Cross-Situational Statistical Learning of Phonologically Overlapping Words

Paola Escudero (paola.escudero@uws.edu.au)
MARCS Institute, University of Western Sydney
Locked Bag 1797, Penrith NSW 2751, Australia

Karen Mulak (k.mulak@uws.edu.au)
MARCS Institute, University of Western Sydney
Locked Bag 1797, Penrith NSW 2751, Australia

Haley Vlach (hvlach@wisc.edu)
Department of Educational Psychology, University of Wisconsin, Madison
1025 W. Johnson Street, Madison, WI 53706, USA

Abstract

Recent research has sought to examine how learners are able to track the co-occurrence of words and objects across moments in time, a behavior commonly termed cross-situational statistical learning. The current experiment was designed to examine if learners can simultaneously determine word-referent pairings while engaging in other cognitive processes that support language learning, such as distinguishing phonologically overlapping words. Participants were presented with a cross-situational statistical learning task with pairs of words in four categories: non-minimal pairs, near minimal pairs, vowel minimal pairs, and consonant minimal pairs. The results revealed that participants were able to simultaneously learn word-referent pairings while distinguishing all four categories of word pairings. However, learners experienced the most difficulty learning vowel minimal pairs. This work demonstrates that learners are able to simultaneously engage in multiple cognitive processes that support language learning.

Keywords: cross-situational statistical learning; statistical learning; word learning; phonologically minimal pairs; bilingualism

Introduction

In one moment in time, the world presents learners with a seemingly infinite amount of information. Across several fields of study, including cognitive psychology, developmental psychology, computer science, and linguistics, a large research pursuit has been to characterize how it is that learners acquire, store, and later retrieve such a large data set of information. Indeed, this task has historically been characterized as theoretically impossible (e.g., Quine, 1960), but yet learners appear to acquire a great deal of information with ease.

A more recent trend in research has been to examine how it is that learners acquire, store, and later retrieve information across several moments in time. For example, in the domain of language learning and development, research has sought to determine how learners resolve ambiguity in word-referent pairings across moments in time. This phenomenon is most commonly termed *cross-situational* or *statistical word learning* (e.g., Fazly, Alishahi, & Stevenson, 2010; Frank, Goodman, & Tenenbaum, 2009; Smith & Yu, 2008; Vlach & Sandhofer, 2011; Yu & Smith, 2007, 2011).

In a typical experiment, learners are presented with a series of ambiguous learning events, which include multiple words and multiple objects. After a series of learning events, adult participants are presented with a forced-choice test in which they are asked to infer object-label pairings, while infants are presented with a preferential-looking task. This body of work has revealed that infants (e.g., 12- and 14-month-olds; Smith & Yu, 2008) and adults (e.g., Yu & Smith, 2007) are able to learn word mappings by tracking co-occurrence probabilities across learning events.

Cross-situational statistical learning research has focused on questions examining learners' ability to determine word-referent pairings. However, in real-world language learning environments, learners are faced with the challenge of determining word-referent pairings while simultaneously engaging in other cognitive processes that support language learning. For example, learners must simultaneously determine word-referent pairings while parsing words that overlap phonologically.

To date, experiments have primarily used words that contain gross phonological differences, that is, words that differ in multiple sound segments, such as "beat" and "rule". However, many words, especially in English, contain the same sounds with the exception of one segment, either a vowel or a consonant. In other words, they form phonologically minimal pairs such us "beat"-"bit" or "bet"-"debt". Consequently, it is unknown whether learners are able to simultaneously learn cross-situational statistics while distinguishing phonologically minimal pairs.

Adults have difficulty in learning phonologically minimal pairs. For example, Dutch and Spanish listeners were presented with a word learning task in which they were explicitly taught twelve pseudo-words together with their corresponding visual referents (Escudero, Broersma, & Simon, 2012). The words followed Dutch phonotactic

probabilities and were produced by a Dutch female speaker. Their visual referents were pictures of novel objects. At test, the native Dutch listeners made more errors for words that formed a minimal pair (e.g. "pax"-"pix") than when they formed a non-minimal pair (e.g. "beeptoe"-"pix"). Spanish listeners demonstrated an even greater difficulty in this task for minimal pairs that contained Dutch vowel contrasts that are not present in Spanish (e.g. "piex"-"pix", "pax"-"paax").

Can learners simultaneously learn cross-situational statistics and distinguish phonologically overlapping words? The current study examined whether phonologically overlapping words or minimal pairs can be successfully learned within a typical cross-situational statistical learning paradigm. In this experiment, learners were exposed to eight novel English words and eight picture referents with no explicit instructions. To examine the effect of word-pair similarity on word learning, the experiment presented learners with monosyllabic words such as "bon" and "deet" that when paired, formed four different levels of phonological overlap: (1) *non-minimal pairs* (nonMP), (2) *near minimal pairs* (nearMP), (3) *vowel minimal pairs* (vowelMP), and, (4) *consonant minimal pairs (*consMP).

Method

Participants

Participants were 71 undergraduates at the University of Western Sydney. A language background questionnaire revealed that 31 participants were monolingual English speakers, whose age range was 17.85 years to 52.19 years (M = 26.52 years, SD = 10.21 years; 10 males), while 40 participants spoke two or more languages and ranged in age from 17.73 years to 28.94 years (M=20.70 years, SD = 3.18; 7 males). English was the dominant language of all participants.

Stimuli

Eight monosyllabic nonsense words were recorded by a female native speaker of Australian English. Figure 1 shows the eight spoken words (in phonetic symbols) together with their randomly assigned picture referents. Four of the words were minimally different in their first consonant (left), while the other four differed in their vowel (right).

The novel words followed English phonotactic probabilities and were chosen from those included in previous studies with infant learners (see Curtin et al., 2009 for the words differing in vowels, and; Fikkert, 2010 for those differing in consonants). The female speaker produced a number of tokens of each word with child-directed intonation contours. These words and speech style were chosen to enable direct comparison of adult and infant responses to the same stimuli (Escudero, Mulak & Vlach in preparation-a).

Two tokens of each of the eight spoken words were selected to be used in the experiment such that intonation contours were comparable across words. The visual referents for the words were colorful pictures of novel items previously used in studies of cross-situational word learning (e.g., Vlach & Sandhofer, 2011).

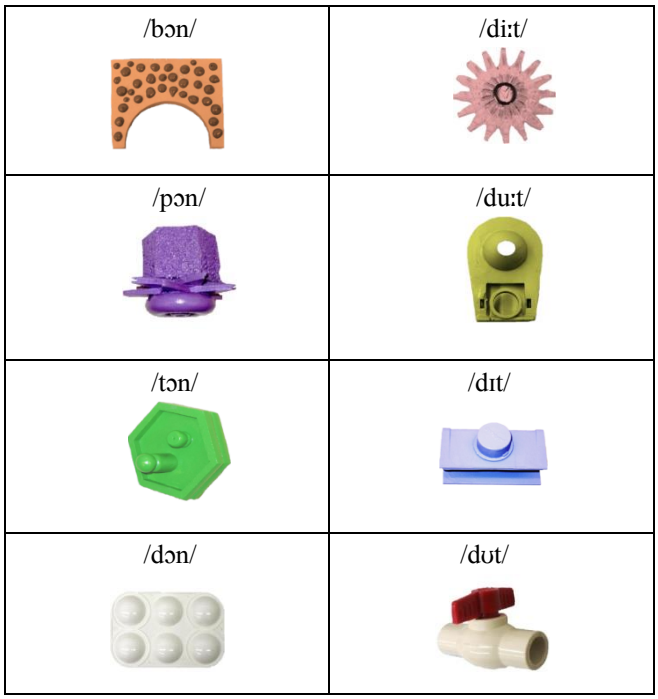

Figure 1. The eight novel words and their novel object referents.

Stimuli were presented in pairs, with four types of phonological overlap between the two spoken words that were the names of the pictures within a pair: (1) *non-minimal pairs* (nonMP), where the two words in the pair differed in all three sounds (e.g. /dɪt/-/pɔn/); (2) *near minimal pairs* (nearMP), where the words overlapped in one sound (e.g. /dɔn/-/dɪt/); (3) *vowel minimal pairs* (vowelMP), where the words only differed in their vowel (e.g. /dɪt/-/dit/), and, (4) *consonant minimal pairs (*consMP), where the words differed in only their consonant (e.g. /bɔn/-/dɔn/).

Procedure

Participants were presented with the cross-situational learning tasks in two phases: a learning phase and a subsequent test phase.

During the learning phase, stimuli were presented via Tobii Studio on a 17-inch screen, and the spoken words played from two speakers positioned below the screen. In each learning trial, two of the eight pictures of novel items appeared on the screen while two novel words for the pictures were spoken, such that pictures were either named left to right, or right to left. The pairings of the words and pictures were randomly assigned. The word for each picture was played once with 500 ms between them.

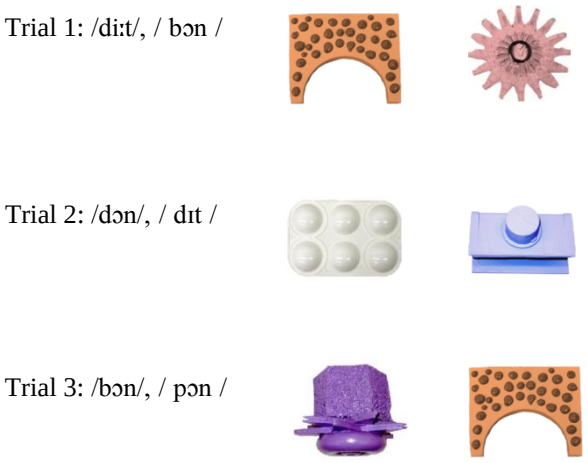

Figure 2. Examples of the word learning trials.

Participants were instructed to watch the pictures and listen to the words and were not told that the words were names for the pictures, nor were they asked to try and discover which word was associated with which picture.

Across the learning phase, there were a total of 36 learning trials, presented in a counterbalanced order. As mentioned above, stimuli were presented in pairs, with each trial consisting of two different pictures and two different words. There were 18 nonMP trials, and 6 each of nearMP, vowelMP, and consMP trials presented during the learning phase. Figure 2 shows examples of nonMP, nearMP and consMP trials, respectively.

After the learning phase, two testing phases were presented, though only one is reported here. In the first testing phase, stimuli were presented through Tobii Studio, as in the training phase. This phase followed immediately after the testing phase and participants were not given any additional instruction. Participants' eye-gaze was recorded without them having to make any overt response. This was done to later compare these adult data to infant eye-gaze to the same training and testing trials (Escudero, Mulak & Vlach in preparation-a).

Here we report the results of the second testing phase which was performed immediately after the first. During the second test phase, participants performed a forced-choice inference test, which required learners to infer word-picture pairings by clicking on the corresponding computer key. Stimuli were presented through a laptop computer with a 15-inch monitor, which was set up next to the monitor for the training and first test phase. Stimuli presentation was controlled with E-Prime and participants listened to the stimuli through headphones.

During the test phase, participants saw a pair of pictures and heard four repetitions of the word that always co-occurred with one of the pictures during the learning phase. The word was presented using two alternating repetitions of the same two tokens of that word used in the training phase, with a 500 ms interval between repetitions. Participants were asked to select whether the word corresponded to the left or right picture. There were 36 test trials in total with the same picture pairs as in the training, but the left/right positions of the pictures were randomized once for the test trials.

Results

The current experiment sought to determine if learners could simultaneously acquire cross-situational statistics in order to learn word-referent pairings, and parse phonologically minimal pairs. Figure 3 shows the percentage of correct word-referent pairings chosen during the testing trials, separately for the four different types of phonologically overlapping pairs. Percentages for monolingual and multilingual participants are presented separately. This is because it has been shown that bilingualism affects language processing, especially word retrieval (Fennell, Byers-Heinlein, & Werker, 2007; Bialystok, Craik, & Luk, 2008)

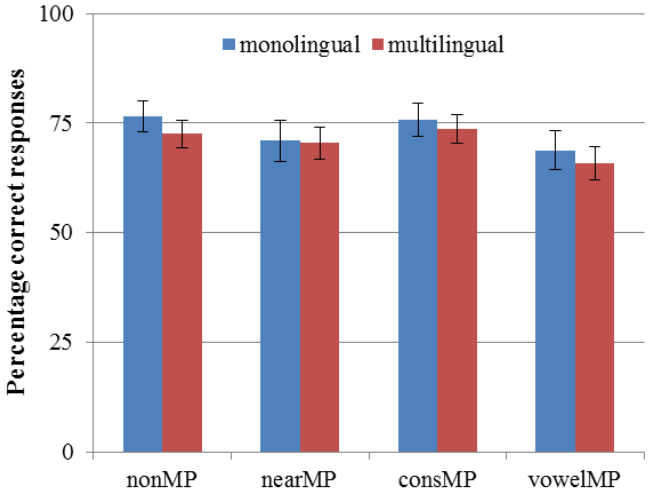

Figure 3. Percentage of correct word-referent pairings for the different pair types in monolingual and multilingual listeners.

The first set of analyses examined learners' overall performance on the testing trials. Accuracy was above chance for all pair types and in both participant groups (M = 65-76%, t = 15-23, all $ps < .001$). These results suggest that, despite the additional challenge of distinguishing phonologically overlapping words, learners

are able to learn and infer word-referent pairings during cross-situational statistical learning.

Further, a repeated measures ANOVA on the percentage of correct word-referent pairings chosen on testing trials, with pair type as the within-subject factor and language group as a between-subjects factor, revealed a main effect of pair type ($F(3, 69) = 3.009$, $p = .031$), which indicates that both groups of listeners had lower performance for some pair types than for others. Neither the main effect of language group ($F(1, 69 = 0.360$, $p = .550$) nor the interaction of pair type * language group ($F(1, 69 = 0.31$, $p = .942$) yielded statistical significance. These results suggest that, although there were not differences across groups of learners, overall participants' performance differed across the word pair types.

To follow up the effect of pair type, a series of planned comparisons were conducted between the pair types, with Bonferroni corrections for multiple comparisons: four comparisons (2-tailed). The results of these tests revealed that participants were less accurate on vowelMP than on nonMP ($t(70) = -2.53$, $p = .014$) and consMP ($t = -2.44$, $p = .017$), while no difference was found between consonants and nonMP (t (70) = 0.189, $p = .850$) or nonMP and nearMP ($t(70) = 1.468$, $p = .147$). In sum, learners demonstrated the lowest performance on the vowel minimal pairs.

Discussion

The results demonstrate that young adults can successfully learn monosyllabic nonsense words in a statistical cross-situational paradigm and without explicit instruction of word-referent pairings. Specifically, learners are able to simultaneously acquire cross-situational statistics and parse phonologically minimal pairs when learning novel words. Thus, the present study extends the findings of previous cross-situational studies (e.g., Fazly et al., 2010; Frank et al., 2009; Smith & Yu, 2008; Vlach & Sandhofer, 2011; Yu & Smith, 2007) by demonstrating that a more challenging set of word-referent pairings can still be learned through the tracking of co-occurrence and statistical probabilities.

The current experiment also demonstrates that vowel minimal pairs are more difficult to learn because participants' accuracy for vowel minimal pairs was lower than that of non-minimal and consonant minimal pairs. This finding is consistent with the numerous studies that demonstrate that consonant information is more important than vowel information for lexical processing (e.g., Berent & Perfetti, 1995; Lee, Rayner, & Pollatsek, 2001; Perea & Carreiras, 2006; Perea & Lupker, 2004) and lexical acquisition (Bonatti, Peña, Nespor, & Mehler, 2005; Nazzi & New, 2007; Nazzi, 2005; Nespor, Peña, & Mehler, 2003; Peña, Bonatti, Nespor, & Mehler, 2002).

The above line of research has proposed that the main role of consonants is to signal word meaning, while vowels enable the identification of rhythm and syntactic structure (Nespor et al., 2003). Additionally, consonant information is more critical in accessing the whole word form (see Berent & Perfetti, 1995; Carreiras, Vergara, & Perea, 2007; Lee et al., 2001; Lee, Rayner, & Pollatsek, 2002; Perea & Carreiras, 2006; Perea & Lupker, 2004). For example, in an experiment using response time and electrophysiological measures, Carreiras et al. (2009) demonstrated that a delay in the presentation of consonant information is more detrimental for lexical processing than a delay in presentation of vowel information.

However, studies have also shown vowel information to be more important than consonant information when identifying words in fluent speech (Cole, Yan, Mak, Fanty, & Bailey, 1996; Kewley-Port, Burkle, & Lee, 2007). In Kewley-Port et al. (2007), the vowels or consonants were removed from sentences produced in fluent speech and it was found that vowel information had a 2:1 benefit over consonant information for both young normal-hearing listeners and elderly hearing-impaired listeners.

The authors argue that the reason why they find opposite results to those of the studies described above is because linguistic processing of monosyllables relies on sound-by-sound, bottom-up information, while sentence intelligibility tasks incorporate considerable predictive information from top-down processing. Thus, in the context of fluent speech, we may have observed a different pattern of results. Future research should examine how acquiring cross-situational statistics and distinguishing minimal word pairs may differ in the context of fluent speech streams.

In that respect, Curtin et al. (2009) demonstrated that in lexical acquisition, infants can learn some vowel minimal pairs earlier than consonant minimal pairs, which suggests vowels may have a more lexical role than consonants in early word learning. However the authors explain that different task demands may cause the contradictory results. For example, Nazzi (2005) and Nazzi and New (2007), who found contrasting results, used a task in which infants were presented information from a real speaker, with multiple labels in the interactive communication. These task demands may thus be very different from the ones in the explicit word-referent association task in Curtin et al. (2009). Interestingly, Giezen, Escudero & Baker (under review) suggest that these divergent results may have a developmental nature, since they found more successful vowel than consonant minimal pair learning in children, while adults exhibited the opposite bias.

Ongoing research (Escudero, Mulak & Vlach, in preparation-a) examines infant word learning abilities using the same cross-situational word learning task as that of the present study. The results of this new study will likely shed light on the differential processing of vowels versus consonants across development.

The lack of group effects in the present study suggests that multilingualism does not influence cross-situational word learning and word retrieval immediately after learning. Interestingly, it runs contrary to studies that have demonstrated a negative influence of multiple language activation on word learning (Fennell et al., 2007) and

retrieval (Bialystok, Craik, & Luk, 2008), and a positive influence on cognitive control (Bialystok et al., 2008; Bialystok & Martin, 2004). Given that a bilingual processing advantage has been shown across a wide range of problem types, including both verbal and nonverbal domains, the null effect in the present study may come as a surprise. However, although word learning within a cross-situational paradigm involves intricate statistical computations and a high load on short-term memory (see Vlach & Sandhofer, 2011, for a discussion), the 2x2 (two pictures and two spoken words per trial) learning condition may have not provided enough challenge to observe differences. It may be the case that, in the context of a cross-situational learning task with higher working memory demands, the differences between monolingual and multilingual learners will emerge.

Ongoing research (Escudero, Mulak & Vlach, in preparation-b) is being conducted using tasks that present many words and objects in each learning event, in turn taxing working memory (e.g., in the context of 3x3 and 4x4 learning conditions). The results of this new study will likely reveal the influence of multiple language activation and cognitive control on the learning of phonologically overlapping word pairs.

On a final note, it is important to highlight that this study demonstrates the incredible capacity that human learners possess for learning language. Mapping new words to referents in the world has historically been characterized as a theoretically impossible task (e.g., Quine, 1960). However, the results of the current work demonstrate that learners can map words to referents in the world while simultaneously distinguishing phonologically overlapping sounds into words. Indeed, learners appear to accomplish multiple challenging cognitive tasks at the same time. Future research should continue to examine the cognitive processes that operate in parallel in order to support language learning and development.

Acknowledgments

This research was funded by MARCS Institute start-up funds awarded to the first author. We would like to thank the staff at the MARCS BabyLab for their support and help during testing.

References

Berent, I., & Perfetti, C. A. (1995). A rose is a REEZ: The two-cycles model of phonology assembly in reading English. *Psychological Review*, *102*(1), 146–184. doi:10.1037/0033-295X.102.1.146

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 859–873. doi:10.1037/0278-7393.34.4.859

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Developmental Science*, *7*(3), 325–339. doi:10.1111/j.1467-7687.2004.00351.x

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, *16*(6), 451–459. doi:10.1111/j.0956-7976.2005.01556.x

Carreiras, M., Vergara, M., & Perea, M. (2007). ERP correlates of transposed-letter similarity effects: are consonants processed differently from vowels? *Neuroscience letters*, *419*(3), 219–224. doi:10.1016/j.neulet.2007.04.053

Carreiras, M., Vergara, M., & Perea, M. (2009). ERP correlates of transposed-letter priming effects: The role of vowels versus consonants. *Psychophysiology*, *46*(1), 34–42. doi:10.1111/j.1469-8986.2008.00725.x

Cole, R. A., Yan, Y., Mak, B., Fanty, M., & Bailey, T. (1996). The contribution of consonants versus vowels to word recognition in fluent speech. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings* (Vol. 2, pp. 853 –856 vol. 2). doi:10.1109/ICASSP.1996.543255

Curtin, S. A., Fennell, C., & Escudero, P. (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Developmental Science*, *12*(5), 725–731. doi:10.1111/j.1467-7687.2009.00814.x

Escudero, P., Broersma, M., & Simon, E. (2012). Learning words in a third language: Effects of vowel inventory and language proficiency. *Language and Cognitive Processes*, *iFirst*, 1–16. doi:10.1080/01690965.2012.662279

Escudero, P., Mulak, K. & Vlach, H. (in preparation-a). Infants learning of phonological overlapping words through cross-situational statistics: An eye-tracking sudy.

Escudero, P., Mulak, K. & Vlach, H. (in preparation-b). Cross-situational learning of phonologically overlapping words with different levels of cognitive demand: Monolingual versus multilingual learners.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, *34*(6), 1017–1063. doi:10.1111/j.1551-6709.2010.01104.x

Fennell, C. T., Byers-Heinlein, K., & Werker, J. F. (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child Development*, *78*(5), 1510–1525. doi:10.1111/j.1467-8624.2007.01080.x

Fikkert, P. (2010). Developing representations and the emergence of phonology: evidence from perception and production. In C. Fougeron, B. Kühnert, & M. D'Imperio (Eds.), *Laboratory Phonology 10: Variation, Phonetic Detail and Phonological Representation* (Vol. 10, pp. 227–258). Berlin: De Gruyter Mouton.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early

cross-situational word learning. *Psychological Science*, *20*(5), 578–585. doi:10.1111/j.1467-9280.2009.02335.x

Giezen, M., Escudero, P., & Baker, A. E. (under review). Rapid learning of minimally different words in five- to six year old children: Effects of hearing impairment and sound perception.

Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *122*(4), 2365–2375. doi:10.1121/1.2773986

Lee, H.-W., Rayner, K., & Pollatsek, A. (2001). The Relative Contribution of Consonants and Vowels to Word Identification during Reading. *Journal of Memory and Language*, *44*(2), 189–205. doi:10.1006/jmla.2000.2725

Lee, H.-W., Rayner, K., & Pollatsek, A. (2002). The processing of consonants and vowels in reading: Evidence from the fast priming paradigm. *Psychonomic Bulletin & Review*, *9*(4), 766–772. doi:10.3758/BF03196333

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: differences between consonants and vowels. *Cognition*, *98*(1), 13–30. doi:10.1016/j.cognition.2004.10.005

Nazzi, T., & New, B. (2007). Beyond stop consonants: Consonantal specificity in early lexical acquisition. *Cognitive Development*, *22*, 271–279. doi:10.1016/j.cogdev.2006.10.007

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, *2*.

Pater, J., Stager, C., & Werker, J. F. (2004). The perceptual acquisition of phonological contrasts. *Language*, *80*(3), 384–402. doi:10.1353/lan.2004.0141

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*(5593), 604–607. doi:10.1126/science.1072901

Perea, M., & Carreiras, M. (2006). Do transposed-letter effects occur across lexeme boundaries? *Psychonomic bulletin & review*, *13*(3), 418–422. doi:10.3758/BF03193863

Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, *51*(2), 231–246. doi:10.1016/j.jml.2004.05.005

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. doi:10.1016/j.cognition.2007.06.010

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381–382.

Vlach, H.A., & Sandhofer, C. M. (2011). Retrieval dynamics of in-the-moment and long-term statistical word learning. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 789-794). Boston, MA: Cognitive Science Society.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*(1), 1–30. doi:10.1207/S15327078IN0301_1

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420. doi:10.1111/j.1467-9280.2007.01915.x

Yu, C., & Smith, L. B. (2011). What you learn is what you see: using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*(2), 165–180. doi:10.1111/j.1467-7687.2010.00958.x