Data-driven learning and the acquisition of Italian collocations: from design to student evaluation

Luciana Forti[1]

Abstract. This paper looks at how corpus data was used to design an Italian as an L2 language learning programme and how it was evaluated by students. The study focuses on the acquisition of Italian verb-noun collocations by Chinese native students attending a ten month long Italian language course before enrolling at an Italian university. It describes how an Italian native corpus, the Perugia Corpus (PEC), and an Italian learner corpus, the Longitudinal Corpus of Chinese Learners of Italian (LoCCLI), were used to build a data-driven learning programme for an eight week long Italian language course. The paper shows how different kinds of data can make a contribution not only to the creation of learning materials, but also to the definition of learning aims and the construction of assessment tools, and it presents the results of an end-of-course student questionnaire.

Keywords: Italian, data-driven learning, collocations.

1. Introduction

The integration of authentic corpus data in second language teaching was first reported by McKay (1980) and further developed by Johns (1991). When Johns (1991) formulated the expression Data-Driven Learning (DDL), the reference was also to a precise teaching methodology based on the guided-discovery of patterns in concordance lines. Since then, DDL has seen a plethora of versions in terms of teaching strategies and tools used (Boulton, 2017).

1. University for Foreigners of Perugia, Perugia, Italy; luciana.forti@unistrapg.it

How to cite this article: Forti, L. (2017). Data-driven learning and the acquisition of Italian collocations: from design to student evaluation. In K. Borthwick, L. Bradley & S. Thouësny (Eds), *CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017* (pp. 110-115). Research-publishing.net. https://doi.org/10.14705/rpnet.2017.eurocall2017.698

The view of language as largely formulaic and primed has risen from the increasingly powerful analysis of large corpora containing instances of real language use in a variety of contexts. Knowing a word entails knowing the company it keeps (Firth, 1957), with clear implications for second language pedagogy.

This paper presents a pilot method to design a language learning programme tailored for the acquisition of Italian verb-noun collocations, and the effect it had on students. Two corpora were used: an Italian native reference corpus, the PEC (Spina, 2014), and an Italian non-native corpus, the LoCCLI. The corpus data was used to select learning aims, design learning activities, and build a proficiency test.

2. Method

2.1. Selecting learning aims

The PEC is used through the DICI-A, a PEC-based dictionary of collocations built for learners of Italian L2 (Spina, 2010), in order to identify the list of verb-noun collocations that are mostly used in Italian.

The LoCCLI is used to analyse the errors made in verb-noun collocations, and to serve as a basis for the creation of classroom activities based on error correction, as well as for the selection of distractors in the multiple-choice section of the collocational proficiency test.

Each weekly lesson focused on a set of eight collocations. A list of 32 collocations more frequently used with errors in LoCCLI was made. Errors were tagged according to whether they involved the noun, the determiner, the verb, or the whole combination (Nesselhauf, 2005; Wang, 2016). This initial list was then grouped into eight topics, corresponding to the general weekly topics that each lesson was based on. The missing spots for each weekly set were then filled by selecting collocations from DICI-A and according to the following three main criteria: highest frequency and dispersion values; thematic relevance to the identified topics; and presence of a delexicalised verb.

This two-stage selection process resulted in a list of thematically linked collocations sets. Each set was used to create experimental and traditional activities, as well as devise an appropriate take-home assignment.

2.2. Designing learning activities

Data from LoCCLI was used in error correction activities where learners needed to decide, in their groups, whether the sentences shown contained an error or not. Most activities, however, drew on data extracted from PEC, for designing both traditional as well as concordance-based DDL activities on paper.

Being a sample of multiple instances of a single collocation, the concordance allows the construction of a variety of guided-discovery activities aimed at fostering the interiorisation of a verb-noun collocation, in its specific context of occurrence and in relation to its structural and semantic pattern.

2.3. Building a proficiency test

In order to try to capture both definitional and transferable knowledge of collocations in a balanced manner, the proficiency test was divided into three parts:

- 32 multiple choice items, using the language and the errors found in LoCCLI as distractors;
- 32 gap-fill items, with sentences adapted from PEC, with the omission of the verb collocate;
- a collocational table like the one designed by Gyllstad (2005).

Similarly to Supatranont's (2005) work, the first set of 32 items was aimed at eliciting definitional knowledge, while the second set of 32 items looked at transferable knowledge. The table was aimed at assessing decontextualised transferable knowledge.

3. Results and discussion

An end-of-course questionnaire, composed by closed and open questions, was administered to all eight classes who were exposed to the data-driven experimental lessons. Here we will focus on the 50 questionnaires collected from the experimental classes, and particularly on the closed questions dealing with the specifics of concordance-based materials, all of which were based on a 6-point Likert scale (see Table 1).

An even-numbered scale was chosen in order to avoid a neutral middle option, thus guiding the students to make an accurate choice (Dörnyei & Taguchi, 2010, pp. 28, 114). A balanced mix of both positively and negatively worded items were formulated in order to avoid the tendency of the respondents to select options from only one side of the scale (Dörnyei & Taguchi, 2010, p. 43).

Table 1. Questionnaires collected from the experimental classes; mean based on 6-point scale

Item 1. Reading groups of sentences containing the same combination confused me				
ANSWER		%	MEAN	SD
1	Totally disagree	10%	3.60	1.56
2	Disagree	22%		
3	Partially disagree	10%		
4	Partially agree	26%		
5	Agree	20%		
6	Totally agree	12%		
Item 2. The observation of groups of sentences containing the same combination has helped me to understand how to use that combination in the future				
ANSWER		%	MEAN	SD
1	Totally disagree	2%	5.20	1.14
2	Disagree	4%		
3	Partially disagree	2%		
4	Partially agree	6%		
5	Agree	36%		
6	Totally agree	50%		
Item 3. The groups of sentences will help me make less errors in the future				
ANSWER		%	MEAN	SD
1	Totally disagree	0%	5.08	0.92
2	Disagree	2%		
3	Partially disagree	4%		
4	Partially agree	12%		
5	Agree	42%		
6	Totally agree	40%		
Item 4. A new smartphone application with groups of sentences for word combinations would be useless				
ANSWER		%	MEAN	SD
1	Totally disagree	28%	2.64	1.55
2	Disagree	30%		
3	Partially disagree	14%		
4	Partially agree	14%		
5	Agree	6%		
6	Totally agree	8%		

The proposal of looking at groups of sentences showing how a single combination is used turned out to be, to some extent, challenging, with 60% of respondents stating it was somewhat confusing, although the mean and *SD* values show that answers fall into an evenly distributed in-between area.

On the other hand, responses are more polarised when most students indicate that the groups of sentences they initially found confusing did in fact help them to understand how to use the combination in future, decreasing the perceived likelihood of producing errors. Furthermore, the respondents appear to largely favour the idea of a smartphone app based on concordance lines.

The data from the questionnaire clearly indicates a need to improve the concordance-based activities. One major issue, in fact, is to make concordance lines and pattern hunting tasks more effective for learners. These kinds of improvements would minimise the chances of causing confusion, while strengthening the positive outcomes that have already been observed.

4. Conclusions

Despite the growing body of research in the field of DDL, practicalities related to how corpus data can actually be selected and integrated in second language pedagogy are still often overlooked. This paper attempted to provide a contribution in this direction, by describing a method followed to ease the acquisition of Italian verb-noun collocations, through concordance-based work. The questionnaire results seem to show some promise, especially in relation to possible mobile-assisted language learning applications.

Chinese learners are one of the largest learning populations of Italian as a second/foreign language learning. As a result, the challenges they face have become central in the debate concerning educational effectiveness and innovation in methods and materials design. The position of Italian as an underrepresented language in corpus-based pedagogical research makes it an ideal candidate for future work and development in this sense.

References

Boulton, A. (2017). Research timeline. Corpora in language teaching and learning. *Language Teaching, 50*(4), 483-506. https://doi.org/10.1017/S0261444817000167

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research. Construction, administration, and processing* (2nd ed.). Routledge.

Firth, J. R. (1957). *Papers in linguistics 1934-1951*. Oxford University Press.

Gyllstad, H. (2005). Words that go together well: developing test formats for measuring learner knowledge of English collocations. *The Department of English in Lund: Working Papers in Linguistics, 5*, 1-31.

Johns, T. (1991). Should you be persuaded - two examples of data driven learning materials. *Classroom Concordancing, English Language Research Journal 4*, 1-13.

McKay, S. (1980). Teaching the syntactic, semantic and pragmatic dimensions of verbs. *TESOL Quarterly, 14*(1), 17-26. https://doi.org/10.2307/3586805

Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins Publishing Company. https://doi.org/10.1075/scl.14

Spina, S. (2010). The dictionary of Italian collocations: design and integration in an online learning environment. In *LREC 2010 Proceedings, Malta* (pp. 3202-3208).

Spina, S. (2014). Il Perugia Corpus: una risorsa di riferimento per l'italiano. Composizione, annotazione e valutazione. In *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & the Fourth International Workshop EVALITA 2014litica* (Vol. 1, pp. 354-359). Pisa University Press.

Supatranont, K. (2005). *A comparison of the effects of the concordance-based and the conventional teaching methods on engineering students' English vocabulary learning*. Doctoral dissertation. Chulalongkorn University, Thailand.

Wang, Y. (2016). *The idiom principle and L1 influence. A contrastive learner-corpus study of delexical verb+noun collocations*. John Benjamins Publishing Company. https://doi.org/10.1075/scl.77

Research-publishing.net

Published by Research-publishing.net, not-for-profit association
Contact: info@research-publishing.net

CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017
Edited by Kate Borthwick, Linda Bradley, and Sylvie Thouësny

Typeset by Research-publishing.net

Cover design based on © Josef Brett's, Multimedia Developer, Digital Learning, http://www.eurocall2017.uk/, reproduced with kind permissions from the copyright holder.

Cover layout by © Raphaël Savina (raphael@savina.net)
Photo "frog" on cover by © Raphaël Savina (raphael@savina.net)

Fonts used are licensed under a SIL Open Font License

ISBN13: 978-2-490057-04-7 (Ebook, PDF, colour)
ISBN13: 978-2-490057-05-4 (Ebook, EPUB, colour)
ISBN13: 978-2-490057-03-0 (Paperback - Print on demand, black and white)
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit: Bibliothèque Nationale de France - Dépôt légal: décembre 2017.