

Effects of Corpus-based Instruction on L2 Recognition, Recall, and Use of Signal Markers

Adnan Ajsic

Northern Arizona University

### **Abstract**

Although they are a crucial type of lexical discourse structure, signal markers are neglected in the EAP literature on vocabulary instruction in general and on corpus-based vocabulary instruction in particular. A total of 69 advanced intermediate L1 Arabic, Chinese, and Korean ESL learners in four intact IEP classes were recruited to participate in this study. Based on a specialized pedagogic corpus compiled from class readings, ten single- and multiword signal markers were selected and taught over a period of four weeks using three experimental treatment conditions (direct corpus-based, indirect corpus-based, and traditional). The effects of the treatment conditions on the learners' ability to recognize, recall, and use signal markers were tested using three tasks in a pre-test, post-test study design. The results corroborate the findings of previous studies that all explicit methods are effective for the teaching of text structure (Grabe, 2009). However, they also suggest an aptitude-treatment interaction effect, whereby the learners' L1s combined with the treatment conditions and tasks to produce L1-based differences in treatment effects. Contrary to findings in much educational research, a survey of the learners' perceptions of treatment effectiveness shows that they are not necessarily orthogonal to the actual treatment effects. The pedagogical implications of the findings are discussed.

### Background

Although they are a crucial type of lexical discourse structure, signal markers (relational markers, Degand & Sanders, 2002; transition signals, Oshima & Hogue, 2006; signal words and phrases, Grabe & Stoller, 2011) are neglected in the EAP literature on vocabulary instruction in general and on corpus-based vocabulary instruction in particular. Unlike formulaic sequences, signal markers (SM) can be single- or multiword items (e.g., *however*, *as a result*), and they are often also multifunctional (e.g., *while* can be used both as a time and as a contrastive signal marker). This is an important, if somewhat fuzzy, category of vocabulary for at least two reasons: first, it plays a crucial role in the understanding and interpretation of discourse structure (see further below for details), and second, it includes both single- and multiword items which not only makes for a more natural category, but can also potentially contribute to our understanding of effective methods to teach formulaic sequences by a contrastive analysis of the effectiveness of various methods for the teaching of single- and multiword signal makers.

Although many L2 textbooks for English do address SMs to a greater or lesser extent (for a good example, see, e.g., Oshima & Hogue, 2006; see also the discussions of textbook coverage of formulaic sequences in Martinez & Schmitt, 2012 and Jones & Haywood, 2004), there is little guidance on effective teaching techniques. Furthermore, although SMs are obviously important for a correct and efficient interpretation of the discourse structure of any text and thus also the development of critical reading (and writing) skills (see Grabe, 2009), they have not received much attention. The only recent study that examined SMs specifically is Degand and Sanders (2002). Using a traditional, experimental crosslinguistic (Dutch/French) design, they analyzed the effects of both single- and multiword SMs on the comprehension of expository texts. Their findings indicate that such markers clearly “help readers construct a coherent cognitive

representation of the information in the text” (Degand & Sanders, 2002, p. 751) in both L1 and L2.

As mentioned above, despite the apparent salience of SMs and the corpus-based origins of interest in formulaic language in current phraseology, available empirical, pedagogically relevant research often deals with idioms rather than signal phrases (e.g., Alali & Schmitt, 2012; Boers and Lindstromberg, 2008), mostly relying on more traditional as opposed to corpus-based materials (e.g., idiom/flash cards, fill-in-the blank sheets); further, available corpus-based studies mostly deal with writing rather than reading (e.g., Ohlrogge, 2009; Park 2012; Todd, 2001). In addition, available research exhibits little effort to link reading and writing (see Zhang & Grabe, forthcoming, for a critique).

### **Research Questions**

RQ1: How do ESL students score on (a) a general vocabulary test, and (b) tests of ability to recognize, recall, and use SMs before the treatment?

RQ2: Are there differences in the effects of the different treatment conditions (corpus-based online, corpus-based offline, traditional) on the experimental groups’ recognition, recall, and use of SMs?

RQ3: What are ESL students’ perceptions of the effectiveness and usefulness of the different treatment conditions and how do they compare with the actual findings?

### **Methods**

#### **Participants**

A total of 69 ESL learners from four intact classes participated in this study. The sample included 38 Arabic L1 speakers, 30 Chinese L1 speakers, and 1 Korean L1 speaker.<sup>1</sup> All learners were enrolled full time at a terminal undergraduate level in an IEP at a mid-sized Southwestern

university during the time of instruction and assessment. They were classified as advanced intermediate learners by the IEP; however, their scores on a program-internal placement test ranged from 10 to 21 out of 30 on the reading portion of the test, and between 9 and 19.5 out of 30 on the writing portion of the test.<sup>2</sup>

## **Materials and Design**

### **Instructional materials and procedures**

The selection of SMs to be included in the instruction and assessment, as well as the instruction itself, was based on a frequency analysis of SMs in a specialized pedagogic corpus compiled specifically for this class (cf. Reppen, 2010). This included raw single-word frequencies in the case of single-word SMs and two- and three-word n-grams in the case of multi-word SMs (see below for the entire set; cf. Coxhead, 2000; Jones & Haywood, 2004; Gilquin et al., 2007; Grabe & Stoller, 2011). The most frequently occurring SMs thus identified were then chosen and grouped on the basis of their mutual similarities and differences, as suggested, for example, by Römer (2011). Instruction was sequenced as follows: Week 1 (also, as well/as), Week 2 (and so on, too), Week 3 (for example, for instance, such as), and Week 4 (during, while, although).

The corpus consisted of ENG 105 class readings,<sup>3</sup> including a relevant textbook chapter, and newspaper and journal articles which students also used as mandatory source material for their writing assignments in ENG 105; corpus size was 32,419 word tokens (4,076 word types). After a set of 10 SMs to be taught had been identified, a separate handout was produced for each of the four class sections and each of the four class sessions, covering subsets of near-synonymous markers, e.g., also, too, as well, as well as, and so on (cf. Cobb, 1997; Römer, 2011; for a sample class handout, see Appendix A). The instructional material reflected the treatment conditions as follows (cf. Jones & Haywood, 2004): Section A, which received a direct corpus-

based, hands-on concordancing treatment (corpus online), was provided with a version of the handout which included instructions for concordance-line searches in the text analysis software AntConc (Anthony, 2011) and signal marker-specific questions and exercises; Section B, which received an indirect corpus-based treatment (corpus offline), was provided with a version of the handout that excluded the concordancing instructions but included AntConc screenshots of the same concordance-line searches run by Section A students and the same signal marker-specific questions and exercises; Section C, which received a traditional treatment based on the use of dictionaries (both print and online), was provided with a handout which excluded all corpus-related information but included the same signal marker-specific questions and exercises as the previous two; Section D, the control group, did not receive any instruction.

### **Results**

Data were screened and tested for parametric assumptions. All statistical analyses relied on the General Linear Model-Univariate, Multivariate, and Repeated Measures (i.e., RM ANOVA) procedures available in SPSS, unless noted otherwise. The study relied on a complex mixed-subjects  $2 \times 4 \times 2$  RM ANOVA design, with the treatment (corpus online vs. corpus offline vs. traditional vs. control) and L1 (Arabic vs. Chinese) as the between-subjects variables, and time (pre-test vs. post-test) as the within-subjects variable.

#### **General Vocabulary Knowledge**

There were no significant differences between the groups on the test of their general vocabulary knowledge,  $F(3, 58) = .551, p > .05, \eta^2 = .028$ . Table 2 shows the mean scores and standard deviations for all groups. The total mean ( $M = 10.24$ , number of unknown words) further suggests that the learners in this group knew about 96.59% of the vocabulary in class readings, which mitigates the high readability level of the assessment texts used in this study to a certain

extent.

### Signal Markers

#### Task 1: Recognition (pre-test and post-test)<sup>4</sup>

On the pre-test, no main effects emerged for treatment,  $F(7, 40) = .437, p > .05, \eta^2 = .032$ , or L1,  $F(7, 40) = .014, p > .05, \eta^2 = .000$ , and no significant interaction effect (Treatment  $\times$  L1),  $F(7, 40) = 1.605, p > .05, \eta^2 = .107$ . Table 3 shows the mean scores and standard deviations for all groups, pre- and post-test, broken down by L1. Please note that L1 groups (Arabic, Chinese) exhibited a larger mutual difference than the treatment groups, as well as that the control group scored the highest mean on this measure on the pre-test.

Table 3

*Means and Standard Deviations Comparing Groups on Ability to Recognize SMs Pre- and Post-test, by Treatment Condition and L1*

Treatment	L1	<i>n</i>	Pre-test		Post-test	
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
CorpusOnline	Arabic	5	3.20	3.27	9.60	1.95
	Chinese	5	6.00	5.79	6.20	4.32
	Total	10	4.60	4.67	7.90	3.64
CorpusOffline	Arabic	9	5.78	5.33	6.78	3.77
	Chinese	5	4.40	5.60	7.20	4.50
	Total	14	5.29	5.25	6.57	3.88
Traditional	Arabic	7	6.71	4.15	7.29	1.98
	Chinese	6	2.17	1.72	7.67	3.01
	Total	13	4.62	3.93	7.46	2.40
Control	Arabic	7	5.29	4.46	10.29	3.82
	Chinese	4	7.75	5.91	6.00	2.95
	Total	11	6.18	4.90	8.73	4.00
Total	Arabic	28	5.43	4.43	8.29	3.35
	Chinese	20	4.80	4.94	6.60	3.53
	Total	48	5.17	4.61	7.58	3.45

The results were similar on the between-subjects post-test: no main effects for treatment emerged,  $F(7, 40) = .425, p > .05, \eta^2 = .031$ , or L1,  $F(7, 40) = 1.049, p > .05, \eta^2 = .026$ , and there was also no significant interaction effect (Treatment  $\times$  L1),  $F(7, 40) = 3.237, p > .05, \eta^2 = .010$ . On the within-subjects post-test, however, a main effect for time emerged,  $F(7, 40) = 13.994, p < .01$ , with a large effect size ( $\eta^2 = .259$ ), as well as a significant interaction effect (Time  $\times$  Treatment  $\times$  L1),  $F(7, 40) = 5.175, p < .01$ , again with a large effect size ( $\eta^2 = .280$ ). Other interaction effects were not significant: Time  $\times$  Treatment,  $F(7, 40) = .602, p > .05, \eta^2 = .043$ ; Time  $\times$  L1,  $F(7, 40) = 2.082, p > .05, \eta^2 = .049$ .

When broken down by L1, the significance is shown to derive from the L1 Arabic within-group differences,  $F(7, 40) = 18.542, p < .01$ , with a very large effect size ( $\eta^2 = .436$ ), whereas the L1 Chinese group showed a substantial but non-significant gain compared to the pre-test,  $F(7, 40) = 1.884, p > .05, \eta^2 = .105$ . This is an indication of an aptitude-treatment interaction (ATI), whereby L1 combined with the treatment conditions to produce a significant difference in mean gains among the L1 groups between the pre-test and post-test. Figures 1-2 show that all three treatment conditions were effective overall, as well as that Arabic speakers achieved larger mean gains.



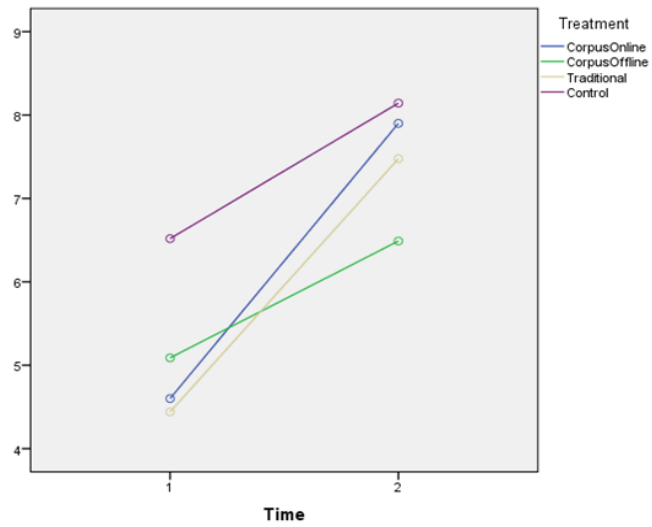


Figure 1. Means comparing groups on ability to recognize SMs pre- and post-test, by treatment condition.

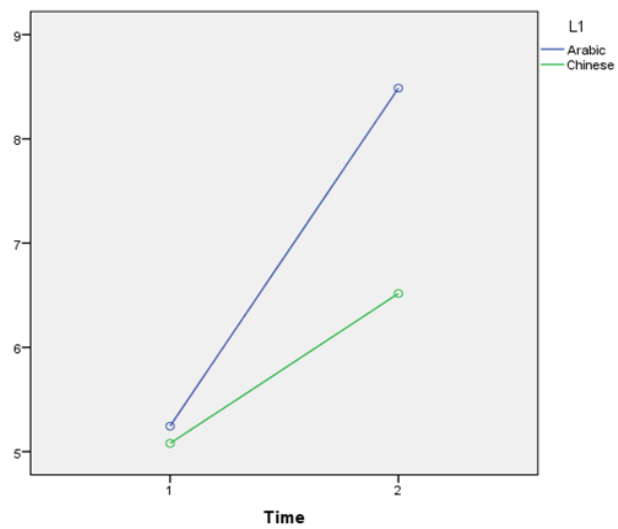


Figure 2. Means comparing groups on ability to recognize SMs pre- and post-test, by L1.

### Perceptions Survey

As mentioned above, the three experimental groups were also asked to respond to three statements about their perceptions of the effectiveness of their respective treatments, by indicating their degree of agreement or disagreement on 5-point Likert scales, as well as one open-ended question. The responses to questions about treatment effectiveness, learning success, and willingness to use treatment in the future were generally positive across the treatment groups.

Interestingly, the only significant difference emerged between the treatment groups' responses to the first statement: "The use of (treatment) has been helpful for my learning of SMs",  $F(5, 31) = 3.538, p < .05$ , with a medium effect size ( $\eta^2 = .186$ ). A Tukey HSD post-hoc analysis showed a significant difference between the CorpusOffline group ( $M = 1.93, SD = .73$ ) and the Traditional group ( $M = 1.31, SD = .48$ ),  $p < .05$ , (lower mean scores indicate more positive perceptions of the treatment; see Figure 3). There were no significant differences between the L1 groups.

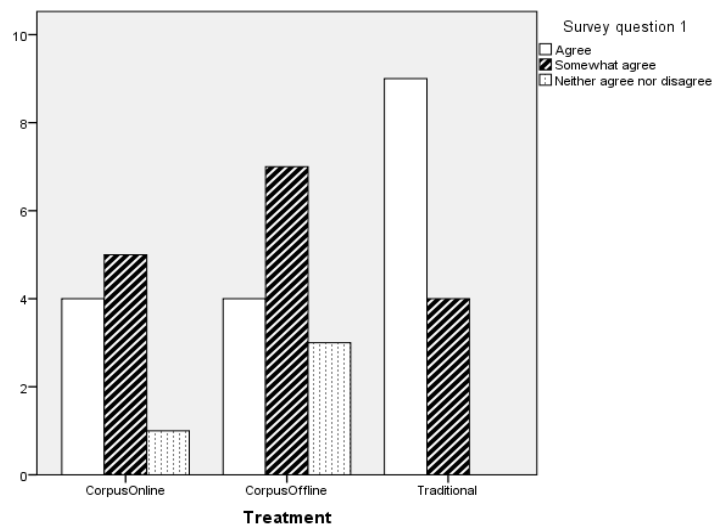


Figure 3. Frequencies and percentage values comparing groups' treatment perceptions.

### **Relevance to PIE and Second Language Learning**

The relevance of this study to both the PIE and second language learning is primarily in its demonstration of an aptitude-treatment interaction effect ( $L1 \times \text{Treatment}$ ) on corpus-based and traditional methods for the teaching of signal markers in reading lab classes. The main pedagogic implication is that since the effectiveness of a method depends on the learners' L1, methods need to be combined because classes are made up of learners with different L1s. It has also been shown that learners' evaluations of method effectiveness can be congruent with the actual method effects.

### **Notes**

---

<sup>1</sup> This is the total number of students in all four sections who received instruction and took either the pre-test or the post-test. The actual totals used in different analyses differ on account of missing data and/or attrition, and the deletion of outliers during assumptions testing. The data from the one Korean L1 student was also excluded from analysis because of a lack of variation required for inferential statistical analyses. The total number of cases included in the statistical analyses was thus 48 for the pre- and post-test, and 37 for the survey.

<sup>2</sup> Not all students were required to take the IEP placement test as some (roughly 40%) automatically promoted to the terminal level (Level 5) having satisfied the minimum requirements for promotion from Level 4. Nevertheless, their proficiency levels and performance in classes at this level appeared to be roughly equal.

<sup>3</sup> ENG 105 is a freshmen composition course that is part of the liberal arts requirement for all students matriculating at this university.

<sup>4</sup> In the interest of limiting length, analysis of only one of the three tasks (Task 1) is included in this report.

### References

- Alali, F. A., & Schmitt, N. (2012). Teaching formulaic sequences: The same as or different from teaching single words? *TESOL Journal*, 3(2), 153-180.
- Anthony, L. (2011). AntConc (3.2.4w) [Software]. Retrieved from <http://www.antlab.sci.waseda.ac.jp/software.html>.
- Boers, F., & Lindstromberg, S. (2008). From empirical findings to pedagogical practice. In F. Boers & S. Lindstromberg (Eds.), *Cognitive linguistic approaches to teaching vocabulary and phraseology* (pp. 375–393). Berlin, Germany: Mouton de Gruyter.
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, 25(3), 301-315.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Degand, L., & Sanders, T. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing: An Interdisciplinary Journal*, 15, 739–757.
- Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6, 319-335.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge: CUP.
- Grabe, W., & Stoller, F. (2011). *Teaching and researching reading*. Harlow: Longman.
- Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 269–292). Amsterdam, the Netherlands: John Benjamins.

- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299-320.
- Ohlrogge, A. (2009). Formulaic expressions in intermediate EFL writing assessment. In R. Corrigan, E. A. Moravcsik, H. Ouali, and K. M. Wheatley (Eds.), *Formulaic language: Acquisition, loss, psychological reality, and functional explanations* (pp. 375–386). Amsterdam, the Netherlands: John Benjamins.
- Oshima, A., & Hogue, A. (2006). *Writing academic English*. Pearson Longman.
- Park, K. (2012). Learner-corpus interaction: A locus of microgenesis in corpus-assisted L2 writing. *Applied Linguistics*, 1-26. doi:10.1093/ams012.
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge: CUP.
- Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205-236.
- Schmitt, N., Jiang, X. and Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, 95, 26–43.
- Todd, R. W. (2001). Induction from self-selected concordances and self-correction. *System*, 29, 91-102.
- Zhang, C. & Grabe, W. (forthcoming). Reading and writing together: A critical component of EAP teaching and learning.