Processing Function Words in Artificial Languages: Effects of Function Word Presence and Perceptual Cues

Dan Hufnagle (dghufnag@ucalgary.ca)
Department of Psychology, 2500 University Drive NW
Calgary, AB T2N 1N4 Canada

Suzanne Curtin (scurtin@ucalgary.ca)
Departments of Psychology and Linguistics, 2500 University Drive NW
Calgary, AB T2N 1N4 Canada

Abstract

Language learners use function words as indicators of structure. Across languages, function and content differ phonologically and acoustically, and learners may use these cues to help organize an emerging lexicon into categories. We tested three versions of an artificial language: no differences between function and non-function words, reduced vowels for function words, and no onsets for function words. Counter to expectations, learners of the language with no cues to category membership performed best on grammaticality judgments. At test, the presence of function words was also manipulated. There is an interaction between presence of function words in test sentences and the type of pattern being tested, indicating that function words help learners process some patterns, but are detrimental for others.

Keywords: Artificial language learning; function words; syntax acquisition.

Introduction

Function words and morphemes have a number of characteristics that make them special in language. They serve as the scaffold around which the meaning in a sentence is built. They organize sentences, providing structure within phrases as well as connections among phrases. Function words are a closed class: new function words rarely enter a language and constitute a small part of the lexicon relative to non-function words. Despite their low type frequency, function words have high token frequency.

Experiments using artificial languages have explored the role of function words in acquisition. Learners can form categories quite rapidly by calculating distributions of co-occurrences (Mintz, 2002) and can use those distributions to determine which elements are predictive of other elements (Saffran, 2001). When predictive elements in artificial languages act to reliably signal whole classes of other elements that occur in the language, they become analogous to function words in natural language. Thus, when learners have access to distributional information about co-occurrence, learners can induce categories in artificial languages that play the role of function words, even when words in those languages are not tied to meaning.

Artificial languages that include function words are learned better than languages that do not include predictive elements or languages with similar elements that are not predictive (Green, 1979). This advantage for predictive elements only holds when the relative frequency of predictive to non-predictive elements is high (Valian & Coulson, 1988): learners of an artificial language with a low ratio of function words to non-function word (mimicking natural language) learned the language faster and at a deeper level than learners who learned a language with a higher ratio of function to non-function words.

Whether or not function words are optional also affects acquisition. Hudson Kam and Newport (2005) exposed learners to a simple, meaning-based artificial language and used a number of tasks to measure the effects of the proportion of exposure sentences that contained function words. While the researchers framed the experiment in terms of regularizing patterns from inconsistent input (i.e., creolization), their experiments also show how people differ in learning languages with required versus optional function words if the input is taken as faithful to the target language. Adults' productions and sentence ratings mirrored the function word characteristics of the input language: when function words were presented in all contexts (i.e., required by the language), learners recognized the requirement. When function words were variable in the input, learners recognized the optionality, either dropping them altogether or used them variably. Clearly people can use function words, even when those words compete with meaningful words for cognitive resources.

Together, this research underscores the importance of the distributional properties of function words in the acquisition of language. Another question involves the use of function words in processing the grammatical structure of sentences. In a language with optional function words, does their presence in a given sentence help or hinder processing? Our hypothesis is that because function words act as markers of structure, sentences with more function words will be easier to process than sentences with fewer function words.

In order to test this hypothesis, we varied the presence and absence of function words within the test phase for each language in the experiment. If function words act as syntactic anchors, and people are able to learn which words of the grammar are function words, then ungrammatical sentences should be easier to detect in the trials that make greater use of function words, leading to higher scores on a grammaticality judgment task.

In addition to distributional information about co-occurrence, function words in natural language can be differentiated from lexical words along a number of phonological and acoustic dimensions (Shi, Morgan & Allopenna, 1995). In an examination of Mandarin and Turkish mothers' speech to their children, Shi and colleagues found that not only were frequency and position information useful distinguishing categories, but a number of phonological and acoustic dimensions, including vowel duration, syllable count and the presence of a syllable coda or diphthong, were informative when determining the category of a word. Monaghan, Chater and Christiansen (2005) found other phonological cues that distinguish function and non-function words in English, including word length in phonemes, presence of stress, and the presence and complexity of initial consonant clusters. However, the existence of cues does not require their use in learning the fundamental distinction between lexical and function words.

Confirmation of the idea that people use these cues lies in research that provides evidence that grammatical categories are distinguished before those categories can be differentiated on the basis of meaning. Early in language acquisition, children can distinguish function and lexical words. Six-month-old infants prefer listening to lexical words over function words, indicating early discrimination (Shi & Werker, 2001). At 8 and 11 months of age, segmentation is facilitated by the use of high frequency function words; however, at 8 months the segmental makeup of the function word is irrelevant (Shi, Cutler, Werker & Cruickshank, 2006). Between 11 and 13 months, infants notice when function words are altered by changing one segment to another, indicating the emergence of richly detailed phonological and grammatical representations of words that do not yet have meanings attached (Shi, Werker & Cutler, 2006).

In order to investigate how phonetic cues to grammatical category influence the manner in which people acquire and use function words, we tested languages that either included cues or lacked such cues. The hypothesis is that languages that include cues to function versus non-function category membership help learners to establish categories and acquire language patterns better. Essentially, the cues allow categories that are available immediately (e.g., words with initial consonants) to act as shortcuts to grammatical categories. To test this hypothesis, we compared how people learned a baseline language that includes no cues to word category, a language with shortened vowels in function words, and a language in which function words lack syllable onsets (initial consonants). These cues were chosen because they have correlates both in English and other languages (Monaghan et al., 2005).

It is crucial to note that methods are not currently available to judge directly whether adults group words together into higher-level categories (e.g., the broad category of function words). Rather, we test a variety of patterns that occur in the languages to determine indirectly how learners process specific function word categories.

Experiment

Method

Participants Seventy-six students enrolled in psychology courses at the University of Calgary participated in this experiment in exchange for course credit.

Materials To test our hypotheses about the acquisition and use of function words, we created 3 varieties of a language that differed in whether and how they included phonetic cues to grammatical category. The Control language was adapted from one used by Saffran (2001) and made to conform to the structure of Korean. All varieties of the language included 4 lexical categories (4 subjects, 4 objects, 3 transitive verbs and 3 intransitive verbs) and 4 function categories (See Table 1).

Table 1: Artificial language lexicon.

Category	Natural Language Counterpart	Words in Control/Short Vowel	Words in No Onset
A	Subject	*meeb rud pell shoke*	*meeb rud pell shoke*
C	Object	*lum neb sheeg tuke*	*lum neb sheeg tuke*
D	Subject Marker	*mun*	*un*
E	Intransitive Verb	*sem gung hep*	*sem gung hep*
F	Transitive Verb	*tupp hool loke*	*tupp ket loke*
G	Object Marker	*doot*	*oot*
H	Postposition	*jeck*	*eck*
&	Conjunction	*ket*	*ool*

Other than the conjunction, function words were deterministic predictors of non-function word type. Unlike English, the language requires non-function words to precede function words. See Table 2 for the phrase structure rules that generated the language. Note that non-function words, particularly C, are not deterministic predictors, as they can occur in multiple sentence roles. No category is anchored to a single serial position within a sentence.

Table 2: Phrase structure rules for the artificial language. Optional elements are in parentheses.

Phrase Type	Rewrites as
Sentence	(HP) AP BP *or* AP (HP) BP
AP	A(D)
HP	CH
BP	GP F *or* E
GP	C(G)
Conjunction	*x*P & *x*P

The complete language results in over 90,000 sentences if conjunctions are limited to one per sentence. Non-function words did not appear more than once in sentences presented to participants (e.g., the *cat* chased the *cat*), though a given function word can occur more than once. Subject and object markers, which are optional in the language, were presented to participants in half of their possible contexts. The optional H-phrase was present in one-third of sentences. Two-thirds of verbs were transitive.

Lists consisting of five repetitions of a word read by a female speaker were recorded. Sentences were constructed by splicing words from the lists according to their serial positions in the sentence. The resulting sentences were characterized by a natural-sounding, list-like intonation, with approximately 250 ms of silence between each word. Sentence final words were characterized by a rapidly falling end-of-list intonation contour and were followed by 500 ms of silence and signaled the end of each sentence. List-like intonation and equal spacing of words were used because they could not serve as cues to grammatical structure. The lexicon was constructed so that phonemes and phonetic features were evenly distributed across all category types.

In the Control language, all words were CVC (consonant-vowel-consonant) single syllables, with no phonetic cues to word category. In the Short Vowel variety of the language, sound editing software was used to shorten the vowels in all function words by 40%. This change mirrors one aspect of a cue found in English: function words are frequently unstressed, resulting in shorter vowels, among other differences. In another variety of the language, sound editing software was used to remove the onset (initial consonant) from all function words. In order to make sure that function words were maximally distinguishable from one another, we switched *ket* with *hool* in the lexicon of the No Onset language, preventing the barely distinguishable *et* and *eck* from occurring in the lexicon. Exposure and test sentences were otherwise identical across languages.

Procedure For the Control language, 24 participants were tested, 27 for the Shortened Vowel language, and 25 for the No Onset language. Before the experiment, participants were told that they would listen to an artificial language and then answer questions to determine their knowledge of the language. During training, participants listened to a list of 280 sentences from one of the artificial languages. Training sentences varied between 3 and 8 words. The list of sentences was identical across languages, with only the characteristics of function words varying between languages. Sentences were presented aurally. In order to encourage implicit learning and alleviate boredom, participants also performed a picture-copying task in which they were to copy a series of four drawings using markers and paper. Participants were told that their drawings would be used in analysis in order to ensure that they were engaged in the task. The training phase of the experiment lasted approximately 24 minutes.

The test phase consisted of 84 trials, with 12 trials for each of 7 patterns tested (see Table 3). Each trial required a grammaticality judgment. On each trial, participants heard two sentences in which the only difference was the pattern violation (i.e., lexical items and overall structure were held constant), and then indicated which of the sentences they thought was correct by button press. For each sentence, either ONE or TWO was displayed on the screen to indicate which sentence was playing. Test sentences were 3 to 6 words long. In addition, within each pattern 6 trials included function words in all present contexts up to the six word limit (Max Function), while 6 other trials presented function words only when required by the grammar (Min Function). Each Max Function trial was created by matching the category structure of a Min Function trial and inserting additional function words. For example, a Min Function trial with the category structure ACF would be reconstructed as ADCGF for the Max Function condition, using different lexical items in order to avoid effects of repetition. Inevitably, sentences that maximized function words were longer: there were 5.2 words per sentence in the Max Function trials versus 4.0 words per sentence in the trials that minimized function words. However, the conditions were controlled for complexity, such that there was no difference in the number of phrases per sentence, with an average of 3.1 phrases per sentence for both trial types. With the exception of the experimentally manipulated characteristics of function words, test and exposure sentences were identical across languages and participants.

Table 3: Patterns tested.

Pattern	Description
AP/BP	Subject phrases precede verb phrases
A exists	Sentences must have subjects
Conjunction	Conjunctions only join phrases of the same type
Double	Non-function categories may not be doubled without a conjunction
E/F exists	Sentences must have verbs
Function	Function words must be matched with the appropriate non-function category
Transitive	Transitive verbs require objects and intransitive verbs must not have objects

Results

A 3 x 2 x 7 mixed-model analysis of variance was conducted with a between-participant factor of Language (Control, Short Vowel, and No Onset) and within-participant factors of Function word presence (Max and Min) and Pattern type (AP/BP, A exists, Conjunction, Double, E/F exists, Function, and Transitivity).

The main effect of Language was marginally significant, $F(2, 73) = 2.50$, $p = .089$. None of the interactions involving language were significant: Pattern x Language, p = .858; Function presence x Language, p = .834; and Pattern x Function presence x Language, p = .729 (all Fs < 1).

Planned follow up comparisons found marginally significant differences between the Control and No Onset languages, p = .087, Bonferroni corrected. The Short Vowel language fell between the other two (see Figure 1). Counter to our hypothesis, participants who heard the control language, in which there was no phonetic distinction between function and non-function words, obtained the best performance on the grammaticality judgment task. Performance was above chance for all languages. Control: $t(23) = 9.26$, $p < .001$; Short Vowel: $t(26) = 5.74$, $p < .001$; No Onset: $t(24) = 4.93$, $p < .001$.

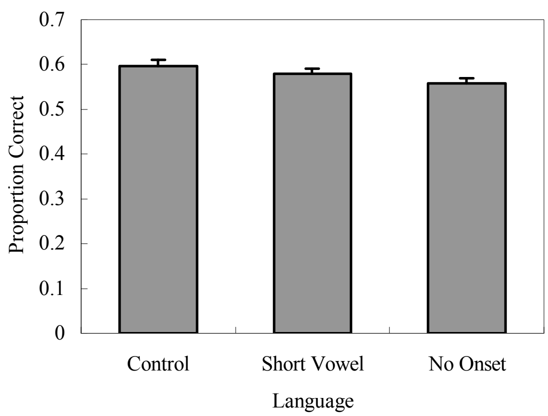

Figure 1: Proportion correct by language type. Error bars represent standard error.

The main effects of the within-participant factors and their interaction were significant. The factors Function Presence: $F(1,73) = 4.28$, p = .042; Pattern $F(6, 438)$ = 27.82, p <.001; and Pattern x Function Presence interaction, $F(6, 438) = 16.36$, p <.001 were all significant. Some patterns were easier to acquire than others. Overall, trials with fewer function words were more accurately judged for grammaticality. For the interaction, minimal function words helped for A exists, Conjunction, and E/F exists, did not significantly help or impede grammaticality judgments for #AP/BP# and Transitivity, and hindered the ability to make judgments for the Double and Function patterns (see Figure 2).

Discussion

Phonetic cues to grammatical category The ability to distinguish function words from non-function words early in the acquisition of a language should facilitate learning because the learner needs to devote fewer resources to uncovering the distinction during learning. Furthermore, this distinction can act as a bootstrap for the acquisition of more complex structures that rely on the distinction between the already-known categories. Our data initially appear to challenge this idea but may better serve to demonstrate the complex dynamics between learning categories and learning relationships among categories.

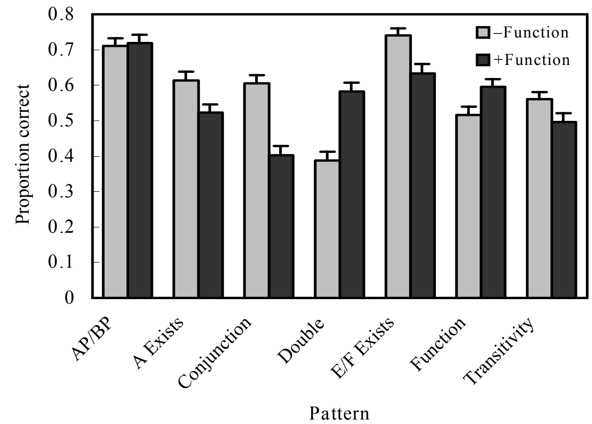

Figure 2: Proportion correct by pattern and minimal or maximal function word presence. Error bars represent standard error.

While the distinction between function and lexical words is an important one that holds across languages, knowing this distinction does not by itself indicate greater knowledge of the patterns that govern a grammar. If our cues encouraged participants to devote resources to learning the broader categories of function and non-function, then fewer resources were available to learn dependencies among words classes at a level below the function versus non-function distinction. Though distributional learning mechanisms are able to take advantage of many kinds of patterns, they are guided and constrained by the information that they are given (Gómez, 2002).

With respect to our manipulations of the phonetic characteristics of function words, it is possible to gain insights from our underwhelming results. The effect of language type was marginally significant, though not in the direction that we predicted. Post hoc analysis determined that learners of the Control language performed significantly better than learners of the No Onset language. Cues to category may have guided learners to construct categories on a broad-based division between function and non-function word categories. Learning this division would likely prove helpful in acquiring the grammar over the long term in terms of generalizing as new lexical items are introduced. This knowledge would not help for the patterns that were used in this experiment, which tested participants' knowledge of the relationships between intermediate-level categories (e.g., A versus C, both members of the broader non-function word category). We raise the possibility that when function words are perceptually distinguishable from non-function words learners are encouraged to attend to different kinds of information than they would in a language that does not cue the distinction between function and non-function words phonetically. Though methods are available

Table 4: Proportion correct by language, pattern and presence or absence of function words. Asterisks denote performance significantly above chance. Daggers indicate significant differences between minimum and maximum function word trials.

Language	Function Presence	AP/BP	A Exists	Conjunction	Pattern Double	E/F Exists	Function	Transitive
Control	Mean	72.2*	59.0*	51.4	48.6	74.7*	58.3*	53.5
	Max	72.9*	54.2	41.7	59.7*	69.4*	63.2*	47.2
	Min	71.5*	63.9*	61.1*	37.5	79.9*	53.5	59.7*
	Difference	1.4	-9.7	-19.4†	22.2†	-10.5	9.7	-12.5†
Short Vowel	Mean	71.3*	55.6	53.4	49.1	67.9*	55.6*	52.5
	Max	74.7*	49.4	42.0	61.1*	61.1*	59.9*	51.2
	Min	67.9*	61.7*	64.8*	37.0	74.7*	51.2	53.7
	Difference	6.8	-12.3†	-22.8†	24.1†	-13.6†	8.7	-2.5
No Onset	Mean	71.0*	56.0	46.7	48.0	63.3*	52.7	52.7
	Max	68.0*	53.3	37.3	54.0	59.3*	55.3	50.7
	Min	74.0*	58.7	56.0	42.0	67.3*	50.0	54.7
	Difference	-6.0	-5.4	-18.7†	12.0	-8.0	5.3	-4.0
	Mean Difference	**4.9**	**-9.1†**	**-20.3†**	**19.4†**	**-10.7†**	**7.9†**	**-6.3**

to determine whether infants distinguish broad function and non-function word categories (e.g., Shi & Werker, 2001), currently there are no comparable methods for adults.

Participants who heard the Shortened Vowel language did not differ from either of the other languages. It is possible that the vowel length differences were not perceptible and, thus, could not contribute to differentiating function words from non-function words. However, performance on the Short Vowel language was numerically between the Control and No Onset languages for five of the seven patterns tested, making it plausible that the less salient cue yielded a smaller effect size that was not captured by the statistical power of this experiment.

On the other hand, listeners may not have perceived the vowel difference: First, in English vowel length is not contrastive, though it contributes to the perception of stress differences in English, along with pitch and intensity. By dissociating vowel length from these other cues in the Shortened Vowel language, it is possible that we made it difficult for participants to perceive function words similarly to the way that they treat unstressed words. In addition to the problems of dissociating vowel length from stress for adult speakers of English, the within-category variation of vowels in the Shortened Vowel language was larger than the between-category variation. For example, while the /E/ in non-function words was longer than the /E/ in function words, it was still shorter than the /u/ in function words. Therefore, vowel length variability patterns were complex (though this variability follows the pattern found in English for stressed and unstressed words), the result of an interaction of grammatical category and vowel type. This, combined with the dissociation of vowel length variability from stress, likely reduced the perceptibility of the vowel length differences.

If the overall effect of phonetic cues to function words is that they are not helpful, than perhaps cues were at least helpful for the Function pattern. This was not the case: learners of the Control language outperformed those who learned the No Onset language by 5.6 percent on this pattern. Moreover, learners of the No Onset language did not perform significantly above chance on the Function pattern. Clearly, while phonetic differentiation of the types used in this experiment was not helpful to learners for the experimental tasks, more research is required, as the phonetic differences between languages did cause differences in learning. If phonetically differentiating function from non-function words impedes learning, then languages would not continue to exhibit this characteristic. Further research is needed to understand exactly how function words help language learners.

Function word presence The significant function word presence by pattern interaction reflects the complex nature of the role of function words in language processing. While our prediction was that more function words would enhance performance across patterns, the data show that learners performed significantly better on Max Function trials on only two patterns: Double and Function. They performed significantly better on Min Function trials for three other patterns: A Exists, Conjunction, and E/F Exists, leaving no differences in the remaining two patterns. This pattern of results was replicated across all three of the languages that were tested (See Table 4).

Though there was a negative overall effect for function words, the significant interaction is the critical finding: predictive elements are helpful in processing some grammatical structures and interfere with processing for other patterns. One possible explanation is that function words require cognitive resources in processing. For some patterns, the results of function word processing will be helpful in determining structure (and therefore yield better grammaticality judgments), and for others it will not. In either case, resources that would have gone to process other words in sentences were used to process function words, reducing the amount or quality of processing of non-function words. If this is the case, then there are two sources of the overall utility, or lack thereof, of the presence of function words. First, there is the direct consequence of function words: they can help processing of other words by reinforcing or assigning category status to other words in a sentence, allowing that knowledge to bolster efforts to determine a sentence's structure. This effect should be positive in cases where structure is ambiguous or should be null when other cues allow the learner to determine structure. Second, including more function words has the indirect consequence of not allowing as many resources for processing other words in a sentence. If a pattern involves subtle differences between grammatical and ungrammatical sentences, then the reduction in resources for non-function words may cause a catastrophic loss of the ability to detect violations. If violations are salient, the negative effect could be negligible. The effects captured in this experiment were the result of a combination of these factors, and future research should attempt to disentangle them and determine how they interact.

We suggest that function words have a complex and dynamic effect on language learning. This effect, for various tasks at certain points in development, is not always positive. In our experiment, designed to replicate aspects of early language development, the presence of function words hurt performance on grammaticality judgments involving some patterns, but helped on others. This finding corresponds to results from natural language research in infants in which non-function words are preferred over function words, even though function words help learners in critical tasks.

In the languages used in this experiment, most of the function words were optional. While this may have made the learning situation more difficult for our participants, the optionality of certain function words was meant to mirror natural languages. Korean, which the current grammar was based on, frequently omits subject and object markers. Even though they were optional, the token frequencies of function words were greater than those of non-function words. Because function word presence appears to interact with a number of other factors, future research should vary the ratio of function to non-function words in the lexicon as well as the proportion of contexts in which a given function word can appear.

References

Gómez, R. L. (2002). Variability and the detection of invariant structure. *Psychological Science, 13*, 431-436.

Green, T. R. G. (1979). The necessity of syntax markers: two experiments with artificial languages. *Journal of Verbal Learning and Verbal Behavior, 18*, 481-496.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child language learners in language formation and change. *Language Learning and Development, 1,* 151-195.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory and Cognition, 30*, 678-686.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential contribution of phonological and distributional cues in grammatical categorisation. *Cognition, 96*, 143-182.

Onnis, L., Christiansen, M. H., Chater, N., & Gómez, R. (2003). Reduction of uncertainty in human sequential learning: evidence from artificial grammar learning. *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. *Proceedings of the Twenty-sixth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*, 493-515.

Shady, M. E. (1996). *Infants' sensitivity to function morphemes.* Unpublished doctoral dissertation, State University of New York at Buffalo, Buffalo, NY.

Shi, R., Cutler, A., Werker, J., & Cruickshank, M. (2006). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *Journal of the Acoustical Society of America, 119*, EL61-EL67.

Shi, R., Morgan, J. L., & Allopenna, P. (1995). Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language, 25,* 169-201.

Shi, R., & Werker, J. F. (2001). Six-month-old infants' preference for lexical words. *Psychological Science, 12,* 70-75.

Shi, R., Werker, J., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy, 10*, 187-198.

Valian, V., & Coulson, S. (1988). Anchor points in language learning: the role of marker frequency. *Journal of Memory and Language, 27,* 71-86.