The Meaning-Sound Systematicity Also Found in the Korean Language

Hana Jee (Hana.Jee@ed.ac.uk)
School of Philosophy, Psychology and Language Sciences,
The University of Edinburgh, Edinburgh, UK

Monica Tamariz (m.tamariz@hw.ac.uk)
Psychology, Heriot-Watt University,
Edinburgh, UK

Richard Shillcock (rcs@inf.ed.ac.uk)
School of Philosophy, Psychology and Language Sciences,
The University of Edinburgh, Edinburgh, UK

Abstract

Recent studies of meaning-sound systematicity have consistently found a small but significant positive correlation between semantics and phonology. The current study adds further evidence from an etymologically distinct language, Korean. Through multiple methods, the study shows that similar sounds tend to have similar meanings in Korean monosyllables. Several cultural aspects of the language are also quantified. Pure Korean words return stronger meaning-sound correlation than Sino-Korean words, which is attributable to the higher portion of homonyms in Sino-Korean. The most frequent words show the strongest systematicity, which permeates all of the monosyllables. Certain types of vowels seem to contribute to this effect.

Keywords: systematicity, meaning-sound mapping, Korean, homonymy

Introduction

The phonology of words seems to correlate with their meaning (Blasi, Wichmann, Hammarstrom, Stadler, & Christiansen, 2016; Dautriche, Mahowald, Gibson, & Piantadosi, 2017; Monaghan, Shillcock, Christiansen, & Kirby, 2014; Tamariz, 2008). The tendency has been observed in various languages (Blasi et al., 2016; Dautriche et al., 2017) and becomes stronger in the case of early acquired words (Monaghan et al., 2014), or universal, basic words (Blasi et al., 2016). The possibility of different cultural influences on individual languages has also been suggested (Tamariz, 2008).

Korean has not been investigated in this regard, though. Korean is a unique language for many reasons. It used to be categorized as one of the Ural-Altaic languages along with Mongolian and Turkish (Ramstedt & Kim, 1979), but is increasingly considered to be a 'language isolate' with no language relatives (Georg, Michalove, Ramer, & Sidwell, 1999; Lee, 1972).

As an agglutinative language, Korean features polysyllabic roots with a complex system of suffixes that express different nuances (Sampson, 1985). The verbs have four different formalisms according to the relationship with the audience, which respectively conjugate in the past tense, suggesting mode, imperative, and willing/decision mode. Furthermore, they conjugate depending not only on the semantic function (for linking, for contrast, for assuming, and for purposes), but also on the part of speech (noun forms and adjective forms). For example, the verb 'go' has 19 different conjugated forms. The number of the cases becomes double if the forms to honour the elders are taken into account.

Korean also features vowel harmony where vowels in the same class co-occur. Not as strictly applied as in Middle Korean (15~16c, Kwon, 2018), vowel harmony is still observed in Modern Korean as phonotactics, in onomatopoeia, in predicate suffixes (Sohn, 2001) and in postpositions (Larsen & Heinz, 2012). Korean vowels are divided into three classes: the light vowels like /a/ or /o/ connote *light*, *bright*, and *small*; the dark vowels like /ʌ/ and /u/ connote *heavy*, *dark*, and *large*; the vowels that correspond to none of these two, like /i/ and /ɯ/, referred to as the neutral vowels (Kim-Renaud, 1976; Larson & Heinz, 2012).

Another historically interesting aspect of Korean is the substantial presence of Sino-Korean words, i.e. words that have Korean pronunciation but originate from Chinese. According to the National Institute of the Korean Language (2016), 57% of Korean lexical items in the dictionary are Sino-Korean, although pure Korean words take up a greater proportion in real usage (pure Korean 54% vs. Sino-Korean 35%; the remaining 11% are loan words). The Korean peninsula has been under Chinese influence for centuries. Although spoken Korean and Chinese are very different languages (each belongs to a different language family), written Chinese had always been the main means of communication among Korean intellectuals until Hangeul, the Korean orthography that was invented and promulgated in 1446. When Chinese words were introduced, their pronunciations had to be modified to suit Korean phonology, where neither the tone system nor the final sound /r/ exist. This eventually resulted in various Korean words with different meanings having the same pronunciation.

Park, Zhang, and Kim (2000) estimated that 90% of Korean words have some sort of ambiguity due to

homonymy. Take '개 /gae/' for instance. According to the standard Korean dictionary, there are 10 different meanings, which is not rare for Korean vocabulary. Each has a different root, but coincidentally sounds the same. The Sino-Korean meanings below are marked with Chinese characters.

개 1: Noun. The area where sea water flows in and out
개 2: Noun. The hive made to save honey and pollen where bees raise their larvae
개 3: Noun. The mammal of the family Canidae, a clever, friendly domestic animal
개 4: Noun. In 'a game of yut', when two flat sides are up
개 5: 介 Noun. A Korean surname
개 6: 疥 Noun. Itch
개 7: 個/箇/介 Bound noun. A unit for counting individual things
개 8: 蓋 Noun. The lid of a food container
개 9: Prefix. Wild low quality, quasi- but different
개 10: Suffix. The tool for doing such an activity

This issue becomes severe when it comes to two-syllable words (Kang, 2005). For example, '사고' /sa-go/, a Sino-Korean word, has 21 different meanings, which can all be distinguished in Chinese: 司庫, 史庫, 四苦, 四顧, 死苦, 私考, 事故, 社告, 思考, 思顧, etc. Another example, '연패' /jʌn-pæ/, has two exactly opposite meanings: successive winning (連霸) and successive losing (連敗).

The current study investigates whether meaning-sound systematicity exists in this etymologically, geographically, and historically unique language. If so, how similar or different is it to English (Monaghan et al., 2014), Spanish (Tamariz, 2008) and other languages (Blasi et al., 2016; Dautriche et al., 2017)? Addressing this question may provide not only more solid ground for exploring linguistic systematicity, but also an opportunity to observe any cultural influence on this systematicity.

Procedure

In line with previous research (Dautriche et al., 2017; Monaghan et al., 2014; Tamariz, 2008), we measured all the pairwise distances between phonological representations of words, and all the pairwise distances between word meanings. The correlation between two lists of distances indicated the level of form-meaning correlation.

Preparation

Using web scraping, the corpus was created based on the Korean internet content reflecting authentic, contemporary language use, and including various styles: spoken and written, short comments and long narration. We collected the data on 22 July 2019 (Jee, Tamariz & Shillcock, *in prep.*). The total number of word tokens was 28,858,796. (Further details are available from the first author)

As mentioned earlier, Korean predicates feature dozens of different suffixes and postpositions with an identical meaning. For language processing, these extremely varied phonological forms need to be categorized into the one relevant semantic form, and distinguished from homonyms. This is why morpheme tagging is particularly important but also very challenging in research on Korean natural language processing. What makes the problem worse is that online writers frequently omit spaces and punctuation, which confuses morpheme taggers. We compared multiple morpheme taggers from *Konlpy* (Park & Cho, 2014; ver. 0.5.2), the Python package for Korean NLP, in terms of performance and decided to use *Open Korean Text Processor* (Okt) which demonstrated moderately satisfying morphemic analysis.

Semantic distance between words

The meaning of a word can be defined by its context (Firth, 1957): the more contexts two words share, the more similar they tend to be. Word-embedding techniques can quantify the meanings of words: each word within its contexts is mapped onto a binary vector. Because one word is assigned to one vector, the vectors tend to be very lengthy and sparse, which has caused problems in Latent Semantic Analysis, for example (Landauer & Dumais, 2008). Applying a neural network approach, recent algorithms seem to work better in this regard. *Word2Vec* (Mikolov et al., 2013) trains the word vectors in such a way that each one updates itself based on the contexts in which it co-occurs. It does not, however, reflect morphology and cannot process novel words that it has not encountered during the training phase. This issue was resolved by *FastText* (Joulin, Grave, Bojanowski, & Mikolov, 2016; Mikolov, Grave, Bojanowski, Puhrsch, & Joulin, 2017). We compared these two word-embedding techniques.

We divided the corpus into identically-sized chunks—five words before and after each word, which makes the window size 11. A particular word's semantics is defined by the other words that appear in its window, for each token of that word. We trained *Word2Vec* and *FastText* with our corpus and calculated cosine similarity between every word pair. The procedure was conducted on *Google Colab* due to the large size of the data.

Phonological distance between words

The Korean basic alphabet set consists of 14 consonants and 10 vowels. However, the total numbers of the possible consonants and vowels are 30 and 21, respectively, when including the cases where those consonants are combined (e.g. ㄴ + ㅎ = ㄶ) or duplicated (e.g. ㄱ + ㄱ = ㄲ), as well as diphthongs. In total, 11,172 monosyllabic combinations are mathematically possible, but not all of them are usable (Choi, 2000). Because Korean phonology only allows 19 consonants for the initial position and 7 consonants for the final position, 3,192 syllables are actually possible to use (Byun, 2003). Among these, 66 (5%) syllables never appeared in Byun's (2003) contemporary spoken language

corpus, where he collected 537,245 syllables from contemporary spoken corpora. It was found that very small number of syllables are highly frequent in use. Only 138 syllables (35%) accounted for 95% of daily conversation and 26 syllables (6.5%) took up 50%. Considering this frequency in use, 315 monosyllabic words were constituted as our sample; 79 of them were CV and 236 were CVC.

We defined Korean phonemes by the location (hard palatal, soft palatal, labial, dental, and throat) and manner of articulation (plosive, affricate, fricative, fortis, lenis, aspirated, nasal, and flow). For vowels, the location of the tongue and roundness were considered. By marking 1 if a phoneme has the feature and 0 if it does not, each phoneme was transformed into a binary vector. For example, /b/ can be represented as [0,0,1,0,0,1,0,0,0,1,0,0,0], where the 1s represent the presence of the feature labial, plosive, lenis, respectively.

The difference between two vectors was measured by: feature edit distance, which counts how many different features there are between two vectors; Euclidean distance, which measures the shortest geometric distance between two vectors; Jaccard similarity, in which the number of shared features is divided by the total number of both features; and cosine similarity, which measures the inner angle between two vectors. It should be noted that the first two measure the distance, so the more different two vectors are, the larger the values are, whereas the last two measure the similarity, so the more similar two vectors are, the larger the values are.

To calculate the phonological distance between two monosyllables, the distance between the first consonants, the distance between the vowels, and the distance between the final consonants were combined (Monaghan, Christiansen, Farmer, & Fitneva, 2010). For all phonological distance measures, textdistance 4.1.4 was used (Python 3.7.1).

Results and Discussion

General findings

Pearson's r was calculated between the lists of distances, and a Monte Carlo permutation test was conducted to estimate the significance values. Small but robust correlations were found between Korean monosyllabic sounds and their meanings (Table 1). In general, the meaning similarity increases as the phonological similarity increases, which indicates that similar sounds tend to have similar meanings, as is the case in English and other languages (Blasi et al., 2016; Dautriche et al., 2017; Monaghan et al., 2014). Our r values are similar to those in previous studies (Dautriche et al., 2017; Monaghan et al., 2014; Tamariz, 2008).

Table 1 also shows that the sound-meaning correlation is consistently stronger for the 119 pure Korean monosyllables (7,140 pairs in total) compared with the Sino-Korean monosyllables. We suggest that this difference reflects the greater number of homonyms in Sino-Korean. When Chinese words were modified to suit Korean phonology, they became phonetically biased because the available phonemes for those Sino-Korean words were limited (Kim, 2001; Park, 2015), and tone was not available, drastically increasing the number of homonyms. These homonyms necessarily reduce systematicity in that identical phonological distances are paired with different multiple semantic distances. The number of homonyms for the pure Korean words were 4.62 on average whereas it was 11.12 for the Sino-Korean words.

Table 1: Meaning-sound correlation (all p-values < .00).

		Total	Pure	Sino
Meaning	Phoneme	r	r	r
Word2Vec	Cosine	.05	.10	.07
	Jaccard	.05	.10	.07
	feature edit	- .10	- .16	- .14
	Euclidean	- .11	- .18	- .15
FastText	Cosine	.04	.08	.05
	Jaccard	.04	.09	.05
	feature edit	- .07	- .14	- .15
	Euclidean	- .10	- .16	- .16

Note: The negative correlations are due to the opposite directions of the semantic and phonological distance measurements (similarity vs. distance).

It should be pointed out that Korean homonyms can be either pure-Korean or Sino-Korean (as shown above, e.g. '개'). This issue was not considered for the current study largely because our semantic measurements did not use WordNet-type definitions. Rather, calculating context vectors involves only the distribution of the set of 'context words' within the 11-word window of text surrounding each of the tokens of the word being studied. This is why the existence of many homonyms potentially disturbs semantic distances. For deeper analysis, the meanings of a pure Korean word need to be distinguished from the meanings of the identical Sino-Korean word, which is a formidable task. Unlike English (Stevenson, 2003), Korean homonyms cannot be successfully distinguished by part of speech tagging (Kang, 2005). This is because Korean homonyms tend to belong to the same part of speech. Frequency can help in distinguishing homonyms (Gahl, 2008, Kang, 2005). Through corpus analysis, Kang (2005) revealed that only one of the meanings is frequently used in reality, so people hardly consider it a homonym. However, there may be individual differences in the canonical or representative meanings (Kreuz, 1987). It should also be noted that the results from corpus linguistics are susceptible to the type of corpus used.

How do Koreans distinguish those homonyms without additional cues like tones? Future research may investigate whether they treat homonyms differently from non-homonyms (cf. Caramazza, Costa, Miozzo, & Bi, 2001). If dependency on the context matters (Li & Yip, 1998), it is useful to observe how much context they need and what are the essential features of the satisfying context. It is likely that

the homonyms hardly share their contexts. It would be informative to explore how Korean children acquire the homonyms.

Distinct role of consonants from vowels

Consonants and vowels contributed very differently to the meaning-sound systematicity. Unlike the consonants, the vowels negatively contributed to the correlations as a whole (Table 2). This is congruent with Tamariz (2008), who found that words sharing vowels tended to have different meanings in Spanish, while words sharing consonants tended to have similar meanings. Table 2 also shows that word-final consonants in Korean contributed the most to a positive meaning-sound correlation.

Table 2: Partial correlation of initial consonants, vowels, and final consonants (all p-value < .00).

Unit	Measure	r
Initial consonant	Word2Vec	.03
	FastText	.04
Vowel	Word2Vec	-.02
	FastText	-.03
Final consonant	Word2Vec	.09
	FastText	.07

Lexical Frequency

The sample was re-grouped by lexical frequency based on Byun's (2003) list. The monosyllables in his list constituted a subset (199 syllables) of our sample (315 syllables). Table 3 indicates that the most frequent Korean monosyllables returned the strongest meaning-sound correlation. The accumulated data consistently showed robust results whereas only the top 25% most frequent words did when analyzed by independent section. It is interesting to note that the systematicity of such a small group of syllables (top 25%) penetrates and influences the whole of the rest of the lexicon even though the systematicity is reduced compared with the most frequent subset. High frequency words act as a nucleus around which meaning-form systematicity coalesces.

Table 3: The meaning-sound correlation of separate and accumulated frequency groups

	By section	N	r	p
Frequent . . . Rare	25%	378	- .26	.00
	25~50%	231	.03	.68
	50~75%	703	- .01	.73
	75~95%	703	- .06	.13
	95~100%	136	.01	.92

	Accumulated	N	r	p
Frequent . . . Rare	25%	378	- .26	.00
	50%	1,225	- .10	.00
	75%	3,828	- .06	.00
	95%	7,875	- .05	.00
	100%	10,153	- .05	.00

Note: Semantic distance was measured by Word2Vec. Phonological distance was measured by feature edit distance. The negative coefficients are, again, attributed to the opposite directions of the methods.

To investigate whether the above tendency is due to the high portion of pure Korean vocabularies in the most frequent words, each frequency group was divided into pure and Sino-Korean. Counter to expectation, the pure Korean and Sino-Korean monosyllables were evenly distributed, at least in this sample (Figure 1), which implies that being pure-Korean is hardly related to frequency of use. This observation contrasts with Choi (2012), who collected all the vocabularies from the elementary school textbooks and reported that the proportion of Sino-Korean words is higher when they are simply counted but that of pure Korean words is higher when it comes to the accumulated frequency and actual use.

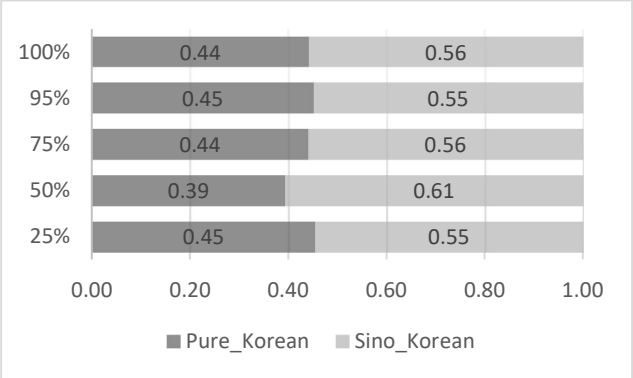

Figure 1: Proportion of accumulated pure and Sino-Korean syllables

Types of Vowels

In line with the vowel categorization of Kim-Renaud (1976), we examined whether the monosyllables have light vowels (/ɛ/, /ø/, /a/, /o/), dark vowels (/e/, /y/, /ʌ/, /u/) or neutral vowels (/i/, /ɯ/). Figure 2 suggests that the systematicity may be related to the nature of the vowels. The most frequent monosyllables consist of light and neutral vowels only, which may contribute to the meaning-form correlation in that the range of phonetic variation becomes narrow and dense. In addition, the fact that there are no diphthongs in this group demonstrates that high frequency words tend to be simple and easy to pronounce.

Naturally, most research on vowel harmony in Korean have dealt with two and three-syllable onomatopoeic words (Kwon, 2018; Larsen & Heinz, 2012), strategically excluding

monosyllables. As shown, however, the nature of vowels may additionally shed light on the relation with lexical frequency and the principle of least effort (Zipf, 1949).

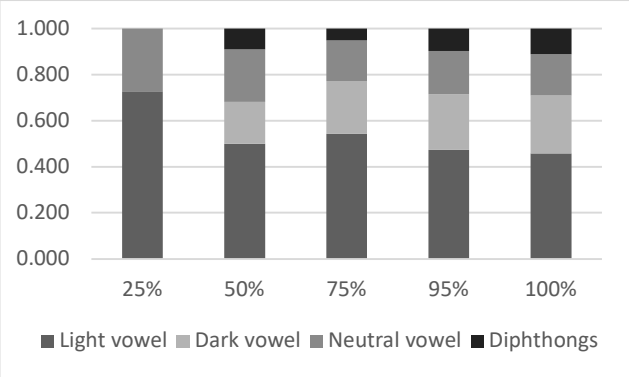

Figure 2. Proportion of light, dark, neutral vowels and diphthongs in each frequency group

Conclusion

Previous studies have found systematic relations between linguistic sub-systems: between syntax and phonology (Fitneva, Christiansen, & Monaghan, 2009; Kelly, 1992; Kelly, Morgan, & Demuth, 1996; Monaghan & Christiansen, 2008; Morgan & Demuth, 1996; Reali, Christiansen, & Monaghan, 2003; Shi, Morgan and Allopenna, 1998; Vendler, 1968); and between semantics and phonology (Blasi et al., 2016; Dautriche et al., 2017; Monaghan et al., 2011; Monaghan et al., 2014; Tamariz, 2008). These studies have demonstrated that each sub-system has its own rules, but, at the same time, there is some sort of order acting across them. The current study adds further evidence from Korean, a language with a unique background in connection with meaning-sound systematicity. Many of the results successfully confirmed those of previous studies, which implies that the positive correlation between semantics and phonology may be a universal feature of phonographic languages. The current study also showed that some cultural influences on languages can be quantified. The pure Korean vocabularies increased the meaning-form systematicity in Korean monosyllables, whereas Sino-Korean words make the language more arbitrary, presumably by producing a substantial number of homonyms. In addition, the positive meaning-sound correlation was strengthened by the distribution of the final consonants. Along with the previous finding in Spanish (Tamariz, 2008), the fact that Korean vowels negatively influence the total sound-meaning correlation suggests a general role of vowels as the meaning distinguisher. The most frequent monosyllabic words seem to act as a magnetic force that organizes the systematicity.

References

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, 113(39), 10818-10823.

Byun, S. W. (2003). Frequencies of Korean Syllables and the Distribution of Syllables of PB Word List. *Korean Journal of Otorhinolaryngology-Head and Neck Surgery*, *46*(9), 737-741.

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1430.

Choi, S. W. (2000). Some statistical properties and Zipf's law in Korean text corpus. *Journal of Quantitative linguistics*, *7*(1), 19-30.

Choi, W. S. (2012). A Study on the Meaning Based on the Ratio o Pure-Korean Words and Sino-Korean Words Occuring in Elementary School Textbooks. *Journal of Korean Language and Literature Education*, 44, 517-547.

Dautriche, I., Chemla, E., & Christophe, A. (2016). Word learning: Homophony and the distribution of learning exemplars. *Language Learning and Development*, 12(3), 231-251.

Dautriche, I., Mahowald, K., Gibson, E., & Piantadosi, S. T. (2017). Wordform similarity increases with semantic similarity: An analysis of 100 languages. *Cognitive Science*, 41(8), 2149-2169.

Firth, J. R. (1957). A synopsis of linguistic theory 1930–1955. *In Studies in linguistic analysis*, 1–32. Blackwell.

Fitneva, S. A., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language*, 36(5), 967-997.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474-496.

Georg, S., Michalove, P. A., Ramer, A. M., & Sidwell, P. J. (1999). Telling general linguists about Altaic. *Journal of Linguistics*, *35*(1), 65-98.

Jee, H., Tamariz, M. & Shillcock, R. (*in prep.*). A corpus of Korean internet text.

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). *Fasttext. zip: Compressing text classification models.* arXiv preprint arXiv:1612.03651.

Kang, B. M. (2005). Aspects of the use of homonyms. *Language Research,* 41(1), 1-29.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological review*, 99(2), 349.

Kelly, M. H., Morgan, J. L., & Demuth, K. (1996). The role of phonology in grammatical category assignments. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 249-262.

Kim, C. S. (2001). Creation of Sino-Korean and Phonotactics of Native Korean. *Journal of Korean Linguistics,* 37, 177-195.

Kim-Renaud, Y. K. (1976). Semantic features in phonology: evidence from vowel harmony in Korean. In *Papers from the... Regional Meeting. Chicago Ling. Soc. Chicago, Ill* (No. 12, pp. 397-412).

Kwon, N. (2018). Iconicity correlated with vowel harmony in Korean ideophones. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 1.

Kreuz, R. J. (1987). The subjective familiarity of English homophones. *Memory & Cognition*, *15*(2), 154-168.

Landauer, T. K., & Dumais, S. (2008). Latent semantic analysis. *Scholarpedia*, *3*(11), 4356.

Larsen, D., & Heinz, J. (2012). Neutral vowels in sound-symbolic vowel harmony in Korean. *Phonology*, 29(3), 433-464.

Lee, K. M. (1972). 改訂 國語史槪說, T'ap ch'ulp'ansa.

Li, P., & Yip, M. C. (1998). Context effects and the processing of spoken homophones. In *Cognitive Processing of the Chinese and the Japanese Languages* (pp. 69-89). Springer, Dordrecht.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *In Advances in neural information processing systems,* 3111-3119.

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017). Advances in pre-training distributed word representations. arXiv preprint arXiv:1712.09405.

Monaghan, P., & Christiansen, M. H. (2008). Integration of multiple probabilistic cues in syntax acquisition. *Trends in corpus research: Finding structure in data*, 139-63.

Monaghan, P., Christiansen, M. H., Farmer, T. A., & Fitneva, S. A. (2010). Measures of phonological typicality: Robust coherence and psychological validity. *The Mental Lexicon*, 5(3), 281-299.

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130299.

Morgan, J. L., & Demuth, K. (1996). Signal to syntax: An overview. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 1-22.

Park, E. L., & Cho, S. (2014, October). KoNLPy: Korean natural language processing in Python. In *Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology* (Vol. 6).

Park, N. Y. (2015). Comparative phonotactics of Native and Sino-Korean. Korean Linguistics 67, 147-192.

Park, S. B., Zhang, B. T., & Kim, Y. T. (2000). Word Sense Disambiguation From Unlabelled Data. *KIISE*, spring, 330-332

Ramstedt, G. J., & Kim, M. S. (1979). *Remarks on the Korean language*. T'ap ch'ulp'ansa.

Reali, F., Christiansen, M. H., & Monaghan, P. (2003, January). Phonological and distributional cues in syntax acquisition: Scaling up the connectionist approach to multiple-cue integration. *In Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 25, No. 25).

Sakuma, N., Sasanuma, S., Tatsumi, I. F., & Masaki, S. (1998). Orthography and phonology in reading Japanese kanji words: Evidence from the semantic decision task with homophones. *Memory & Cognition*, *26*(1), 75-87.

Sampson, G. (1985). Writing systems. *London, U K: Hutchinson*.

Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of child language*, 25(1), 169-201.

Sohn, H. M. (2001). *The korean language*. Cambridge University Press.

Stevenson, M. (2003). *Word Sense Disambiguation: The Case for Combinations of Knowledge Sources*, Stanford: CSLI.

Tamariz, M. (2008). Exploring systematicity between phonological and context-cooccurrence representations of the mental lexicon. *The Mental Lexicon*, 3(2), 259-278.

Vendler, Z. (1968). *Adjectives and nominalizations* (No. 5). Mouton.

Zipf, G. K. (1949). Human behavior and the principle of least effort.