Cognition 124 (2012) 66–71

Contents lists available at SciVerse ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Brief article

Learning individual talkers' structural preferences

Yuki Kamide *

School of Psychology, University of Dundee, Dundee DD1 4HN, Scotland, UK

ARTICLE INFO

Article history:
Received 7 July 2010
Revised 27 February 2012
Accepted 7 March 2012
Available online 11 April 2012

Keywords:
Speaker variability
Talker adaptation
Learning
Relative clause attachment ambiguity
Visual-world eye-tracking

ABSTRACT

Listeners are often capable of adjusting to the variability contained in individual talkers' (speakers') speech. The vast majority of findings on talker adaptation are concerned with learning the contingency between *phonological* characteristics and talker identity. In contrast, the present study investigates representations at a more abstract level – the contingency between *syntactic attachment style* and talker identity. In a 'visual-world' experiment, participants were exposed to semi-realistic scenes depicting several objects (e.g., an adult man, a young girl, a motorbike, a carousel, and other objects) accompanied by a spoken sentence with a structurally ambiguous relative clause (e.g., 'The uncle of the girl who will ride the motorbike/carousel is from France.' In the context of the scene, 'motorbike' suggested the uncle as the agent of the riding, whereas 'carousel' suggested the girl as the agent). For half the experimental items, one version of the sentence was read by one talker, who *always* uttered sentences that resolved, pragmatically, to the high attachment (the uncle as the agent), and the other by another talker, who *always* uttered sentences resolving to the low attachment (the girl as the agent). For the other half of the experimental items, both versions were read by a third talker who produced both high and low attachments. It was found that, after exposure to these stimuli, and for new sentences not heard previously, participants learnt to anticipate the 'appropriate' attachment depending on talker identity (with no attachment preference for the talker who produced both attachment types). The data suggest that listeners can learn the relationship between talker identity and abstract, structural, properties of their speech, and that syntactic attachment decisions in comprehension can reflect sensitivity to talker-specific syntactic style.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Speech is highly variable, even for the same utterances: Different speakers exhibit different phonetic characteristics, such as frequency, voice onset time, and speech rate (e.g., Ladefoged, 1980; Monson & Engebretson, 1977; Peterson & Barney, 1952). However, listeners easily 'normalise' such speaker (or talker)-specific attributes, and generally recognise spoken words effortlessly. Listeners' ability to adapt to different speaker characteristics has been demonstrated repeatedly (e.g., Creel, Aslin, & Tanenhaus, 2008; Eisner & McQueen, 2005; Goldinger, 1996, 1998; Kraljic & Samuel, 2006; McLennan & Luce, 2005).[1]

Recently, Creel et al. (2008) conducted a visual-world eye-tracking experiment that investigated whether and when, during processing, listeners learn the contingency between spoken words and talker identity.[2] When hearing the voice of a talker who only ever uttered 'cow' but never

* Tel.: +44 (0)1382 384614; fax: +44 (0)1382 229993.
E-mail address: y.kamide@dundee.ac.uk

[1] The term 'speaker' seems to indicate focus on the language production system as the term is often used in the production literature. Thus, in this paper, we will use the term 'talker' rather than 'speaker' for the producer of the utterance that listeners comprehend in order to make it clear that our focus is on the comprehension process in the listeners.

[2] In other contexts, the term 'identity' could be used to indicate individuals' high-level personal characteristics, such as personalities, appearances, nationalities. However, in our contexts, the term is limited to meanings akin to talker 'label' (e.g., talkers A and B).

0010-0277/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2012.03.001

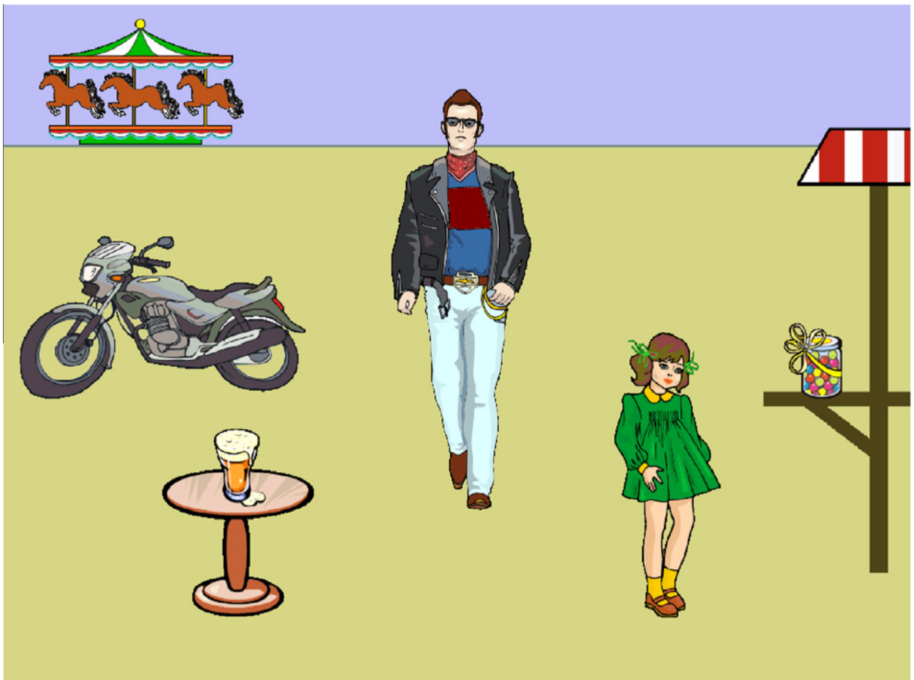

Fig. 1. An example visual stimulus used in the experiment.

the competitor 'couch', participants would not look at a couch when hearing the earliest moments of 'cow'. They *would* look to the couch during 'cow' if listening to a talker who had uttered both during training. Thus, listeners can learn to use phonetic cues associated with talker identity to restrict the domain of reference to just those objects whose names were compatible with the previous experience of the talker-specific vocabularies.

In the context of spoken word recognition, the sequence /kau.../ is ambiguous in respect of whether it will continue one way or another (as in 'cows' or 'couch'), and yet participants in the Creel et al. (2008) study learnt to use talker identity to resolve that ambiguity early on during the word, according to the idiosyncrasies of the individual talkers. There exists a potential parallel in the context of sentence comprehension: The sequence 'the uncle of the girl who...' is ambiguous in respect of whether it will continue one way (the relative clause introduced by 'who' modifying the uncle) or another (the relative clause modifying the girl) – if one talker always modified the first noun phrase (NP1; high attachment), and the other talker always modified the second one (NP2; low attachment), could listeners use this talker-specific information to resolve the ambiguity in advance of any disambiguating information? Can listeners modify their *syntactic* expectations as a function of their experience of individual talkers and their idiosyncratic attachment preferences? Could they modify these expectations from moment to moment depending on the individual talker?

In the present visual-world eye-tracking study, a display with several objects (e.g., Fig. 1) was presented with an auditory sentence (e.g., 1a or 1b) based on the ambiguity first investigated by Cuetos and Mitchell (1988):

- (1a) The uncle of the girl who will ride the motorbike is from France.
- (1b) The uncle of the girl who will ride the carousel is from France.

Syntactically, either the NP1 ('the uncle') or NP2 ('the girl') could be the antecedent of the relative clause in (1a) and (1b). The resolution of the ambiguity can be achieved only by the application of real-world knowledge concerning the relative likelihood of the individuals depicted in the scene performing the different actions (cf. Kamide, Altmann, & Haywood, 2003). Thus, (1a) should be resolved in favour of high attachment (the uncle riding the motorbike), whereas low attachment should be preferred for (1b) (the girl riding the carousel). For the 'cued' stimuli of the study described below, (1a) would be spoken by a male talker, and (1b) by a female talker (counterbalanced across participant groups); thus, speaker identity could in principle cue the appropriate attachment. For the 'uncued' stimuli, both (1a) and (1b) would be spoken by a third talker, and thus the identify of this speaker did not cue the appropriate attachments for the utterances she produced (half the items were cued, and half uncued – participants would thus hear three voices; one always attaching high, another always attaching low, and a third attaching high on some occasions and low on others). At issue is whether, after exposure to all three voices, new ambiguous sentences uttered by these same talkers will be resolved according to the talker-specific contingencies encountered during exposure.[3]

Mitchell and colleagues (Mitchell & Cuetos, 1991; Mitchell, Cuetos, Corley, & Brysbaert, 1995) proposed a linguistic 'tuning' hypothesis to explain why native speakers of different languages prefer different attachment decisions for the same ambiguity. Cuetos and Mitchell (1988) found that Spanish readers ordinarily (i.e., out of context) prefer to interpret 'Someone shot the servant of the actress who was on the balcony' as meaning that the servant was on the balcony ('high attachment'), whereas English readers prefer to attach the relative clause 'low', to mean that the actress was on the balcony, despite the fact that the constituent order is identical in the two languages. According to the tuning hypothesis, readers' attachment decisions are determined by the relative frequency of occurrence of the alternative structures in a given language. Corpus studies provided support for the hypothesis by showing that high attachment structures occur more frequently than low attachments in Spanish corpora, whereas the opposite pattern was found in English corpora (however, see Mitchell & Brysbaert, 1998).

Crucially, whereas the tuning hypothesis focussed on the relative occurrence of alternative structures in a given *language*, here we ask whether listeners are sensitive to the relative occurrence of alternative structures in a given *talker*, and whether listeners can dynamically alter their expectations on a trial-by-trial basis as a function of talker identity.

Such trial-by-trial modification of syntactic expectation would have implications for syntactic priming (e.g., Bock, 1986; Pickering & Branigan, 1998), which obtains between comprehension of one sentence and comprehension of another with the same structure (Arai, van Gompel, & Scheepers, 2007; Ledoux, Traxler, & Swaab, 2007; Scheepers & Crocker, 2004). Although these previous studies, like those associated with the original tuning hypothesis, suggest that comprehenders can adapt to patterns in the linguistic environment as a function of either immediate experience (priming) or longer-term experiential knowledge (tuning), the suggestion that listeners might dynamically change their expectations on a trial-by-trial basis would complicate accounts of syntactic priming that predict that attachment on one trial should predict (the same) attachment on the next.

[3] One might argue that our study deals with learning of less abstract contingencies than are required for syntactic attachment. For example, one might argue that our study merely tests whether or not listeners can learn the association between particular talkers and one or other NP (depending on its position within the sentence) as the agent of the given verb. We do not dispute such a decompositional view – rather, we view it as a core element of syntactic attachment operations. Thus, we argue that selecting one of the two possible NPs as the agent of the action (based on talkers' voice characteristics in our case) is one of the most central parts of the abstract attachment operation, and there is an operational equivalence between learning such contingencies and learning certain attachments patterns.

2. Experiment

2.1. Method

2.1.1. Participants

Forty-eight participants from the University of Dundee student community took part in this study. They participated either for course credit or for £5.00. All were native speakers of English and either had uncorrected vision or wore soft contact lenses or spectacles.

2.1.2. Materials

Twenty scenes similar to Fig. 1 were created based on a subset of Kamide et al.'s (2003) materials. A 254-colour palette was used, and a resolution of 1024×768 pixels. For each scene, four sentences were recorded for the training phase, two based on sentences (1a) and (1b), repeated below, and two based on (2a) and (2b):

- (1a) The uncle of the girl who will ride the motorbike is from France.
- (1b) The uncle of the girl who will ride the carousel is from France.
- (2a) The niece of the man who will ride the carousel is from France.
- (2b) The niece of the man who will ride the motorbike is from France.

(2a) and (2b) are high and low attachment sentences, respectively. These versions were included in addition to (1a) and (1b) to ensure that participants would be exposed to both theme objects ('carousel' and 'motorbike') after each verb, and for each talker, to prevent them from learning individual verb-theme sequences (e.g., In the cued cases, a talker would read both (1a) and (2a), and another talker both (1b) and (2b)). All four versions of the sentences were presented once to each participant during the training phase.

For the test phase, the same pictures were used with different versions of each sentence[4]:

- (3a) The uncle of the girl who will taste the beer is from France.
- (3b) The uncle of the girl who will taste the sweets is from France.
- (4a) The niece of the man who will taste the sweets is from France.
- (4b) The niece of the man who will taste the beer is from France.

The test sentences introduced a new verb ('taste') and theme ('sweets'/'beer'). Thus, participants could not simply learn that a particular sequence (NP1-NP2-verb-theme) was associated with a particular talker. Each participant heard either (3a) and (3b), or (4a) and (4b) (only two versions were presented to each participant in order to mini-

[4] The current design was adopted following Creel et al. (2008), who used the same pictures in the two phases. This had the benefit of allowing a simple within-participant and within-item design in the statistical comparison across the two phases.

mise learning effects within the test phase). It was predicted, for example, that if the participant had heard (2a) read by one talker in the training phase, s/he should anticipate for the same talker that the sweets (target) rather than the beer (competitor) would be referred to after 'The niece of the man who will taste the…' in (4). Such anticipation would be indicated by more saccades towards the target objects (as opposed to the competitor objects) in the test phase than in the training phase for the cued stimuli (the stimuli for which speaker identity reliably correlated with attachment type). For the uncued stimuli (for which speaker identity does not correlate with attachment type), in contrast, such an increase in the eye-movement proportion is not expected, as the talker identity would not be a cue for the eventual attachment.

Forty filler items were also included in the training phase, containing unambiguous single noun phrase antecedents with a relative clause (e.g., 'The woman who will pour the wine into the glass has two Abyssinian cats'). The accompanying pictures contained five objects. Twenty of the fillers were repeated in the test phase.

The experimental sentences were recorded by three different talkers. The profile of each talker was as follows: talker A – male, standard English accent; talker B – female, standard English accent; talker C – female, standard Scottish accent. Talkers A and B served in the cued condition, and talker C in the uncued condition. There were significant differences in the speech rate of each talker (average duration of the experimental sentences: talker A – 4741 ms; talker B – 5184 ms; talker C – 4358 ms; $F(2,318) = 877.772$; $p < 0.001$). The filler sentences were read by only talkers A and B (20 each).

Four lists of presentation items were created, presented to four different participant groups. The four lists enabled all items to be tested in both stimulus conditions (cued vs. uncued), and both attachment sentences to be spoken by both talkers A and B. Within both the training and test phases, trials were divided into smaller blocks of 30 trials (allowing a break if necessary). Within each block, the following were counterbalanced for each participant: (a) the number of sentences read by each talker (10 each); (b) the number of high attachment/low attachment sentences (10 each, plus 10 fillers); and (c) the number of trials in cued/uncued conditions (10 each, plus 10 fillers). The items were pseudo-randomised with the following constraints: (a) the first two trials of each block were fillers; (b) no two consecutive trials belonged to the same combination of the same talker condition. The items in the training phase were ordered differently from the corresponding items (those sharing the same scenes) in the test phase.

In sum, the training phase consisted of four blocks (30 trials in each block; 20 experimental trials and 10 filler trials). The test phase consisted of two blocks (with the same structure as the training blocks). Thus, in training, participants saw each scene 4 times, each of them accompanied by a different version of the corresponding sentence set (for half the items, two spoken by talker A, and two by talker B, and for the other half of the items, all four spoken by talker C). In the test phase, participants saw each scene presented twice (with a different version of the sentence). The makeup of each block is shown in Table 1:

2.1.3. Procedure

Participants were seated in front of a 21 in. display with their eyes approximately 60 cm from the display. They wore an SR Research EyeLink II head-mounted eye-tracker, sampling at 500 Hz from one eye (viewing was binocular). Sentences were presented over loudspeakers. Participants were told that 'we are interested in what happens when people look at these pictures while listening to sentences that describe something that might happen in the picture'. Between each trial, participants were shown a single centrally-located dot to correct for any drift in the eye-track calibration. This dot was then replaced by a fixation cross and participants would press a response button for the next trial. The onset of the visual stimulus preceded the onset of the spoken stimulus by 1000 ms. The trial was automatically terminated after 10 or 12 s, depending on the length of the auditory stimulus. After every eighth trial, the eye-tracker was recalibrated using a 9-point fixation stimulus. There were four practice trials before the main experimental blocks. At the end of each block, there was a display indicating participants had reached the end of the block and were free to have a short break if they wished. The entire experiment lasted approximately 60 min.

2.2. Results

Data collected in the first two blocks ('early-training' phase) and the final two blocks ('test' phase) were entered into the statistical analysis described below. Table 2 shows the percentage of trials in which at least one saccadic eye movement was launched towards the target (motorbike in 1a) or the competitor (carousel in 1a) in each condition, in the Early-training phase and the Test phase (standard deviations in the parentheses). The time-window the data were taken from started at the onset of the relative-clause verb and ended at the onset of its theme object (e.g., 'ride_the_'). The mean duration of the critical temporal region was 528 ms (talker A – 542 ms; talker B – 512 ms)

Table 1
Properties of each block[a].

Trial type	Experimental		Filler
Sentence type	High attachment	Low attachment	Unambiguous relative clauses
Number of items	10	10	10
Talkers (number of items per talker)	Talker A (5) Talker C (5)	Talker B (5) Talker C (5)	Talker A (5) Talker B (5)

[a] There were four participant groups, with talker (A uttering High, B Low, or vice versa) and Stimulus set (one or the other half of stimuli in High or Low attachment version) fully counterbalanced across groups.

Table 2
Percentage of trials with at least one saccade towards the Target or Competitor objects (with standard deviations in parentheses) during 'verb_the_' in the relative clause in the Early-training and Test phases in the experiment.

Phase	Early-training		Test	
Object	Target	Competitor	Target	Competitor
Cued stimuli (talkers A and B)	17.7 (10.29)	19.4 (7.67)	13.1 (9.73)	8.6 (7.19)
Uncued stimuli (talker C)	12.3 (7.72)	11.7 (7.81)	5.2 (4.83)	7.4 (6.52)

for the cued condition, and 365 ms for the uncued condition, respectively. Since the crucial analysis lies in the interaction between Phase (early-training vs. test), Object (target vs. competitor) and Stimulus (cued vs. uncued), the length difference across the three talkers could not confound the results:

Statistical analyses were performed using hierarchical log-linear models. Participants and items were entered, separately, as factors in the computation of partial association Likelihood Ratio Chi-Squares (LRCS1 and LRCS2, respectively) in order to assess the generalisability of the effects across participants and items.

In the early-training phase, there was no significant interaction between Object and Stimulus type (Odds Ratio (Target/Competitor): cued – 0.89; uncued – 1.06). However, in the test phase, the interaction was significant (LRCS1 = 8.005, df = 1, $p = .005$; LRCS2 = 15.675, df = 1, $p < .001$), consistent by participants and items (LRCS1 = 29.917, df = 47, $p = .975$; LRCS2 = 20.887, df = 19, $p = .343$). Planned comparisons indicated more looks to the target objects than the competitor objects for the cued stimuli in the test phase (LRCS1 = LRCS2 = 9.087, df = 1, $p = .003$),[5] consistent by participants and items (LRCS1 = 36.445, df = 47, $p = .867$; LRCS2 = 18.568, df = 19, $p = .485$) (Odds Ratio (Target/Competitor): 1.62). The difference between the two objects was marginally significant for the uncued stimuli *in the opposite* direction to that in the test phase (LRCS1 = LRCS2 = 3.663, df = 1, $p = .056$), consistent by participants and items (LRCS1 = 40.202, df = 47, $p = .748$; LRCS2 = 17.858, df = 19, $p = .532$) (Odds Ratio (Target/Competitor): 0.69).

The overall 3-way interaction – Phase (early-training vs. test) × Object (target vs. competitor) × Stimulus type (cued vs. uncued) – was significant by both participants and items (LRCS1 = 7.217, df = 1, $p = .007$; LRCS2 = 10.943, df = 1, $p = .001$), consistent by participants and items (LRCS1 = 30.980, df = 47, $p = .965$; LRCS2 = 19.310, df = 19, $p = .437$). Main effects of Phase and Stimulus type were both significant (*Phase*: LRCS1 = 72.668, df = 1, $p < .001$; LRCS2 = 72.668, df = 1, $p < .001$; *Stimulus type*: LRCS1 = 46.737, df = 1, $p < .001$; LRCS2 = 46.737, df = 1, $p < .001$).

3. Discussion

The data suggest that listeners *can* learn the association between the talker's identity and his/her tendency to produce one sentence structure over another. Previous data have shown listeners' ability to learn talker-specific acoustic characteristics (e.g., Creel et al., 2008; Eisner & McQueen, 2005; Goldinger, 1996, 1998; Kraljic & Samuel, 2006; McLennan & Luce, 2005); the novel finding here is that talker-specific adaptation can manifest as a learnt relationship between talker identity (extracted from acoustic cues) and an *abstract* non-phonological representation (structural attachment): Thus, and depending on the talker for that specific trial, either one representation, or another, is preferred. Crucially, and in contrast with Creel et al.'s (2008) study, we found that talker-adaptation generalised to novel stimuli, with new verbs and themes (e.g., '... ride the motorcycle/carousel ...' in training vs. '... taste the sweets/beer ...' at test). The present results could further indicate that abstract properties, such as structural preferences, of individual talkers' utterances may serve as a similarly informative role as acoustic properties in auditory language processing.

As discussed in the introduction, the tuning hypothesis needs to be modified to accommodate our findings that listeners are sensitive to the distributional information abstracted across individual talkers; the overall preference for a certain syntactic structure may be the result of the structure being preferred by more talkers in the given language, but listeners can apparently 'flip' their syntactic preference depending on the identity of the talker. Similarly, and in respect of syntactic priming, this ability to expect one structure or another as a function of talker-specific contingencies suggests a more complex picture of syntactic priming than has been presented thus far. Specifically, our data suggest that the expectation of one structure or another is not determined solely by which structures have been heard just beforehand, but in addition, by which structures have been heard from which talker just beforehand (note that, in our study, 'learning' was measured in terms of predictively looking at the appropriate object (according to the given talker) before the object was mentioned in the sentence).

This last observation can be restated more generally as a form of contextual dependence, in which expectation or anticipation is driven by contextual cues. In the original Altmann and Kamide (1999) study, anticipatory eye movements were driven by verb-based information pertaining to what kinds of object might be expected to be referred to after a verb such as 'eat'. In Kamide et al. (2003), they were driven by the combination of the verb and its subject. In Altmann and Kamide (2009), they were driven by the combination of the contents of the current sentence and the discourse-mediated *situation* (event structure). In each case, constraints on anticipation were derived from whichever lexical, sentential, or discourse cues were informative.

[5] The simple effect tests are based on marginal associations that do not distinguish between participant and item analyses – a constraint on the way in which SPSS calculates LRCS.

Here, we show that to the extent that talker identity is informative, it too can function as a constraint on anticipation. Whereas these previous studies explored anticipation in respect of which objects might be referred to next, here, we explore it in respect of which grammatical dependencies might be intended. Anticipation is thus not restricted to words, objects or referring expressions, and nor is it constrained solely by linguistic or discourse information; as observed in Altmann and Kamide (2009), it is constrained also by the situation, and talker identity, as an episodic component of that situation, can also constrain what can or cannot be anticipated.

It remains unclear which acoustic–phonetic features may have contributed to our effects. The three talkers had distinctive acoustic characteristics aside from their voice quality; the two English talkers belonged to different genders, and the two female talkers had distinctive accents (one English, the other Scottish). Also there was considerable variation in speech rate (see above). One interesting question to consider is how these properties interact in the learning process. Are certain properties weighted more heavily than others? A further question is whether the artificial 'attachment styles' we manipulated in this study might be reflected in sociolinguistic differences amongst individuals: Do some individuals tend to produce more high attachments, or even nested structures, for example, than others? And do comprehenders *naturally* learn such styles? Here, we have shown how, if such styles exist, comprehenders *may* be sensitive to them. It remains to be determined whether, 'in the wild', equivalent individual styles, and associated sensitivities exist.

In sum, we have shown that listeners can learn to use acoustic variability across different talkers to disambiguate attachment ambiguity after exposure to idiosyncratic differences in the kinds of attachment each talker produces. Although listeners have been shown to deploy talker-specific characteristics to resolve ambiguities at the phonological level, this is the first demonstration that such characteristics can be deployed in service of abstract syntactic processing.

Acknowledgements

The author would like to thank Susan Hainey for her help in collecting the eye-tracking data reported in this paper. The author is also grateful to Alissa Melinger, and three anonymous reviewers, for their helpful comments on an earlier draft of this paper. The research was partially supported by an award from the Nuffield Foundation (SGS/33694).

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*, 247–264.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111*, 55–71.

Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology, 54*, 218–250.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition, 106*, 633–664.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition, 30*, 73–105.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67*, 224–238.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 22*, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133–156.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*, 262–268.

Ladefoged, P. (1980). What are linguistic sounds made of? *Language, 56*, 485–502.

Ledoux, K., Traxler, M. J., & Swaab, T. Y. (2007). Syntactic priming in comprehension: Evidence from event-related potentials. *Psychological Science, 18*, 135–143.

McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory & Cognition, 31*, 306–321.

Mitchell, D. C., & Brysbaert, M. (1998). Challenges to recent theories of crosslinguistic variation in parsing: Evidence from Dutch. In D. Hillert (Ed.), *Syntax and semantics: A crosslinguistic perspective*. San Diego, CA: Academic Press.

Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research, 24*, 469–488.

Mitchell, D. C., & Cuetos, F. (1991). The origins of parsing strategies. In C. Smith (Ed.), *Current issues in natural language processing*. Austin: University of Texas, Center for Cognitive Science.

Monson, R. B., & Engebretson, A. M. (1977). Study of variations in the male and female glottal wave. *Journal of the Acoustical Society of America, 62*, 981–993.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175–184.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic persistence in written language production. *Journal of Memory & Language, 39*, 633–651.

Scheepers, C., & Crocker, M. W. (2004). Constituent order priming from reading to listening: A visual-world study. In M. Carreiras & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond*. New York: Psychology Press.