-specific morpho-syntactic or phonotactic rule is required. In addition, even though some research suggests that lexical-semantic prediction is intact in bilinguals, there is also evidence suggesting that lexical-semantic prediction is weaker or later in bilinguals comprehending L2 input. We hypothesise that even though lexical-semantic prediction can occur in L2 comprehension, the inconsistent findings above may be due to differences in spreading semantic activation and/or temporal dynamics between L1 and L2, with differences especially arising in more challenging contexts. Here, we will investigate when and how prediction in L2 differs from L1, using targets that vary in predictability, and how spreading semantic activation evolves differently when listening in different languages. More specifically, we expect pre-activation of semantic competitors of expected words to be weaker and/or slower in the L2 than in the L1, especially when the semantic distance between expected words and semantic competitors is larger. That is, we expect prediction to be semantically narrower in the L2. If L2 words are indeed mapped onto fewer semantic features than L1 words (Schoonbaert et al., 2009), they also activate fewer features shared with semantically associated concepts, which should trigger less activation spreading to those concepts in L2.

The present study

In the present experiment, we used the visual world paradigm to test whether prediction of semantic information during auditory speech recognition, based on lexical-semantic information from the sentence context, is weaker and/or slower in the L2 than in the L1. Dutch-English bilinguals listened to sentences in Dutch and in English while they looked at four-picture displays on a screen in front of them. The picture display included three items that were unrelated to the target word and an experimental image: either a depiction of the target word or of a semantically related competitor. The semantic distance between the target word and the semantic competitor varied. This way, we were able to test in a more refined way whether prediction in the L1 vs. the L2 leads to a different degree of spreading semantic activation. If this were the case, one would expect a different effect of semantic distance between targets and competitors in each language. Ito, Corley, and Pickering (2018) also included a semantic competitor in a visual world paradigm experiment in which they compared prediction in the L1 and L2. However, no pre-activation of the semantic competitor was found in either the L1 or the L2. The absence of an effect of pre-activation may have been caused by the fact that the picture displays in that study included both a target object and a semantic competitor, so that the target object attracted looks so strongly that it prevented any looks to the competitor object (Huettig & Altmann, 2005; Huettig, Rommers, & Meyer, 2011). As a more sensitive measure of competitor activation, we therefore opted for a design in which either the target object or the semantic competitor object was present in the display.

Many studies on predictive language processing in the L2 focused on prediction during sentence reading (Foucart et al., 2014; Ito et al., 2017; Martin et al., 2013; Molinaro, Giannelli, Caffarra, & Martin, 2017). However, predictive processing may be particularly challenging for non-native speakers in the auditory modality. Speech unfolds over time and therefore a listener cannot go back to the beginning of a sentence like in reading, where the information remains available. Also, misperceptions and misrepresentations of non-native phonemes, a problem that doesn't exist for bilingual reading in the same alphabet, may increase lexical competition during listening comprehension (Weber & Broersma, 2012). Like Dijkgraaf et al. (2017), Foucart, Ruiz-Tada, and Costa, 2016, Ito, Corley, and Pickering (2018) and Hopp (2015), the current experiment therefore studied predictive processing in the auditory modality.

It is important to note that a comparison of L1 and L2 listening leaves two options: the first is that native listeners are compared with other subjects that listen in the same language, which is however their L2 (e.g. Ito, Corley, & Pickering, 2018). Even when participant groups are matched on a number of variables such as age, education level and socio-economic status, they may have very different cultural, educational, and linguistic backgrounds. Thus, any differences found between groups may be due to such variables, rather than the experimental factor Language.

The other option is to compare listening in different languages, within the same subjects. Here, we compared listening between L1 and L2 within the exact same Dutch-English bilingual participants. This way, we eliminated confounding effects of individual cognitive differences that may affect prediction such as working memory capacity, processing speed (Huettig & Janse, 2016), age (Federmeier & Kutas, 2005), and verbal fluency (Rommers, Meyer, & Huettig, 2015). This also eliminated the high inter-individual variability that characterises eye movements (Bargary et al., 2017; Rayner, 1998) and which may confound between-group differences in visual world paradigms. To account for differences between the two languages used in this within-subject design, we included linguistic factors of stimuli such as sentence length, phoneme count, word frequency, and semantic distance measures in our analyses.

Method

Participants

Bilinguals. 50 native speakers of Dutch took part in the experiment (11 men and 39 women, mean age 19 years, $SD = 2.85$). They were Ghent University students participating for course credit. Dutch was the participants' dominant and most proficient language, and English was their second (49 participants) or third (1 participant) language. On average, participants started acquiring English at age 11 (*SD* = 2.46), mainly in school, on holiday or through (online) media. None of the participants had spent time living in an English-dominant country. The participants reported to be exposed to Dutch an average of 73% of the time, and to English 22% of the time. Forty-seven participants also had knowledge of French, and 24 participants had knowledge of German. Nine participants had knowledge of Spanish, two knew Arabic, one Portuguese, and one Italian (all late learners). Language proficiency in English and Dutch was assessed with the LexTALE vocabulary knowledge test (Lemhöfer & Broersma, 2012) and with self-ratings. The LexTALE is a 60-item lexical decision task (unspeeded). It indicates word knowledge and general language proficiency (Lemhöfer & Broersma, 2012). The bilinguals' mean LexTALE scores and self-ratings are reported in Table 1. The participants were significantly less proficient in their L2 than in their L1.

Materials and design

Three hundred sixty-two trials were included in the experiment. On each trial, participants listened to a sentence and saw a four-item picture display. Fifty further participants filled out a cloze probability test for an initial set of 871 candidate sentences,[1] with the dual purposes of (a) sentence selection and (b) measuring predictability of sentence-final (target) words. The candidate sentences were constructed so that word order was as similar as possible in Dutch and English. Sentences were excluded from the final sentence set if the Dutch and English target provided by the participants were not translation equivalents, and if the provided target word was not depictable or a picture of the word was not included in the normed picture set that we used (Severens, Lommel, Ratinckx, & Hartsuiker, 2005). Also, only one pair of sentences (translation equivalents in Dutch and English) was selected for each target picture. All English sentences were checked for grammaticality by a native speaker of American English. Like the participants in the main experiment, the participants were Ghent University students with knowledge of Dutch (L1) and English (L2). Half of the participants filled out the cloze test for the sentences in Dutch and the other half of the participants filled out the test in English. In the cloze test, participants read each sentence without the sentence-final word and were asked to complete each sentence with the first word that came to mind. For each sentence, the highest cloze probability target was selected in English and in Dutch. The final sentences had varying cloze probabilities (see Figure 1 panel A). The mean cloze probability was .71 (*SD* = .23) in Dutch (L1) and .68 (*SD* = .24) in English (L2).

Table 1. Participants' Mean (SD) L1 and L2 LexTALE Scores and Self-ratings.

	L1 Dutch	L2 English	*p*-value[c]
Lextale[a]	88.72 (7.25)	70.05 (10.59)	<0.001
Rating listening[b]	4.98 (.14)	4.00 (.54)	<0.001
Rating speaking[b]	4.94 (.32)	3.36 (.60)	<0.001
Rating reading[b]	4.94 (.24)	3.78 (.55)	<0.001
Rating general[b] proficiency	4.94 (.24)	3.64 (.55)	<0.001
Category fluency	23.46 (5.23)	14.19 (3.96)	<0.001

[a]Scores consist of percentage correct, corrected for unequal proportion of words and nonwords (Lemhöfer & Broersma, 2012). Due to technical problems one participant's score is missing.
[b]Means of self-assessed ratings on a scale of 1–5 (1 = not at all, 5 = perfect/ mother tongue) for listening, speaking, reading and general proficiency.
[c]Reported *p*-values indicate significance levels of dependent samples t-tests between scores for Dutch and English in bilinguals. Df of t-test on LexTALE scores = 48, Df of t-test on Category Fluency = 47 (due to technical problems one participant's LexTALE score and two participants' Fluency scores are missing). Df of all t-tests on ratings = 49.

Figure 1 panels B and C show the frequency and phoneme count information of the Dutch and English final set of target words (Baayen, Piepenbrock, & Gulikers, 1995; Keuleers, Brysbaert, & New, 2010; Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). The translation equivalents of the words were mostly phonologically dissimilar in English and Dutch (normalised phonological Levenshtein distance ≤.50, *M* = .25, *SD* = .25),[2] but cognates were also included (e.g. L2-L1: *tent-tent, wheel-wiel, nest-nest*), because Dutch and English are related languages and excluding all cognates would lead to unrepresentative word selections. As phonological similarity between the target word and its translation equivalent may affect looking behaviour, target Levenshtein distance was included as a factor in the analyses and we also confirmed that the data excluding all cognates yielded a similar pattern of results.[3] Levenshtein distance between the unrelated picture names and translation equivalents, and between the (auditory) words in the sentences and translation equivalents of each trial may also affect looking behaviour. Given the many English-Dutch cognates and restrictions that had to be taken into account during item construction, we were unable to control for this factor. However, to account for differences in looking behaviour for each item, a random intercept of item was added to the linear mixed models in our analyses.

The pictures in the displays accompanying the sentences were line drawings from the normed database by Severens et al. (2005). Each display accompanying a sentence consisted of either a target picture (the last word in the sentence) or a semantic competitor (a word semantically related to the target word), and

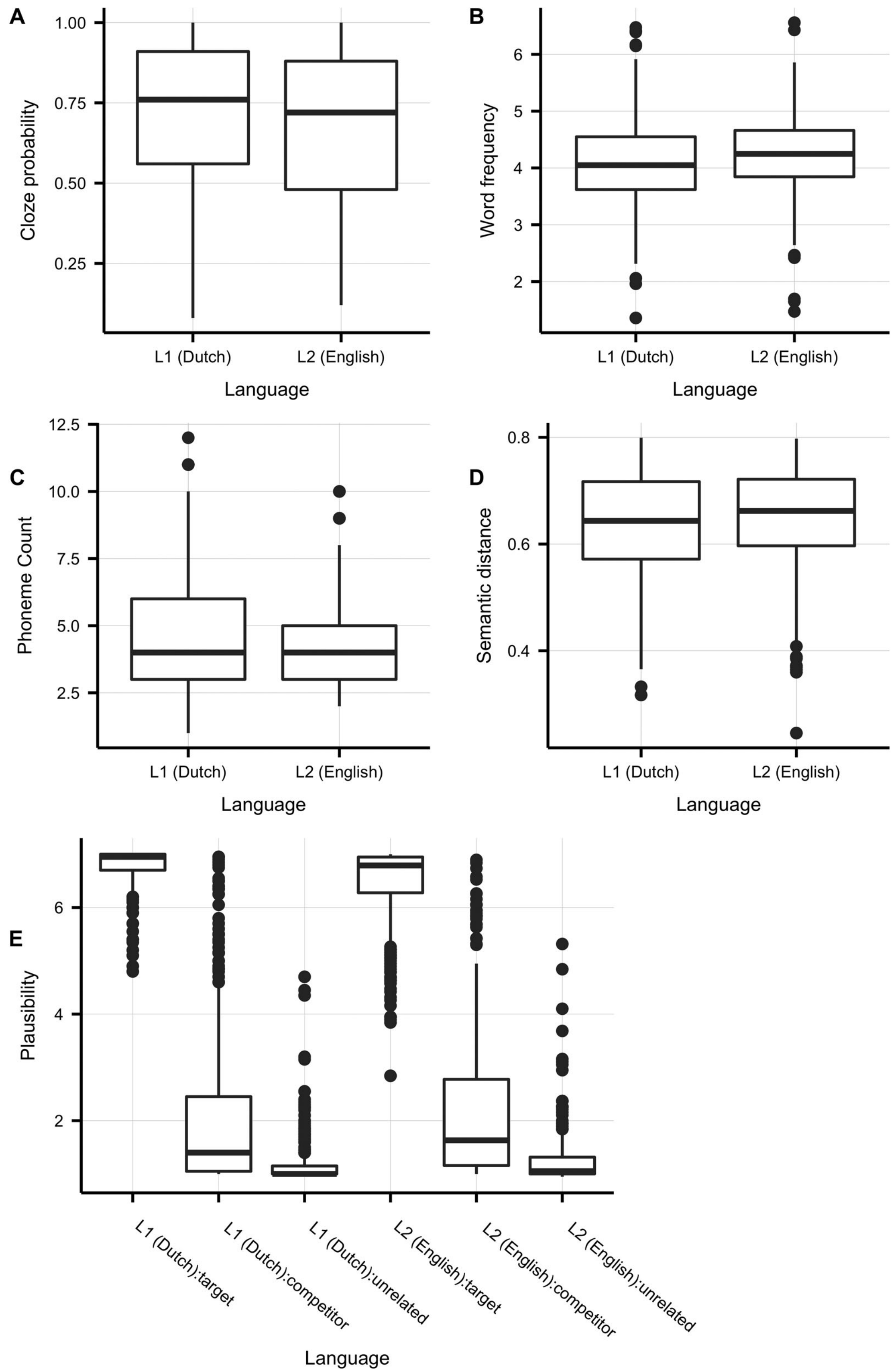

Figure 1. Stimulus information. (A) Stimulus sentence cloze probability. (B) Target word frequency. Zipf value (log10(frequency per million*1000)) retrieved from the SUBTLEX-UK and SUBTLEX-NL databases (Keuleers et al., 2010; Van Heuven et al., 2014). Please note that for six compound nouns no frequency score was available for English. (C) Target word phoneme count retrieved from CELEX database (Baayen et al., 1995). (D) Semantic distance target-competitor pairs extracted from SNAUT (Mandera et al., 2017). (E) Plausibility ratings of target, competitors, and unrelated words as sentence endings. Ratings were given on a 7 point scale ranging from “not likely at all as sentence ending” to “very likely as sentence ending”.

three pictures unrelated to the target word. Whether a sentence was accompanied by a target or competitor image was counterbalanced across participants. To ensure that target pictures did not inherently draw more overt visual attention than competitors or unrelated pictures, each of the 362 target pictures was included as a competitor picture for another sentence and as unrelated picture in three other sentences. The 362 experimental sentences thus belonged to 181 sentence pairs. For each sentence pair, the target of one sentence was the competitor of the other and vice versa.[4] The display of an experimental trial never included the same picture more than once.

The competitor picture for each target word was selected based on semantic distance scores extracted from the SNAUT database (Mandera, Keuleers, & Brysbaert, 2017).[5] The distance score is based on word co-occurrences in large text corpora.[6] The smaller the semantic distance score for a word pair, the more related they are. The score varies between 0 and 1. We included a large range of distance scores for the semantic competitors (see Figure 1 panel D), but the distance score for target-competitor pairs was always smaller than .8. The target-unrelated pairs always had a distance score of more than .8. This cut-off point was chosen because we required a large range of semantic distance scores, and because it was the lowest cut-off point for which it was still possible to pair each target word with the same competitor word in Dutch and in English. Mean semantic distance score was .63 in Dutch ($SD = .11$) and .64 in English ($SD = .10$). Mean cloze probability for the competitors was $M = .01$, $SD = .03$ in the L1 and $M = .01$, $SD = .03$ in the L2. The competitor word never occurred in the accompanying sentence.[7] Target and competitor words never started with the same phoneme (except for one pair in Dutch, *orange-lemon*, *sinaasappel-citroen*). There were the target trials where a target and three unrelated pictures were presented, and there were competitor trials where a competitor was presented instead of the target, leading to five possible picture "positions" (target, competitor, unrelated 1, unrelated 2, unrelated 3). As the picture set was limited and each picture had to be used once in every position, it was not possible to take phonetic overlap between unrelated and experimental pictures into account when constructing the picture sets.

Plausibility ratings were generated by 40 further unbalanced Dutch-English bilingual participants (20 in English and 20 in Dutch) for each sentence ending with a target word, a competitor word, and with an unrelated word (L1 target: $M = 6.8$, $SD = .33$, L1 competitor: $M = 2.08$, $SD = 1.51$, L1 unrelated: $M = 1.19$, $SD = .48$, L2 target: $M = 6.46$, $SD = .75$, L2 competitor: $M = 2.20$, $SD = 1.40$, L2 unrelated: $M = 1.25$, $SD = .49$ on a 7 point scale ranging from "not likely at all as sentence ending" to "very likely as sentence ending", see Figure 1, panel E).[8] The participants were recruited from the same Ghent University participant pool, but none of them participated in the cloze probability test nor in the actual experiment. Plausibility was measured after targets were paired with competitors and did not play a role in competitor selection. Competitor plausibility was taken into account in the analyses. Figure 2 shows an example stimulus set, and Appendix A contains the sentences and object names of the target and competitor pictures for each stimulus set.

Every twelve experimental sentences were followed by a visually presented simple yes/no question about the preceding sentence to ensure the participants would continue to pay attention to the sentences. To ensure that there were no carry-over effects from answering the question in the data for analysis and to ensure that not every trial would have a target or competitor in the display, we added a filler sentence after each question. The four pictures are shown on a filler trial never included a picture of the target word of the accompanying sentence. Unlike the experimental sentences, the filler trials were presented to each participant in Dutch (mean cloze probability = .64) and in English (mean cloze probability = .57). There was no significant difference between the cloze probabilities of the Dutch and English fillers ($t(11) = 1.08$, $p = .30$). The sentences were selected from the same initial candidate sentences as the experimental sentences. The pictures used for the filler trials were not used for the experimental trials.

Recordings. The sentences for the experiment were recorded in a sound attenuating room. A Dutch-English bilingual (female, 21 years old) from Flanders who had lived in England from age five to twelve recorded the sentences. The participants in the experiment rated her accent in English as 3.6 and her accent in Dutch as 4.6 on a scale from 1 (very foreign accent) to 5 (native accent). The speaker was asked to pronounce the sentences clearly at a relaxed but natural rate. Each sentence was recorded three times (sampling frequency 48 kHz); the recording that we judged to have the clearest pronunciation and most neutral prosody was selected for the experiment. The average speech rate was 220 words per minute.

The target word onset in each sentence was marked using Praat (Broersma & Weenink, 2014). The average target length was 507 ms (range 224–942 ms) in English and 511 (240–1168 ms) in Dutch. The mean length of the sentence leading up to the target word was 1977ms in English (range 708–4557 ms) and 2164 ms in Dutch (range 764–4764 ms). Sentence length up to the target was included as factor in the analyses.

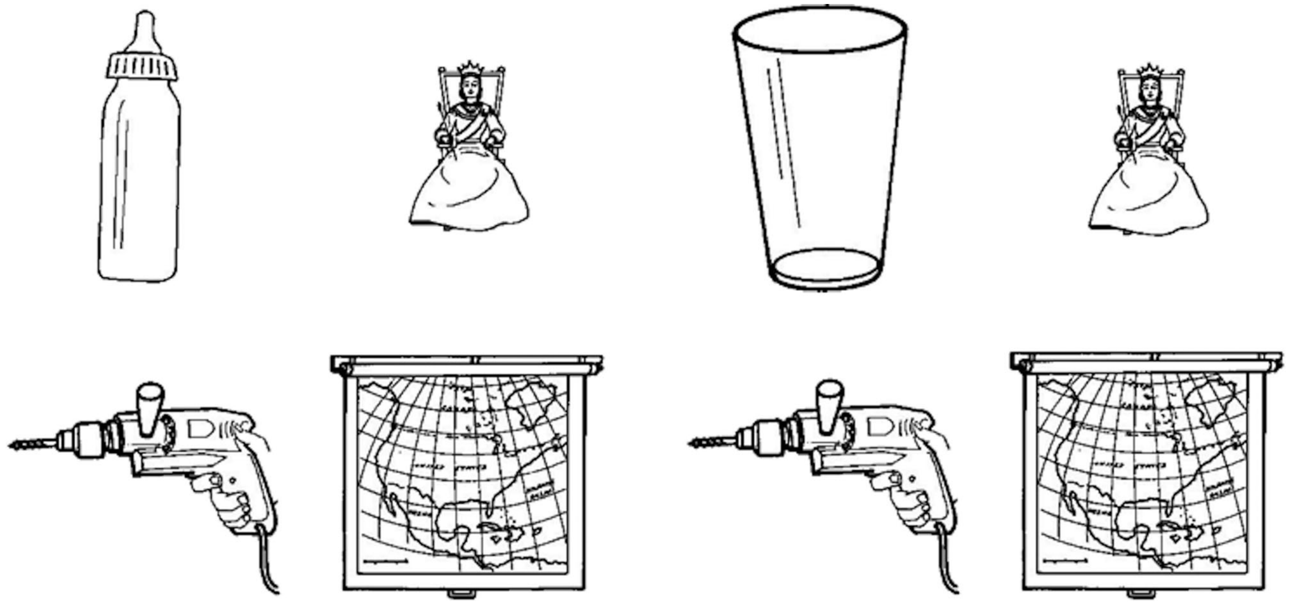

Figure 2. Example stimulus displays. Each participant was presented with one of these two displays with the sentence "*Her baby doesn't like drinking from a bottle*". The left display includes a picture of the target word for prediction (*bottle*) and the right display includes a picture of a semantic competitor (*glass*). Each display also included 3 unrelated images.

Procedure

Participants followed written and oral instructions to listen carefully to Dutch and English sentences and to look at pictures on the screen. They were instructed to look wherever they wanted as long as their gaze did not leave the screen (Huettig & Altmann, 2005; McQueen & Huettig, 2012). In addition, participants were asked to answer the occasional yes/no question about a preceding sentence by pressing "j" for yes and "f" for no. The questions were included to ensure participants continued to listen to the sentences attentively. Participants were presented with 24 questions throughout the experiment (twelve in Dutch and twelve in English). Eye movements were recorded from the right eye with an Eyelink 1000 eye-tracker (SR Research) (1000 Hz) in tower mount.

A fixation cross appeared on the screen for 500 ms, followed by the presentation of a sentence over headphones. Following the procedure in Rommers et al. (2013), the four pictures were presented only 500 ms before the onset of the target word in the sentence. This was done to avoid visual priming of the target or competitor word semantics by the target or competitor picture. Picture location was randomised. After sentence offset, the pictures remained on the screen for 1000 ms. After drift check the next trial started.

The sentence pairs (where one sentence's target was the other sentence's competitor and vice versa) were split into two lists (list A and list B). Each sentence could be presented in Dutch and in English with either a target or a competitor picture. The participants were presented with one block in English and one in Dutch, with each block consisting of a list of 181 sentences (and 12 fillers). Language order, list (A or B), and condition (target or competitor) were counterbalanced, resulting in eight presentation lists with a fixed random order. Between the two blocks, eye-tracker calibration was repeated. The eye-tracking part of the experiment took approximately one hour.

After the eye-tracking experiment, participants completed the following additional tests: a digit span task, a verbal fluency task, LexTALE Dutch, LexTALE English (Lemhöfer & Broersma, 2012) (see Table 1 for results), and a language background questionnaire based on LEAP-Q (Marian, Blumenfeld, & Kaushanskaya, 2007). The verbal fluency task was performed in Dutch and in English. The participants were asked to name as many words as they could within the categories "food" and "animals" within 1 min. The categories were counterbalanced across languages between participants. Completion of the additional tests took approximately 40 min.

Analyses

Our data set was analysed with linear mixed effects models in R (3.3.2) (R Core Team, 2013) with lme4 (version 1.1-12) (Bates, Mächler, Bolker, & Walker, 2015). The *p*-values for the fixed effects in our models were obtained using the lmerTest package (version 2.0-33) (Satterthwaite degrees of freedom approximation) (Kuznetsova, Brockhoff, & Christensen, 2016). Post-hoc contrasts were performed with the lsmeans package (Kenward-Roger's approximation to degrees of freedom). Our dependent variable was the empirical logit (a quasi-logit transformation suitable for probabilities that are near 0 or 1) of the proportion of eye-data samples in which there was a fixation to a picture over the total number of samples (Barr, 2008). The proportions of looks to the three unrelated pictures were averaged. We ran separate analyses for the trials in which the display featured the target, and trials in which the display featured a competitor. This was done because the competitor model included the semantic distance factor (semantic distance between the competitor picture name and the target for prediction), whereas the target model did not. We also added an analysis of the combined data set, excluding the semantic distance factor.

We first analysed the data of the prediction time frame, without taking into account the time course for prediction. As planning and executing a saccade takes approximately 200 ms (Matin, Shao, & Boff, 1993; Saslow, 1967) the prediction time frame included the eye-data samples starting from 200 ms after the onset of the picture display, to 200 ms after target word onset. We also analysed the data in the time frame starting 200 ms after display onset and ending 1000 ms after target offset (display time frame) to see whether any differences in semantic activation between languages persisted after hearing the target word of the sentence. For these analyses, we first constructed a full model including all theoretically relevant fixed effects and interactions for the prediction time frame (Table 2). The model also included random intercepts of participant and item. All continuous predictors were scaled and centeed. We then used a backward fitting procedure for the fixed effects, followed by forward fitting the random slopes and then backward fitting fixed effects again to find the optimal model (following Cop, Keuleers, Drieghe, & Duyck, 2015; Dirix & Duyck, 2017). To be more specific, we started backward fitting by excluding the fixed model term that was contributing the least to the goodness of fit of the current model. Then we used model comparisons to confirm that the newly constructed model was not significantly lower in goodness of fit than the previous model. We always kept the fixed effects of main experimental interest in the model (see Table 2). When arriving at the restricted model, we added random slopes starting with the factor with the largest t-value. We tested the contribution of each of the random slopes with model comparisons.

Table 2. Factors and interactions included in the full model for the Target trials and Competitor trials.

Fixed factors	Two-way interactions	Three-way interactions
Language (L1 Dutch vs. L2 English)	*Language : Image type*	Language: Image type: Target word onset time
	Language : Target word onset time	Language: Image type: Cloze probability
	Language : Cloze probability	Language : Image type : English LexTALE score
	Language : English LexTALE score	
Image type (experimental vs. unrelated)	Image type : Target word onset time	
	Image type : Cloze probability	
	Image type : English LexTALE score	
	Image type : experimental image frequency	
	Image type : experimental image phoneme count	
	Image type : experimental image phonetic levenshtein distance	
Target word onset time (sentence length upto the target word in ms)		
Cloze probability		
Presentation list		
English LexTALE score		
Experimental image frequency		
Experimental image phoneme count		
Experimental image phonetic levenshtein distance (between L1 and L2 translation equivalents)		
Additional terms competitor model		
Fixed factors	Two-way interactions	Three-way interactions
Semantic distance (between competitor and target, continuous variable)	*Language : Semantic distance*	*Image type : Language : Semantic distance*
	Image type : Semantic distance	
Plausibility (plausibility rating of competitor word as sentence ending)	Image type : Plausibility	

Note: Main experimental terms that were never excluded from the model during backward fitting are printed in italics.

We used this data-driven approach for determining the random effects structure because the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013) resulted in convergence errors. After adding all of the contributing random slopes, we again excluded non-significant fixed interaction effects one by one, until we arrived at the optimal model. An effect or interaction was excluded if a Chi-square test comparing the model with and without the effect was not significant. Backward and forward fitting were performed in the order of the lowest or highest t-value of the model terms, respectively. We report the results for the optimal model. The optimal models we found for the full prediction time frame for the target and competitor data were then used for a time course analysis, in which we fitted the model for each 50 ms time bin in the display time frame (200 ms after display onset up to 1000 ms after target word offset). The data sets and scripts used for the analyses are available online at Open Science Framework (https://osf.io/wy9tm/).

Results

Figure 3 shows the time course of fixations to target, competitor, and unrelated pictures in L1 (Dutch) and L2 (English). The graph shows raw fixation proportions.

Visual inspection of the graph suggests that participants were more likely to fixate on target objects than on competitor objects, and also more likely to fixate on competitor objects than on unrelated objects. Fixation proportions for the target, competitor, and unrelated pictures started to diverge well before the target word onset time both in Dutch and in English.

Analyses full prediction time frame

Target trials. The optimal model for the prediction time frame (200 ms after the onset of the picture display to 200 ms after target word onset) included the factors language, image type (target versus unrelated), target

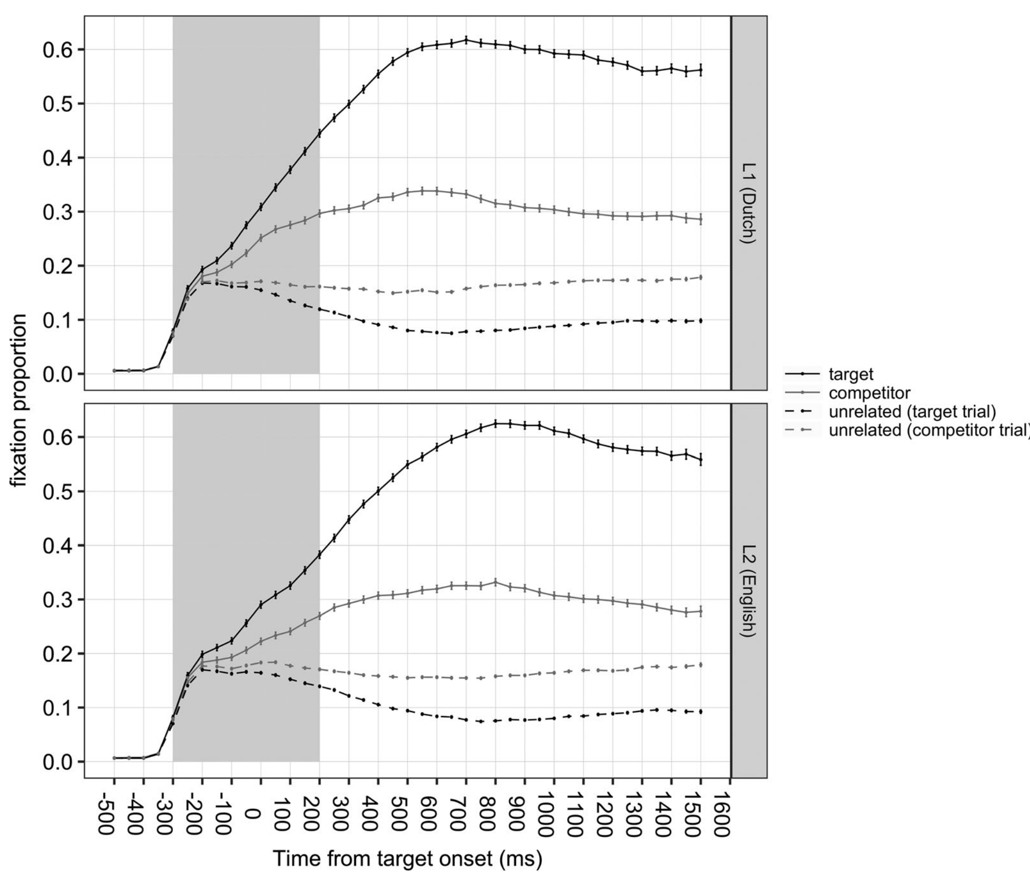

Figure 3. Time course of fixations to target, competitor, and unrelated pictures in the L1 (Dutch) and the L2 (English) relative to target onset. Display onset was 500 ms before target onset. Proportions are based on proportion of samples in which there was a fixation to the picture, aggregated in 50 ms time bins. Proportions for unrelated images were averaged. The area shaded grey is the prediction time frame, in which bottom-up information from the target word could not yet affect looking behaviour (but top-down information from the preceding sentence could). The prediction time frame included the eye-data samples starting from 200 ms after the onset of the picture display to 200 ms after target onset. Whiskers indicate the mean ± standard error.

word onset time (sentence length upto the target word in ms), and presentation list, as well as the interaction between image type and language, and the interaction between image type and target word onset. A random slope of image type was included for each participant and sentence (full results are presented in Table B1 of Appendix B). There was a significant effect of image type (Figure 4, panel A). Importantly, image type also interacted with language. During the prediction time frame, participants were more likely to fixate target images than unrelated images in both the L1 (target raw fixation probability: $M = .26$ $SD = .30$, unrelated raw fixation probability: $M = .14$, $SD = .10$) and the L2 (target raw fixation probability: $M = .24$ $SD = .29$, unrelated raw fixation probability: $M = .15$, $SD = .11$), and this effect was larger in the L1 than in the L2 ($\beta = .26$, $SE = .08$, $t = 3.40$, $p < .001$).

The interaction between image type and target word onset time was also significant ($\beta = -.38$, $SE = .09$, $t = -4.42$, $p < .0001$). As the length of the sentence leading up to the target word increased, so did the difference between fixations to the target and unrelated images. The interaction between image type and cloze probability did not contribute significantly to the model ($\chi^2(2) = .28$, $p = .87$), suggesting that the bias in looks toward the target picture in the prediction time frame did not increase when the cloze probability of the sentence increased. Also, the interaction between L2 LexTALE score, language, and image type did not contribute significantly to the model ($\chi^2(4) = 4.46$, $p = .35$), thus there was no evidence suggesting that relatively proficient bilinguals predicted more than less proficient bilinguals.

Competitor trials. The optimal model included the main effects of language, image type (competitor versus unrelated), semantic distance (between competitor and sentence target, as continuous factor), target word onset time, and presentation list. The model also included the two-way interactions between image type and language, image type and target word onset, image type and semantic distance, and language and semantic distance. Additionally, the model included the three-way interaction between image type, language, and semantic distance. A random slope of image type

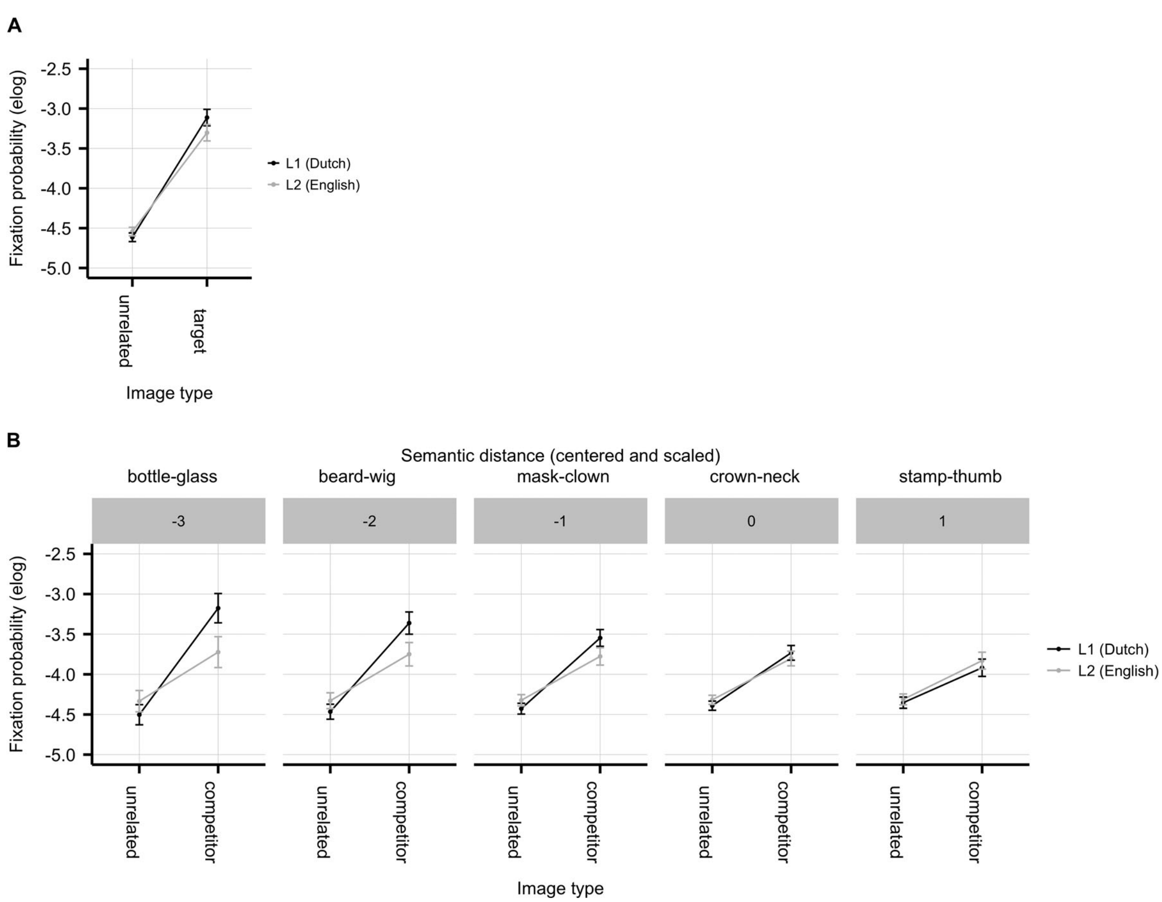

Figure 4. A. Interaction between image type and language for target trials (model predicted means). B. Interaction between image type, language and target-competitor semantic distance (model predicted means). The word pairs above each semantic distance facet are example competitor word pairs in that semantic distance category.

was included for each participant and sentence (full results are presented in Table B2 of Appendix B).

There was a significant main effect of image type (competitor vs. unrelated) ($\beta = -.66$, $SE = .10$, $t = -6.35$, $p < .001$). As shown in Figure 4 panel B, there was a stronger fixation bias to the competitor (versus unrelated images) when the semantic distance between target and competitor was smaller (e.g. *bottle-glass*) ($\beta = .22$, $SE = .07$, $t = 3.04$, $p = .002$). This interaction effect was larger in L1 than in L2 ($\beta = -.19$, $SE = .08$, $t = -2.49$, $p = .013$). Post-hoc tests reveal that the interaction between semantic distance and image type was significant in both languages (L1 Dutch: $\beta = .66$, $SE = .10$, $t = 6.35$, $p < .0001$, L2 English: $\beta = .51$, $SE = .10$, $t = 4.97$, $p < .0001$).

As in the target image data analysis, the interaction between image type and target word onset time was significant ($\beta = -.29$, $SE = .08$, $t = -3.57$, $p < .001$). Longer sentences before the target words yielded larger differences between fixations to the competitor and fixations to the unrelated images. As in the target image data, the interaction between image type and cloze probability did not contribute significantly to the model ($\chi^2(2) = 1.33$, $p = .51$). Also, the interaction between L2 LexTALE score, language and image type did not contribute significantly to the model ($\chi^2(4) = 2.36$, $p = .67$), so that relatively proficient bilinguals did not predict competitors more than less proficient bilinguals.

Individual cognitive differences

Forward digit span score ($M = 9.53$, $SD = 1.83$) and fluency (English and Dutch) (Table 1) and their interactions with image type and language did not contribute to the optimal model fit for the competitor and target trials (all *ps* >.1).[9]

Time course analyses

A time course analysis was carried out to test whether the language effects found in the analyses of the prediction time frame were caused by a delay in fixation bias in the L2 relative to the L1, rather than by an overall weaker fixation bias in L2. With this goal, the data were aggregated in 50 ms time bins starting from the prediction time frame (200 ms after the onset of the picture display). The optimal model for the target trials was run for each 50 ms time bin in the target trial data, and the optimal model for the competitor trials was run for each 50 ms time bin in the competitor trials.[10] We continued to run the models for the 50 ms time bins after the prediction frame, up to 1500 ms after target word onset (the average target word duration was 509 ms and pictures were left on screen for 1000 ms after target offset). In those time bins, looking behaviour could be influenced by hearing the target word. Therefore, we do not interpret the effects in this time window as prediction effects but as effects of ease of integration of information from the auditory target and sentence and the semantic information from the picture display. This type of time course analysis increases the likelihood of Type I errors, and therefore the differences reported here only include those differences that were found consistently in multiple (>1) time bins (following Ito, Corley, & Pickering, 2018). In addition, we plotted the *p*-values in each time bin of the most relevant effects with horizontal lines indicating alpha and corrected alpha (Bonferroni style) in Figure C1 and Figure C2 of Appendix C.

Figure 5 shows the time course of fixations on the target and unrelated objects in the L1 and L2. The solid circles at the top of the graph indicate a significant interaction between language and image type ($p < .05$).

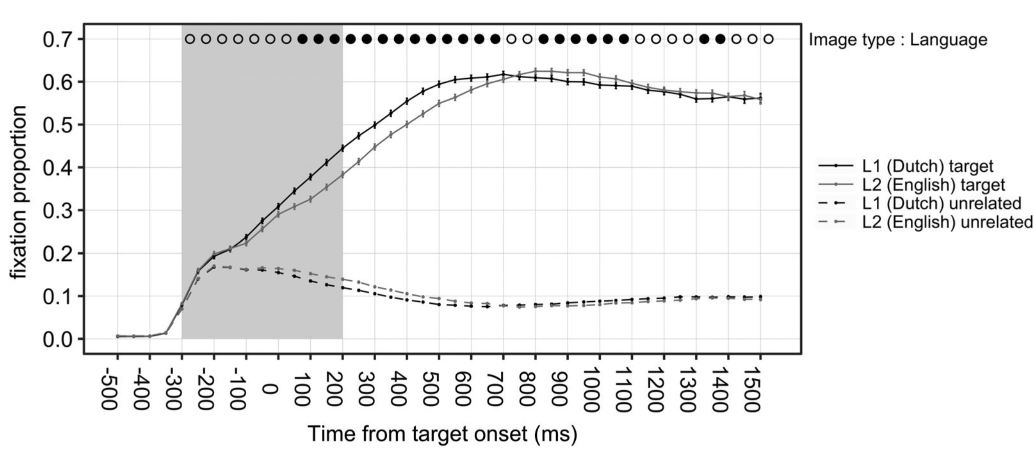

Figure 5. Time course of fixations to the target image and unrelated images in the L1 and L2 relative to target onset. Display onset was 500 ms before target onset. Proportions are based on proportion of samples in which there was a fixation to the picture, aggregated in 50 ms time bins. Proportions for unrelated images were averaged. The area shaded grey is the prediction time frame. Whiskers indicate the mean ± standard error.

In the prediction frame of the target trials, the image type by language interaction was significant only in the last three time bins (50–200 ms after target word onset). The main effect of image type (target vs. unrelated) was already significant at 250 ms before target word onset. After the prediction time frame, at 700 ms, the bias towards the target did reach the same level in the L2 as in the L1 and from 800 to 1100 ms after target word onset the bias towards the target was even larger in the L2 than in the L1.

Figure 6 shows the time course of fixations on the competitor and unrelated objects in the L1 and the L2. The solid circles at the top of the graph indicate a significance of the effects listed on the right ($p < .05$).

First, the main effect of image type became significant 100 ms before target word onset in the competitor trial data set. The interaction between language and image type was significant from −50 ms to 200 ms in the prediction frame and continued to be significant for 50 ms (200–250 ms) in the post-prediction time frame. The bias towards the competitor object was weaker in the L2 than in the L1. The image type effect also became significant at 100 ms before target word onset in both languages separately.

Within the prediction time frame, the interaction between semantic distance and image type was modulated by language from 300 ms before target word onset until 150 ms after target word onset; the effect of semantic distance on the bias towards the competitor was larger in the L1 than in the L2 in those time bins. Figure D1 of Appendix D shows that the interaction effect of semantic distance on the bias towards the competitor gradually increased in the L2 until the three-way interaction with language was no longer significant at 150 ms after target word onset. The effect of semantic distance on bias towards the competitor continued to grow in the L2 after the prediction time frame, and from 450 to 550 ms, the three-way interaction with language was significant again. This time, the effect of semantic distance on the bias towards the competitor was larger in the L2 than in the L1. There are four more later time bins in which the three-way interaction was significant. Again, the effect was larger in the L2 than in the L1 in those time bins. Interestingly, post-hoc tests with lsmeans show that the interaction between image type and semantic distance became significant 300 ms later in the L2 (English) data than in the L1 (Dutch) data (see Figure 6).

Overall time course analysis

In order to compare the time course for target and competitor pre-activation in both languages we ran an additional time bin analysis on the entire data set, including both target and competitor trials, for the bins in the prediciton time frame. All factors included in both the competitor final model and the target final model were included in the model for the overall analysis. The factor trial type (target vs. competitor) was added as well. Semantic distance was not included as factor as it applied only to the competitor trials. A random slope for image type was added by items and by participants. Further random slopes did not contribute to the model fit (as determined by model comparisons with and without each slope for the model applied to the full prediction time frame data set). The image type effect was significant from 250 ms before target word onset ($ps < .05$), and this effect was modulated by trial type from 150 ms before target word onset ($ps < .05$). The bias towards the experimental image was larger on target trials than on competitor trials. The image type effect

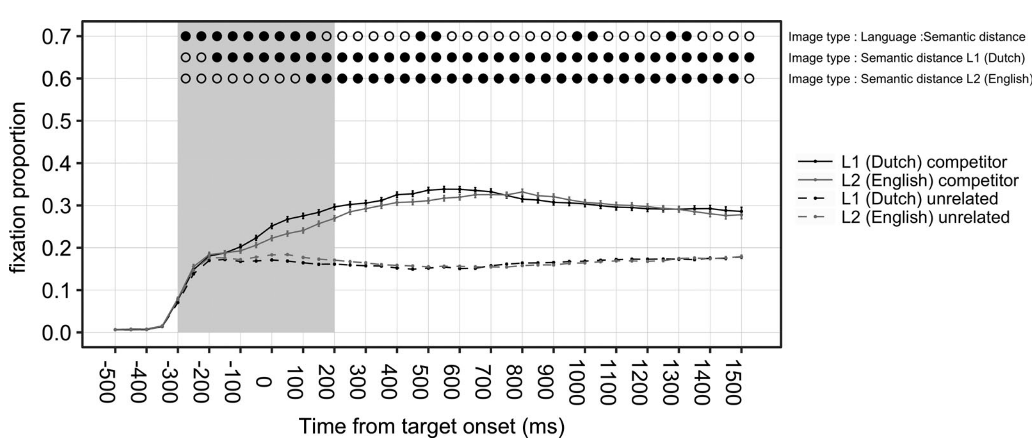

Figure 6. Time course of fixations to the competitor image and unrelated images in the L1 and the L2 relative to target onset. Display onset was 500 ms before target onset. Proportions are based on proportion of samples in which there was a fixation to the picture, aggregated in 50 ms time bins. Proportions for unrelated images were averaged. The area shaded grey is the prediction time frame. Whiskers indicate the mean ± standard error.

interacted with language from time bin 0 onwards, with a larger bias towards the experimental image in L1 than in L2. The three-way interaction between image type, trial type, and language did not reach significance until the final bin of the prediction time frame. Post-hoc tests reveal that on target trials the effect of image type became significant from 250 ms before target word onset onwards in L2, and from 200 ms before target word onset in L1. On competitor trials, the effect of image type was significant from 100 ms before target word onset onwards in both languages.

Discussion

In the present study, we tested whether prediction of meaning during speech comprehension is affected by language (native versus non-native). We found that bilinguals predicted the semantics of target words both in the L1 and the L2; participants were more likely to focus on target objects than on unrelated objects before the auditory target could affect eye-movements. We found a larger prediction effect when bilinguals listened in the L1 than when they listened in the L2. Bilinguals were also more likely to look at semantic competitor objects than at unrelated objects, in both languages. This shows that semantic pre-activation during listening in both languages is strong enough to spread to related concepts, at least when a picture of the related concept is present on the screen. The strength of the competitor fixation bias depended on the semantic distance between target and competitor (the smaller the distance, the larger the bias) and language: the effect of semantic distance on bias to competitor objects was larger in the L1 than in the L2, with an especially strong competitor effect in the L1 for the most strongly related competitors. Time course analyses showed that there was a significant prediction of target word semantics in the L1 and the L2 250 ms before target word onset, and that the prediction effect was larger in the L1 than in the L2 from 150 ms before auditory exposure to the target word could influence looking behaviour. The difference remained significant for 500 ms afterwards. The effect of semantic distance on the bias to competitor objects was larger in the L1 than in the L2 throughout almost the entire prediction time frame. After the prediction time frame, the effect of semantic distance on the bias to the competitor object was the same in the L1 and the L2, and it even became larger in the L2 than in the L1 for a brief period (6 time bins in total).

In this study, differences were found when directly comparing prediction between the L1 and the L2 of the same individuals when both the cues and information to be predicted are of a lexical-semantic nature. The results indicate that semantic prediction in the L2 does not always occur as efficiently as in the L1.

Target prediction

The finding that the effect of pre-activation of the target was smaller in the L2 than in L1 could be due to weaker and/or slower pre-activation in L2. Target pre-activation became significant at approximately the same time in English and Dutch, suggesting that predictive pre-activation of the target was weaker, rather than slower in L2 than in L1. However, these two explanations cannot be teased apart enequivocally in this paradigm. The finding that the effect of pre-activation of the target was smaller in the L2 than in L1 differs with earlier findings on semantic prediction in the L2 (Dijkgraaf et al., 2017; Hopp, 2015; Ito, Corley, & Pickering 2018). Dijkgraaf et al. directly compared predictive looking behaviour in the L1 and the L2 in bilinguals and found no significant difference. Hopp found predictive looking behaviour in L2 like in L1, but only when the cues used for prediction were lexico-semantic and not when predictions were to be based on case-marking information. No direct comparison of prediction in the L1 and L2 was reported for lexico-semantic prediction. Ito et al. found predictive looking behaviour in the L1 and the L2 but they did not report a direct comparison of the strength of the prediction effect in each language. Instead, they reported a similar effect of cognitive load on predictive processing in the L1 and L2.

It is of course possible that the difference between our findings and previous findings is driven by the greater statistical power in the current study. After all, we had 4525 observations per condition in the current study, 270 in Dijkgraaf et al. (2017), 768 in Ito, Corley, and Pickering (2018) (ignoring the cognitive load factor), and 360 and 96 observations for the L2 and L1 groups in Hopp (2015), respectively. But more interestingly, the differences between our findings and the findings of Dijkgraaf et al, Ito, Corley, and Pickering, and Hopp can be attributed to contextual factors or to individual differences between our participants and theirs. The sentences used in the current experiment were longer and often syntactically more complex (e.g. compound sentences) than the simple sentences used in previous studies. This may have to lead the participants to use the routes to prediction to a different extent. Specifically, as predictions in Dijkgraaf et al., Ito et al., and Hopp were based mainly on information from only one word (the verb), low-level lexical associations may have played a large role. The present study used longer, more naturalistic sentences and therefore predictions were likely at least partly based on higher level meaning. The latter may require more cognitive resources unavailable to the L2-comprehenders than prediction via low-level lexical associations (e.g. Huettig, 2015; Pickering & Gambi, 2018), hence the diverging findings. In line with this hypothesis, Ito, Pickering, and Corley (2018) also found an L2 disadvantage in semantic prediction, similar to the current study. These authors also used longer, more naturalistic sentences (e.g. *The tourists expected rain when the sun went behind the **cloud***). Both English native speakers and Japanese-English bilinguals showed anticipatory eye-movements to predictable targets (e.g. *cloud*), but the L2-listeners did so later than the L1-listeners.

Further, in Dijkgraaf et al. (2017), Hopp (2015), and Ito, Corley, and Pickering (2018) the picture display appeared before sentence onset. Pre-activation of target word semantics may have been increased greatly because of the visual presentation of a plausible target object. This may be especially true for bilinguals, as they may rely strongly on visual information during language processing (Navarra & Soto-Faraco, 2007). Therefore, in order to maximise sensitivity for language differences in the current experiment, the pictures appeared only 500 ms before the onset of the target word. This was also done to minimise effects of priming by the visual context.

Besides task and stimulus differences, individual differences could also contribute to differences across studies, but this does not seem to be the case here. For example, prediction in the L2 is thought to approach prediction in the L1 as L2 proficiency increases (Kaan, 2014). However, participants in Ito, Corley, and Pickering (2018), and Dijkgraaf et al. (2017) were highly proficient like the participants in the current experiment, which makes proficiency an unlikely explanation for the diverging results. Also, like in Ito, Corley, and Pickering, Hopp (2015) and Dijkgraaf et al., no effect of proficiency on semantic prediction in L2 was found in the current experiment. The range of proficiencies was possibly too small to detect such an effect.

Competitor prediction

Our finding that the semantic distance effect on competitor prediction was smaller in the L2 than in the L1 in the prediction time frame indicates that spread of semantic activation started later in the L2 than in the L1, that activation spreading was weaker (especially for the most strongly related concepts), or both. The first explanation receives support from the time course analyses of competitor trials, which indicated that the effect of spread of semantic activation became significant later in the L2 than in the L1. When we compared looking behaviour in the L1 and L2 in later time bins (including time bins where hearing the target word could affect looking behaviour) the effect of semantic distance on the bias to the competitor was the same in both languages, or even bigger in the L2. The later significant effect in the L2 suggests a delay in activation. This would be consistent with the temporal delay assumption of the BIA+ model of bilingual visual word recognition (Dijkstra & van Heuven, 2002). This assumption states that due to lower subjective L2 word frequency, activation of word forms and, as a consequence, semantic codes is somewhat delayed in the L2 compared to the L1, while activation patterns themselves are the same.

We also obtained evidence supporting the second explanation above, namely that of weaker lexico-semantic activation in the L2. We observed that the semantic distance effect in the competitor trials was stronger in the L1 than the L2, even though the prediction effect itself became significant in the same time bin in both languages. We predicted such an effect from the assumption that L2 words are mapped onto fewer semantic features than L1 words (Schoonbaert et al., 2009; Van Hell & De Groot, 1998), and that therefore spreading semantic activation should be narrower in the L2 than in the L1. We expected that the difference between the L1 and L2 would be particularly large for less strongly related competitors, because L2 concepts should map onto the core semantic features (shared by strongly related concepts), but perhaps not onto the more remote ones (shared by weakly related concepts). Somewhat surprisingly, the difference between the competitor effects in L1 and L2 was most pronounced for the most strongly related competitors, with very strong semantic pre-activation of closely related concepts especially from L1 words. This suggests that stronger spreading semantic activation for the L1 is determined by the strength of mappings between word forms and semantics, rather than by the number of mapped semantic features. Our interaction effect between language, image type, and semantic distance suggests that L1 words have stronger links with the underlying concepts than L2 words, which then leads to stronger semantic pre-activation for very related concepts. Such an explanation is consistent with for instance the weaker links account, which assumes that divided language practice across languages leads to weaker links between representations in the bilingual language system (Gollan et al., 2008; Gollan, Montoya, Fennema-Notestine, & Morris, 2005). Because L2 exposure is far less frequent for our bilinguals, mappings from L2 word forms onto semantics are weaker.

Finally, in this paradigm, we cannot distinguish between competitor activation through target word pre-activation, followed by spreading activation to the competitor on the one hand, and competitor activation via passive resonance of the semantics of semantically related words in the sentence on the other hand. Both mechanisms may also be additive. Future studies could be aimed at pinpointing the exact locus of the delay in/weaker effect of spreading semantic activation in L2 compared to L1. In any case, the present results show that L2 yields slower and/or weaker semantic prediction overall.

Other potential modulating factors

As less cognitive resources may be available during L2 than during L1 processing (e.g. Francis & Gutiérrez, 2012; McDonald, 2006) we expected that participants with a larger working memory capacity would have less of a disadvantage in L2 prediction. However, we found no effects of working memory span (forward digit span) on prediction in L1 and L2, suggesting that working memory resources may not drive the current differences between L1 and L2. Consistent with our finding, Ito, Corley, and Pickering (2018) found that a cognitive load during speech comprehension affects prediction in L1 and L2 to the same extent. On the other hand, the sample of 50 participants in our study may not have been large enough to detect an effect of individual differences in working memory capacity, or there may not have been sufficient variation in resources given that all participants were university students. Future research using a more sensitive design could be aimed at testing whether working memory resource limitations in L2 may underlie the L2 disadvantage in prediction.

For both the target and the competitor data we found that target word onset time (the length of the sentence leading up to the target word) affected prediction. The longer the sentence, the larger the prediction effect. This may be due both to the increased time for pre-activation in longer sentences and the increased amount of context information to serve as cue for prediction. The effect of sentence length on predictive looking behaviour was not modulated by language (L1 vs. L2). Apparently, even though semantic pre-activation was weaker in the L2 than in the L1, the length of the sentence did not differentially affect pre-activation in the L1 and the L2. A limitation of the current study is that the Dutch sentences were slightly longer than the English sentences, possibly contributing to the L2 disadvantage in prediction.

Somewhat unexpectedly, we found no effect of sentence cloze probability on target or competitor pre-activation, even though we included sentences with a rather large range of cloze probabilities (0.08–1). The cloze probability test was filled out with the sentences as context only. The presence of a picture display with a target or competitor word may have increased the probability of the sentence ending with the target word, thereby eliminating the cloze probability effect. Furthermore, participants listened to 362 experimental sentences with an average cloze probability of .68 for English and .71 for Dutch. The exposure to so many predictable sentences may have further enhanced the likelihood of predictive behaviour overall (Lau, Holcomb, & Kuperberg, 2013), and thereby reduced the chances of finding an effect of cloze probability.

Conclusion

In sum, even in an experimental setting with many relatively high cloze sentences and additional visual information, we find differences in the strength and time course between L1 and L2 semantic prediction. Therefore, language dominance (L1 versus L2) can not only affect prediction based on (morpho-)syntactic cues but also prediction of semantic information based on semantic context information, if more fine-grained measures of semantic activation are targeted. The difference between prediction in the L1 and the L2 is compatible with the hypothesis that lexico-semantic mappings are weaker for L2 than for L1 (Gollan et al., 2008, 2005), and with slower word form activation and, as a consequence, slower spread of semantic activation in L2 than in L1, due to smaller subjective word frequency in the L2 (Dijkstra & van Heuven, 2002). As working memory (digit span score) did not affect prediction, an explanation in terms of limited cognitive resources in L2 (Francis & Gutiérrez, 2012; McDonald, 2006) is less likely. We suggest that there is no qualitative difference between lexico-semantic prediction in the L2 and the L1, but that subtle quantitative differences arise when graded semantic relations are assessed, like in the present paradigm. The differences between our findings and previous research in which no language effect on semantic prediction was found, illustrate again that prediction during language comprehension is a highly flexible process. Future studies should be aimed at testing which exact contextual factors and individual differences, best explain the diverging findings on predictive behaviour in L2 comprehension.

Notes

1. Out of the 871 sentences, 54 were from the Block and Baldwin (2010) sentence set, and 31 from Hamberger, Friedman, and Rosen (1996). Another 39 were adapted from Block and Boldwin, and 31 were adapted from

Hamberger, Friedman and Rosen. These sentences were adapted so that they could be translated to Dutch without changing the sentence final word.

2. 0 = no overlap, 1 = identical (Schepens, Dijkstra, Grootjen, & van Heuven, 2013).
3. We applied the optimal models to the prediction time frame data excluding trials in which the experimental image was a cognate (phonological levenshtein distance >.5; following Schepens et al., 2013). For the target, the language by image type interaction remained significant ($\beta = 35$, $SE = .08$, $t = 4.19$, $p < .001$). For the competitor data, the three-way interaction between language, image type and semantic distance also remained significant ($\beta = -.21$, $SE = .08$, $t = -2.54$, $p = .01$).
4. The target/competitor words sometimes had false friends in the other language (e.g. *map*, meaning *folder* in Dutch). We applied the optimal models to the prediction time frame data excluding trials in which the experimental image (target or competitor) had (identical) false friends in the other language. Both words with identical orthographic false friends (85 out of 724 words) and words with identical phonological false friends (25 out of 724 words) were excluded (106 in total). For the target, the language by image type interaction remained significant ($\beta = .24$, $SE = .09$, $t = 2.77$, $p = .006$). As for the competitor, competitor semantic distance still interacted with image type ($\beta = .28$, $SE = .08$, $t = 3.49$, $p < .001$), but the three-way interaction with language was no longer significant ($\beta = -.13$, $SE = .09$, $t = -1.54$, $p = .12$). To investigate whether the three-way interaction disappeared because of loss of power or because false friend status actually affected looking behaviour we compared the final model with the final model plus the factor false friend status (false friend in the other language yes or no) and the interaction between false friend status and image type. False friend status did not contribute to the model fit ((2) = 1.73, $p = .42$).
5. Competitors were sometimes ungrammatical as sentence ending (e.g. because of a gender mismatch with the preceding determiner) and/or they could violate a phonotactic rule (due to a mismatch with preceding indefinite article *a* or *an*). To test whether competitor grammaticality affected our results we applied the optimal models to the prediction frame data excluding trials in which the competitor was ungrammatical or violated a phonotactic rule. Fifty (out of 362) English sentences and 43 (out of 362) Dutch sentences were excluded. For the target, the language by image type interaction remained significant ($\beta = .25$, $SE = .09$, $t = 2.89$, $p = .004$). For the competitor data, the two-way language by image type interaction remained significant ($\beta = .22$, $SE = .08$, $t = 2.68$, $p = .007$), as did the interaction between image type and semantic distance ($\beta = .27$, $SE = .08$, $t = 3.45$, $p < .001$). The three-way interaction between language, image type and semantic distance approached significance ($\beta = -.15$, $SE = .08$, $t = -1.87$, $p = .06$). In addition, adding competitor grammaticality and the interaction between grammaticality and image type to the optimal model for the prediction time frame (competitor data set) did not improve the model fit ($\chi(2) = 1.63$, $p = .44$).
6. The English corpora used were UKWAC (Ferraresi, Zanchetta, Baroni, & Bernardini, 2008) (containing texts from the .uk internet domain) and a subtitle corpus (Mandera et al., 2017) (downloaded from http://opensubtitles.org). For Dutch Sonar-500 text corpus (Oostdijk, Reynaert, Hoste, & van den Heuvel, 2013) (texts from conventional and new media) and another subtitle corpus (Mandera et al., 2017) were used.
7. In 8 sentences (out of 362 Dutch and 362 English sentences) either the target word or the competitor word was present in the sentence, either with the same meaning or a slightly different meaning (e.g. *She locked her bicycle to a fence with a lock*, *Ivory is derived from an elephant or a rhino*-> competitor: *elephant*). A picture of the target or competitor word also present in the sentence was likely to attract more fixations in these sentences than in other sentences. The random slope for item in the analyses ensured that this possible confound did not affect the results. In addition, an analysis of the target and competitor data of the full prediction time frame without these 7 sentences did not change the results.
8. Due to an error in the test plausibility ratings for three (out of 724 sentences) were missing.
9. Due to technical problems the scores for fluency (Dutch and English) is missing for two participants, and the score for digit span is missing for one participant.
10. It is possible that any of the factors excluded from the final models had a significant effect in some of the time bins. We used the final model for the time bin analyses in order to investigate whether different languages showed a different time course of effects of the relevant variables, as observed in the full prediction time frame analysis. The alternative of running a separate backfitting procedure for each time bin could not fulfill this goal, as this would lead to models with different factors in each bin, so that the results for each time bin would not have been directly comparable.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by a Concerted Research Action (GOA grant number BOF13/GOA/032) from the Special Research Fund (Bijzonder Onderzoeksfonds), Ghent University.

ORCID

Aster Dijkgraaf http://orcid.org/0000-0002-2420-4150

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. doi:10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. doi:10.1016/j.jml.2006.12.004

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, *28*(4), 525–560. doi:10.1080/01690965.2012.658072

Baayen, H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database. Release 2 (CD-ROM)*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. D. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Research*, *141*, 157–169. doi:10.1016/j.visres.2017.03.001

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. doi:10.1016/j.jml.2007.09.002

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods*, *42*(3), 665–670. doi:10.3758/BRM.42.3.665

Broersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer (Version 5.3.78). Retrieved from www.praat.org

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and Affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 687–696. doi:10.1037/0278-7393.30.3.687

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, *22*(5), 1216–1234. doi:10.3758/s13423-015-0819-2

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394–20120394. doi:10.1098/rstb.2012.0394

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition*, *20*(5), 917–930. doi:10.1017/S1366728916000547

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(03), 175–197. doi:10.1017/S1366728902003012

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkänen, L. (2010). Early Occipital sensitivity to syntactic category is based on form Typicality. *Psychological Science*, *21*(5), 629–634. doi:10.1177/0956797610367751

Dirix, N., & Duyck, W. (2017). The first- and second-language age of acquisition effect in first- and second-language book reading. *Journal of Memory and Language*, *97*, 103–120. doi:10.1016/j.jml.2017.07.012

Duyck, W., Assche, E. V., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 663–679. doi:10.1037/0278-7393.33.4.663

Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, *42*(2), 133–141. doi:10.1111/j.1469-8986.2005.00274.x

Ferraresi, A., Zanchetta, E., Baroni, M. G., & Bernardini, S. (2008). Introducing and evaluating ukWaC, a very large web-derived corpus of English. In *Proceedings of the 4th Web as Corpus Workshop (WAC-4)* (pp. 47–54). Marrakech, Morocco: LREC.

Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language*, *51*(1), 1–22. doi:10.1016/j.jml.2004.01.004

Foucart, A. (2015). Prediction is a question of experience. *Linguistic Approaches to Bilingualism*, *5*(4), 465–469. doi:10.1075/lab.5.4.04fou

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461–1469. doi:10.1037/a0036756

Foucart, A., Moreno, E., Martin, C. D., & Costa, A. (2015). Integration of moral values during L2 sentence processing. *Acta Psychologica*, *162*, 1–12. doi:10.1016/j.actpsy.2015.09.009

Foucart, A., Ruiz-Tada, E., & Costa, A. (2015). Anticipation processes in L2 speech comprehension: Evidence from ERPs and lexical recognition task. *Bilingualism: Language and Cognition*, 1–7. doi:10.1017/S1366728915000486

Foucart, A., Ruiz-Tada, E., & Costa, A. (2016). Anticipation processes in L2 speech comprehension: Evidence from ERPs and lexical recognition task. *Bilingualism: Language and Cognition*, *19*(01), 213–219. https://doi.org/10.1017/S1366728915000486

Francis, W. S., & Gutiérrez, M. (2012). Bilingual recognition memory: Stronger performance but weaker levels-of-processing effects in the less fluent language. *Memory & Cognition*, *40*(3), 496–503. doi:10.3758/s13421-011-0163-3

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis⋆. *Journal of Memory and Language*, *58*(3), 787–814. doi:10.1016/j.jml.2007.07.001

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*(7), 1220–1234. doi:10.3758/BF03193224

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, *140*(2), 186–209. doi:10.1037/a0022256

Hamberger, M. J., Friedman, D., & Rosen, J. (1996). Completion norms collected from younger and older adults for 198 sentence contexts. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 102–108. doi:10.3758/BF03203644

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. doi:10.1177/0267658312461803

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching*, *53*(3), 277–306. doi:10.1515/iral-2015-0014

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. doi:10.1016/j.brainres.2015.02.014

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32. doi:10.1016/j.cognition.2004.10.003

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93. doi:10.1080/23273798.2015.1047459

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. doi:10.1016/j.actpsy.2010.11.003

Ito, A., Corley, M., & Pickering, M. J. (2018). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition*, *21*(2), 251–264. doi:10.1017/S1366728917000050

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*, 157–171. doi:10.1016/j.jml.2015.10.007

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 635–652. doi:10.1037/xlm0000315

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11. doi:10.1016/j.jml.2017.09.002

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*(2), 257–282. https://doi.org/10.1075/lab

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650. doi:10.3758/BRM.42.3.643

Kohlstedt, T., & Mani, N. (2018). The influence of increasing discourse context on L1 and L2 spoken language processing. *Bilingualism: Language and Cognition*, *21*(01), 121–136. doi:10.1017/S1366728916001139

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. doi:10.1080/23273798.2015.1102299

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). *lmerTest: Tests in Linear Mixed Effects Models*. Retrieved from https://CRAN.R-project.org/package=lmerTest

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2013). Interlingual lexical competition in a spoken sentence context: Evidence from the visual world paradigm. *Psychonomic Bulletin & Review*, *20*(5), 963–972. doi:10.3758/s13423-013-0405-4

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484–502. doi:10.1162/jocn_a_00328

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

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. doi:10.1016/j.jml.2016.04.001

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. doi:10.1044/1092-4388(2007/067)

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588. doi:10.1016/j.jml.2013.08.001

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372–380. doi:10.3758/BF03206780

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381–401. doi:10.1016/j.jml.2006.06.006

McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America*, *131*(1), 509–517. doi:10.1121/1.3664087

Mitsugi, S., & Macwhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition*, *19*(01), 19–35. doi:10.1017/S1366728914000881

Molinaro, N., Giannelli, F., Caffarra, S., & Martin, C.. (2017). Hierarchical levels of representation in language prediction: The influence of first language acquisition in highly proficient bilinguals. *Cognition*, *164*, 61–73. https://doi.org/10.1016/j.cognition.2017.03.012

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research*, *71*(1), 4–12. doi:10.1007/s00426-005-0031-5

Nieuwland, M., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N. … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, *7*, e33468.

Oostdijk, N., Reynaert, M., Hoste, V., & van den Heuvel, H. (2013). *SoNaR user documentation*. Retrieved from https://ticclops.uvt.nl/SoNaR_end-user_documentation_v.1.0.4.pdf

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*, *8*(1), 89. doi:10.1186/1471-2202-8-89

Peters, R., Grüter, T., & Borovsky, A. (2015). Anticipatory and locally coherent lexical activation varies as a function of language proficiency. In *Proceedings of the 37th Annual*

Meeting of the cognitive Science Society (pp. 1865–1870). Austin, TX: Cognitive Science Society.

Phillips, C., & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Linguistic Approaches to Bilingualism*, *5*(4), 409–453. doi:10.1075/lab.5.4.01phi

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. doi:10.1037/bul0000158

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329–347. doi:10.1017/S0140525X12001495

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi:10.1037/0033-2909.124.3.372

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria.: R Foundation for Statistical Computing,. Retrieved from http://www.R-project.org/

Rommers, J., Meyer, A. S., & Huettig, F.. (2015). Verbal and nonverbal predictors of language-mediated anticipatory eye movements. *Attention, Perception, & Psychophysics*, *77*(3), 720–730. https://doi.org/10.3758/s13414-015-0873-x

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, *51*(3), 437–447. doi:10.1016/j.neuropsychologia.2012.12.002

Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica*, *137*(2), 172–180. doi:10.1016/j.actpsy.2010.09.010

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America*, *57*(8), 1030–1033. doi:10.1364/JOSA.57.001030

Schepens, J., Dijkstra, T., Grootjen, F., & van Heuven, W. J. B. (2013). Cross-Language Distributions of high frequency and Phonetically similar cognates. *PLoS ONE*, *8*(5), e63006. doi:10.1371/journal.pone.0063006

Schoonbaert, S., Duyck, W., Brysbaert, M., & Hartsuiker, R. J. (2009). Semantic and translation priming from a first language to a second and back: Making sense of the findings. *Memory & Cognition*, *37*(5), 569–586. doi:10.3758/MC.37.5.569

Severens, E., Lommel, S. V., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica*, *119*(2), 159–187. doi:10.1016/j.actpsy.2005.01.002

Van Bergen, G., & Flecken, M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. *Journal of Memory and Language*, *92*, 26–42. doi:10.1016/j.jml.2016.05.003

Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, *1*(3), 193–211. doi:10.1017/S1366728998000352

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. doi:10.1080/17470218.2013.850521

Weber, A., & Broersma, M. E. (2012). Spoken word recognition in second language acquisition. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Hoboken, NJ: Wiley. doi:10.1002/9781405198431