

CONTRASTIVE PERSPECTIVES ON THE RELIABILITY OF GRAMMAR TESTING: THE CASE OF THE VERB PHRASE IN L2 ENGLISH

In this article we present the preliminary results of empirical research on the reliability of grammar testing. The following testing techniques are examined: multiple choice, translation, gap-filling and, finally, performance. In particular, for the purposes of this paper we set up four different tests to identify knowledge of the English verb phrase among native speakers of Dutch. Drawing from a detailed examination of selected results it is concluded that a renewed interest in a contrastive approach to language testing and practice is to be recommended.

Key words: grammar testing, contrastive analysis, L2 English, the English verb phrase.

1. Introduction

This article reports on the preliminary results of an empirically-based investigation into selected grammatical issues in advanced-level non-native English. In particular, we set out to examine the reliability of the various techniques that we used to identify knowledge of the English verb phrase among Dutch-speaking first-year undergraduates at the Faculties of Applied Economics at two Belgian (i.e. Flemish) universities. For our purposes, reliability can be defined as the consistency of results from one test to another (cf. Bachman & Palmer 1997).

Within the framework of error analysis, the following testing techniques will be dealt with:

- multiple choice
- translation from Dutch (L1) into English (L2)
- gap-filling (contextualized completion assignments)

- performance: i.c. information-gap and decision-making dyads between testees¹.

The first three techniques are based on written tasks and they were used in different tests with different sets of between 400 and 500 testees. As far as the results are concerned, we will be presenting percentages, some of which are based on representative subcorpora. No statistical analysis has been carried out. The fourth and final test is based on oral communication: we used audio-recordings of dyads involving some 35 testees.

Crucially, the present article is meant to provide no more than a first exploration into what the pros and cons of the different techniques could be. That's why, from the wide range of data available, we will only focus on the following four specific features of the grammar of the English verb phrase, all of which tend to be problematic for speakers of Dutch: viz.

- information questions (or wh-questions)
- the past and past participle inflections of irregular verbs
- the continuative use of the present perfect
- absence of obligation

For all of these features of the grammar of the English verb phrase, the problems for native speakers of Dutch have been sufficiently described (cf. Zandvoort's 1975 *Handbook of English Grammar* as well as, more recently, Vandenberg, 1991 and Lemmens & Parr, 1995). Therefore, the purpose of our tests, while diagnostic, was not to make an inventory of problems but rather to point to a possible hierarchy of problems as well as formulating tentative hypotheses for refining the description.

2. Information questions

To start with, we shall look at information questions in which the question word is not the subject, as in translation-based multiple choice test (1). Since this is a part of basic grammar, few problems are to be expected, except perhaps for the use of the operator, which is unknown in Dutch.

¹ Egýud & Glover (2001) argue that such dyads lead learners to produce highly varied language. In addition, Garcia Mayo & Pica (2000) have shown that they are not significantly different from dyads between learners and native speakers.

(1) Wat betekent dat Engelse woord?

- a. What does that English word mean? (82%)
- b. *What means that English word? (4%)
- c. *What does that English word means? (14%)

The results are predictable and surprising at the same time. Predictable, because most of the testees opted for *a*. But surprising, because the major source of mistakes was not the use of the operator, but the inflection of the main verb.

Interestingly, translation (2) – with no possible answers suggested to the testees – yielded rather different results:

(2) Wat betekent dat Engelse woord waarover je gisteren sprak?

[What does that English word mean you talked about yesterday?]

First of all, it should be pointed that one out of six testees came up with a rather free translation, viz. a paraphrase, in which the problem of the use of the operator was avoided:

(2a) What is the meaning of that English word you talked about yesterday? (17%)

Of course, from a communicative point of view this answer is perfectly adequate but we should keep in mind that it does point to a first restriction in using translation to test students' knowledge of specific items of grammar (cf. Ellis 1998 for an overview of research on avoidance strategies in language learning and language testing).

Let's return to (2) now. The interesting thing is that this time only few testees came up with:

(2b) *What does that English word means you talked about yesterday? (4%)

In contrast, many more failed to use the operator, which points to the L1 transfer that we were expecting:

(2c) *What means that English word you talked about yesterday? (18%)

Indeed, it looks as if, by offering the right alternative, multiple choice seems to underestimate the testees' real problems (in this case the use of the operator). Its role seems to be limited to that of checking whether testees recognize the right form rather than testing active grammar knowledge. At the same time, the question can be raised whether presenting ungrammatical alternatives does not lead testees to make mistakes that they would not normally make (-s form).

Alternatively, it remains to be seen if in (2) – as for the use of the operator - L1 transfer was not induced by the translation format. The results for gap-filling test (3) seem to suggest that this is not the case, since they are almost identical to those for the translation test:

(3) What observable truth _____ (CONTAIN, IT)?

(3a) *What observable truth contains it? (14%)

(3b) *What observable truth does it contains? (4%)

Apparently, even in the gap-filling assignment, testees seem to resort to a translating strategy. This is in line with Mattar (1999), whose research actually reveals less mother tongue interference with translation tasks than with gap-filling.

Let's return to the use of the operator. Interestingly, in contrast with the results for information questions, our data show that the testees don't have a problem with the use of the operator in simple yes-no questions, like "Do you agree?" Hardly anyone comes up with "Agree you?" Why this difference? We believe that the answer may lie in the second type of information questions, viz. those in which the question word is the subject. For these no 'do' is required, as in gap-filling test (4):

(4) Who _____ (BELIEVE) that wars can really solve humanitarian problems?

Still, quite some testees used do-support here:

(4a) *Who does believe that wars can really solve humanitarian problems? (4%)

Note that on the basis of these criteria Quirk et al (1985) have seven classes. For the purposes of this article we have brought Quirk et al's first and second classes together in our first class and we have renumbered their third class as our second, their fourth class as our third, etc.

LEND	44%
SEND	54%
Class 2: suffixation, V-ed identity	
BUY	8%
FLEE	47%
KEEP	0%, 3%
MEAN	18%
SEEK	18%
Class 3: suffixation	
BREAK	4%
DRIVE	5%
FORGET	0%
KNOW	5%
Class 4: V-ed identity, vowel identity	
SET	5%
Class 5: V-ed identity	
GET	4%
LEAD	75%
MEET	1%
READ	6%
UNDERSTAND	4%
Class 6: /	
BEGIN	12%
RING	4%

Apparently, class 1 is the only class that poses major problems throughout. Certainly – and perhaps surprisingly -, those irregular verbs that have vowel variation (or so-called ‘ablaut’) seem far more simple, except for relatively unusual verbs like ‘seek’ and ‘flee’. The high scores for ‘mean’ and ‘lead’ can be related to confusion about spelling. Generally speaking, though, vowel variation seems relatively easy.

But why then do the verbs in class 1 prove so problematic? We think the problem lies in devoicing the final phoneme: while vowel variation is a

loud and clear indicator, one which speakers of Dutch also know in their L1, devoicing the final phoneme is more subtle and completely unfamiliar to them. It follows that Dutch-speaking learners of English should be made aware of the different classes of irregular verbs in general and of the voiced-voiceless opposition for a number of the class 1 irregular verbs in particular. There are two interesting points to be made about this contrastive issue: it was revealed by a gap-filling assignment (and not by translation) and it falls beyond the reach of translation-based practice.

4. The continuative use of the present perfect

The continuative use of the present perfect refers to the use of the present perfect (simple or progressive) 'to invoke a state of affairs that began in the past, has continued until the speaker's now and may be expected to endure into the future' (Mackenzie: 1997: 44). Dutch traditionally uses a simple present here (Lemmens & Parr: 1995: 36). This continuative use of the English present perfect was tested in the final two sentences of translation (6):

(6) Toen hij jong was, heeft hij een jaar als kelner gewerkt. Toen is hij verhuisd naar Londen. Daar studeert hij nu al sinds twee jaar aan de universiteit. Hij heeft er een nieuwe flat sinds april.

[When he was young, he worked as a waiter for a year. Then he moved to London. There he has studied/been studying at university for two years. He has had a new flat since April.]

The results are predictable:

(6a) *There he studies at university for two years. (42%)

(6b) *There he is studying at university for two years. (24%)

(6c) *He has a new flat since April. (67%)

Again the question can be raised whether or not this type of error was induced by the translation format of the test. So let's also look at gap-filling exercise (7):

(7) Since Nato's air campaign against Slobodan Milosevic got underway, Macedonia – a small, impoverished state that broke away from Yugoslavia in 1991 – ____ (FACE) a nightmare.

Present perfect	30%
Simple	15%
Progressive	15%
Present	49%
*Simple	17%
*Progressive	32%
Past	18%

Here, still a lot of students opt for the present. Just like for information questions, it looks as if the effect of translation should not be exaggerated. But there are two further remarks to be made.

First, quite some students decided to select a past, which indicates that the extract was misunderstood as representing a past time frame. This seems to be a recurring problem with discourse-based gap-filling exercises. While contextualization is more important for some grammatical issues (e.g. tenses) and less so for others (like inflection of irregular plurals), it should be pointed out that it can also complicate matters.

Second, it should be noted that in (7) the present progressive is more frequently used than the simple present. We would suggest that this can be linked up with the results for (8), which is the translation-based multiple choice version of (6).

(8) Toen hij jong was, heeft hij een jaar als kelner gewerkt. Toen is hij verhuisd naar Londen. Daar studeert hij nu al sinds twee jaar aan de universiteit. Hij heeft er een nieuwe flat sinds april.

When he was young, he worked as a waiter for a year. Then he moved to London. There he (1) ____ at university for two years. He (2) ____ a new flat since April.

- (1)
- a. *studies (4%)
 - b. *is studying (24%)
 - c. has studied (6%)
 - d. has been studying (66%)

- (2) a. *has (72%)
 b. *is having (12%)
 c. has had (12%)
 d. *has been having (4%)

(7) and (8) together seem to suggest that the testees are least likely to use a present perfect with state verbs, like 'have'. The reason could well be that, just like native speakers of English (cf. Mackenzie 1997: 44), they seem to prefer the present perfect progressive and that they tend to resort to the L1 (simple) present-tense equivalent as soon as the (present perfect) progressive variant is excluded. Compare also (9) and (10):

(9) My family ____ (OWN) this land for almost three centuries now. So I wouldn't dream of selling it.

Present perfect	20%	
Simple		18%
*Progressive		2%
Present	72%	
*Simple		70%
*Progressive		2%
*Past	8%	

(10) Tony Blair is an extraordinary politician, but the past few weeks he ____ (WORK) under extraordinary pressure.

Present perfect	58%	
Simple		10%
Progressive		48%
Present	6%	
*Simple		2%
*Progressive		4%
Past	36%	

Finally, also the results for (11) seem to confirm the state verb hypothesis.

(11) I ____ (LOVE) eating crisps since I ____ (BE) a kid.

LOVE		
Present perfect	12%	
Simple		10%
*Progressive		2%
Present	82%	
*Simple		80%
*Progressive		2%
*Past	6%	
BE		
Present perfect	8%	
*Simple		8%
*Progressive		0%
Present	2%	
*Simple		2%
*Progressive		0%
Past	90%	

It should be noted that, just like with the time phrase in the final sentence of (6) and (8), the time clause introduced by ‘since’ fails to induce a present perfect. Indeed, the fact that 8% used a present perfect in the subclause seems to indicate that ‘since’ seems to typically confuse the testees and not, as Petrovitz (1997) warns, that like other so-called cueing devices it serves to mislead them into believing that the imagined co-occurrence restrictions are uniquely defined.

5. Absence of obligation

In translation few testees made a mistake:

(12) Je hoeft niet zo hard te werken.

[You don't need/have to work so hard.]

(12a) *You must not work so hard. (1%)

(12b) *You should not work so hard. (3%)

Our intuition in interpreting these results was that there would have been more tokens of 'must not' if the Dutch version in our test had not been 'hoeft niet', which is a grammatically correct way of expressing absence of obligation in Dutch, but which is a marked form for the testees, almost all of whom are of Flemish origin. Under normal conditions they would start from 'moet niet' and this, through L1 transfer, might well lead them to choose 'must not'.

This is to some extent confirmed by the results for the multiple choice and gap-filling tests:

(13) For EU residents there are interesting tax regulations: they practically _____ any taxes.

- a. *mustn't pay (14%)
- b. *shouldn't pay (6%)
- c. don't have to pay (80%)

(14) You _____ (BELIEVE / add a modal verb: 'it is not necessary') in science fiction.

(14a) *You mustn't believe in science fiction. (9 %)

(14b) *You shouldn't believe in science fiction. (11%)

It should be noted that the prompt in (14), viz. 'it is not necessary', may have misled the participants into using 'mustn't' and 'shouldn't' in the same way that the translation was probably misleading in (12). Certainly, the prompt was unclear because a massive 37% of the testees came up with a modal other than any of those expressing prohibition or absence of obligation. This is surprising to say the least: 'it is not necessary' seems to come close to the so-called 'youuser-friendly metalanguage' of which Berry (2000) shows that it is not clear whether it's preferable to traditional metalanguage. It remains to be seen here what the score for a gap-filling test with 'absence of obligation' as a prompt would have been.

Let's return to the 'niet moeten/niet hoeven' question with translation (15), based on the Flemish 'niet moeten':

(15) Dat moet je niet doen. Je hebt al voldoende je best gedaan.

[You don't have to do that. You've done your best already.]

(15a) *You mustn't do that. (12%)

(15b) *You shouldn't do that. (30%)

Contrary to what we had expected, not many more testees opted for 'mustn't', while a large group were apparently confused into expressing negative advice, which is another meaning of standard Dutch 'niet moeten'. Once again, it looks as if the use of translation-based testing techniques does not lead to more L1 interference

It is high time now to turn to the performance test. From the wide range of data we have available here, we would like to select a single exchange. Clearly, the point we set out to make is not that the kind of mistake that we report on here is in any way representative but simply that the analysis of actual learner performance – in this case through dyads – constitutes a powerful tool for testing active knowledge of L2 grammar.

Here's the extract, somewhere towards the end of a two-minute simulated phone call between two learners, who were being supervised by a teacher as part of an oral exam and both of whom were using a card with background information on the situation at hand

- A Could you call me back then when you exactly know uhm
(0.5) when I can expect the goods to arrive?
- B hehh yes Sir I could (0.5) uhm call you back but I think I
haven't (0.5) hehh I mustn't call you back because I'm sure
it will be the thirtieth of this month.

First of all, unlike with the other techniques referred to above, there can be no doubt that student B was trying to express absence of obligation. Indeed, after the pairwork, the two interactants discussed this exchange in Dutch and student B said:

Ik moest toch helemaal niet meer terugbellen. Awel ja, ik zei van, eh, ik moet eigenlijk niet meer terugbellen want het is zeker den dertigsten van deze maand.

[I didn't have to call back at all. Yes, I said like hehh I don't have to call back actually because it is certainly going to be the thirtieth of this month.]

Also, the exchange from the dyad seems to be perfect evidence of the fact that absence of obligation is a problem for student B. The data are very explicit in this respect: B first seems to go for a combination with 'have', then starts doubting about the negative form (she probably realizes that 'I haven't to call' is not appropriate) and finally turns to 'must not'. This seems to point to a possible source of the problem, viz. that students find it difficult to put 'have' in the negative. Clearly, this kind of data seems to go further than providing an inventory of testees' knowledge; it goes some way to revealing what lies beneath.

On the other hand, it should be emphasized that this is only a single-case analysis. As we have pointed out before, it is difficult to know to what extent it can be generalised. In addition, the analysis of real interaction for grammar-testing purposes tends to foreground what learners don't do correctly at the expense of what they do correctly (the analyst typically focuses on mistakes) as well as of what they avoid doing altogether: if learners don't know about questions tags, they will not use them and the analysis of their interaction will have very little to show about this (cf. Stenson 1991).

6. Conclusions

The present exploration into the field of grammar testing has shown that L2 course design in the domain of English grammar can benefit from a renewed interest in the contrastive perspective. This includes using translation as well as making learners aware of differences between L1 and L2 that cannot be documented through translation. Clearly, our research confirms the findings of previous work in that we have shown that translation does not lead to more L1 interference than other techniques. We have also demonstrated that it is important to introduce patterns of real social interaction in grammar testing as well as classroom practice, with but a minimal dependence on meta-grammatical terminology.

References

- Bachman, L.F. & Palmer, A.S. 1997. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: OUP.
- Berry, R. 2000. "'Youser-friendly' Metalanguage: What Effect Does it Have on Learners of English?" *IRAL* 38: 195-211.
- Együd, G. & Glover, P. 2001. "Oral Testing in Pairs". *ELT Journal* 55: 70-76.
- Ellis, R. 1998. *Second Language Acquisition*. Oxford: OUP.
- Garcia Mayo, M. & Pica, T. 2000. "L2 Learner Interaction in a Foreign Language Setting: Are Learning Needs Addressed?" *IRAL* 38: 35-58.
- Lemmens, M. & Parr, T. 1995. *Handboek voor de Vertaler Nederlands-Engels*. Leuven: Wolters.
- Mackenzie, J.L. 1997. *Principles and Pitfalls of English Grammar*. Bussum: Coutinho.
- Mattar, H. 1999. "Translation Elicitation Techniques and Mother-Tongue Interference: Any Significant Connection?" *IRAL* 37: 307-320.
- Nunan, D. 1998. "Teaching Grammar in Context". *ELT Journal* 52: 101-109.
- Petrovitz, W. 1997. "The Role of Context in the Presentation of Grammar". *ELT Journal* 51: 201-207.
- Stenson, N. 1991. "Induced Errors" in B.W. Robinett & J. Schachter (eds.) *Second Language Learning: Contrastive Analysis, Error Analysis, and Related Aspects*. Ann Arbor: The University of Michigan Press.
- Vandenberg, A.-M. 1991. *Exploring English Grammar*. Leuven: Garant.
- Zandvoort, R.W. 1975. *A Handbook of English Grammar*. 7th edn. London: Longman.