Language dominance and lexical diversity: how bilinguals and L2 learners differ in their knowledge and use of French lexical and functional items

Jeanine Treffers-Daller (UWE Bristol)

1. Introduction

From the psycholinguistic literature we know that monolinguals and bilinguals differ from each other in how they process language and that bilinguals can therefore not be seen as two monolinguals in one person (Grosjean, 1997: 167). We also know that perfect bilinguals are extremely rare and that most bilinguals are dominant in one or the other language (Fishman, 1971; Grosjean, 1997; Romaine, 1995). Therefore, there are probably important differences between bilinguals in the command they have of their languages, depending on the frequency with which they use each language, and the purposes for which they need them As Grosjean (1998) has pointed out, there is a lot of confusion around the concept of bilinguals and researchers use widely differing operationalisations of this concept. Few researchers attempt to assess the knowledge bilinguals have of either language in any detail, although it is legitimate to question how one can differentiate between different types of bilinguals or between bilinguals and second language learners. Some researchers are reluctant to engage in precise assessments of bilinguals' proficiency profiles because this often leads to negative views of bilinguals or L2-users (see Cook, 1997 on the monolingual bias that is built into SLA research). Obtaining precise information about the proficiency of bilinguals is however

important because language proficiency has an impact on language processing and thus it affects bilinguals' performance on lexical decision tasks or any other tasks that involve informants' language processing mechanisms.

According to Kroll, Bobb and Wodnieczka (2006: 128) we do not yet have a comprehensive overview of how language proficiency and relative language dominance affect the processes engaged during the planning of spoken utterances, but they point out that this is an important variable that researchers need to take seriously. Many researchers have shown that bilinguals are slower in picture naming tasks or lexical decision tasks, probably because using two languages has the consequence of lowering the functional frequency of each (Kroll et al, 2006). The bilinguals' disadvantage may however disappear if one controls for vocabulary size. Bialystok, Craik and Luk (in press) have recently shown that bilinguals whose lexical knowledge is matched to that of monolinguals outperform monolinguals on a task of letter fluency and word naming, because bilinguals have an advantage over monolinguals in tasks that involve executive control. Their study illustrates the importance of obtaining precise measurements of informants' vocabulary knowledge: instead of reinforcing existing negative views of bilinguals, such measurements can contribute to the discovery of exciting new information about the advantages of being bilingual.

These results also illustrate that it is very important to get a better understanding of the notion of language dominance. Most bilinguals are dominant in one or the other language, but most researchers use the term *language dominance* without providing any measurements of their subjects' knowledge of either language. It therefore remains unclear what language dominance means in linguistic terms, that is to say, whether this

mainly affects the lexicon or whether other areas of the language system are also more developed in one language than in the other of the bilinguals under study.

This chapter reports a follow-up to an earlier study in which the language dominance among different groups of Turkish-German bilinguals was investigated, with a particular focus on lexical richness (Daller, Van Hout and Treffers-Daller, 2003). In this study we showed that the proficiency profiles of Turkish-German bilinguals differ significantly from each other depending on whether they lived in Germany or in Turkey. The Turkish-German bilinguals in Germany were clearly dominant in German in that they obtained higher scores on various measures of lexical richness in German but lower scores in Turkish and the opposite was true for Turkish-dominant bilinguals who had returned to Turkey eight years prior to the recording. Further analyses of the use of Turkish syntactic embeddings among all groups showed that German-dominant bilinguals used simpler syntactic embeddings than Turkish-dominant bilinguals (Treffers-Daller, Özsoy and van Hout, 2007). These studies demonstrate that it is possible to measure language dominance in bilinguals using different syntactic and lexical variables. The current study aims to contribute further to our understanding of variation in lexical knowledge and use among different groups of bilinguals and how these groups differ from L2 learners in this respect.

For a number of reasons it is particularly important to focus on lexical issues. First of all because the lexicon plays a central role in the latest versions of generative grammar (e.g. Minimalism) and in psycholinguistic models such as Levelt's (1989) speech production model. Most models are lexically driven, that is to say, the grammar, morphology and phonology are determined by the lexical items selected by the speaker.

Under this view, vocabulary is the key to learning (Bialystok, 2001: 48). Bates and Goodman (1997) even argue that the emergence of grammar depends directly on vocabulary size. In the second place, psycholinguistic research often focuses on lexical access in production or reception, and much less on syntactic structures. Third, it is reasonable to assume that there is important variability in the number of words individuals (monolinguals) know and the knowledge they have about these words, as lexical knowledge is clearly dependent on a range of sociolinguistic variables, in particular education. Achieving full grammatical competence is normal for individuals, at least in L1 acquisition, but it is difficult to define what full competence means in relation to the lexicon. Monolinguals as well as bilinguals are likely to vary considerably in their knowledge and use of lexical items, but because the latter use their two languages for different purposes, the variability in lexical knowledge among bilinguals is probably even greater than among monolinguals.

Bialystok (2001) summarizes the evidence concerning the existence of variation in lexical knowledge among children but few researchers have attempted to measure variation in adult bilinguals' knowledge and use of lexical items in any detail. In those studies which do consider vocabulary, the focus is most often on receptive knowledge of vocabulary, in particular vocabulary size, as measured with the Peabody Picture Vocabulary Test (Dunn and Dunn, 1959; 2006) or the Mill Hill Vocabulary Scale (Raven, 1960). Studies which make use of these tests often show that bilinguals obtain lower scores than comparable monolinguals (Craik and Bialystok, 2006), but we know little about bilinguals' use of vocabulary in productive, more naturalistic tasks.

The aim of the current chapter is to obtain a clearer picture of variability in adult bilinguals' knowledge and use of vocabulary and how they differ from L2 learners. The focus is in particular on lexical diversity as measured with different tools that have recently been proposed in the literature and that are available under CLAN, the computerized data analysis tools developed by MacWhinney and colleagues (2000). The main hypothesis of the study is that indices of lexical diversity are excellent tools to measure the lexical proficiency of bilinguals and L2 learners, and to reveal the existence of differences in their use of lexical items. However, only detailed qualitative analyses can reveal the subtle differences in the ways in which Dutch-dominant and French-dominant bilinguals use functional items.

2. Measuring lexical richness: lexical items and function words

As Nation (2001: 27) has shown, vocabulary knowledge is multidimensional and therefore most researchers will agree with Richards and Malvern (2007: 82) that no "single index can represent competence or performance in relation to vocabulary, or for that matter, any other linguistic domain." Attempting to characterize the vocabulary used by learners with the help of a single measure of lexical richness is therefore necessarily a simplification, and it will be useful to complement this with additional analyses, which can give insights into qualitative aspects of vocabulary knowledge and use.

Previous studies have demonstrated that generic measures such as the Index of Guiraud (Guiraud, 1954) and D (Malvern and Richards, 1997; Malvern, Richards,

Chipere and Durán, 2004) give a good overall impression of the differences in lexical diversity between texts from different sources, including learner language (see Van Hout and Vermeer, 2007 for an overview and a critical discussion of the different measures). These measures do not, however, reveal what the relative contribution of lexical and functional categories is to the lexical diversity of texts. In addition, further analyses need to be carried out if one wants to obtain qualitative information about the lexical knowledge of informants, for example whether they differ in their knowledge of lexical items or function words, or whether there are any particular issues with the ways in which these words are being used. As is well-known, learners often overuse particular words or structures that are simpler (Ellis, 1997) or avoid those that they are less familiar with (Schachter, 1974) but the above-mentioned generic measures cannot reveal this.

In order to address those issues, I have carried out analyses of the diversity of lexical categories, in particular nouns and verbs, although adjectives will also be discussed briefly. As nouns and verbs are the main lexical categories in French corpora (Gendner and Adda-Decker, 2002), one might expect that they contribute most to the variability of texts. According to Laudanna and Voghera (2002: 8) the frequency of nouns and verbs in English corpora depends on the amount of dialogue and the amount of planning, in that nouns are generally more frequent in monologues and planned texts, while verbs are more frequent in dialogues and spontaneous texts (see also Biber, 1995; Biber, Johansson, Leech, Conrad and Finegan, 1999). It will be interesting to see to what the proportion of nouns and verbs is our French corpus, and what these parts of speech contribute to the lexical diversity of the texts.

After having studied lexical items, we focus our attention on the ways in which learners and bilinguals differ from each other in their use of function words. Relativisers were chosen because their usage is relatively complex: L2 learners of French need to acquire many different forms, some of which (lequel/laquelle/lesquels/lesquelles) agree in gender and number with their antecedent, distinguish between different syntactic functions of these forms and learn how to use them for different purposes in discourse. In addition, relative clauses can be embedded in a variety of ways into sentences, which adds to their complexity.

The literature on the L1 acquisition of French relative clauses is rather limited but the available evidence suggests that subject relative clauses are relatively early acquired and used frequently, but for a limited number of functions (Jisa and Kern, 1998). In addition, Jisa and Kern show that *que* is used much less frequently than *qui* by children as well as adults. In his study of the L2 acquisition of French relativiser morphology, Hawkins (1989) shows that the subject form *qui* is easier than the object form *que* because the former is closer to its extraction site (indicated with a _____) in the examples below, that is, the site from which the WH-word has been moved to COMP, as (1) and (2) illustrate. The form *dont* (which is used for genitive relative clauses) is the most difficult one because it is furthest away from its extraction site (see 3).

(1) L'homme *qui* ____ connaît Pierre

"The man who knows Pierre." (Hawkins, 1989: 162)

(2) L'homme *que* Pierre connaît _____

"The man who Pierre knows." (Hawkins, 1989: 162)

(3) Le visiteur *dont* j'avais oublié le nom _____

"The visitor whose name I had forgotten." (Hawkins, 1989: 163)

It is the relative proximity of the relativiser to its extraction site that explains why first and second year students who are studying French for their degree course make more errors with *que* than with *qui* and most errors with *dont* (Hawkins, 1989). These findings form an excellent point of comparison for the use of relativers by our three groups.

If our hypothesis is correct, measures of lexical diversity should be able to reveal interesting differences between French-dominant and Dutch-dominant bilinguals, as well as between bilinguals and L2 learners. Quantitative analyses can however not uncover more subtle differences between French-dominant and Dutch-dominant bilinguals in their use of functional items. Qualitative analyses are therefore needed if we want to obtain a better understanding of the ways in which bilinguals differ from each other in their use of function words.

3. Methods

Three groups of subjects participated in the study. The first group consists of 25 adult bilinguals from Brussels who have always lived in Anderlecht, the south-western part of the Brussels agglomeration, in which a relatively large proportion of speakers of Dutch can be found. Participants are all speakers of Brussels Dutch, the local variety of Dutch and Brussels French, and some of them, but not all of them also speak the standard varieties of either language. From the interviews held with participants in 2006 it is clear that most of these speakers are dominant in Dutch but they use French on a daily basis in

everyday life as is common in Brussels which has a predominantly French-speaking population.

The second group are 25 eighteen-year-old Flemish students of French from Aalst, who were recorded by a team of researchers led by Housen in the framework of a project on the simultaneous learning of two foreign languages (French and English). The data for this project are available on the website of the French Learning Language Oral Corpora (FLLOC): http://www.flloc.soton.ac.uk/brussels.php.

The third group consists of French-English bilingual students from a business school in Paris, who grew up with French only but learnt English (and other languages) at secondary school. One student indicated to have spoken Spanish in addition to French in early childhood. They were taught in Paris through the medium of English and they were enrolled in an English course at UWE Bristol in 2006, during which they took part in this study. This group is clearly French-dominant, as is obvious from their language history, even though they use English on a daily basis for all subjects of their studies.

A controlled productive task was chosen rather than a free productive task to ensure the comparability of the content across the three groups, which is particularly important in studies which focus on lexical items. Mayer's (1969) storybook *Frog Where are you?* was used to elicit semi-spontaneous speech from all individuals. This story has frequently been used to study language use of monolinguals and bilinguals (e.g. Berman and Slobin, 1994) which makes it relatively easy to obtain comparable data sets, such as the Brussels corpus on the FLLOC database. Because the Brussels bilinguals regularly use French in conversation but are not necessarily biliterate, written language tests were not considered appropriate for the target group. All informants were given some

preparation time before telling their story individually to the researcher, either in their own home (the participants from Brussels) or in the school/university they attended. The bilinguals from Brussels also told another Frog story (*Frog goes to dinner*, Mayer, 1974) in Dutch. The Parisian students told this story in English, but these stories are not being analysed for this chapter, which focuses on variation in French. Some Flemish students were offered help by their interlocutor if they did not know a particular word, but all words that students learned from the researchers were discarded from the analysis. Two students who received a disproportionate amount of feedback from their interlocutor were excluded from the study altogether.

All data were transcribed in CHAT format (MacWhinney, 2000), and subsequently a morphosyntactic coding tier (the mor tier) was added to the transcripts, with the help of the MOR and POST commands under CLAN. Any remaining ambiguities, errors or inconsistencies in the resulting MOR tier were corrected by hand. In addition, all proper names, filled pauses and other hesitation markers, exclamations as well as words from other languages (mainly Dutch or English) were excluded from the analysis.

For several reasons, using the mor tier for analyses of lexical richness is particularly useful. In the first place because the mor tier makes it possible to distinguish between homophones (e.g. *tu* "you" as a personal pronoun and *tu* "was silent" as the past participle of the verb *se taire* "to be silent") which is only possible on the main tier by adding disambiguation codes by hand. In the second place, on the mor tier all entries are lemmatized. In a previous study, we lemmatised the data on the main tier in a way that is described in detail in Tidball and Treffers-Daller (2007) but if all researchers who work

on French lemmatise their data in slightly different ways, this reduces the comparability of results significantly (see also the discussion in Richards and Malvern, 2007, on the effects of different lemmatisation strategies on D). The mor tier offers a standard that can be used by everyone. In the third place, new switches that can be used with the frequency command FREQ have recently become available under CLAN. These make type/token analyses of individual syntactic categories on the mor tier possible, which is extremely useful for studies of lexical richness. An example of a file with a mor tier is given in Appendix 1.

The only problem encountered using the mor tier in CLAN is that the French mor tier distinguishes different subcategories of verb forms (infinitival, participial, progressive and other forms) in the information to the left of the pipe separator (|) which separates the syntactic category information from the word itself. Thus, the verb *trouver* "to find" is categorized on the mor tier in many different ways: v:pp|trouver; v:inf|trouver; v:prog|trouver and v|trouver (see Appendix 1 for examples). This means that FREQ counts these different forms of *trouver* as different types rather than as different tokens of the type *trouver*, which results in inflated indices of lexical richness. This problem also exists, but to a lesser extent, for other syntactic categories such as pronouns. Using switches such as +s"*-% %", which tell CLAN to ignore form variants, does not solve the problem, because these switches only look at information *after* the pipe separator, not before. For the purposes of this paper I therefore decided to erase the above-mentioned subcategories of verbs with the help of the change string command (CHSTRING), leaving only the codes for subcategories of modal verbs (v:mdl| and v:mdllex) and auxiliairies (v:aux) in place, as the distinction between lexical uses of

verbs such as *avoir* (*il a un livre* "he has a book") and auxiliary uses of this verb (*il a acheté un livre* "he bought a book") are obviously important for analyses of lexical richness. If the information about different verb forms could be coded after the pipe separator on the French mor tier, together with other morphological information regarding person and tense, this problem would not exist.

4. Results

4.1 Generic measures of lexical richness: the Index of Guiraud and D

In the first instance the differences between the three groups were investigated by calculating two generic measures of lexical richness, the Index of Guiraud and D. As Table 1 shows, there are significant differences between the groups for both measures, in that the French-dominant bilinguals obtain the highest scores and the L2 learners the lowest scores, whereas the scores of the Dutch-dominant bilinguals fall between those of the other two groups. The results of the ANOVA and Tukey post hoc tests show that all groups are significantly different from each other for the Index of Guiraud (ANOVA, $F(2,64) = 50.58$, $p < .001$) as well as for D (ANOVA, $F(2,64) = 56.9$, $p < .001$), but D is a bit more powerful in that it discriminates slightly better between the groups, as can be seen from the Eta Squared values. Group 3 also produces significantly fewer types and tokens than groups 1 and 2, but groups 1 and 2 do not differ significantly from each other in their use of tokens, and only marginally in their use of types. Therefore more sensitive measures such as D or the Index of Guiraud are needed to demonstrate the existence of

differences between the groups. Both measures correlate very strongly and significantly with each other (r = .951; N= 69; p <.01), which gives a clear indication that they are measuring similar aspects of lexical richness.

Table 1 Descriptive statistics for tokens, types, D and Guiraud for each group, and effect sizes for differences between groups (Eta squared)

groups	N	Tokens	Types	D (SD)	Guiraud (SD)
1 = Business students Paris	19	571	175	68.7 (13.5)	7.4 (.85)
2 = Bilinguals from Brussels	25	500	143	50.3 (15.3)	6.5 (1.0)
3 = Flemish L2 learners	25	283	83	28.7 (7.3)	4.9 (.58)
Total – mean scores	69	441	130	47.5 (20.2)	6.2 (1.3)
Eta Squared				.633	.612

.

The D-values in Table 1 are relatively high in comparison with those reported for French in Tidball and Treffers-Daller (2007), in which first year students obtained mean scores of 18.78, final year students 26.46 and French native speakers from the same Parisian business school obtained mean values of 34.87 for oral descriptions of cartoon strips. The results for the Index of Guiraud are however only marginally higher than in our previous study in which the two student groups and the Parisian Business students

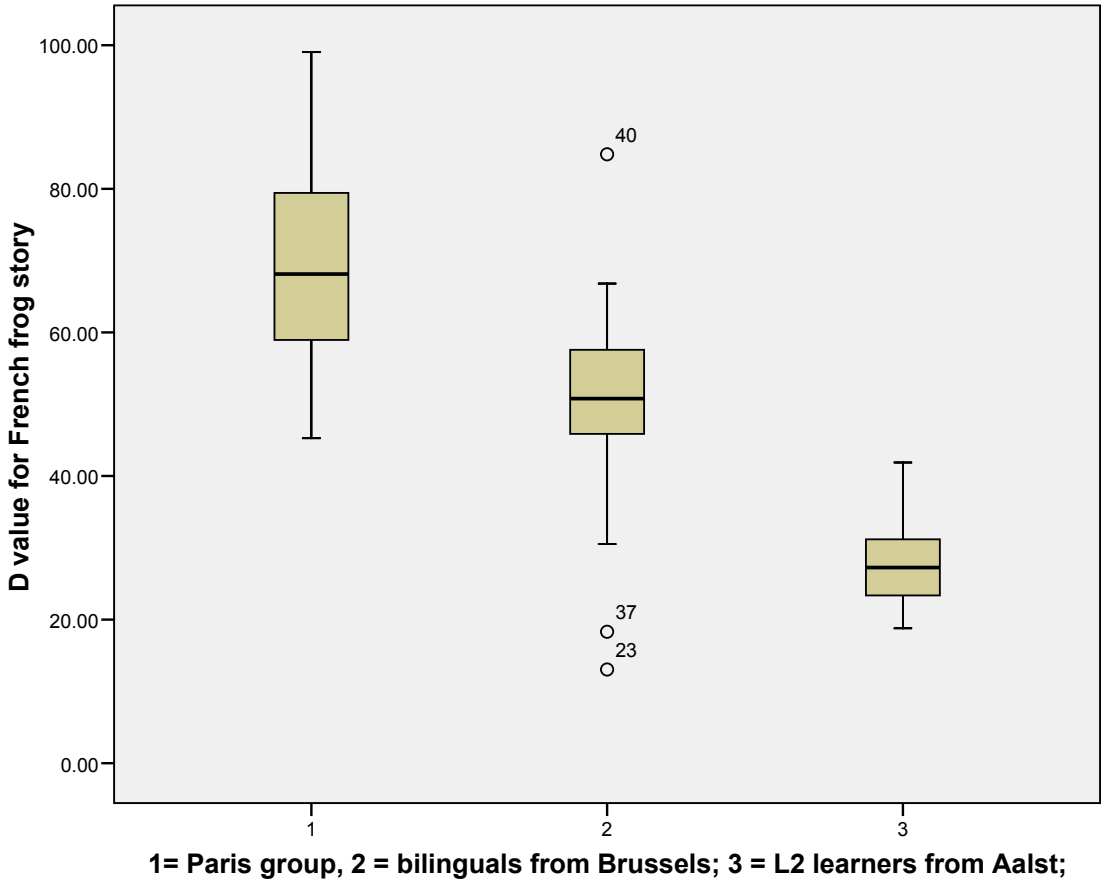
obtained scores of, respectively, 4.30; 5.25 and 6.27. There are several potential explanations for these differences, but it is most likely that the main reason for the differences between the two studies should be sought in the fact that different elicitation materials were used. It possible that the relatively complex story line of the frog story invites informants to produce more detailed narratives than the father-and-son comic strips used in the earlier research. In the former there is a wide range of activities involving many different participants, whereas the latter revolves around a small number of actions involving two protagonists with one or two additional characters. Evidence for this explanation can be found in the relatively large number of types (130) and tokens (441) the subjects in the current study produced in comparison with the students and the native speakers in Tidball and Treffers-Daller (2007), who produced 97 types and 327 tokens on average in the father-and-son story telling task. Second, lemmatization was done on the main tier in a slightly different way in our previous study, whereas the mor tier was used for this purpose in the current study. The mor tier distinguishes between different uses of function words such as *qui* "who", which can either be an interrogative pronoun or a relativiser. The same applies to function words such as *le/la/les* "the/him/it/her/them", which function not only as determiners but also as object pronouns. CLAN programs consider the different uses of these words as different types, which results in slightly higher D values and slightly higher scores on the Index of Guiraud, if these measures are calculated on the mor tier.

Given the differences in the elicitation task and the lemmatization issues mentioned above, it is remarkable that the values of the Index of Guiraud are relatively similar in both studies. This could be an indication that this measure is slightly more

robust in that it is less sensitive to task effects or lemmatization strategies. A comparison of absolute D-values or scores on the Index of Guiraud remains however very difficult if the elicitation materials are not the same across studies and if there is no standard way to lemmatize French (see also David, 2008, who makes a similar point). Using the mor tier for measurements of lexical richness could however offer a solution to the latter problem.

Figure 1 illustrates the extent to which the standard deviations (given in Table 1) are higher for the bilingual group than for the other two groups. This is to be expected as bilinguals inevitably vary in the amount of use they make of their two languages, with some using French on a daily basis for a range of purposes whereas others make use of French in much more limited ways. The higher standard deviations and the presence of three outliers in the bilingual group form a good illustration of the variability in vocabulary knowledge and use among bilinguals.

Figure 1. D scores for all three groups



4.2 Lexical diversity of nouns and verbs

In the second part of this study the focus is on lexical diversity in two lexical categories and one functional category. While it would have been interesting to compare the results

for D and the Index of Guiraud in this part of the study too, this turned out to be impossible because D can only be calculated if a minimum of 50 tokens is available. Hardly any of the L2 learners from Aalst and only half of the bilinguals from Brussels produced a sufficient number of verb tokens. A similar problem exists for the nouns and the relativisers (of which no informant produces more than 25 tokens). For this reason, D and its derivative the Limiting Relative Diversity Index could not be calculated for these individual syntactic categories, and only the Index of Guiraud was used.

With the help of new switches under CLAN, which are available with FREQ but not with VOCD, it is possible to obtain lemmatized frequency lists per syntactic category (see Appendix 2 for an example). This allowed us first of all to establish that there are more nouns (2045 types and 4662 tokens) than verbs (1677 types and 2954 tokens) in the current corpus, which provides evidence for the Laudanna and Voghera's (2002) claim that nouns are generally more frequent than verbs in monologues and planned texts. On the basis of the output of FREQ, the diversity of nouns and lexical verbs was calculated in two different ways: first the ratio of noun types over the square root of noun tokens (Guiraud nouns 1), and then the ratio of noun types over the square root of all tokens (Guiraud nouns 2). The same procedure was followed for the verbs. The two calculations of Guiraud differ only marginally from each other but the second way of calculating Guiraud may be preferable, as the same denominator is used for all calculations (nouns and verbs).

The three groups differ in predictable ways from each other in their use of nouns as well as verbs: the business students from Paris obtain the highest scores and the L2 learners the lowest scores, and the scores of the bilinguals from Brussels fall in

between those two (see Tables 2 and 3). It is interesting to see that Eta Squared for the verbs is higher than for nouns, which is an indication that the diversity of verbs as measured with Guiraud discriminates between the groups to a greater extent than the same measure for nouns. In order to find out whether verbs contribute more to the diversity of the texts than nouns, a paired t-test was carried out on the pooled data in which the mean values for Guiraud nouns 1 and Guiraud verbs 1 were compared. The differences between the mean Guiraud for the verbs (3.96) and the mean Guiraud for the nouns (3.82) approach significance with a two-sided t-test ($t = 1.7$; $df = 68$, $p = .093$) and they are significant in the predicted direction with a one-sided t-test. Thus, verbs may indeed contribute somewhat more to the diversity of the texts than nouns in this data set.

If noun and verb types are counted together, the calculation of Guiraud (verb types+noun types / √all tokens) discriminates even better between the groups (ANOVA, $F (2,66) = 41.2$, $p < 001$; Eta Squared .555). This result can be improved only slightly by adding adjective types to the calculation (ANOVA, $F (2,66) = 42.1$, $p < .001$; Eta Squared of .560). As these effect sizes are very close to those obtained by D (.633) and the Index of Guiraud (.613), which are based on *all* types and tokens, words belonging to categories other than nouns or verbs contribute probably very little to the between-group differences.

Table 2 Mean scores for noun types, noun tokens and the Index of Guiraud

group	Noun types	Noun tokens	Guiraud nouns 1 (Noun types/√noun tokens)	Guiraud nouns 2 (Noun types/√all tokens)
1 = Business students Paris	49.8	104.2	4.93	2.1
2 = Bilinguals from Brussels	38.8	98.1	3.93	1.75
3 = Flemish L2 learners	20.8	52.7	2.86	1.23
F-value (2,66)	-	-	31.0 (p <.001)	30.3 (p<.001)
Eta squared	-	-	.484	.479

(All groups differ significantly from each other; Tukey post hoc analysis, p< .01)

Table 3 Mean scores for verb types, verb tokens and the Index of Guiraud

groups	Verb types	Verb tokens	Guiraud verbs 1 (Verb types/√ verb tokens)	Guiraud verbs 2 (Verb types/√all tokens)
1 = Business students Paris	42.2	70.6	4.98	1.77
2 = Bilinguals from Brussels	31.9	58.9	4.12	1.42
3 = Flemish L2 learners	16.6	29.7	3.04	.98
F-value (2,66)			30.8 ($p < .001$)	35.54 ($p < .001$)
Eta Squared			.483	.510

(All groups significantly different; Tukey post hoc analysis, $p < .01$)

4.3 Relative frequency of relativisers

As there are very few different relativisers (*qui*, *que*, *dont*, *où* and *leque/laquelle/lesquels/lesquelles*), calculating the Index of Guiraud for relativisers is not very meaningful because informants differ very little in the types they use. The number of tokens does however vary considerably per individual. Therefore a calculation of the proportion of all tokens that are relativisers can give interesting information about the differences between the groups.

Table 4 Frequency of relativisers in all three groups

Groups	All relativiser tokens /all tokens (SD)	All relativiser tokens not triggered by " il y a" /all tokens (SD)
1 = Business students Paris	.012 (.005)	.012 (.005)
2 = Bilinguals from Brussels	.015 (.01)	.011 (.005)
3 = Flemish L2 learners	.004 (.008)	.009 (.007)
F-values / Tukey post hoc analyses	F (2,66) = 15.4 (p<.001)	F (2,66) = 9.97 (p<.001)
1 and 2	ns	ns
1 and 3	*	*
2 and 3	*	*

As Table 4 shows, there are no significant differences between the bilinguals from Paris and the bilinguals from Brussels in their use of relativisers, but the L2 learners use significantly fewer relativisers than the two other groups. Contrary to expectations, Dutch-dominant bilinguals from Brussels obtained slightly higher scores than the French-dominant bilinguals from Paris. Even though this result was not significant and thus not generalisable to the wider population, I wanted to explore this finding in this particular

sample, to see if there was any indication of overuse of particular structures by the bilinguals from Brussels.

This analysis revealed that the bilinguals in Brussels use the relativiser *qui* "who" very frequently in combination with *il y a* "there is", in utterances such as (4) whereas the other groups do not do this.

(4) Allez et en une fois y a un hibou qui sort (bilingual informant JEA from Brussels)
"Well and all of a sudden there is an owl who comes out."

With the help of COMBO[1] it was established that among the Brussels group 57 of the 184 uses of *qui* as a relativiser occur in structures such as (4). The students from Paris however used the structure *il y a un X qui Verb* only once (out of 120 uses of *qui* as a relativiser) and the Flemish L2 learners employed it only once out of 43 uses of the relativiser *qui*. If the relativisers which are triggered by the occurrence of *il y a* are excluded from the calculations, the unexpectedly high frequency of relativisers disappears (see Table 4, final column). This does not however affect the overall results: the differences between the two groups of bilinguals in their use of relativisers are not significant.

Guillot's (2005) detailed comparative analyses of this structure across a range of written and oral sources can help to throw new light on its frequency in the data. Guillot

[1] The command used was: combo +s"y^*^qui", which tells CLAN to look for an occurrence of *y*, followed immediately or eventually by *qui*. The output then needs to be checked to see whether these occur within the same clause or not.

shows that the occurrence of the prefabricated formula *il y a NP relative clause* is not only frequent in L2 learners' spoken and written language but also in unplanned native speaker speech and it is thus not an indication of non-nativeness (Guillot, 2005: 120). The fact that the L2 learners in the current study were not exposed as much to spoken French as the bilinguals from Brussels can probably explain why they did not use this structure frequently. The students from Paris however, who were in daily contact with French, did not use this structure frequently either, which is somewhat puzzling. Jisa and Kern's (1998) analysis of the functions of relative clauses can help to throw light on this issue. They show that children use relative clauses more for general discourse functions (mainly to establish and introduce new referents) whereas adults use these for a much wider variety of functions. Although a detailed analysis of the functions for which the bilinguals from Brussels use relative clauses is beyond the scope of this paper, bilinguals frequently use relative clauses to introduce new referents, as example (4) illustrates. This usage is very similar to the examples discussed in Jisa and Kern (1998). French-dominant bilinguals however hardly make use of this strategy to introduce new referents.

Table 5 gives further details of the qualitative differences in the uses of relativisers by the three groups. The L2 learners use only the subject relativiser *qui*, but the two other groups also use the object pronoun *que* and a small number of other relativisers. As the two main types of relativisers are used in roughly the same proportion, this is another indication that the groups from Paris and from Brussels do not differ significantly from each other on this point, but the stories of the L2 learners display less diversity on this variable. The data thus confirm the findings of Hawkins (1989) and Jisa

and Kern (1998) that subject relative clauses are the most common (and probably the easiest), followed by object relative clauses, whilst other types are less frequent.

Table 5: Frequency of each relativiser in each of the three groups

	Qui (subject)[1]	que	où	dont	lequel (+ form variants)	Total (100%)
1 = Business students from Paris	120 (80%)	18 (12%)	3 (2%)	2 (1.3%)	7 (4.7%)	150
2 = Bilinguals from Brussels	184 (82.5%)	18 (8.1%)	17 (7.6%)	2 (0.9%)	2 (0.9%)	223
3 = L2 learners from Aalst	43 (100%)	0	0	0	0	43

There were no occurrences of oblique uses of *qui* (i.e. *qui* following a preposition) in the data.

5. Conclusion

In this chapter we have seen that there are important differences in the lexical diversity of stories told by bilinguals and L2 learners, and that D and the Index of Guiraud are excellent tools in demonstrating the existence of those differences. D proved

to be somewhat more powerful than the Index of Guiraud, in that the former discriminated more strongly between the groups than the latter.

As one of the aims of the study was to find out which syntactic categories contribute most to the diversity of the stories, separate analyses were carried out of the diversity of two lexical categories (nouns and verbs) and one functional category (relativisers) with the help of tools that have recently become available under CLAN. The Index of Guiraud was employed for the analysis of nouns and verbs, because D could not be used for reasons explained in section 4.2. As nouns and verbs are the word categories which have most members it is not surprising that we found that these two categories contribute most to the total between–groups variance in the data. The Eta Squared values obtained for analyses based on nouns and verbs approached those based on all the words in the stories. Adding adjectives to the computation contributed very little to this result.

There were also significant differences between the L2 learners and the bilinguals in their use of relativisers, in that the L2 learners used fewer and a more limited range (only subject relativisers) than the bilinguals. Although there were no significant *quantitative* differences between Dutch-dominant and French-dominant bilinguals in their use of relativisers, a detailed *qualitative* analysis demonstrated that the Dutch-dominant group overuse of relativisers in prefabricated formulae to introduce new referents in the story. These subtle differences in the bilinguals' use of functional items could not be revealed with the help of generic measures of lexical diversity.

The main differences between the Dutch-dominant bilinguals from Brussels and the French-dominant bilinguals from Paris resided however in the diversity of the lexical items they used, in particular nouns and verbs, and not in differences in their use of the

functional items studied here. The L2 learners in our study, on the other hand, differed significantly from the two groups of bilinguals in the diversity of lexical as well as functional items they used in the stories.

It is of course possible that language dominance manifests itself in some bilinguals in their use of lexical as well as functional items (see for example Treffers-Daller, Özsoy and Van Hout, 2007). Therefore we need further insight into the ways in which bilingual competence can vary in individuals, in other words, we need a typology of bilingual competence and an operationalisation of the notion of language dominance in terms of the different language levels. The main contribution of the current study to our understanding of these issues is perhaps that it has shown that key aspects of language dominance can be measured with the help of indices of lexical diversity. These need to be complemented, however, with qualitative analyses of the ways in which functional items are being used if one wants to reach an in-depth understanding of language dominance.

Bibliography

Bates, E., and J.C. Goodman (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language and Cognitive Processes* 12, 507-584.

Berman, R. A. and D.I. Slobin (1994). *Relating events in narrative: a crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates

Bialystok, E. (2001). *Bilingualism in Development: Language, Literacy, and Cognition*. Cambridge: Cambridge University Press.

Bialystok, E., F. I.M. Craik and G. Luk (in press). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics* (to appear in 2008).

Biber, Douglas (1995). *Dimensions of register variation*. Cambridge:
Cambridge University Press.

Biber, D., S. Johansson, G. Leech, S.Conrad and E. Finegan (eds.) (1999). *Longman Grammar of Written and Spoken English*. London: Longman.

Cook, V. (1997). Monolingual Bias in Second Language Acquisition Research. *Revista Canaria de Estudios Ingleses* 34, 35-50. Also available from: http://homepage.ntlworld.com/vivian.c/Writings/Papers/MonolingualBias.htm

Craik, F. I.M. and E. Bialystok (2006). Planning and task managements in older adults: Cooking breakfast. *Memory and Cognition* 34 (6), 1236-1249.

Daller, H., R. van Hout, R. and J.Treffers-Daller, J. (2003). Lexical richness in spontaneous speech of bilinguals. *Applied Linguistics* 24, 197–222.

David, A. (2008). A developmental perspective on productive lexical knowledge in L2 oral interlanguage. *Journal of French Language Studies* 18(3), pagenumbers tbc

Dunn, L. M. and D. M. Dunn (1959, 2006). *Peabody Picture Vocabulary Scale*. Pearson Assessments.

Ellis, R. (1997). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Fishman, J. (1971). *Advances in the Sociology of Language*. The Hague: Mouton.

Gendner, V. and M. Adda-Decker (2002). Analyse comparative de corpus oraux et écrits français: mots, lemmes et classes morpho-syntaxiques.XXIVièmes Journées d'Étude sur la Parole, Nancy, 24-27 juin 2002. Available from : ftp://tlp.limsi.fr/public/jep02POS.pdf

Grosjean, F. (1997). The bilingual individual. *Interpreting* 2 (1/2), 163-187.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1, 131-149.

Guillot, M.-N. (2005). Il y a des gens qui disent que ...'there are people who say that ...' Beyond grammatical accuracy in FL learners' writing: Issues of non-nativeness. *International Review of Applied Linguistics in Language Teaching* 43, 109–128.

Guiraud, P. (1954). *Les caractéristiques statistiques du vocabulaire*. Paris: Presses Universitaires de France.

Hawkins, R. (1989). Do second language learners acquire restrictive relative clauses on the basis of relational or configurational information? The acquisition of French subject, direct object, and genitive restrictive clauses by language learners. *Second Language Research* 5, 156-188.

Hout, R. and A. Vermeer (2007). Comparing measures of lexical richness. In H. Daller, J Milton and J. Treffers-Daller (eds.) *Modelling and assessing vocabulary knowledge* (pp. 93 – 115). Cambridge: Cambridge University Press.

Jisa, H. and S. Kern (1998). Relative clauses in French children's narrative texts. *Journal of Child Language* 25, 623-652.

Kroll, J., S. C. Bobb and Z. Wodnieczka (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism, Language and Cognition* 9 (2), 119-135.

Laudanna, A. and M. Voghera (2002). Nouns and verbs as grammatical classes in the lexicon. *Italian Journal of Linguistics* 14, 9–26.

Levelt, W. (1989). *Speaking: From Intention to Articulation*. Cambridge Mass: MIT Press.

MacWhinney, B. (2000). *The Childes Project. Tools for Analyzing* Talk. Lawrence Erlbaum Associates. The most recent version is available on this website: http://childes.psy.cmu.edu/

Malvern, D. D. and B.J. Richards (1997). A new measure of lexical diversity. In A. Ryan and A. Wray (eds.). *Evolving models of language.* Papers from the annual meeting of the

British Association of Applied Linguistics held at the University of Wales, Swansea, September 1996 (pp. 58-71). Clevedon: Multilingual Matters.

Malvern, D. D., Richards, B. J., Chipere, N. and Durán, P. (2004). *Lexical Diversity and Language Development: Quantification and Assessment*. Houndmills, Basingstoke: Palgrave Macmillan.

Mayer, M. (1969). *Frog Where are You?* New York: Dial Books for Young Readers.

Mayer, M. (1974). *Frog goes to dinner*. New York: Dial Books for Young Readers.

Nation, I.S.P (2001). *Learning Vocabulary in Another Language*. Cambrige: Cambridge University Press.

Raven, J. C. (1960). *Guide to the standard progressive matrices*. London: H. K. Lewis.

Richards, B.J. and D.D. Malvern (2007). Validity and threats to the validity of vocabulary assessment. In: H Daller, J Milton and J Treffers-Daller (eds.) *Modelling and assessing vocabulary knowledge* (pp. 79-92). Cambridge: Cambridge University Press,.

Romaine, S. (1995). *Bilingualism*. Second Edition. Oxford: Blackwell

Schachter, J. (1974). An error in error analysis. *Language Learning* 27: 205-214.

Tidball, F. and J. Treffers-Daller (2007). Exploring measures of vocabulary richness in semi-spontaneous speech of native and non-native speakers of French: a quest for the Holy Grail? In: Daller, H., J. Milton and J. Treffers-Daller (eds.) *Modelling and assessing vocabulary knowledge* (pp. 133-149). Cambridge: Cambridge University Press.

Treffers-Daller, J., S.A. Özsoy and R. van Hout (2007). (In)complete acquisition of Turkish among Turkish-German bilinguals in Germany and Turkey: an analysis of complex embeddings in narratives. *International Journal of Bilingual Education and Bilingualism*, 10 (3), 248-276.

Appendix 1

Excerpt of a transcript of the frog story as told by one of the bilingual informants from Brussels

@Begin

@Languages: fr

@Participants: DEM 003 Informant, JTD Jeanine Investigator

@ID: fr|AND|DEM|||||Informant||

@ID: fr|AND|JTD|||||Investigator||

@Date: 06-APR-2006

@Coder: JTD

*DEM: ça c' est le garçon qui avec son chien regarde le la grenouille dans un bocal .

%mor: pro:dem|ça pro|ce/ces&SING v:exist|être&PRES&3SV det|le&MASC&SING n|garçon&_MASC pro:rel|qui prep|avec det:poss|son&MASC&SING n|chien&_MASC v|regarder-PRES&_3SV det|le&MASC&SING det|la&FEM&SING n|grenouille&_FEM prep|dans det|un&MASC&SING n|bocal&_MASC&_SING .

*DEM: ici le garçon est en train de dormir et son pantalon reste droit et la grenouille sort du bocal et se dirige vers les pantoufles .

%mor: adv:place|ici det|le&MASC&SING n|garçon&_MASC v:exist|être&PRES&3SV prep:art|en n|train&_MASC prep|de

v:inf|dormir&INTRANS conj|et det:poss|son&MASC&SING n|pantalon&_MASC v|rester-PRES&_3SV adj|droit&MASC conj|et det|la&FEM&SING n|grenouille&_FEM v|sortir&PRES&3SV det|du&MASC&SING n|bocal&_MASC&_SING conj|et pro:refl|se&3SP v|diriger-PRES&_3SV prep|vers det|les&PL n|pantoufle&_FEM-_PL .

*DEM: &oh dit le garçon et son chien où est passé notre grenouille ?

%mor: v:pp|dire&_MASC&_SING det|le&MASC&SING n|garçon&_MASC conj|et det:poss|son&MASC&SING n|chien&_MASC pro:int|où v:aux|être&PRES&3SV v:pp|passer&_MASC&_SING det:poss|notre&_SING n|grenouille&_FEM ?

(transcript continues)

Appendix 2: Example of command used to extract verbs from the transcript of one of the students of the Business School in Paris

freq +t%mor -t* +s@r-*,|-v,o-% +f

Thu Aug 07 15:22:45 2008

freq (09-Jul-2008) is conducting analyses on:

ONLY dependent tiers matching: %MOR;

From file <c:\DOCUMENTS AND SETTINGS\J-TREFFERSDALLER\MY DOCUMENTS\JEANINE\VOCABULARY STUDIES IN L1 AND L2 ACQUISITION\FINAL POST 18 JULY 08 NEGOCIA\NEGOCIA VERBS SIMPLIFIED\F03.mor.pst.str.str.cex> to file <C:\DOCUMENTS AND SETTINGS\J-TREFFERSDALLER\MY DOCUMENTS\JEANINE\vocabulary studies in L1 and L2 acquisition\final post 18 July 08 Negocia\Negocia verbs simplified\verb freq\F03.mor.pst.str.str.frq.cex>

****

3 v|amuser	1 v|blesser
1 v|apercevoir	1 v|cacher
2 v|appeler	1 v|contrarier
1 v|arriver	1 v|coucher
1 v|assurer	2 v|courir
4 v|attaquer	1 v|dire
2 v|attraper	1 v|disparaître
1 v|avertir	2 v|dormir

2 v|découvrir

1 v|entourer

2 v|essayer

2 v|grimper

1 v|immiscer

1 v|manquer

1 v|parler

1 v|partager

1 v|perdre

2 v|prendre

1 v|profiter

1 v|préoccuper

1 v|préparer

1 v|rechercher

3 v|regarder

2 v|repartir

1 v|retenir

2 v|revenir

1 v|réussir

1 v|réveiller

1 v|sortir

2 v|soulever

4 v|tenter

4 v|tomber

4 v|voir

1 v|échapper

2 v|éviter

---

43 Total number of different word types used

71 Total number of words (tokens)

0.606 Type/Token ratio