Verb Constructions in Learners' Dictionaries

Paul Bogaards, Willem A. van der Kloot
Department of French, Department of Psychology
Leiden University
POB 9515
NL-2300 RA Leiden
The Netherlands
P.Bogaards@let.leidenuniv.nl

Abstract

Bogaards and Van der Kloot (2001) did not find evidence that differences between the systems used for giving information about verb completion in three learners' dictionaries produced consistent differences in usability and correctness. Nevertheless, it is not satisfactory to conclude that the differences that exist in the ways dictionaries present syntactical information about possible or obligatory constructions with verbs make no difference at all for the user. The research to be reported in this paper again tries to investigate differences among dictionaries, but focuses more on aspects of usability, excluding aspects of findability, which may have been one of the confusing factors in the former experiment. Four sources of information about verb constructions are distinguished and tested for their frequency of use and for their usefulness. Three groups of high intermediate and advanced students of English (total N = 117) underlined the information used in manipulated parts of lemmas and completed the translation of twelve Dutch sentences. The results of this experiment indicate that different groups of subjects show different patterns of preference for different types of information, and that all types of information lead generally to high proportions of correct translations.

Introduction

Recently Bogaards and Van der Kloot [2001] studied the usefulness of grammatical information that is given with verbs in English learners' dictionaries. In the first part of their study they described the evolution of this type of information from about 1970 onwards. It appears that each learners' dictionary and each new edition of the existing ones offers new schemes for verb completion. Broadly speaking there is a tendency to give ever more explicit information on verbal constructions: lists of codes are replaced with more mnemotechnical indications and word class information tends to disappear to make place for prepositions or other verb completion elements that are obligatory with a given verb.

In their empirical study Bogaards and Van der Kloot [2001] set out to investigate the comparative usefulness of the rather different systems of grammatical information that can be found in the latest editions of three learners' dictionaries [CIDE 1995; Cobuild2 1995; LDOCE3 1995]. They asked subjects to finish the translation of a number of Dutch sentences using the relevant information on verb construction that could be found in entries taken from three different learners' dictionaries. This study researched two aspects of dictionary use: findability, measured as the number of seconds needed to find the necessary information, and usability, measured as the correctness of the translations given by the subjects.

The outcomes of this study were rather inconclusive. No clear-cut evidence was found regarding overall differences between the three dictionaries used in the experiment. Although there were indications that suggested a slight preference for Cobuild2 over LDOCE3 as far as correctness is concerned, and a preference for both of these dictionaries over CIDE, the differences were not statistically significant.

As this study did not yield concrete suggestions for the improvement of the existing systems, we decided to take a further step into the discovery of any differences there might be in spite of the above, not very satisfactory conclusions. One of the confusing factors in the earlier experiment could be the fact that the subjects had a rather complex, albeit realistic, task to fulfil. Apart from the translation task as such, they had to browse through an entire dictionary entry in order to find the relevant meaning, to interpret definitions, to understand examples and to interpret and adapt the grammatical information given. As entries are very differently structured and do not contain the same information in different dictionaries, the task set for the subjects may not have been sufficiently similar in the three different dictionaries used in the experiment. Although length of entries and place of the relevant information in the entries had been controlled, there remain many other aspects that have nothing to do with grammatical information as such but that could possibly explain the absence of significant differences between the dictionaries.

In the experiment to be reported in this paper, we preferred to focus more exclusively on the aspect of usability, excluding the aspect of findability. The subjects again had to finish the translation of a number of Dutch sentences having relevant dictionary information at their disposal. However, in this case only the relevant meaning of each verb was presented. In addition, the information given about this meaning was standardised so as to permit systematic manipulation.

Analysis of a number of verb entries makes it clear that information about the possible or obligatory constructions to be found in the recent editions of learners' dictionaries is given in four different ways:

- grammatical information can be given in the form of *grammatical codes* like **V n**, **V prep n**; this is what can be found in the extra column of Cobuild2. We will call this type of information **C** (for grammatical code);
- grammatical information can also be given in a more *explicit* way, as is done, for instance in LDOCE3, where we find things like **be charged with** or **reward sb with sth**. We will call this Info Type **E** (for explicit grammatical information);
- grammatical information can be given *implicitly* in the context of a definition, as is done in Cobuild2 and, but not as systematically, in CIDE; this leads to formulations like "When the police **charge** someone, ..." (to be called Info Type **D**, for definition);
- grammatical information is given most of the time in the context of *examples*, as in *She's been charged with fraud* (Info Type **X**, for example).

All of the current learners' dictionaries combine several of these possibilities, so as to give their users a choice as to what (combination of) information suits them best. The main aim of the experiment to be presented here was to find answers to the following questions:

1. which type of information is used most?

2. which type of information is most useful?

Method

Materials

The same twelve verbs in the same contexts have been used as in the experiment run by Bogaards and Van der Kloot [2001], that is to say: frequent verbs in rather unfamiliar constructions. Each verb was presented with a specific combination of types of information, but each verb had in all cases the same two examples. Codes, explicit and implicit information were standardised. Two versions of the test were prepared. The first item in version A presents a combination of implicit information in the definition, grammatical code and two examples (Info Type DCX):

Hij wordt nu officieel van diefstal beschuldigd.

CHARGE
...
• When the police **charge** someone, they formally accuse them of having done something illegal. | **V n, V n *with* n** *Police have charged Mr Bell with murder. ... She's been charged with fraud.*

He ..... now ..... officially ..... theft.

This same item has the following form in the B version, where the grammatical code has been replaced with explicit information (Info Type DEX):

Hij wordt nu officieel van diefstal beschuldigd.

CHARGE
...
• When the police **charge** someone, they formally accuse them of having done something illegal. | **be charged with** *Police have charged Mr Bell with murder. ... She's been charged with fraud.*

He ..... now ..... officially ..... theft.

An example of an item with explicit information only is (Info Type EX):

Hij wordt nu officieel van diefstal beschuldigd.

CHARGE
...
• to be accused of having done something illegal. | **be charged with** *Police have charged Mr Bell with murder. ... She's*

been charged with fraud.

He ..... now ..... officially ..... theft.

The verbs were in the same order in both versions. The distribution of the different types of information over the two versions was as follows:

	Version 1	Info Type	Version 2	Info Type
sentence 1 - 4	definition + codes	DCX	definition + explicit	DEX
sentence 5 - 8	definition + explicit	DEX	explicit only	EX
sentence 9 -12	definition only	DX	definition only	DX

As already said, all items had always two examples which were the same in both versions. This design makes it possible to make a number of interesting comparisons between types of information as well as between types of students.

Subjects

There were three groups of subjects. The first group consisted of 33 secondary-school students who had taken English courses during at least six years. The second group consisted of 56 first year university students of English; they had taken English courses during at least eight years. The third group consisted of 28 third year university students of French or Spanish. This latter group was included because their level in English is comparable to that of the second group, but as students of other languages they may be taken to be more familiar with several types of coded grammatical information. All subjects had Dutch as their mother tongue.

Procedure

The tests were presented in the context of a normal class period. They took about 15 minutes to be completed. Subjects were presented with booklets consisting of four pages. The first page explained the procedure and contained an example. Students were asked to write down some factual information such as name of school, class or year, sex, age and native language. The three remaining pages contained four items each; all items on a given page were of the same combination of Info Types. The subjects had to underline the information or any combination of information that had helped them to find the solution they believed to be correct, and then had to write down that solution.

Results

Use of Dictionary Information

Table 1 presents the relation between Info Type and Student Type on the one hand and the choices the subjects have made (Info Use) on the other hand, for Version 1 and 2 respectively. The leftmost entries within each cell are the row-wise percentages of responses falling in one of the categories that denote the element or combination of elements that were underlined in the information presented with each verb. The responses were aggregated over the participants of each Student Type and over the four verbs that were presented within each category of Info Type. It should be noted that several cells in the table are empty. Those cells correspond with so called 'structural zeroes', that is, cells that are necessarily empty because of the manner in which the study was designed. For instance, the Info Use

Info Type	Student Type	Gramcode	gramcode + example	Explicit	explicit + example	definition	definition + example	example	other combinations	Total[a] observed
					Version 1					
DCX										
1-4	Sec. School	0	0 -	-		30.3 *65.2*	31.6 *79.2*	38.2 *69*	0	76
	English 1	3.7 *100*	5.5 *83.3*	-	-	20.2 *95.5*	19.3 *95.2*	40.4 *88.6*	11 *91.7*	109
	Fr./Sp. 3	5.8 *100*	3.8 *100*	-	-	17.3 *100*	28.8 *86.7*	30.8 *81.3*	13.5 *85.7*	52
DEX										
5-8	Sec. School	-	-	5.5 *50*	6.8 *100*	13.7 *30*	23.3 *70.6*	47.9 *57.1*	2.7 *50*	73
	English 1	-	-	25 *96.3*	10.2 *100*	7.4 *87.5*	8.3 *77.8*	38.9 *90.5*	10.2 *100*	108
	Fr./Sp. 3	- -	-	27.5 *100*	9.8 *40*	5.9 *100*	7.8 *100*	25.5 *92.3*	23.5 *91.7*	51
DX										
9-12	Sec. School	-	-	-	-	27.8 *75*	16.7 *75*	55.6 *82.5*	-	72
	English 1	-	-	-	-	45.7 *81.3*	15.2 *75*	39 *78*	-	105
	Fr./Sp. 3	-	-	-	-	31.4 *87.5*	39.2 *85*	29.4 *80*	-	51
					Version 2					
DEX										
1-4	Sec. School	-	-	17.9 *90*	17.9 *80*	10.7 *83.3*	23.2 *100*	25 *92.9*	5.4 *66.7*	56
	English 1	-	-	41.7 *82.2*	19.4 *90.5*	8.3 *77.8*	1.9 *50*	23.1 *88*	5.6 *100*	108
	Fr/Sp. 3	-	-	32.1 *100*	26.8 *100*	10.7 *100*	3.6 *100*	14.3 *100*	12.5 *71.4*	56
EX										
5-8	Sec. School	-	-	24.5 *100*	20.8 *90.9*	-	-	54.7 *86.2*	-	53
	English 1	-	-	45 *71.4*	26.6 *86.2*	-	-	28.4 *74.2*	-	109
	Fr./Sp. 3	-	-	32.7 *83.3*	29.1 *93.8*	-	-	32.7 *72.2*	-	55
DX										
9-12	Sec. School	-	-	-	-	18.2 *90*	30.9 *88.2*	50.9 *71.4*	-	55
	English 1	-	-	-	-	38.7 *87.8*	11.3 *75*	50 *73.6*	-	106
	Fr./Sp. 3	-	-	-	-	32.1 *88.9*	17.9 *90*	50 *82.1*	-	56
Total[b]		3.0 *100*	3.4 *87.5*	34.9 *85.4*	21.7 *89.4*	20.6 *82.3*	16.0 *83.5*	40.7 *79.6*	9.2 *88.9*	

Table 1. Info Use and Correctness as a function of Info Type and Student Type for Version 1 and 2. Entries in normal typeface are row percentages indicating the number of responses in each Info Use category relative to the total number of responses within each combination of Version, Info Type and Student Type. Entries in italics indicate the percentages of correct translations within each cell. [a]Due to missing values, the totals differ from the maximum possible values. [b]Normal face entries in the row 'Total' are percentages of the total number of cases in which the column category was possible.

Version 1. Table 1 shows that for Version 1 the vast majority of the responses are in the Info Use categories 'example', 'definition', and the 'definition + example'. 'Example' has the highest frequencies within Info Type DCX for all student groups; the same holds within Info Type DEX and DX for the secondary school participants. The first year students of English, show a higher frequency for the use of 'definition', whereas the third year students of French or Spanish had a higher frequency of using a combination of 'definition + example'. Under Info Type DEX, when explicit information on verb usage was presented, only the university students used this type of information to an appreciable degree. Grammatical codes (Info Type DCX) were hardly used by any of the participants. These conclusions, reached from an intuitive inspection of the Version 1 data, suggest that two separate effects can be delineated: (a) an interaction between Student Type and Info Use, and (b) an interaction between Info Type and Info Use.[1]

Version 2. Inspection of the Version 2 results, leads to the conclusion that university students tend to prefer 'explicit' information on verb usage to mere 'examples' of usage, when this type of information is presented. Secondary school students, on the other hand, more often use 'examples' and combinations of 'definition + example', even when other information is available. These findings suggest (a) that there is an interaction between Student Type and Info Use, and (b) that there is no interaction between Info Type and Info Use[2].

Correct Translations

For both versions of the questionnaire each participant's proportions of correct translations were computed within the three sets of four items that differed with regard to the information given. The averages of these proportions for the various combinations of Info Type and Student Type are represented in Figure 1, for each version separately. This figure indicates that there are some marked differences between the secondary school students on the one hand and the university students on the other hand. First of all, the university students have overall higher proportions of correct translations than the secondary school students. The performances of the university students are closer to the maximum attainable proportion of 1.00 and therefore have smaller variations (smaller ranges) than the secondary school students. Secondly, the performances of the first year students of English and the third year students of French or Spanish hardly show any differences.

To test the effects of Info Type, Student Type and their interaction we have run analyses of variance (ANOVA) after applying an inverse sine transformation [cf. Kirk, 1995: 106].[3] Although one of the aims of this transformation is to homogenize the within group variances, its effect was not sufficient[4]. Because of the heterogeneity of the within group variances and covariances we have performed two sets of separate ANOVAs on the data of the secondary school students and on the data of the university students.

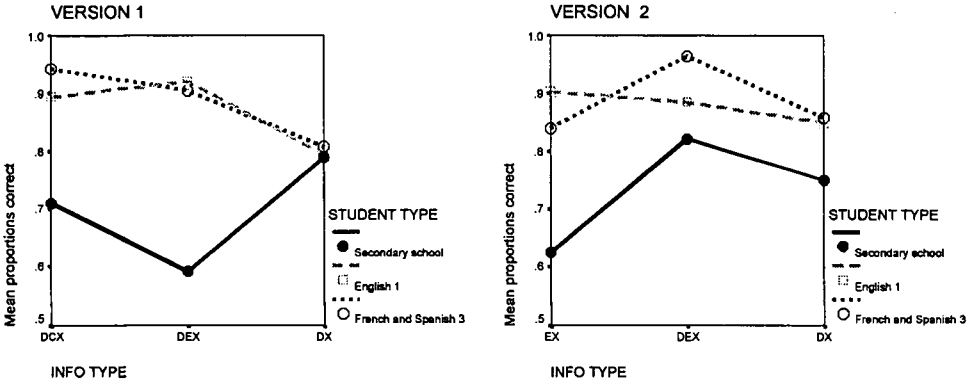

Figure 1. Effects of Info Use and Student type on the proportions of correct translations (left figure: Version 1; right figure: Version 2).

University students. Neither in Version 1 nor in Version 2 differences were found between the two groups of university students. As a matter of fact the only significant effect in the data of the university students was for Info Type in Version 1 ($F = 7.523$; $df = 1.87, 73.05$; $p = .001$). This effect is caused by a significantly smaller proportion of correct applications for Info Type DX as compared to Info Types DCX and DEX ($F_{1\text{-}3} = 9.495$; $df = 1, 39$; $p = .004$; $F_{2\text{-}3} = 11.288$; $df = 1, 39$; $p = .002$), whereas the difference between the latter two Info Types was negligible and nonsignificant ($p \approx 1.00$). Therefore, when the information in Version 1 consisted only of the definition of the verb and an example of its usage, the university students produced a smaller number of correct applications than when this basic information was supplemented by additional grammatical elements (C or E). However, this result was not replicated in Version 2, as Info Type DEX and Info Type DX were not found to differ. This indicates that the effects of Info Type depend on the particular sets of verbs with which the DEX information was combined. It is unlikely that this version effect is caused by differences between the groups of participants who completed the different versions: between-subjects ANOVA on the transformed proportions correct of Info Type DX neither showed a significant difference between the two versions ($p = .610$) nor a significant interaction between Version and Student Type ($p = .595$).

Overall, the data of the university students appear to suffer from a 'ceiling effect', that is, the vast majority of these participants are able to apply the verbs correctly, regardless of the information presented and the language they happen to study.

Secondary school students. For the secondary school students no significant Info Type differences were found in Version 2, whereas Version 1 yielded a significant difference between Info Types DEX and DX ($F_{2\text{-}3} = 4.690$; $df = 1, 18$; $p = .044$). This effect was caused by a smaller proportion of correct applications with Info Type DEX than with Info Type DX. No significant differences occurred between Info Types DCX and DEX ($p = .182$) and between DCX and DX ($p = .208$). Again this difference was not replicated with the items of Version 2.

Info Type	Student Type	gramcode	gramcode + example	explicit code	explicit + example	definition	definition + example	example	other combinations	Total observed
					Version 1					
DCX										
1-4	Sec. school	0	0	-	-	23 *65.2*	24 *79.2*	29 *69.0*	0	76 *71.1*
	University	7 *100*	8 *87.5*	-	-	31 *96.8*	36 *91.7*	60 *86.7*	19 *89.5*	161 *90.7*
DEX										
5-8	Sec. school	-	-	4 *50.0*	5 *100*	10 *30.0*	17 *70.6*	35 *57.1*	2 *50.0*	73 *58.9*
	University	-	-	41 *97.6*	16 *81.3*	11 *90.9*	13 *84.6*	55 *90.9*	23 *95.7*	159 *91.8*
DX										
9-12	Sec. school	-	-	-	-	20 *75.0*	12 *75.0*	40 *82.5*	-	72 *79.2*
	University	-	-	-	-	64 *82.8*	36 *80.6*	56 *78.6*	-	156 *89.0*
					Version 2					
DEX										
1-4	Sec. school	-	-	10 *90.0*	10 *80.0*	6 *83.3*	13 *100*	14 *92.9*	3 *66.7*	56 *89.3*
	University	-	-	63 *87.3*	36 *94.4*	15 *86.7*	4 *75.0*	33 *90.9*	13 *84.6*	164 *89.0*
EX										
5-8	Sec. school	-	-	13 *100*	11 *90.9*	-	-	29 *86.2*	0	53 *90.6*
	University	-	-	67 *74.6*	45 *88.9*	-	-	39 *92.3*	3 *100*	164 *78.7*
DX										
9-12	Sec. school	-	-	-	-	10 *90.0*	17 *88.2*	28 *71.4*	-	55 *80.0*
	University	-	-	-	-	59 *88.1*	22 *81.8*	81 *76.5*	-	162 *81.5*
Total		7 *100*	8 *87.5*	198 *85.4*	123 *89.4*	249 *82.3*	194 *83.5*	509 *79.6*	63 *88.9*	1351 *83.0*

Table 2. Number of responses and percentages of correct translations (italics) within each combination of Version, Info Type and Info Use for secondary school and university students.

Relation of Info Type Used and Correctness.

In Table 1 each cell contains a number in italics, which indicates the percentage of the responses in the cell that led to a correct translation. For instance, the first italic number on the first line of the table body, indicates that of the responses in the category 'definition' given by the secondary school students on the DCX items of Version 1, 65.2% were correct. Apart from overall lower values for the secondary school students, these percentages do not seem to show consistent patterns of differences between the student groups. As the university students do hardly differ from each other we have aggregated the data over the university students, which leads to the cross tabulation presented in Table 2. The figures in normal typeface indicate the number of responses in each combination of Version, Info Type and Info Use. The numbers in italics are the percentages of correct responses within the cells. For instance: in Version 1, Info Type DEX, 41 responses of the university students fell into the category 'explicit'; 97.6% of these responses corresponded with a correct translation.

Again, Table 2 does not reveal a distinct pattern of conditions with systematic differences in correct translations. The differences that are present, however, probably must be explained by differences among the items used as stimulus material.

Discussion and Conclusions

Our experiment seems to lead to some broad conclusions. First, traditional grammar codes in terms of word classes seem to be only very rarely used by dictionary users, even by those who may be assumed to have a certain level of linguistic schooling like the third year students of French or Spanish. Examples are widely used to gather information about verb constructions, especially by high intermediate learners like our secondary school students, but also by advanced learners like our first year students of English or third year students of French or Spanish. The groups of university students tend however to prefer grammatical information that is given in a less abstract way whenever this is available. This type of information is not used very often by less advanced learners. The implicit information that is given in Cobuild-style definitions does not seem to be used as it could be, that is to say as examples of the use of the verbs, but most of all as definitions that give information about the meaning of these verbs. It should be remarked, however, that in many cases the choice of 'definition' and 'definition + example' taken together yield rather high percentages. The groups of university students, and especially the students of English, show high proportions of choices for 'definition' or 'definition + example', and even more so when no explicitly formulated grammatical information is available.

Grammar codes seem to be chosen only by those who feel confident with this type of information, as is clear from the high levels of correct translations for this category (see bottom rows Tables 1 and 2). A high proportion of correct solutions is found also with the category 'other combinations', where subjects often underlined more than two different types of information. For the other categories the correct scores are also generally high, that is, between 80 and 90%. Generally speaking, one can say that once the users have found the information they need, they manage to work out the right solution in the majority of cases. The percentages of correct solutions are higher for the advanced learners than for the high

intermediate ones. Whereas the university students generally found correct solutions in more than 4 out of 5 cases, the secondary school students did so in approximately 3 to 4 out of 5. Even this latter proportion is a satisfactory result. The high level of correct use attained by the university students suggests that they have learnt to handle dictionaries.

As learners' dictionaries are meant for advanced learners, it is important to ascertain what is meant by 'advanced'. It is clear from this study that different groups of subjects show different patterns of preference: secondary school students like examples most (as did the subjects of several experiments mentioned by [Tono 2001 : 35]), whereas university (language) students prefer explicitly given grammatical information (but not traditional grammar codes; cf. [Bogaards & Van der Kloot 2001 : 118]). Even though the secondary school subjects have somewhat lower correct scores than the university students, they reach acceptable levels and are able to use dictionaries meant for advanced learners. As very few subjects did choose traditional grammar codes, it is questionable whether learners' dictionaries should present this type of information. As to the Cobuild-style definitions, one may wonder whether they are recognised as an important source for productive information.

Let us return, finally, to the two research questions that were asked at the end of the introduction. The first question, about the type of information that is used most, has received a rather clear answer, even though the answer has to be differentiated for more advanced and less advanced users. As to the second question, which asked what is the most useful information, the answer is less clear: no one type of information yields significantly higher correct scores than any other. It seems as if users intuitively know how to gather the information that suits them best. Once they have found this information, they manage to use it correctly in most of the cases.

Acknowledgements

We would like to thank Cora Vis, Dr. Laura Perez Rosales and Dr. Robert Lankamp for their kind collaboration.

Endnotes

[1] A loglinear analysis of Table 1 that took account of the structural zeroes, showed that the above two interactions were indeed significant ($G^2_{ST*IU} = 66.089$, $df = 14$, $p = < .001$; $G^2_{IT*IU} = 32.165$, $df = 9$, $p < .001$). Only these two interactions were necessary to account satisfactorily for the data: the three-way interaction between Info Type, Student Type, and Info Use, turned out to be nonsignificant ($G^2_{IT*ST*IU} = 16.474$, $df = 10$, $p = .09$).

[2] Loglinear analysis (taking the structural zeroes into account) supports the above conclusions. A significant interaction between Student Type and Info Use ($G^2_{ST*IU} = 54.878$, $df = 10$, $p < .001$) was found, together with a nonsignificant interaction of Info Type and Info Use ($G^2_{IT*IU} = 5.056$, $df = 4$, $p = .282$) and a nonsignificant three-way interaction of Student Type, Info Type and Info Use ($G^2_{IT*ST*IU} = 5.650$, $df = 8$, $p = .69$).

[3] The inverse sine transformation is suitable when the data are proportions correct in a fixed number of trials. The formula is $Y = 2\arcsin\sqrt{prop}$ with *prop* = the proportion of correct responses. If *prop* = 1.00 or *prop* = 0, the values .9375 and .0625 were substituted, respectively.

[4] Box *M* tests showed that there was still significant heterogeneity of the within group variances (Version 1: $F = 2.461$, $df = 12, 7424.3$, $p = .003$; Version 2: $F = 2.399$, $df = 12, 6907.4$, $p = .004$), notably because of the differences between the university and the secondary school students.

References

Dictionaries

[CIDE 1995] Procter, P. (ed.), 1995. *Cambridge International Dictionary of English.* Cambridge University Press, Cambridge UK.

[Cobuild2 1995] Sinclair, J., P. Hanks, G. Fox, R. Moon & P. Stock (eds.), 1995. *Collins Cobuild English Dictionary* (Second edition). HarperCollins, London.

[LDOCE3 1995] Summers, D. (ed.) 1995. *Longman Dictionary of Contemporary English* (Third edition). Longman, London.

Other References

[Bogaards & Van der Kloot 2001] Bogaards, P. & W.A. van der Kloot, 2001. The use of grammatical information in learners' dictionaries. in: *International Journal of Lexicography 14 (2)*, 97 – 121, Oxford University Press, UK.

[Kirk 1995] Kirk, R.E., 1995. *Experimental design: Procedures for the behavioral sciences* (Third edition). Brooks Grove, Pacific Grove, Ca., USA.

[Tono 2001] Tono, Y. 2001. *Research on Dictionary Use in the Context of Foreign Language Learning. Focus on Reading Comprehension.* Tubingen: Max Niemeyer Verlag, Germany.