Evaluating Data-Driven Learning Effects in the Italian L2 Classroom
Etic and Emic Perspectives Combined

Luciana Forti
Università per Stranieri di Perugia, Italia

Abstract   In this article we outline the results of an evaluation of Data-driven learning (DDL) effects in relation to the development of Italian L2 phraseological competence. We do this on the basis on empirical data, by combining an external and objective perspective based on data elicited through a competence test, and an internal and subjective perspective based on data elicited through a student questionnaire. In the first case we refer to etic data, while in the second case we refer to emic data. Overall, the results indicate mild positive effects in terms of etic data, but stronger positive effects in terms of emic data. The article concludes by stressing the importance of combining two perspectives such as the ones adopted in this study, in order to be able to observe some of the many different aspects of educational effectiveness within a single, integrated framework.

Keywords   DDL. Italian L2. Etic. Emic. Collocations.

Summary   1 Introduction. – 2 Background. – 3 Method. – 3.1 Study Design. – 3.2 The Collection of Etic Data Through a Phraseological Competence Test. – 3.3 The Collection of Emic Data Through an End-of-Course Questionnaire. – 4 Results. – 4.1 The Etic Perspective. – 4.2 The Emic Perspective. – 5 The Usefulness of an Integrated View of DDL Effects. – 6 Conclusions.

1   Introduction

Research methods in second language learning are numerous. In order to gain insight into how second language learning works, researchers may analyse learner production and comprehension, learner errors, and the development of learner proficiency, or they may analyse the experience of learning from the learners' perspective, using focus groups with learners or by administering student questionnaires. The first group of methods elicits etic data, while the second group of methods elicits emic data. Both kinds of data can be analysed both/either quantitatively and/or qualitatively.

The etic vs. emic dichotomy is claimed to have been introduced in cultural anthropology by Kennedy Pike in 1954, modelling the phonetic vs. phonemic dichotomy that was already productive in linguistics (Pike 1967). A working definition of this dichotomy can be found in the *Concise Dictionary of Social and Cultural Anthropology*, where we read that:

> an emic representation of the ideas or actions of the members of a culture is drawn from the views of its own participants; an etic one is drawn from outside. For example, the eternal observer may regard certain phenomena as symptoms of a disease – this is an etic judgment. But the cultural group in question may recognise other symptoms as characteristic of a particular illness that is not recognised elsewhere – this would be called and emic explanation. (Morris 2012, 80)

Although the boundaries between these two kinds of perspectives have been debated (Headland, Pike, Harris 1990), they are both seen as desirable in second language research, particularly within mixed-methods research designs (Riazi 2017).

In the context of Data-driven learning (DDL), the etic perspective can be found in studies where data is collected by means of a competence test or by analysing learner production. The emic perspective, on the other hand, can be found in studies based on the collection of data via student questionnaires, where the learners being exposed to DDL express their views and feelings in relation to the approach. The following paragraph outlines the background related to the present study, and concludes by identifying the research gaps in the literature and by formulating the research questions addressed in this article.

2 Background

We can trace the state of the art of the etic perspective on DDL on the basis of three meta-analyses published in recent years. The earliest is the one by Mizumoto and Chujo (2015), based on 32 studies referring to EFL learning in Japanese speaking contexts. Learning gains are measured on the basis of effect size (i.e. the standardised difference between two means), in relation to four different learning areas: lemma, category, phrase and proficiency. DDL appears to be most effective in relation to learning at the level of the lemma, and least effective when looking at changes in proficiency (Mizumoto, Chujo 2015, 9). Observable changes in proficiency when exposing learners to DDL are challenging to obtain if we consider, as the authors note, that in the case of the TOEIC (Test of English for International Communication) at least 100 hours of language training are needed before changes in proficiency are likely to be observed (Mizumoto, Chujo 2015, 10).

A second meta-analysis was published by Boulton and Cobb (2017), including 88 unique samples drawn from 64 separate studies. For the first time, a wide range of moderator variables was factored into the analysis: 25 in total and a total of 40 different levels for both within and between groups designs. Overall, this meta-analysis found that 60% of the moderator variables produce large effect sizes, while 25,5% produce medium effect sizes (Boulton, Cobb 2017, 39). The smallest effect sizes are found in between groups designs, in studies with participant samples exceeding 50 students and in cases where the proficiency level of participants is lower-intermediate.

The most recent meta-analysis published on the topic is contained in Lee, Warschauer and Lee (2018). This time, the focus is restricted to DDL studies on vocabulary learning, and to studies based on a controlled design. It also includes different dimensions of vocabulary knowledge. The main findings indicate that the largest effect sizes are to be found in relation to in-depth knowledge and higher proficiency levels.

If we pull together the findings from all three meta-analyses, we see that DDL is most effective in the following cases: with vocabulary, in within-groups designs, in a foreign language context, at higher proficiency levels, with mixed paper/computer-based modalities, for in-depth knowledge of vocabulary and with more than 10 sessions.

The emic perspective on DDL, on the other hand, can be found in a few selected key publications, where accounts of learners' attitudes and reactions towards working with corpora and/or corpus-based materials are provided. To the best of our knowledge, the first and only comprehensive review on learner attitudes can be found in Chambers (2007). Here, the author lists 10 studies using questionnaires to elicit student attitudes, and reviews them in terms of positive and

negative attitudes. In the first case, positive attitudes are related to the perceived relevance and authenticity of the data, and the inductive nature of the learning process that corpus consultation involves (Chamber 2007, 11). In the second case, the negative attitudes found are mostly related to the fact that corpus consultation can be difficult and time consuming, and this indicates a possible need for preliminary corpus training so that any potential obstacles in consulting a corpus can be removed. In her concluding remarks, however, Chambers notes not only the great variety of data collection methods used in these studies, but also the different kinds of DDL treatments about which the learners are asked to provide their views. This, naturally, calls for greater homogeneity in both aspects, so that a comparison among different studies can be possible.

More recently, Mizumoto, Chujo and Yokota (2016) made a considerable step forward in this direction by developing and validating a student questionnaire tailored to measure DDL effects at the emic level. The questionnaire is divided into 18 items eliciting learners' perceived benefits in relation to DDL and the degree of usefulness that the approach is perceived to bring in the short and long term. This tool considerably improves the area of research tools that are needed to investigate the DDL attitudes among, although it implies a specific kind of DDL pedagogical treatment that is able to match it.

What seems evident in the current literature on the etic and emic perspectives related to DDL effectiveness is that there is limited evidence related to the two perspectives combined in a single study. In other words, the samples of learners are generally the object of DDL studies either from an etic or from an emic perspective only.

In the following paragraph we will describe the method adopted in the present study, aimed at combining etic and emic perspectives in order to evaluate DDL effects in the context of Italian L2 learning and teaching. We are guided by the following two research questions:

1) How does phraseological competence develop over time when comparing a DDL versus a non-DDL approach?
2) What are the learners' attitudes towards DDL?

3 Method

The following paragraphs outline the methods adopted in the present study, in relation to its design, and the two research tools developed in order to collect the etic and emic data, namely the phraseological competence test and the end-of-course questionnaire.

3.1 Study Design

The study is based on a controlled between-groups design. It involved the participation of eight intact classes of Chinese learners of Italian at lower-intermediate proficiency level, attending a 10-month long Italian language course at the University for Foreigners of Perugia, Italy. Of these classes, four were randomly assigned to the experimental condition, and the other four randomly assigned to the control condition. In the experimental condition, learners were exposed to paper-based DDL materials developed with data derived from the *Perugia Corpus* (Spina 2014), while in the control condition, learners were exposed to traditional activities. In both cases, each lesson was characterised by the same learning aims referring to a specific phraseological unit: verb + noun collocations. The DDL materials were all concordance-based, thus including multiple-sentence gap-fill, matching, guessing, and guided-discovery activities. The control activities, on the other hand, were all based on single sentences, one for each target collocation within a given activity. Each class received a 1-hour lesson per week for 8 weeks. Both the DDL and the control lessons were taught by the author of the study. Detailed lesson plans with clear stages and sequences were followed in order to minimise bias. The phraseological competence test was administered at 4 four-week intervals. The aim of the last administration of the test was to elicit retention rates 4 weeks after the end of the lessons.

Table 1 Summary statistics of participant samples

	Control	Experimental
No.	61	62
Gender		
(F / M)	38 / 23	47 / 15
Age		
(range / mean / median / SD)	18-26 / 21,01 / 2,33	18-27 / 21,43 / 2,50
Months learning Italian before coming to Italy		
range / mean / median / SD	0-12 / 3,72 / 3 / 2,75	0-24 / 4,29 / 3 / 3,93
English as an FL		
(0 / B / I / A)	0 / 34 / 24 / 3	0 / 35 / 23 / 4
Note: 0 = no English; B = beginner; I = intermediate; A = advanced		

3.2   The Collection of Etic Data Through a Phraseological Competence Test

The aim of the phraseological competence test was to elicit language gains over time. The items in the phraseological competence test were matched with the weekly sets of learning aims identified at the beginning of the study. Each weekly lesson was designed on the basis of 8 verb + noun collocations, all of which were linked thematically so as to provide each lesson with a unified communicative context.

The test was divided into two main parts: the first, based on 32 multiple-choice items, and the second, based on 32 gap-fill items. Each item contained one verb + noun collocation from the overall list of learning aims identified for the lesson planning, reaching a total of 64 items.

The multiple-choice items were developed on the basis of data extracted from the LOCCLI (*Longitudinal Corpus of Chinese Learners of Italian*; Spina 2017): frequent errors extracted from the corpus were used to inform the options provided in the multiple choices forming the item.

The gap-fill items were constructed on the basis of data derived from the native Italian reference corpus PEC (*Perugia Corpus*; Spina 2014), by omitting the verb collocate so that the students would have to decide which one would fit in the sentence provided. Answers to the gap-fill part of the test were treated as a binary variable, i.e. correct/incorrect. Responses deemed acceptable within the context of the test item, though different from what was expected, were marked as correct. These items were 7 out of the total of 32.

3.3   The Collection of Emic Data Through an End-of-Course Questionnaire

The aim of the questionnaire was to elicit learner attitudes in relation to their exposure to corpus-based pedagogical materials. The questionnaire was divided into two parts: the first containing likert-scale items and the second containing open-ended questions. All likert-scale items were based on a 6-point scale, with values ranging from "totally disagree", valued at 1, to "totally agree", valued at 6. This choice follows the recommendations found in Dörnyei (2010) indicating that an even-scaled range prevents respondents from choosing a neutral, middle option. The items were also evenly worded either positively or negatively, in order to avoid respondents marking only one end of the scale (Dörnyei 2010). The open-ended questions were aimed to give students more freedom to express their views on the lessons, and provide them with an opportunity to make suggestions. In order to avoid possible difficulties in understanding the

items and questions, and considering that the aim of the question-naire was not that to elicit language competence, all items and questions were followed by a translation in Chinese.

4 Results

In the following paragraphs, we describe the results obtained within both the etic and emic perspectives.
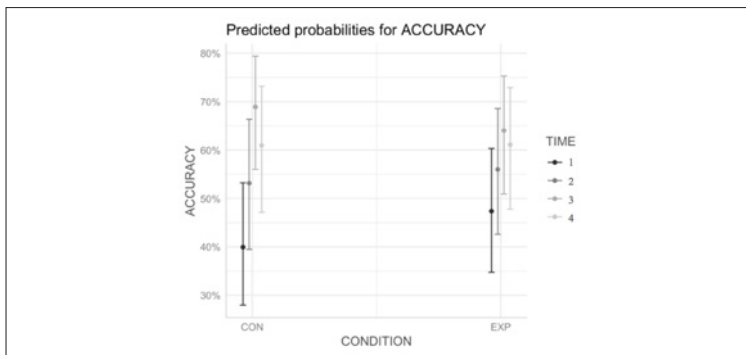
4.1 The Etic Perspective

As mentioned in § 2.2, the etic perspective related to the effects of DDL on language gains was elicited by means of a phraseological competence test. The collected tests include data from 123 students, 62 for the experimental and 61 for the control condition. The dataset was analysed through generalised mixed-effects modelling (Linck, Cunnings 2015), with successive differences contrast coding (Venables, Ripley 2002). The independent variable (i.e. teaching approach) had two levels: DDL vs. non-DDL. The outcome variable (i.e. accuracy) had also two levels: correct vs. incorrect. The formula of the final model looking at the overall effects of DDL on accuracy over time was: ACCURACY ~ CONDITION * TIME + (1 | STUDENT_ID) + (1 | CLASS) + (1 + CONDITION | ITEM_ID). Table 2 contains the coefficients for fixed effects and interactions.

As can be seen, the predicted probability of accuracy is higher in the experimental condition, though this value is non-significant (*Estimate* = 0.05129, *SE* = 0.27487, *p* = 0.851973): learning patterns in the two groups were not significantly affected by the differences in pedagogical treatment. However, all time contrasts (test 2 vs. test 1, test 3 vs. test 2, and test 4 vs. test 3) are highly significant. In particular, time contrasts between tests 2 and 1 (*Estimate* = 0.53413, *SE* = 0.06658, *p* = 1.04e-15), and between tests 3 and 2 (*Estimate* = 0.66896, *SE* = 0.06658, *p* = <2e-16) exhibit significant positive estimates, meaning that accuracy increases progressively, while it decreases in test 4 compared against test 3, where we observe a significant negative estimate (*Estimate* = -0.35048, *SE* = 0.07446, *p* = 2.51e-06). With regards to the interactions, all were significant, particularly the one between condition and the test 4 vs. test 3 contrast (*Estimate* = 0.22534, *SE* = 0.09917, *p* = 0.023069. This contrast is connected to the timeframe of four weeks, where no lessons were held, and this was used to analyse retention rates.

Table 2   Overall DDL fixed effects: table of coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.23963	0.27444	0.873	
CONDITIONEXP	0.05129	0.27487	0.187	
TIME2-1	0.53413	0.06658	8.022	***
TIME3-2	0.66896	0.06770	9.881	******
TIME4-3	-0.35048	0.07446	-4.707	*
CONDITIONEXP:TIME2-1	-0.18781	0.09429	-1.992	***
CONDITIONEXP:TIME3-2	-0.33348	0.09198	-3.626	*
CONDITIONEXP:TIME4-3	0.22534	0.09917	2.272	

In figure 1, we see a visual representation of the predicted probabilities for accuracy in the two different conditions. The first thing that we notice is that in both groups we have two U-shaped learning patterns. The second thing that we notice is that the two U-shaped learning patterns have different characteristics: the variation in the control group is much larger in comparison to the experimental group. The predicted values in the experimental condition, in fact, are much closer, especially in relation to the difference between points c and d, which correspond to the differences between tests 3 and 4, this indicating retention rates (cf. § 3.1.). This leads to conclude that although there are no significant differences between the two conditions in relation to overall learning patterns over time, there are differences in terms of retention rates: the experimental condition seems to be characterised by an increased probability of determining better retention rates.

Figure 1 Overall DDL effects: plot of fixed effects in each condition

4.2 The Emic Perspective

We elicited the emic perspective on DDL effects by means of a set of likert-scale items and open-ended questions. In relation to the first ones, the mean and standard deviation values, partly reported in Forti (2017), can be found in table 2. The total number of questionnaires collected from the experimental groups was 50.

The first four likert-scale items focused on general aspects of the pedagogical treatment, while the second group of four likert-scale items focused specifically on aspect related to the DDL component of the pedagogical treatment.

In the first item, we look at learner attitudes in relation to working with word combinations: how useful did they find it? We see that a vast majority agrees or totally agrees on the idea that working on word combinations is useful, and the responses have the third lowest SD value (M = 5.42; SD = 0.94). We then asked whether students thought that the group work, based on collaborative pattern hunting, was useful or rather slowed down their learning. Most respondents either disagree or partially disagree on the idea that collaborative work slowed them down (M = 2.86), but the distribution of the responses is quite varied (SD = 1.30).

Third, we looked at whether the comments provided on the homework were seen as useful. The homework that was given at the end of each lesson typically involved writing a specific kind of text (e.g. a letter, a dialogue, a story, etc.) using the eight verb + noun collocations that were seen during the lesson. In this case, we notice that a large majority (M = 5.48) found the comments useful, and the distribution is the lowest out of all the results (SD = 0.88), indicating the highest degree of agreement.

We then asked whether working on eight collocations in an hour was too challenging. Here we see that the mean value (M = 2.12) sits between "disagree" and "partially disagree", though closer to the former. The variety of answers is higher than in other cases (SD = 1.05) indicating some degree of disagreement among respondents.

The fifth item is the first one where we seek to elicit learner attitudes towards one key aspect of the DDL pedagogical intervention: whether reading groups of sentences containing the same combination was confusing. Here we obtained a mean value (M = 3.60) with the highest degree of standard deviation (SD = 1.56), sitting between "partially disagree" and "partially agree". This would indicate that the approach did indeed produce some confusion.

This, however, does not seem to prevent the learners from seeing concordances as a useful resource to work with. In the following item, in fact, we ask whether the learner saw some usefulness in engaging in activities based on concordances, in terms of understanding how a combination is used. Here, the learners seem to general-

ly "agree" or "totally agree" about the usefulness of the approach in developing greater awareness in relation to word combination usage (M = 5.20; SD = 1.14).

Then we went a step further, asking whether working on concordances was perceived as something that could help learners make fewer errors in the future. This time, the responses are not as varied as in item 6 (SD = 0.92) and learners generally agree that this would be the case (M = 5.08).

Finally, we asked whether the learners thought that a smartphone application with groups of sentences instead of definitions would be useful. Here, the responses were quite varied (SD = 1.55), with most learners mildly agreeing that it would be useful (M = 2.64).

Table 3 Likert-scale item results

Item no.		M	SD
1	Learning word combinations was useful	5.42	0.94
2	Working in groups with my peers slowed down my learning	2.86	1.30
3	The comments on my homework helped me to improve my writing	5.48	0.88
4	Engaging in activities on 8 word combinations in one hour was too challenging	2.12	1.05
5	Reading groups of sentences containing the same combination confused me	3.60	1.56
6	The observation of groups of sentences containing the same combination has helped me to understand how to use that combination in the future	5.20	1.14
7	The groups of sentences will help me make fewer errors in the future	5.08	0.92
8	A new smartphone application with groups of sentences for word combinations would be useless	2.64	1.55
Scale: 1 = totally disagree; 2 = disagree; 3 = partially disagree; 4 = partially agree; 5 = agree; 6 = totally agree.			

If we group the "totally agree/agree" and "totally disagree/disagree" attitudes, and normalise them according to whether they were worded positively or negatively, we can obtain an overall picture of which aspects of the DDL treatment determined the most and least favourable attitudes. This picture is visible in figure 2.
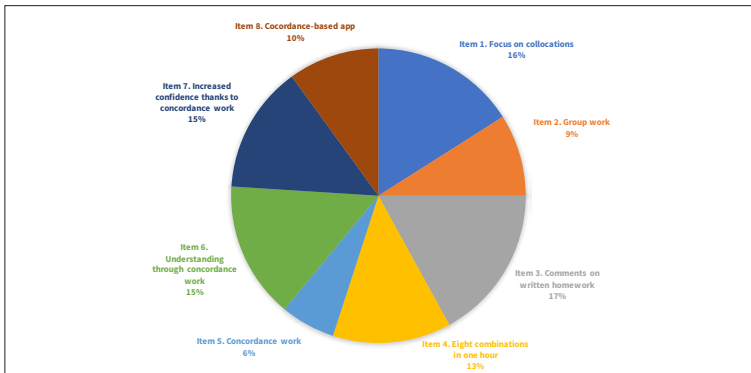
Figure 2 Summary of favourable learner attitudes towards the overall DDL treatment

As we can see, the largest favourable attitudes are observable in the comments on the written homework (17%), while the least favourable responses are seen in the concordance work (6%). This may have been due to the fact that the learners were not provided with preliminary training on concordance-based materials. Despite this, the perception of increased understanding and confidence in language use seems to be shared by a considerable proportion of learners (15% and 14% respectively).

We now move on to the open-ended questions. Table 4 shows the top three responses given for each question. In the first question we asked the learners to indicate what they liked most of the lesson: a vast majority of them, 18, liked working with word combinations, 8 enjoyed playing games and being part of different class teams, and 7 liked working with peers. When asked about what they least liked about the lessons, the first aspect was related to having to sit four tests over 12 weeks, and the second was that time was too short. And when asked to describe the lessons with three adjectives, the most frequent ones were: *interesting*, *useful* and *energic*. Finally, when asked to provide ideas or make suggestions to improve the course, of the few who responded, the main idea emerging apart from wanting fewer tests and more time was to be able to extend the activities into a story or dialogue with other students, so that the patterns observed through the concordances could be remembered better.

Table 4 Open-ended question results

Question 1: What did you like most of the lessons?	
themes	*occurrences*
Learning word combinations	18
Playing games / competitions among class teams	8
Speaking / working with my peers	7
Question 2: What did you like least of the lessons?	
themes	*occurrences*
Nothing / I liked everything	16
Too many tests	8
Time was too short	5
Question 3: Describe the lessons with three adjectives	
themes	*occurrences*
Interesting	35
Useful	30
Energic / Active	7
Question 4: Any ideas or suggestions?	
themes	*occurrences*
Extend activities into a story or a dialogue with other students	3
Less of the same tests	3
More of this course	3

The overall attitudes of learners expressed through the open-ended questions are therefore widely positive, with a dislike for tests emerging in questions 2 and 4, a largely positive attitude towards word combinations in question 1. With more time, concordance-based activities would have definitely been used as a basis to develop further non-DDL activities so as to foster the mechanisms of recycling and interiorisation. The fact that this is suggested by three different students shows their sensitivity and awareness in relation to what could work for their language learning.

5 The Usefulness of an Integrated View of DDL Effects

So how can the emic and etic perspectives be combined into an overall view of DDL effectiveness? If we look at the etic perspective, we may conclude that DDL seems to work only in relation to retention rates, since no significant differences are seen in terms of probabilities of accuracy in comparison to a control, non-DDL condition. The possible reasons for these findings are many. We will mention three.

First, the learners had limited exposure to the DDL approach, both in terms of lesson duration and in terms of overall course length: the amount of sessions that has been seen as necessary for seeing poten-

tial gains in language competence is 10 (Lee et al. 2018). Second, the most significant language gains in relation to DDL have been found in studies with experimental samples involving less that 50 participants (Boulton, Cobb 2017): in our case, the experimental sample of participants reached a total of 62 learners. Third, the study adopted a between-groups design, which makes it harder to detect differences in the participant samples involved, in comparison to within-groups designs where a single sample is exposed to different treatments (Boulton, Cobb 2017; Lee et al. 2018). In addition, it should be mentioned that the non-significant results related to DDL effectiveness in this study could be due to the approach not being particularly effective in this particular context.

When we look at the emic perspective, we see that overall learners find the approach interesting and useful. They particularly enjoy working on word combinations and are able to think about how the concordance work can be extended and used within non-DDL activities. The fact that the concordance work itself was found to be difficult, despite its perceived usefulness, is important for us because it provides insight into another possible aspect influencing the result on the etic perspective, namely the design of the DDL pedagogical materials. It is possible that they did not perfectly match the learners' needs, or that they needed to be preceded by preliminary training sessions. Unfortunately, neither of these two possibilities were an option in the current study but could be considered in future research.

What we are keen to highlight here is the potential of how an emic finding, such as the one related to concordance lines, can inform the interpretation of an etic finding, but more importantly how the interplay between the two can inform study designs for future research on DDL effects. Furthermore, the integration of the two perspectives into a single study means that the risk of 'comparing apples and oranges' diminishes considerably.

6 Conclusions

This article presented an example of how an empirical study of DDL effects can be conducted by integrating two different perspectives: one related to the development of language competence, and the other related to the attitudes of the learners engaging in it.

In the context of a specific study involving Chinese learners of Italian, we found that the results emerging from two perspectives tell two different stories about DDL and its effects: apart from improving retention rates, DDL is unlikely to determine better language gains over time in comparison to non-DDL approaches, but it produces positive attitudes amongst the learners, who are even able to envision how the concordance work can be incorporated into other stages of a

lesson. These stories are of course bound to this particular study: the etic finding on language gains could be compared to a longer study in order to see whether in this case it was simply time that was not sufficient to observe changes in language gains, and the emic finding can be compared to cases where the learners belong to language groups other than Chinese.

But what is evident in this study is how the two stories can speak to one another, informing the way in which they can be interpreted in relation to their single components and in relation to the study as a whole.

Acknowledgement

Bibliography

Boulton, A.; Cobb, T. (2017). "Corpus Use in Language Learning: A Meta-Analysis". *Language Learning*, 67(2), 348-93.

Chambers, A. (2007). "Popularising Corpus Consultation by Language Learners and Teachers". *Language and Computers*, 61(1), 3-16.

Dörnyei, Z. (with Tatsuya Taguchi) (2010). *Questionnaires in Second Language Research. Construction, Administration, and Processing*. 2nd ed. New York; London: Routledge.

Forti, Luciana (2017). "Data-Driven Learning and the Acquisition of Italian Collocations: from Design to Student Evaluation". Borthwick, K.; Bradley, L.; Thouësny, S. (eds), *CALL in a Climate of Change: Adapting to Turbulent Global Conditions – Short Papers from EUROCALL 2017*. Research Publishing, 110-15. https://doi.org/10.14705/rpnet.2017.eurocall2017.698.

Headland, T.N.; Pike, K.L.; Harris, M. (eds) (1990). *Emics and Etics. The Insider/Outsider Debate*. London: SAGE Publications.

Lee, H.; Warschauer, M.; Lee, J.H. (2018). "The Effects of Corpus Use on Second Language Vocabulary Learning: A Multilevel Meta-Analysis". *Applied Linguistics*, 40(5), 721-53.

Linck, J.A.; Cunnings, I. (2015). "The Utility and Application of Mixed-Effects Models in Second Language Research: Mixed-Effects Models". *Language Learning*, 65(S1), 185-207.

Mizumoto, A.; Chujo, K. (2015). "A Meta-Analysis of Data-Driven Learning Approach in the Japanese EFL Classroom". *English Corpus Studies*, 22, 1-18.

Mizumoto, A.; Chujo, K.; Yokota, K. (2016). "Development of a Scale to Measure Learners' Perceived Preferences and Benefits of Data-Driven Learning". *ReCALL*, 28(02), 227-46.

Morris, M. (2012). *Concise Dictionary of Social and Cultural Anthropology*. Malden (MA); Oxford: Wiley-Blackwell.

Pike, K.L. (1967). "Emic and Etic Standpoints for the Description of Behavior". *Language in Relation to a Unified Theory of the Structure of Human Behavior*. Berlin, Boston: De Gruyter, 37-72.

Riazi, A.M. (2017). *Mixed Methods Research in Language Teaching and Learning*. Sheffield; Bristol: Equinox.

Spina, S. (2014). "Il Perugia Corpus: una risorsa di riferimento per l'italiano. Composizione, annotazione e valutazione". *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & the Fourth International Workshop EVALITA 2014* (Pisa, 9-11 December 2014), vol. 1. Pisa: Pisa University Press, 354-9.

Spina, S. (2019). "The Development of Phraseological Errors in Chinese Learner Italian: a Longitudinal Study". *Widening the Scope of Learner Corpus Research. Selected Papers from the Fourth Learner Corpus Research Conference*. Louvain-la-Neuve: Presses Universitaires de Louvain, 95-119.

Venables, W.N.; Ripley, B.D. (2002). *Modern Applied Statistics with S*. 4th ed. New York: Springer.