

Instruction in Semantic Prosody and Collocation: Two Approaches for Deepening Lexical
Knowledge

Anna M. Gates Tapia

Northern Arizona University

Abstract

This paper reports on a study that examined the effects of extensive reading, and inductive versus deductive instructional treatment on four aspects of lexical knowledge: semantic prosody, collocation, definitional meaning, use in sentences. Two groups of low-intermediate students (N=23) enrolled in a 15-week university intensive English program participated in the study. Knowledge of strongly evaluative vocabulary selected from a pedagogical corpus of textbook chapters and graded readers was tested using an adapted Vocabulary Knowledge Scale in a pretest/posttest quasi-experimental design. Two sixty minute instructional sessions were administered to each group of participants, during which they used either inductive or deductive corpus based materials to develop the four aspects of lexical knowledge. Pretest results showed that incidental exposure to vocabulary led to a 58% overall mastery, but that use of collocation was very low. Repeated measures Analysis of Variance was applied to testing data and showed a significant positive affect of instruction on all aspects of lexical knowledge, however there was no difference between treatment groups.

Keywords: semantic prosody, collocation, corpus, lexical knowledge

Instruction in Semantic Prosody and Collocation:
Two Approaches for Deepening Lexical Knowledge

Background

Language learners face a tremendous challenge in mastering the words necessary for successful communication in their new language. Vocabulary not only plays an essential role for productive communication, but for understanding both written and oral texts. Corpus research examining vocabulary knowledge thresholds for adequate text comprehension has suggested that between 95-98% of words need to be known for satisfactory understanding of a text (Hu and Nation, 2000; Laufer, 1989, 1992; Laufer and Ravenhorst-Kalovski, 2010; Schmitt, Jiang and Grabe, 2011; Webb and Rogers, 2009). However, not only are language learners challenged by the quantity of words they must acquire, several aspects of words also need to be known. Broadly speaking, learners need to have some sense of a word's form, meaning and use (Nation, 2001). One area of word knowledge that has been shown to be especially problematic for learners is the use of appropriate collocations. Without explicit instruction in collocation, learners must rely on language input for knowledge about the co-occurrence of words. There is nonetheless strong evidence that input alone is not sufficient for the acquisition of collocational knowledge, as learner difficulty in this area has been well documented over the past twenty years (Altenberg and Granger, 2001; Bahns and Eldaw, 1993; Howarth, 1998; Nesselhauf, 2005). Less well studied however is learners' knowledge of evaluative meanings (i.e. semantic prosody) as demonstrated through appropriate collocate choices. The current study aims to explore low-intermediate second language learners' vocabulary knowledge, especially evaluative and collocational knowledge, and the effects of instruction on this knowledge.

Research Questions

Although extensive research has been conducted in several areas of vocabulary learning and teaching, very little has focused on low intermediate proficiency levels or on L2 lexical

knowledge related to semantic prosody (McGee, 2012). Therefore, the current study aims to provide insight in these areas by answering the following research questions:

1. What is the existing semantic knowledge of low intermediate English language learners of select vocabulary introduced incidentally through reading; specifically in the areas of semantic prosody, definitional meaning, and the ability to produce meaningful sentences with collocations?
2. Is this semantic knowledge affected by instructional treatment?
3. Do corpus based/informed inductive learning materials show any advantage over deductive learning materials on the semantic knowledge of select vocabulary?

Methods

Participants

Twenty-three (23) students of two intact low-intermediate (32-44 TOEFL iBT) reading classes at the Program for Intensive English (PIE) participated in the study. Eleven (11) students comprised Group One (G1). The second group (G2) consisted of twelve (12) students.

Groups were randomly assigned to instructional treatment condition, with G1 being assigned to inductive learning materials and G2 to deductive materials.

Procedures

Word selection

A 36,904 word corpus of the reading passages, instructions and activities students would encounter by the end of their 10th week was built in order to evaluate frequencies of each word students had been exposed to through their reading.

Content words were grouped into word families and their frequencies were determined. Words and word families appearing a minimum six times were evaluated for semantic association and grouped into three categories: negative, positive and neutral semantic association. A random sample of six negative, five positive and three neutral words was taken to

select the words to be tested for study participants' semantic knowledge. Additionally, two verbs appearing in the corpus, *cause* and *commit* that had been identified in previous studies (Stubbs, 1995; Partington, 1998) for their semantic prosodies were included.

Pretest development

An adapted Vocabulary Knowledge Scale (Paribakht and Weche, 1997, as cited in Read, 2000) was developed to measure participants' semantic knowledge of the selected vocabulary. The VKS and the corresponding scoring rubric were modified to be more sensitive to differing aspects of word knowledge by including a question about test-takers knowledge of the semantic prosody of the word. It also separates questions related to a word's meaning from those designed to elicit the word's use in a sentence.

Pretest and posttest administration

A 16-item pretest was administered during 50 minutes of a class period during week 11 of the semester. Participants received five minutes of directions, during which two sample items were modeled. The posttest and its administration occurred on the day immediately following instruction and were identical to the pretest with the exception that it only included the ten instructed words.

Materials development and instruction

The ten words with the lowest mean scores on the pretest were selected for materials development and instruction. Three worksheets per group were developed to promote the learning of four aspects of vocabulary knowledge. The materials developed for G1 were inductive in nature, whereas materials promoting deductive learning were developed for G2. In order to be eligible for worksheet inclusion the text of the concordance lines samples needed to meet three criteria: be complete sentences, provide enough context to allow meanings to be guessed, and contain high frequency collocates.

Participants received two 60-minute instructional treatments over the course of one week. These included completion of the three worksheets per group. G1 (inductive) was also introduced to an on-line concordancing program. Careful modeling of the activities was done to guard against one of the drawbacks associated with corpus based learning, lack of adequate mediation (Boulton, 2010).

Results

A one-way ANOVA was conducted to compare the pretest results of the two study groups. There was not a significant difference among the two groups on overall pretest scores at the $p < .05$ level for the two treatment conditions ($F(1, 21) = 1.838, p = .190$) demonstrating that the two study groups were comparable.

Table 1. *Descriptive statistics for pretest by group*

Group	Mean	Std. Deviation	N
Inductive materials	107.273	27.997	11
Deductive materials	120.167	16.694	12

Total N=23

In answer to the first research question, which examined participants' semantic knowledge of select vocabulary introduced incidentally through course reading, of a possible 208 points, a mean of 114.000 ($SD = 23.216$) was found. Scores ranged from 58.000 to 159.000. These results represent overall word knowledge; however they do not provide a clear picture of how much was known about each aspect of lexical knowledge tested. Table 2 provides descriptive information for each category of semantic knowledge. Additionally, four one-way ANOVA were conducted to compare the groups' pretest results for each of the categories of semantic knowledge. At the $p < .0125$ level, ($df = 1, error = 21$), there was no significant difference between the groups before instructional treatment.

Results from the pretest also suggest that incidental exposure to select vocabulary afforded participants approximately 58% overall mastery of the select vocabulary in the categories measured.

Table 2. *Descriptive statistics and ANOVA for pretest for each category of semantic knowledge*

Category of semantic knowledge	Total possible score (16 items)	Mean	SD	Min.	Max.	F	Sig
Semantic association	16	8.174	2.015	4.00	11.000	0.000	.986
Definitional meaning	64	43.522	7.959	24.00	56.000	3.249	.086
Use	96	60.609	14.437	30.00	88.000	1.267	.273
Use with collocation	32	1.696	1.259	.00	4.000	.765	.392

N=23, alpha, $p < .0125$

One aspect of semantic knowledge for which participants demonstrated poor results was collocational knowledge. Overall, students only had 5% mastery, with most students demonstrating knowledge for only a single collocation of one target item.

The second research question, the effects of instruction on semantic knowledge, was addressed by comparing the overall results of the pretest scores of the ten items of instruction with posttest scores. The total possible score was 130. A repeated measures ANOVA was conducted to compare the effect of instructional treatment on semantic knowledge of select vocabulary from pretest to posttest at the $p < .05$ level. The result of the analysis was $F(1, 21) = 117.311$, $p = .000$, indicating that there was a significant difference in mean across the tests: (1) pretest ($M = 59.739$, $SD = 18.577$), and (2) posttest ($M = 96.341$, $SD = 13.195$). The effect size was $\eta^2 = .848$. The observed post hoc power was 1.000. The mean differences across the four categories were also found to be significant, suggesting that instruction had a positive effect on all categories of semantic knowledge.

The third research question aimed to assess whether corpus based inductive learning materials showed any advantage over corpus based deductive learning materials. A one-way ANOVA was conducted to compare the effect of instructional treatment on gain scores in

inductive learning materials and deductive learning materials conditions. There was not a significant effect of treatment condition, (1) inductive learning materials ($M=33.770$, $SD=12.410$), and (2) deductive learning materials ($M=36.792$, $SD=17.820$), on gain scores at the $p<.05$ level, [$F(1, 20) = .214$, $p = .649$].

When the four categories of semantic knowledge were assessed for significant difference between the two treatment conditions, an analysis of variance found that there was no significant difference between the groups except in the category of collocation. The ANOVA found that the deductive materials instruction groups significantly outperformed the inductive group $F(1, 20)=10.584$, $p=.004$. A post hoc analysis showed that 34.6% of the variance between the groups could be accounted for by instructional treatment type.

Relevance to PIE and Second Language Learning

The present study calls to reflection the practices of vocabulary teaching at the PIE. Results have suggested that much vocabulary knowledge can be gained not only by providing students with ample resources for extensive and intensive reading, but also by dedicating time to vocabulary instruction. Of course, it would be inefficient to spend the time required in this study for gaining depth of knowledge on 10 vocabulary items during normal class time. However, providing students with training and practice with the use of corpus-based activities as well as online corpora seems to have the potential for increasing vocabulary knowledge independently. Therefore, it may be beneficial to teach PIE students about these resources and strategies and require that substantial amounts of time be spent using them for vocabulary development.

References

- Altenberg, B. & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native Student Writing. *Applied Linguistics*, 22(2), 173-195
- Bahns, J. & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21, 101-114.
- Boulton, A. (2010). Data-driven learning: On paper and in practice. In T. Harris & M. Moreno Jaen (eds.) *Corpus Linguistics in Language Teaching*, 128 (pp. 17-52). Bern: Peter Lang.
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review*, 61(3), 355-382.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24-44.
- Hu M. & Nation, I.S.P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403-430.
- Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special Language: From Humans Thinking to Thinking Machines* (pp. 316-323). Clevedon: Multilingual Matters.
- Laufer, B. (1992). Corpus-based versus lexicographer examples in comprehension and production of new words. *EURALEX '92 Proceedings*, 71-76.
- Laufer & Ravenhorst-Kalovski, 2010: Lexical threshold revisited: Lexical text coverage, learner' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15-30.
- McGee, I. (2012). Should we teach semantic prosody awareness? *RELC Journal*, 43(2), 169-186.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.

- Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching*. Philadelphia, PA: John Benjamins.
- Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words of words know in a text and reading comprehension. *The Modern Language Journal*, 95, 26-43.
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford, UK: Oxford University Press.
- Stubbs, M. (1995). Collocations and cultural connotations of common words. *Linguistics and Education*, 7, 379-390.
- Webb, S. & Rogers, M. P. H. (2009). The vocabulary demands of television programs. *Language Learning*, 59(2), 335-366.