Using lexical diversity measures to operationalise language dominance in bilinguals

Using lexical diversity measures to operationalise language dominance in bilinguals.

Jeanine Treffers-Daller (University of Reading) and Tomasz Korybski (University of Warsaw)

1. Introduction

There is a growing consensus in the academic literature of the importance of measuring language dominance in bilinguals, because the comparability of research results in the field of bilingualism is often hampered by individual differences between participants across studies. Bilinguals are a rather heterogeneous lot: they differ with respect to their language ability in both languages, their language acquisition background, including exposure to different languages, the purposes for which they use their languages, the frequency with which they use their languages, etc. (see e.g. Grosjean, 1998; Bedore et al., 2012). A precise quantification of participants' level of ability in their languages is therefore not enough to ensure bilinguals in different studies are comparable. As Wang (2013) demonstrates, how frequently bilinguals use their languages needs to be considered as well, because bilinguals who use one language more frequently than another perform differently in online tasks from those who use both languages equally often.

We agree with Wang (2013: 738) that language dominance cannot simply be equated with language ability, but is a "global measure of relative frequency of use and proficiency in each language". It is this definition we will use in our chapter. Because language dominance is a multi-faceted construct, measuring it with a single index is very difficult. In fact, as Flege (2002: 569) points out, there is no generally accepted method for measuring language dominance; researchers use a wide variety of techniques to measure it, including self-ratings,

fluency tasks, reading speed, lexical decision tasks, word classification tasks (e.g. animacy decision), vocabulary size tasks, etc. (see the introduction to this volume for more details). While many researchers use the term "language dominance" in describing the bilinguals or multilinguals in their studies, it often remains unclear what it means in linguistic terms when participants are claimed to be "balanced bilinguals" or "dominant" in one or the other language in their repertoire (Treffers-Daller, 2011).

In the current study we propose new Indices of Language Dominance (ILDs) to operationalise and measure language dominance, which is based on an analysis of the range of lexical items used by bilinguals in oral production tasks in each of their languages. We also propose a criterion for distinguishing between balanced bilinguals and those who are dominant in one or the other of their languages.

We have chosen vocabulary as the domain to study language dominance, first of all because recent models of grammar, such as the minimalist program (Chomsky, 1995), are lexically driven in that the properties of lexical items shape the sentence; second, because "the richness and diversity of linguistic forms within any particular language are now captured almost entirely by the lexicon" (Bates & Goodman, 1997: 508); third, because differences in lexical knowledge are known to affect performance on online tasks. Bialystok, Craik and Luk (2008) show, for example, that bilinguals whose lexical knowledge is matched to that of monolinguals (using the Peabody Picture Vocabulary Task, PPVT, Dunn & Dunn, 2007) outperform monolinguals on a task of letter fluency and word naming. This bilingual advantage in executive control tends to appear only when vocabulary has been controlled for. Finally, vocabulary is important because lexical knowledge is one of the main prerequisites for academic achievement of monolingual and bilingual children (Daller, 1999; Dickinson & Tabors, 2001).

Although there is a wide variety of vocabulary tests available (see Read, 2007 for an overview), standardized vocabulary tests such as the PPVT or X-lex (Meara & Milton, 2003) exist for a handful of languages only. It is unlikely that language tests will be developed for the majority of the languages of the world, simply because developing such tests is costly and there is no market for such tests for most languages. Therefore, we propose to measure language dominance on the basis of automated indices of lexical diversity which provide a precise measurement of the range of words deployed in oral or written texts, can be used with any language, are freely available, and have ecological validity because they are based on semi-spontaneous data collected in naturalistic settings (Treffers-Daller, 2011).

The Indices of Language Dominance are operationalized in four different ways. The first operationalization involves the use of a traditional index of lexical diversity: the Index of Guiraud (Guiraud, 1954), and the second one is based on a new measure of diversity, HDD, which was developed by McCarthy and Jarvis (2007). The third and the fourth are computational variants of the first two: we apply a formula proposed by Birdsong (this volume) to Guiraud-based and the HDD-based index, and compare the different operationalizations against each other.

We are fully aware that comparing measurements from two languages is notoriously difficult because of typological differences between these. In the Polish-English language pair under study, this is particularly important because Polish is highly inflected, whereas English is not. This poses particular challenges for measuring language dominance across these two languages. In the current study, we ensure the measurements from both languages are made comparable through careful lemmatization of the data.

It is of course crucially important to provide evidence for the validity of any index of language dominance. According to Flege, Mackay and Piske (2002: 572) the best way to validate a measure of bilingual dominance is to show that aspects of performance in either L1

or L2 can be predicted with this measure. We therefore provide evidence for the validity of the ILD by correlating scores on this index with four other variables, namely length of residence in the United Kingdom (UK), participants' attitudes towards English and life in the UK, reported frequency of use of English at home and the frequency of intrasentential code-switching, and use regression to investigate to what each index can predict scores on some of these.

The structure of our paper is as follows. First we discuss the available literature on the measurement of lexical aspects of language dominance (section 2). In section 3 we present the indices of lexical diversity we have chosen as well as an overview of studies in which these measures were used. The data collection method and the computation of the ILD are discussed in section 4. We then present the results of different computations of the ILD, and correlations with participants' code-switching behaviour and language attitudes (section 5). Finally, in section 6 we discuss the implications of our findings for current approaches to language dominance and formulate questions for further research.

2. Measuring lexical aspects of language dominance

Many authors include lexical measures in their study of language dominance, possibly because words appear to be more easily countable than other aspects of language (Milton, 2009: 7). While it is not possible to review all the available literature here, we will try to illustrate the different approaches taken by summarizing a few of the key studies in which researchers have operationalized language dominance by focusing on the lexicon. The reader is referred to the final chapter to this volume for more details on other ways to measure language dominance.

One of the earliest studies is probably Lambert, Havelka and Gardner (1959), who measure language dominance on the basis of a range of measures, several of which are based

on receptive or productive word knowledge. These include word recognition, word completion and word detection tests with items in two different languages. The authors propose to distinguish between dominant bilinguals and balanced bilinguals by calculating difference scores (subtracting scores in one language from those in another) on a range of variables. Unfortunately, the authors do not provide information about the cut-off point chosen, but in their sample of 43 bilinguals, 30 were found to be balanced and thirteen dominant in either French or English.

In their study of measures that could be used to study degree of bilingualism,[1] Fishman and Cooper (1969) included a word naming task where respondents had to give in one minute as many words as possible belonging to a particular domain (for each language separately). The domains chosen were family, neighborhood, religion, education, and work. While it is no doubt important to consider the domains in which languages are used (Grosjean, this volume, and Carroll & Luna, 2011), it is difficult to decide on the domains that need to be measured. Fishman and Cooper also used a word frequency estimation task in their test battery in which respondents were asked to rate on an 8 point scale how frequently they had heard or used 75 Spanish and 75 English words. Another vocabulary task used in the study was a word association task in which participants were asked to provide associations with five words in each language.

More recently, Cromdal (1999) used 48 pictures (24 in each language) from the Peabody Picture Vocabulary task (PPVT, Dunn, 1965) to classify 38 Swedish-English children into highly bilingual (balanced) and partially bilingual (dominant) groups. As in the

[1] The authors did not explicitly aim to measure language dominance with these tasks but it is clear the measures they propose could be used for this purpose.

Lambert et al. (1959) study, the author calculated the difference between the scores on each language to measure language dominance. In this case, the median difference score was used to separate the two groups. Thus, nineteen children ended up being classified as balanced bilinguals and an equal number as partially bilingual. For the latter group Swedish was found to be the weaker language.

Reyes and Azuara (2008) also used the PPVT to classify twelve Mexican Spanish-English bilingual children in the US. They found that ten of these children obtained higher scores on the Spanish version of the test (TIVP, Dunn, Padilla, Lugo & Dunn, 1986), and concluded that these ten children were Spanish-dominant. Of the remaining two, one was classified as English-dominant and one as balanced. Unfortunately, the authors do not explain how they decided upon the cut-off point between the different categories of bilinguals. Another important issue is the comparability of the two versions of the Peabody vocabulary task. According to Prewitt Díaz (1988, cited in Bedore et al., 2012: 618), the Spanish version does not meet standards of test translation and validation.

Finally, Gollan, Weissberger, Runnqvist, Montoya and Cera (2012) developed a new test to measure language dominance in bilinguals, the Multilingual Naming test, which consists of a set of 68 black-and-white line drawings. Participants are required to name all items first in one language and then in the other. Items were carefully selected on the basis of frequency and items for which cognates existed between the languages under study. Again, difference scores were used to compute language dominance, although the cut-off point used was much lower than in some of the other studies mentioned above. The authors allowed for only 5 percent difference between the scores in the two languages. If the difference was larger than 5 percent (in either direction), participants were no longer considered to be balanced. On the basis of this criterion, only 20 percent of the participants were classified as balanced and 80 percent as dominant in one or the other language in this study. While we agree with Gollan

et al. that cut-off points are arbitrary, setting such a restrictive criterion makes it very unlikely for a bilingual to be classified as balanced, and reinforces the widely held assumption that this goal is very difficult to achieve.

As the above discussion of a few selected publications on language dominance illustrates, a wide variety of different vocabulary measures have been used to study language dominance. A first point we would like to make about these measures is that they clearly tap into different components of word knowledge (e.g. receptive versus productive knowledge) and make different kinds of demands on the language processing system, because some tests are timed and some are not. It is therefore not surprising that the tests produce very different results when used to calculate language dominance. We agree with Bedore et al. (2012), who compare a wide range of measures and tests, that it is not useful to try and decide which one is "the best" to measure language dominance. Bahrick, Hall, Goggin, Bahrick & Berger (1994: 265) take a more fruitful approach in this matter by raising awareness of a number of "task-specific language dominance effects". They claim, for example, that an individual's recent language environment is more likely to affect tasks which require a speeded retrieval of words, such as a category generation task or a lexical decision task, than sentence comprehension or word recognition tasks, because "frequency and recency of retrieval strongly affect access to the lexicon (particularly speeded access) on the basis of a cumulative and generalized priming effect but do not affect the availability of the lexicon" (Bahrick et al., 1994: 265). They also assume that self-paced tasks more accurately reflect the size of the lexicon in either language, as frequency of recent retrieval is not so important in that case. This does not mean, in our view, that timed tasks are necessarily less appropriate for the purpose of measuring language dominance, but it is important to be aware of the fact that timed and non-timed tasks will lead to rather different classifications of bilinguals.

A second issue in the operationalization of language dominance that needs to be carefully considered is the issue of typological differences between languages. As is well known, there are important phonetic, morphosyntactic, and semantic/conceptual differences between languages. In his overview of the ways in which researchers have measured the language proficiency of bilinguals, Hulstijn (2012: 428) notes, for example, that fluency measurements are not easily comparable across languages because of differences in the duration of even a single sound in two languages. However, it is possible to adjust scores for overall typological differences between languages, as is done in Daller, Yıldız and De Jong (2011), who found that Turkish stories told by bilinguals were on average shorter than the German ones by a ratio of .89. This means that if bilinguals told stories in German which consisted of 100 words, their Turkish stories would be 89 words long. They therefore adjusted the number of tokens in the German stories by this ratio to make the measurements more comparable. For researchers who use measures of lexical diversity which are calculated on samples of spoken or written language, differences in inflectional or derivational morphology between the two languages of a bilingual can make it very difficult to compare scores across languages (see section 3). In such cases, careful lemmatization can help to reduce discrepancies that are due to typological differences, as is done in Treffers-Daller (2011) and in the current study, but it is unlikely that the resulting scores are completely comparable.

A third issue is the comparability of tests across different languages. Creating vocabulary tests that are equally difficult in each language is very complicated, not only because translation equivalents can belong to different frequency layers in the lexicons of each language, but also because there is no one-to-one relationship between constructs and words in two languages. Thus, for example, Dutch *kom* "bowl" cannot easily be translated into French because the objects contained in the Dutch category *kom* are systematically distributed over two categories in French, namely *plat* and *bol* (Ameel, Malt, Storm & Van

Assche, 2009). Any parallel tests will therefore need to be constructed and piloted very carefully to ensure they are comparable.

A fourth issue that emerged from the studies discussed above concerns the ways in which a single score or index of language dominance can be calculated on the basis of test scores from two languages. Several researchers choose to calculate difference scores by subtracting scores in one language from the other. If this approach is adopted, the question that remains to be resolved then is how to decide on a cut-off point that will allow the researcher to distinguish between balanced and dominant bilinguals. Some researchers opt for a strict criterion, allowing for only 5 percent deviance from zero for the balanced group (Gollan et al., 2012), whilst others choose the median difference score to divide the participants into balanced and dominant groups (Cromdal, 1999) or divide the participants in five different groups on the basis of difference scores which ranged from -100 to 100 (Bedore et al., 2012). In this case, the balanced bilingual group consisted of bilinguals who obtained scores between -20 and 20, while those whose scores fell outside this range were considered to be either dominant in one or the other language. A possible solution to the issue of the comparability of cut-off points is to convert scores from different tests to Z-scores, so that measures are all standardized in the same way. This way we can more easily compare scores from different measures and have a more focused discussion about appropriate cut-off points. The choice of a particular cut-off point will, of course, depend to a large extent on the purpose of the study and the amount of variability in language dominance that is allowable for that purpose. The key point is that researchers need to make it clear which decisions have been taken, as this is the only way in which research in this field can progress.

3. Using lexical diversity indices to measure language dominance

Lexical diversity measures are often used to investigate the range of words that are deployed in spoken or written language. Lexically diverse texts are those which consist of a relatively high number of different words whilst in less diverse texts the same words are repeated more frequently. The basic distinction on which most measures of lexical diversity are based is that between types (V) and tokens (N), whereby types stands for the number of different words, and tokens represents the total number of words in a text. Thus, the sentence *the cat sat on the mat* consists of five types and six tokens, because the word *the* is repeated once. One of the most basic ways to compute the lexical diversity of a language sample is to compute the type-token ratio (TTR), which means the number of types divided by the number of tokens. For the above example, the TTR is 5/6 or 0.833. A well-known problem with this measurement is, however, that its values drop with increasing text length. Thus, for the sentence *the black cat sat on the black mat*, the TTR is 6/8 or 0.75. This example illustrates that the TTR is generally lower for longer texts, because most people tend to repeat the words they use the longer they speak. Therefore, researchers have developed alternatives which are less dependent on text length, such as the Index of Guiraud (Guiraud, 1954), which is the ratio of types over the square root of the tokens ($V/\sqrt{N}$), and other derivatives of TTR (see Tweedie & Baayen [1998] for an overview). More recently, Malvern, Richards, Chipere and Durán (2004) proposed the D value, which is widely used in L1 and L2 research. The D-value is computed through a series of computations of the Type-Token Ratio (TTR) (Templin, 1957) on samples of different text lengths (typically ranging from 35-50 tokens) after which a random sampling TTR curve is computed (but see McCarthy & Jarvis, 2007, for a critical appraisal of D. McCarthy and Jarvis (2010) also proposed an alternative measure of lexical diversity, namely HDD, which calculates, for each lexical type in a text, the probability of encountering any of its tokens in a random sample of 42 words drawn from the text. While these measures

constitute an improvement on TTR, all existing measures continue to be dependent on text-length (Treffers-Daller, 2013).

As Malvern et al.(2004) point out, measures of lexical diversity are often used as general-purpose measures of spoken and written language development in either L1 or L2, but there are only very few studies in which they are deployed to measure language dominance across two languages. One example of a study in which language dominance is measured with lexical diversity scores is Dewaele and Pavlenko (2003), who compared lexical diversity scores in Russian and English among Russian-English bilinguals, L2-learners of English, and native speakers of English and Russian. They used the Uber-Index (Dugast, 1980), a logarithmic transformation of the TTR, to measure lexical diversity and found that the scores were generally lower for English than for Russian, possibly because of the morphological differences between the languages, but did not compute difference scores.

Daller, Treffers-Daller and Van Hout (2003) also used lexical diversity measures to investigate language competence in both languages of bilinguals, namely among two groups of Turkish-German bilinguals who lived in Germany and in Turkey. They used the Index of Guiraud (Guiraud, 1954) and introduced a new measure of lexical sophistication (The Advanced Guiraud), which is the ratio of the advanced types (that is, types not found in a basic vocabulary list) over the tokens (advanced types/√tokens). These were used to analyse transcripts of Turkish and German stories elicited from study participants with the help of different cartoons. The authors found that the participants' lexical diversity scores were asymmetrical across the two groups of bilinguals and across the two languages: the bilinguals in Germany obtained significantly higher scores in German than in Turkish and for the bilinguals in Turkey the opposite was true. Instead of using difference scores, they converted the scores for each language into Z-scores and computed whether there were significant

differences between the Z-scores for each language. This conversion to z-scores made it possible to compare scores between two languages that are structurally very different.

Finally, Treffers-Daller (2011) analysed language dominance in French- English and Dutch-French bilinguals, using the Index of Guiraud (Guiraud, 1954) and the D-measure (Malvern et al., 2004). She calculated an overall group median for all D-values from all three languages and used this as the cut-off point between lower and higher proficiency groups for each language. Each participant could then be allocated to one of four different groups of bilinguals: a higher proficiency level group of balanced bilinguals (with scores in the upper range for both languages) and a lower level group of balanced bilinguals (with scores in the lower range for both languages), or one of two groups of dominant bilinguals (who had a high score in one but a low score in the other language). The predictive value of the classification was then tested in relation to different variables, e.g. directionality of code-switching.

In the current study, we have chosen not to use vocabulary tests to measure language dominance, because comparable tests do not exist for Polish and English, but to analyse the vocabulary deployed in stories told in both languages (see section 1). We use one traditional measure of lexical diversity, the Index of Guiraud, and one of the more recent measures, HDD to investigate language dominance. The latter measure was chosen because McCarthy and Jarvis (2010) call for further validation of this measure and we hope the current paper will therefore also contribute to the development and testing of measures of lexical diversity in different languages.

4. Methodology

4.1 Participants

All 26 participants in this study were Polish-English bilinguals (ten males and sixteen females) living in the UK. The age of the participants ranged from 18 to 64 (mean age 33.19, SD 11.41) at the time of recording. While the majority of the participants were in their twenties and thirties when the data for this study were collected, older generations of migrants are also represented in the data but to a lesser degree; this reflects the approximate proportions of the Polish migrant community in the UK, now dominated by the recent arrivals.[2] The majority (22) were late bilinguals: they arrived in the UK as young adults, generally in their early or late twenties, often after Poland joined the European Union in 2004. On average the participants had stayed in the UK for 12.05 years (SD 18.12) by the time the study started, although there are large differences between the participants: the majority (17) had been in the UK for four years or less at the time of recording, and three of these had only stayed in the UK for one year. There were four long term residents in the group: four had been in the UK for over 40 years, and of these four, three were born in the UK, and one arrived at the age of five. All four are early bilinguals in that they had used English as well as Polish from early childhood onwards. For the current study this wide range in length of residence in the host culture is interesting because the participants will likely differ in their proficiency in English as well as Polish. Thus, it is expected that there will be differences in their language dominance profiles to which the proposed indices of language dominance should be sensitive.

[2] Census data released on November 11, 2012 provides the most recent official number of Poles living in the UK (579,000) and indicates a ten-fold growth since 2001 (when the corresponding figure was 58,000).

The participants were asked to indicate in percentages how often they used English and Polish in daily life at home and at work (see Figure 1). As might be expected, and predicted by the Complementarity Principle (Grosjean, this volume), they reported using English more often at work than at home (Wilcoxon, Z = -4.16, p <.001), and Polish more at home than at work (Wilcoxon, Z = - 4.133, p <.001). Twelve participants (46 percent) reported the ability to speak another language apart from Polish and English, but only one (TIK) reported actually using this language (Spanish) on a daily basis at home as well, namely for 20 percent of the time. Only four out of 26 participants reported using Polish more than 50 percent of the time at work. The others used mainly English and fourteen participants reported using exclusively English at work. In total, the participants clearly use more English (62 percent of the time) than Polish (37 percent of the time) and these overall differences in usage of English and Polish were also significant (Wilcoxon, Z = -2.20, p < .05). This reported preference for English is probably related to the fact that English is the predominant language in the UK society (for the term "predominant language", see the introduction to this volume and Silva-Corvalán, 2014), but does not warrant the conclusion that the majority of participants are English-dominant. Because most participants had moved to the UK as adults and seventeen had been there for four years or less, it is unlikely that the majority had become English-dominant in such a short period of time.

=====Figure 1 approximately here======

As far as education is concerned, the group comprises participants with secondary or incomplete secondary education as well as graduates of universities and polytechnics, both in Poland and the UK. The participants' professional background was rather varied; the range of

professions represented in the sample was relatively broad and included: an accountant, a baker, two cleaners, six managerial staff at different levels, four catering staff/waiters, two manual labourers, a part-time shop assistant, an IT specialist, two nurses, a hospital administrator, a freelance photographer, a teaching assistant, two office clerks, and one retired individual. The participants came from different locations across Poland and the UK (in the case of second-generation participants). The data were collected as part of the second author's PhD thesis on Polish-English code-switching (Korybski, 2013). A social network-based recruitment as described by Milroy (1980) was chosen to recruit participants, as random sampling would be highly unlikely to result in the kind of informal data one needs to analyse code-switching. The starting point for the recruitment of participants was a group of a few Poles known to the second author, who were able to subsequently put him in touch with further participants. In order not to limit the scope of participants to a relatively homogeneous group in terms of age (which can be a consequence of the "snowball effect" in social networking), the researcher developed contacts with the Polish Catholic Church in Bristol at Cheltenham Road. This enabled him to conduct a few interviews with speakers aged over 45 and second generation speakers, either on church premises or during a weekend trip of the Polish Club organized by the church in June 2009. All interviewees were volunteers and did not receive any reward for taking part in the study.

4.2 Instruments and procedure

The data collected for this project consisted of transcripts of an informal interview, questionnaire data and narratives which were elicited with the help of two cartoon books which have been used very frequently in research in first and second language acquisition, (see Berman & Slobin, 1994; Treffers-Daller, 2011), namely the frog stories *Frog Goes to Dinner* (Mayer, 1974) and *Frog, Where Are You?* (Mayer 1969). The cartoon books portray adventures of a little frog which runs away to the forest (*Frog, Where Are You)* or slips into

its owner's pocket to go to a restaurant (*Frog Goes to Dinner*). Both story books are 32 pages long and therefore comparable in length, even though it is difficult to ascertain that they contain exactly the same amount of content. No text is used in the book (apart from pictures with the signs "Fancy Restaurant" and "Fire Exit"). Participants needed to reconstruct the narrative from looking at the pictures. Participants told *Frog, Where are you?* in Polish and *Frog goes to Dinner* in English. Although originally written for children, the cartoon stories are suitable for elicitation with adults with a wide range of different educational backgrounds, as shown in Treffers-Daller (2011), where this method was successfully used to measure language dominance in adult French-Dutch and French-English bilinguals.

Using oral story elicitation as a technique to measure language dominance has a number of advantages because participants who cannot read or write in one of their two languages are not disadvantaged. Another advantage is that *both* languages of a bilingual can be studied in the same way, which is not always possible if one uses standardized tests, because such tests are often only available in one of the languages under study. Furthermore, all the vocabulary used in the language sample can be analysed (not just a selection), and vocabulary is measured in context (not just isolated words) in a naturalistic setting (not a laboratory). Finally, in our experience, participants enjoy telling stories, and find this kind of task much more enjoyable than filling in a test, which is very important if one wants to obtain a valid measurement of participants' language ability.

There can also be disadvantages of this technique: some participants may simply not be good at storytelling or may not feel comfortable with a storyboard designed for children. In the case of the present research the advantages seemed to outweigh the disadvantages, and participants appeared to enjoy carrying out the task. The frog stories told by each participant in Polish and in English were recorded and then transcribed. Their length varies from just over two minutes to more than six minutes. The data were recorded during the second

author's stays in England in 2007 (November), 2008 (June and July), 2009 (June and July), 2010 (May and June) and 2011 (April). Prior to the recordings, he established contacts within the Polish communities in and around Bristol and in London through a network of friends who had migrated to the UK after Poland's accession to the European Union.

It is also important to mention here that the first story to be administered was always *Frog, Where Are You?* Since the interviews were held in Polish, it was perceived as more reasonable and participant-friendly to enable the speakers to get accustomed to the exercise with Polish first. All participants who were interviewed for the data set were asked to tell the stories after the recorded conversations they had with the investigator. Additionally, the participants were asked to interpret the stories freely and to assume the listener was an adult native speaker of the language. In all interviews, frog story recordings were followed by administration of a questionnaire which was used to obtain information about participants' language backgrounds and attitudes towards English and life in the UK.

All of the recordings were made in naturalistic settings in locations across the South-West of England and London. Before the story telling task participants were recorded for an informal interview which lasted between 10 and 40 minutes. While these interviews were held in Polish, the second author, who recorded the interviews, made it clear to participants before the interviews that he understood English well so as not to impede code-switching. However, no mention was made of code-switching, bilingualism or mixing the two languages; the information about the interviewer's knowledge of English was usually passed on as an informal comment during the initial meeting with prospective participants. Data from these recordings were used to investigate participants' code-switching behaviour, as we wanted to know to what extent the ILD could predict the code-switching behaviour of our participants. We counted all intrasentential code-switches in the data (mainly insertions of single English words into Polish) and created a variable (CSTot) which represented the total number of

intrasentential code-switches per participant in the data set of spontaneous conversations. On average, participants produced 16.34 switches. In addition we looked at participants' production of different types of code-switching, namely single nouns, NP insertions, adjectives, verbs and "other switches". This group of other switches consisted of switches of acronyms and place names which could clearly be identified as either English or Polish (on average, 32.2 per participant).

Finally, participants filled in a questionnaire with a list of attitudinal questions and statements which they graded on a seven-point Likert scale. These were used to find out participants' attitudes towards English and towards the UK, because we wanted to know to what extent attitudinal variables correlated with the proposed indices of language dominance. For this purpose, we created a variable EngAtt from the mean scores of four questions out of fifteen which tapped into participants' attitudes towards learning and using English as well as towards life in the UK, see (1) for details. The remainder of the questions focused on attitudes towards Polish, Poland or code-switching and therefore tapped into different constructs. For further information regarding the development of the questionnaire, see Korybski (2013).

(1) Questions used for the measurement of attitudes towards learning English and life in the UK.

1. Poles who have come to the UK should feel obliged to learn English.
2. I like English and I enjoy speaking it.
3. I like England, I feel fine here and I am going to stay.
4. I read the British press regularly, watch the news and am up to date with the country's socio-economic situation.

4.3 Data analysis

The stories were transcribed in CHAT format and analysed with CLAN tools (MacWhinney, 2000). The transcriptions were carefully edited by the second author and proper names given to characters in the story removed prior to data analysis. This is standard procedure in studies of lexical diversity (Malvern et al., 2004), because keeping place names and proper names in calculations of lexical diversity leads to an inflation of scores: If a participant uses proper names such as *Peter* or *Mary*, and these are included in the analysis of lexical diversity, this increases the participant's lexical diversity scores although their use does not necessarily indicate that participants possess a rich and diverse vocabulary in English Overregularisations of irregular past tenses such *throwed* for *threw* and similar mistakes were corrected as we wanted to recognise the fact that their use indicated participants had at least partial knowledge of these words. The current project focused on lexical rather than grammatical issues, and a detailed analysis of whether or not words were used correctly from a grammatical point of view was beyond the scope of the current analysis Unintelligible words or unrecognisable words were omitted from the analysis.

In order to measure lexical diversity we first needed to establish how many different types and tokens were found in each story. Because nouns, verbs and adjectives are inflected in English, it is important to ensure different inflected forms of a verb (e.g. *do, doing, does, did, done,* etc.) are counted as different tokens of one type (the verb *to do*) rather than as different types (Jarvis, 2002). It is therefore necessary to calculate lexical diversity on a lemmatized file (Treffers-Daller, 2013). Two CLAN commands were used to facilitate the calculation of types and tokens in the English stories. Firstly, the MOR command created a morphosyntactic tier in the transcript, linked to the main tier with the participants' utterances. Subsequently, the POST command disambiguated between homographs (e.g. *can* as a modal

verb and *can* as a noun). An example of an English CHAT transcript with the main tier and the dependent morphosyntactic tier is given in (2). The three letter code *BEA indicates the speaker's name as is common in CHAT coding.

(2) Example of a sentence transcribed in CHAT with a main tier and a dependent morpho-syntactic tier.

*BEA: I'm going to have a big big big huge meal and mum looks so nice .
%mor: pro:sub|I~aux|be&1S part|go-PROG inf|to v|have det|a adj|big adj|big adj|big adj|huge n|meal coord|and n|mum v|look-3S adv:int|so adj|nice .

The CLAN command freq +t%mor –t%* +s@r-*,o-% +d3 was used to calculate the final number of lemmatized types and tokens on the morphosyntactic tier. The +d3 switch sends the output to an excel spreadsheet.

In the case of the Polish transcripts the process was not as straightforward. CLAN does not have a dictionary or a MOR tier for Polish, and measurements on the main tier would have led to highly inflated values because Polish is a highly inflected language with seven cases. Therefore, an alternative to the procedure described above was necessary. As a first step all transcriptions of frog stories into Polish were lemmatized on the main tier, as in (3).

(3a) Fragment of a frog story CHAT transcription in Polish with lemmatized words

*EWA: ten słoik nie był [: być] mały i żabka była [: być] bardzo zadowolona [: zadowolony] i któregoś [: któryś] wieczoru [: wieczór] siedział [: siedzieć] i patrzył [: patrzeć] na żabkę [: żabka] razem z psem [: pies] a w nocy [: noc] jak już poszedł [: pójść] spać i jego

pies leżał [: leżeć] razem z nim [: on] na łóżku [: łóżko] żabka wystawiła [: wystawiać] nogi [: noga] i wyszła [: wyjść] z tego [: ten] słoika [: słoik].

(3b) Glosses and free translation

Ten	słoik	nie	był	mały	i	żabka	była
This	*jar*	*not*	*was+masc.+3SG*	*small*	*and*	*frog*	*was+fem+3SG.*

bardzo	zadowolona	i	któregoś	wieczoru	siedział
very	*happy+fem.*	*and*	*one+Gen..*	*evening+Gen.*	*sat+masc.+3SG.*

i	patrzył	na	żabkę	razem	z	psem	a	w
and	*looked+masc.+3SG*	*on*	*frog+Acc.*	*together*	*with*	*dog+Ins..*	*and*	*at*

nocy	jak	już	poszedł	spać	i	jego	pies
night+Ins..	*when*	*already*	*went+masc+3SG*	*sleep*	*and*	*his*	*dog*

leżał	razem	z	nim	na	łóżku	żabka
lay+masc.+3SG	*together*	*with*	*he+Ins.*	*on*	*bed+Loc.*	*frog*

wystawiła	nogi	i	wyszła	z	tego	słoika.
put out+fem+3SG	*legs+Acc.*	*and*	*went out+fem+3SG*	*from*	*this*	*jar+Gen.*

**EWA: this jar was not small and the frog was very happy and one evening the boy was sitting and looking at his frog together with his dog and at night when he went to bed and his dog was lying next to him on the bed the frog put its legs out and went out of the jar.*

We used CLAN to calculate the number of types and tokens on the main tier of the Polish transcripts, because CLAN reads the information in brackets, and thus counts *był* "was"

(masculine form) and *była* "was" (feminine form) in (3a) as tokens of the type *być* "to be" rather than as two different types.

As we wanted to calculated HDD with the Gramulator[3] (McCarthy, Watanabe & Lamkin, 2012), and this software cannot read the CHAT coding, we also created a plain text lemmatized version of the English and the Polish transcripts with the help of the change string command (CHSTRING), as in (4a), from which we extracted a plain text lemmatized file without coding (4b) with FLO.

(4a) Lemmatized version with coding on the main tier

I am [: be] going [: go] to have a big big big huge meal and mum look [: looks] so nice

(4b) Plain text lemmatized version without coding

I be go to have a big big big huge meal and mum look so nice.

After the lemmatization was completed, the number of types and tokens was computed for all 26 transcriptions in both languages. As expected, the bilinguals produced stories which greatly varied in text length. In Polish, the stories varied in length between 164 and 712 tokens and in English they ranged from 134 to 972 tokens, (see Table 1 for details). It is remarkable that the mean number of tokens used for both stories is larger for English than for Polish, although English is the second language of the participants. These differences are statistically significant (t=6.0; df = 25, p <.001). Differences in text length cannot be due to participant fatigue, because participants told the English story after completing the Polish one. If participant fatigue had played a role, the English stories should have been shorter. While it

[3] The gramulator subtracts 35 from the HDD values obtained. We added 35 onto the output of the gramulator in order to ensure the original HDD values could be used.

is theoretically possible that there was more to say about *Frog goes to dinner* than about *Frog where are you?*, it is not a likely explanation because the books have exactly the same number of pages, namely 32. We think that the lower number of tokens for Polish is probably due to typological differences between the languages: there are no articles in Polish, subject pronouns are often dropped and the past tenses are not formed with the help of auxiliaries but with suffixes (Franks, 1995), as is illustrated in (5) and (6).

(5) Prac+uje
Work+3sg
"He/she/it works."

(6) Prac+ował
Work+past tense 3sg masculine
"He was working."

Thus, the same content is, at least in these cases, expressed with fewer tokens in Polish than in English. These examples illustrate the difficulty of comparing data from two different languages. While the lemmatization ensures that differences between Polish and English with respect to inflections are reduced, it is not possible to completely eliminate all differences between the two languages in this process.

It is important to note, however, that the number of types used in the stories in both languages is virtually the same, which indicates that, although the English stories are longer, they contain roughly the same number of types as the Polish stories. As pointed out above, this is probably due to the fact that there are more function words in English than in Polish. The second reason why the English stories are generally longer than the Polish ones may be that the participants needed more words to express themselves in their second language than in their first language. For example, one of the participants used the phrase *bees' house* to

replace the more specialised *beehive*, while another one chose the paraphrase *makes big surprise* to replace the verb *surprise.* Similar results have often been found (Treffers-Daller, 2013), and this phenomenon has been described as the "waffle phenomenon", a term coined by Edmondson and House (1991).

=====

Table 1 approximately here

=====

The fact that there were significant differences among participants in the number of words they needed to tell the stories made it necessary to control for text length in the calculation of lexical diversity, because all measures of lexical diversity are dependent on text length (Treffers-Daller, 2013). As the shortest story contained only 134 words, we could have decided to use segments of 134 words from each transcript. This was not the best solution, because calculating measures of lexical diversity on very small samples is problematic (McCarthy & Jarvis, 2010; Treffers-Daller, 2013). We decided instead to use larger segments and opted for 190 tokens as the cut-off point for each transcript, which led to the exclusion of four participants who did not produce enough words in either English or Polish. The analyses in this chapter are therefore based on calculations of data from 22 out of 26 participants.

We used the Gramulator (McCarthy et al., 2012) to segment the data. After selecting segments of 190 words from the middle of each lemmatized transcript we computed HDD for all English and Polish transcripts with this software. The participants' scores on the Index of Guiraud were computed with the *compute* command in SPSS. We then calculated the Index of Language Dominance (ILD) in two ways, see (7) and (8). First, we subtracted the Polish Guiraud values from the English ones to compute ILD1, and then the Polish HDD scores from the English ones to calculate ILD2. In addition we used the Edinburgh formula proposed in

Birdsong (this volume) to compute a third and a fourth index of language dominance (see 7-10). This formula is based on a ratio of (Difference in Scores) / (Sum of Scores).

(7) ILD1 = English Guiraud – Polish Guiraud

(8) ILD2 = English HDD – Polish HDD

(9) ILD3 = (English Guiraud – Polish Guiraud)/(English Guiraud + Polish Guiraud)*100

(10) ILD4 = (English HDD – Polish HDD)/(English HDD + Polish HDD)

These indices are to be interpreted as follows: positive scores are found for participants whose English scores were higher than their Polish ones and negative scores for those who obtained higher scores on the Polish measures than on the English ones. We also needed to decide on a cut-off point to distinguish between the balanced bilinguals and the ones who were dominant in either English or Polish. While it would have been in principle possible to consider only participants who obtain a score of 0 as balanced, it would be extremely unlikely for anyone to obtain exactly the same score in each language on the story telling tasks. Indeed, none of the participants would have been classified as balanced if this criterion had been adopted. This outcome also seemed improbable on the basis of the participants' language biographies and their reported use of each language, as displayed in Figure 1. Gollan et al. (2012) used difference scores of 5 percent in either direction as the cut-off point, which seems very strict too. If we adopted this criterion only two participants would be classified as balanced on the basis of the ILD1 and none if we use the ILD2. Because the Index of Guiraud and HDD are measured on different scales and we wanted to use the same cut-off point for each measure of language dominance, we converted the scores on both indices to Z-scores. This way all indices could be displayed on the same scale, with a mean of zero, and we could set the same cut-off point for both indices to determine which participants would classify as balanced

bilinguals and which ones as dominant in either Polish or English. We decided to choose one standard deviation below or above the mean as the critical value, as this cut-off point is very commonly used in statistics to identify significant differences. This way participants with Z-scores higher than 1 would be classified as English-dominant and those with Z-scores below -1 would be Polish-dominant, and the remaining ones (with Z-scores between -1 and +1) as balanced bilinguals, but we will also illustrate how the use of a more restrictive cut-off point of .5 affects the classification.

5. Results

Figure 2 displays the distribution of the four Indices of Language Dominance (transformed to Z scores). The distributions for the first three are roughly comparable, but on the ILD4 there are three outliers. Importantly, however, the distributions of the scores on the four indices are not significantly different from the normal distribution as tested with Kolmogorov-Smirnov tests.

=== figure 2 approximately here ====

As explained in section 4, we had chosen the cut-off point of one standard deviation to distinguish between balanced and dominant bilinguals. If this criterion is used as the critical value, sixteen participants are classified as balanced based on the ILD1 scores with three being classified as Polish-dominant and three as English-dominant. For the ILD2 and the ILD4 the result is the same, but for the ILD3 there are seventeen balanced bilinguals, three Polish-dominant and two English-dominant participants (see Figure 3). If we choose a more restrictive cut-off point, namely .5 standard deviations, only eleven participants are classified

as balanced, five as Polish-dominant and six as English-dominant (with ILD1 as criterion) and the corresponding outcome for the ILD2 is eight balanced bilinguals, eight English-dominant, and six Polish-dominant participants. For the ILD3 it is the same as for the ILD2, and for the ILD4 there are twelve balanced bilinguals, five English dominant and five Polish-dominant bilinguals.

As we have explained above, it is not possible to take a principled decision regarding the most appropriate cut-off points at the moment, as more research into this issue is needed, but we can investigate to what extent the scores on the four indices correlate with other variables.

==========

Figure 3 approximately here

==========

Table 2 shows that the four indices of language dominance correlate significantly with each other. The ILD1 and the ILD3 (both based on Guiraud) correlate almost perfectly with each other, which was to be expected as they are mathematical transformations of each other. The same is true for the ILD2 and the ILD4 (both based on HDD). Although the Guiraud-based indices and the HDD-based indices both measure lexical diversity, the correlations between the ILD1 and the ILD2 are not as strong as one might have expected (.557), but the correlations between the ILD3 and the ILD4 are slightly stronger (.662).

=====

Table 2 approximately here

====

We validated our measures of language dominance, by investigating to what extent these can predict scores on other variables, as recommended by Flege, Mackay and Piske (2002). An added advantage of using correlations and regression is that the measure of dominance can be used as a continuous variable: there is no need for categorization in groups of dominant or balanced bilinguals (see Birdsong, this volume). We also agree with Nicoladis (this volume), who points out that these other variables need to be independent from the measure of language dominance. We decided to correlate the scores on the language dominance indices with four different variables (all measured independently of the measures of language dominance): Length of residence in the UK (LOR), mean score on the attitude questions (EngAtt), reported use of English at home (Enguse) and code-switching behaviour (CSTot).

The rationale for our choice is that we expect English-dominant bilinguals to have spent a longer time in the UK than Polish-dominant bilinguals, to express more positive attitudes towards English and life in the UK, to use English more at home and to engage more in code-switching.

The correlations between the indices of language dominance and the four variables mentioned below can be found in Table 3. In this table we report Pearson correlations, except for correlations with LOR and CSTot. We used Spearman for correlations with these two variables as they are not normally distributed. We begin with a presentation of the results for LOR, followed by EngAtt, EngUse and finally CSTtot.

Correlations with LOR

Table 3 shows that the ILD2 and the ILD4 correlate strongly with LOR, whilst the ILD1 does not correlate with LOR, and there is a mid strength correlation between LOR and the ILD3. The correlation between the ILD4 and LOR is shown in Figure 4.

=====

Table 3 approximately here

=======

As might be expected LOR correlates significantly with EngAtt (.789**) and EngUse (.516*), but correlations between LOR and CSTot were not significant.

A more detailed perspective on the role of LOR can be obtained from looking at the profiles of two participants with the longest and two with the shortest LOR. The two participants who had stayed in the UK for the longest period of time when the recordings were made, namely over 50 years (KRY and STA) were both classified as English-dominant, whilst the two with the shortest period of residence (BAR and XOR), who had only been in the UK for a year when the data were collected, were clearly Polish-dominant. It is important, however, to realise that participants with long periods of residence are not necessarily English-dominant: two participants who had stayed in the UK for 41 and 49 years, (KAR and TER, respectively) were both classified as balanced bilinguals. Their ILD1 scores (transformed to Z-scores) were positive, which means their English scores were slightly higher than their Polish scores, but not above 1. These examples illustrate that whilst LOR is clearly important, there are of course many other factors which exert an influence on a person's language ability in the two languages.

==== Figure 4 approximately here =======

Correlations with attitudes to English and life in the UK

As explained above, we expect informants who hold more positive attitudes to English and life in the UK to obtain higher scores on the indices of language dominance. This was indeed the case for three out of four of the indices. The ILD1 and the ILD3 correlate most strongly with EngAtt (see Table 3). The ILD2 correlates slightly less strongly with EngAtt, and the ILD4 does not correlate with this variable at all. The correlation between ILD1 and EngAtt can be seen very clearly in Figure 5.

====Figure 5 approximately here====

Correlations with English usage at home

One factor which is likely to be related to the participants' language dominance is their frequency of usage of English at home. In fact, if frequency of use is a key component of language dominance, we would expect there to be strong correlations with measures of usage frequency. We did indeed find strong positive correlations for all indices, except the ILD1. The ILD4 correlates most strongly with EngUse (.687**), followed by the ILD2 (.644**) and the ILD3 (.423*). As one might expect, LOR also correlates fairly strongly with EngUse (.516**).

Correlations with code-switching

As many researchers assume there is a link between language dominance and code-switching (Genesee, Nicoladis & Paradis., 1995; Schmeisser et al, this volume; Treffers-Daller, 2011), we also investigated whether scores on the indices of language dominance were correlated

with code-switching behaviour. We found that only the ILD3 correlates significantly with CSTot ($r_{s=}$ .454, p =.034).

Because there are many different types of code-switching in the data (see Korybski, 2012, for further details), we also investigated whether there were any correlations between the Indices of Language Dominance and particular types of code-switching. The analyses revealed there were no correlations between the ILD1 and insertions of nouns, NPs or adjectives, but there was a positive correlation with code-switching of verbs (.464*) and other switches (.687**). For the ILD2, there was a correlation with the other switches (.508*) only but not with other categories of switches. The ILD3 correlated significantly with the other switches (.667**) and CSTot (.454**), but the ILD4 did not correlate with any type of code-switching.

Finally we investigated whether the indices of language dominance could predict English use at home. We chose this variable because we see language dominance as a "global measure of relative frequency of use and proficiency in each language" (Wang, 2013). Therefore a valid index of language dominance should be able to predict language use among the bilinguals in a study.

We carried out a series of simple linear regressions as the different indices are too strongly correlated to be entered together in a multiple regression. We found that all indices were significant predictors of use of English at home, except the ILD1. The strongest predictor was the ILD4 (Beta .687, SE = .029, p <.001, adjusted R^2 = .445), followed by the ILD2 (Beta = .683, SE = .29, p <.001, adjusted R^2 = .440), and finally the ILD3, which was marginally significant (Beta =.423, SE = .36, p =.05, adjusted R^2 = .138). Other variables (e.g. LOR and EngAtt) were not found to be significant predictors of EngUse in a regression model.

It was not possible to predict CSTot with the help of the indices of language dominance but the ILD1 was a significant predictor in a regression analysis were the number of switches of verbs was the dependent variable (Beta =.439, SE = .13, p =.041, adjusted R^2 = .152). All language dominance indices had approximately equal power in predicting the number of switches in the "other" category (switches of acronyms and place names).[4]

6. Discussion and conclusion

In this paper we have seen that it is possible to calculate language dominance on the basis of measures of lexical diversity when the data are carefully cleaned and lemmatized. This was particularly important for the language pair under study, Polish and English, where the first is heavily inflected and the second is not. In addition, text length was kept constant to avoid problems related to text length dependency. We computed four Indices of Language Dominance (ILD) on the basis of two different measures of lexical diversity, the Index of Guiraud (Guiraud, 1954) and HDD (McCarthy & Jarvis, 2007). The first Index of Language Dominance, called ILD1, was computed by subtracting the Polish Guiraud scores from the English Guiraud scores. The second Index of Language Dominance, called ILD2, was computed in similar way by subtracting the Polish HDD scores from the English HDD scores. For the third Index (ILD3) we used Guiraud again but applied the Edinburgh formula to its calculation, as proposed by Birdsong (this volume). This formula is a ratio of (Difference in Scores) / (Sum of Scores). The same formula was applied in the calculation of the ILD4, but this time we used HDD instead of Guiraud.

[4] For these the Beta values were around .5, R^2s were between .26 and .28 and p values between .001 and .008.

Positive scores on each of these Indices of Language Dominance mean that informants are more English-dominant and negative scores that they more Polish-dominant (but see below for a discussion of the cut-off point). We then investigated the validity of these indices by correlating the scores with four other variables, namely Length of Residence in the UK (LOR), Attitudes towards English and Life in the UK (EngAtt), frequency of usage of English at home (EngUse) and frequency of code-switching (CSTot). We found that the indices correlated significantly with most of these variables, but there were clear differences between the Guiraud-based indices (ILD1 and ILD3) and the HDD-based indices (ILD3 and ILD4). The HDD-based indices correlated more strongly with LOR and EngUSe than the Guiraud-based indices. The ILD4 correlated slightly more strongly with LOR and EngUse than the ILD2, but the former did not correlate significantly with either EngAtt. Neither of the HDD based indices correlated significantly with CSTot. The ILD1 correlated more strongly with EngAtt, whilst the ILD3 slightly less strongly with EngAtt, but significantly with LOR and EngUse. Only the ILD3 correlated significantly with CSTot.

To test the validity of our indices we also used a regression analysis with EngUse as the dependent variable and the different indices as predictors. We found that the ILD4 and the ILD2 explained a significant proportion of the variance in EngUse (R^2s were .445 and .440 respectively), while the ILD3 was marginally significant, with an R^2 of .138).

It was not possible to predict informants' overall code-switching behaviour (CStot) using any of the above-mentioned variables as predictors in a regression analysis, despite the existence of mid strength correlations between the ILD1 and CStot, but the ILD1 was a significant predictor of switches of verbs and all indices could predict the frequency with which informants used switches in the "other" category. The absence of a model which can predict overall code-switching behaviour is regrettable, but perhaps not unexpected because

informants' code-switching frequency depends on a range of situational factors which were beyond our control (Muysken, 2000; Treffers-Daller, 1992).

One might wonder, of course, whether the calculation of Indices of Language Dominance represents added value by comparison with the individual scores on the two measures of lexical diversity which were used to compute the ILD1 and the ILD2, namely Guiraud and HDD. It could in principle be the case, for example, that the English lexical diversity scores correlate strongly with LOR, EngUse, EngAtt and CSTot, and that computing the ILD is not needed because this does not add any new information. As it turns out, the English Guiraud values do not correlate significantly with LOR or EngAtt, whilst the ILD1 correlates significantly with EngAtt (.550**). However, the English Guiraud does correlate more strongly with EngUse (.732**) than the ILD1 (n.s.) or the ILD3 (.423*). In the case of the ILD1 computing the index constitutes added value by comparison with using English Guiraud values only for one out of three variables studied here.

For HDD, the situation is different. The English HDD values do correlate significantly with LOR (.426*) but the ILD2 (which is based on HDD) correlates more strongly with LOR (.704**). The English HDD value does not correlate significantly with EngAtt, but the ILD2 does correlate significantly with this variable (.442*). The English HDD values are only slightly superior to the ILD2 in their correlations with EngUse. The strength of the correlation coefficient between EngUse and the English HDD value is .761*, whilst for the ILD2 it is .644** and for the ILD4 it is .687**. As the ILD4 correlates more strongly with two out of the three variables studied here, we conclude that it does indeed represent added value by comparison with the individual HDD values for English.

A comparison of the simple indices (ILD1 and ILD2) with the complex ones (ILD3 and ILD4) which are based on the Edinburgh formula as proposed in Birdsong (this volume) shows that the latter correlate more strongly with LOR and EngUse than the former. The

differences between the results for the ILD1 and the ILD3 are particularly striking, but the correlations are also slightly stronger for the ILD4 than for the ILD2. It is only for EngAtt that the more complex formula performs slightly less well, as there are slightly lower correlations with the ILD3 than with the ILD1, and no correlations with the ILD4.

As explained in the introduction to this chapter, there are clear advantages of calculating language dominance in the way done here. These advantages include the fact that this method is available for any language, does not require the purchase of expensive tests or equipment, can be used for languages that are typologically very dissimilar in, for example, their morphology and is not only ecologically valid, but also generally enjoyable for participants. The fact that this operationalization clearly taps into language *use* also needs to be mentioned here because language dominance is not just a matter of proficiency, but also of use (Dunn & Fox Tree, 2009; Grosjean, this volume; Wang, 2013). There are, however, also drawbacks to this method, as there is no guarantee that the content of the stories on which the indices were calculated and the vocabulary used are entirely comparable. Data that are elicited under much more tightly controlled situations, such as those collected with the help of vocabulary tests which are administered under experimental conditions, will provide potentially more rigorous results, but constructing such tests is time-consuming and only possible if valid frequency data are available for each language. In addition, even for widely used parallel tests, such as the English and the Spanish versions of the Peabody Picture Vocabulary Task, the comparability of the tasks remains questionable (Prewitt Díaz, 1988). Another potential drawback of this method is that the issue of the different domains for which bilinguals use their different languages has not been taken into account in this approach, as Grosjean (this volume) argues is important.

We do not think that our measure of language dominance is necessarily better than other measures discussed in this chapter, and agree with Bedore et al. (2012) that the issue of

"the best measure" is not at stake. Language dominance is a multi-faceted construct and therefore we probably need a range of good indices rather than a single one. A possible way forward in research in this field is probably to try out a range of subjective and objective measures on different data sets, so that we can explore the interrelationships between different indices and their correlations with other variables in greater depth.

The fact that the Guiraud-based indices did not correlate that strongly with the HDD-based indices, and that the two groups of indices showed a different pattern of correlations with variables such as LOR, EngUse, EngAtt and CStot may indicate that the indices tap into slightly different dimensions of lexical diversity. This issue cannot be pursued here but is worth exploring in future research.

Another issue that will need to be investigated further concerns the issue of an appropriate cut-off point to distinguish between balanced and dominant bilinguals. As Birdsong (this volume) explains, a good way forward may be *not* to impose a categorization on the data but to use dominance as a continuous independent variable in a regression analysis, where it can be used to predict variance in the researcher's chosen dependent variable. This is also the approach taken by Flege et al. (2002), who posit that the only way to validate a measure of bilingual dominance is to demonstrate it can predict variance in another variable. In the current study, we used regression to predict variance in reported English usage among our Polish-English informants and found that this approach worked well for several of the indices (see above for fuller discussion).

Some researchers do, however, feel the need to divide their sample into different groups of balanced and dominant bilinguals, based on a measurement of language dominance. In the strictest interpretation of the term "balanced", only those participants who obtain exactly the same score in each language would be balanced, but this outcome is highly unlikely. In addition, researchers who chose this operationalization do not take into account

any measurement error in their calculations. Therefore, a more appropriate cut-off point needs to be found. Admittedly, our choice to use one standard deviation from the mean (set at zero when we computed Z-scores) as the cut-off point led to a rather generous interpretation of the notion "balanced bilinguals" because in our sample 64 percent were classified as balanced based on the ILD1 and 73 percent were balanced according to the ILD2. These percentages are quite high by comparison with, for example, Gollan et al.'s (2012) study where only 20 percent of the participants were classified as balanced. We agree with Gollan et al. that the choice of the cut-off point is to a certain extent arbitrary, although the choice will probably depend on the aims of the study: for some studies it may be necessary to adopt a more restrictive cut-off point and for others a more generous one is allowable. In this chapter we have suggested that converting raw scores to Z-scores will make it easier to compare cut-off points among studies, and to have a fruitful discussion about the classification differences which result from the range of cut-off points. Trying to achieve a consensus on this issue is important if we want to establish whether being a balanced bilingual is indeed as rare as is generally assumed. However, different cut-off points may be appropriate depending on the aim of the study.

It is important to further test the Edinburgh formula on other data sets, as the results from the current studies show using this formula often produces better results in that indices calculated on the basis of this formula correlate more strongly with other variables or explain more variance in other variables when entered as a predictor variable in a regression analysis.

Another issue that needs to receive more attention is whether measurements of language dominance are applicable to multilinguals. How dominance can be measured in multilinguals has hardly been explored so far. Clearly, the issue of language dominance becomes more complicated if participants speak three, four or five languages, and being balanced in so many languages seems hardly possible. Using the method we have described

in this chapter with multilinguals is certainly possible, and it would no doubt be very interesting for the research community to explore this further.

References

Ameel, E., Malt, B., Storms, G. &Van Assche, F. (2009). Semantic convergence in the bilingual lexicon. *Journal of Memory and Language, 60,* 270–290.

Bahrick, H.P., Hall, L.K., Goggin, J.P., Bahrick, L.E. & Berger, S.A. (1994). Fifty years of language maintenance and language dominance in bilingual Hispanic immigrants. *Journal of Experimental Psychology: General, 123*, 264-283.

Bates, E. & Goodman, J. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. In G. Altmann (Ed.), Special issue on the lexicon, *Language and Cognitive Processes, 12*, 507–586.

Bedore, L.M., Peña, E.D., Summers, C.L., Boerger, K.M., Resendiz, M.D., Greene, K., Bohman, T.M. & Gillam, R.B. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism: Language and Cognition*, *15*, 616-629.

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

Bialystok, E., Craik, F. I. M. & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics, 21*, 522–538.

Carroll, R. & Luna, D. (2011). The other meaning of fluency: Content accessibility and language in advertising to bilinguals. *Journal of Advertising*, *40*, 73-84.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Cromdal, J. (1999). Childhood bilingualism and metalinguistic skills: Analysis and control in young Swedish–English bilinguals. *Applied Psycholinguistics, 20,* 1–20.

Daller, H. (1999). *Migration und Mehrsprachigkeit. Der Sprachstand türkischer Rückkehrer aus Deutschland*. Frankfurt am Main, etc.: Peter Lang.

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

Daller, M. H., Yıldız, C. & de Jong, N. H. (2011). Language dominance in Turkish–German bilinguals: Methodological aspects of measurements in structurally different languages. *International Journal of Bilingualism, 15*, 215–236.

Dewaele, J.-M. & Pavlenko, A. (2003). Productivity and lexical diversity in native and non-native speech. A study of cross-cultural effects. In Cook, V. (Ed.) *Effects of the second language on the first* (pp. 120-141). Clevedon: Multilingual Matters.

Dickinson, O. K. & Tabors, P. O. (Eds.) (2001). *Beginning Literacy with Language: Young Children Learning at Home and School*. Baltimore: Paul H. Brookes.

Dugast, D. (1980). *La statistique lexicale*. Genève: Slatkine.

Dunn, A. & Fox Tree, J. (2009). A quick and gradient bilingual dominance scale. *Bilingualism: Language and Cognition, 12*, 273-289.

Dunn, L.M. & Dunn, D.M. (1965; 2007). *Peabody Picture Vocabulary Test*. Pearson.

Dunn, L.M., Padilla, E.R., Lugo, D.E., & Dunn, L.M. (1986). *Test de vocabulario en imágenes Peabody (TVIP)* [*Peabody picture vocabulary test*]. Circle Pines, MN: American Guidance Service.

Edmondson, W. & House, J. (1991). Do learners talk too much? In R. Phillipson, E. Kellerman, R. Phillipson, L. Selinker, M. Sharwood Smith, & M. Swain (Eds.), *Foreign/second language pedagogy research* (pp. 273–86). Clevedon, UK: Multilingual Matters.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Los Angeles, etc.: Sage.

Fishman, J.A. & Cooper, R.L. (1969). Alternative Measures of Bilingualism. *Journal of Verbal Learning and Verbal Behavior, 8*, 276-282.

Flege, J.E., Mackay, I.R.A. & Piske, Th. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, *23*, 567–598.

Franks, S. (1995). *Parameters of Slavic morphosyntax*. New York and Oxford: Oxford University Press

Gollan, T.H., Weissberger, G., Runnqvist, E., Montoya, R.I. & Cera, C.M. (2012). Self-ratings of spoken language dominance: A multi-lingual naming test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism: Language and Cognition, 15,* 594-615.

Grosjean, F. (2008). *Studying Bilinguals*. Oxford: Oxford University Press.

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

Hulstijn, J.H. (2012). The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition*, *15,* 422-433.

Jarvis, S. (2002). Short texts, best fitting curves, and new measures of lexical diversity. *Language Testing, 19,* 57-84.

Korybski, T. (2013). *Polish – English code-switching and other language contact phenomena in the speech of Polish migrants in the United Kingdom*, PhD dissertation, University of the West of England, Bristol.

Lambert, W.E. J. Havelka, J. & Gardner, R.C. (1959). Linguistic Manifestations of Bilingualism. *The American Journal of Psychology*, *72*, 77-82.

MacWhinney, B. (2000). *The CHILDES Project. Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates. The most recent version is available on this website: URL (accessed 23 November 2010): http://childes.psy.cmu.edu/.

Malvern, D., Richards, B., Chipere, N. & Durán, P. (2004). *Lexical richness 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.

McCarthy, P. M. & Jarvis, S. (2007). A theoretical and empirical evaluation of vocd. *Language Testing*, *24*, 459–488.

McCarthy, P. M. & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*, 381-392.

McCarthy, P.M., Watanabe, S. & Lamkin, T.A. (2012). The Gramulator: A tool to identify differential linguistic features of correlative text types. In P.M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing: Identification, investigation, and resolution* (pp. 312-333). Hershey, PA: IGI Global.

Meara, P. & Milton, J. (2003). *X_Lex, the Swansea Levels Test*. Newbury: Express.

Milroy, L. (1980). *Language and social networks*. Oxford: Blackwell.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.

Prewitt Díaz, J. O. (1988). Assessment of Puerto Rican children in bilingual education programs in the United States: A critique of Lloyd M. Dunn's monograph. *Hispanic Journal of Behavioral Sciences, 10*, 237–252.

Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *International Journal of English Studies*, *7*, 105-125.

Reyes, I. & Azuara, P. (2008). Emergent biliteracy in young Mexican immigrant children. *Reading Research Quarterly, 43*, 374–398.

Silva-Corvalán, C. (2014). *Bilingual language acquisition. Spanish and English in the first six years*. Cambridge: Cambridge University Press.

Templin, M. **(**1957). *Certain language skills in children*. Minneapolis, MN: University of Minnesota Press.

Treffers-Daller, J. (1992). French-Dutch codeswitching in Brussels: social factors explaining its disappearance, *Journal of Multilingual and Multicultural Development, 13(1&2)*, 143-156.

Treffers-Daller, J. (2011). Operationalizing and measuring language dominance. *International Journal of Bilingualism, 15*, 147-163.

Treffers-Daller, J. (2013). Measuring lexical diversity among L2 learners of French: an exploration of the validity of D, MTLD and HD-D as measures of language ability. In Jarvis, S., & Daller, M. (Eds). *Vocabulary knowledge: Human ratings and automated measures* (pp. 79-103). Amsterdam: John Benjamins.

Tweedie, F.J. & Baayen, R.H. (1998). How variable may a constant be? Measures in lexical richness in perspective. *Computers and the Humanities, 32*, 323-352.

Wang, X. (2013). Language dominance in translation priming: Evidence from balanced and unbalanced Chinese–English bilinguals. *The Quarterly Journal of Experimental Psychology, 66*, 727-743.

Tables

Table 1 Mean length of the stories in English and Polish (lemmatized types and tokens)

	minimum	maximum	mean	SD
Polish types (lemmatized)	91	272	140.19	46.00
Polish tokens	164	712	306.23	146.00
English types (lemmatized)	38	267	144.89	49.46
English tokens	134	972	440.39	186.09

Table 2. Correlations between different Indices of language dominance

	ILD1 (Guiraud)	ILD2 (HDD)	ILD3 (Guiraud, Edinburgh)	ILD4 (HDD, Edinburgh)
ILD1 (Guiraud)	1	.557**	.969**	.563**
		.007	.000	.006
ILD2 (HDD)		1	.636**	.991**
			.001	.000
ILD3 (Guiraud, Edinburgh)			1	.662**
				.001

Table 3. Correlations between Indices of Language Dominance and other variables

	LOR	EngAtt	EngUse	CStot
ILD1	n.s.	.550**	ns	.454*
ILD2	.704**	.443**	.644**	n.s.
ILD3	.549**	.526**	.423*	n.s.
ILD4	.712**	n.s.	.687**	n.s.
LOR		.789**	.516**	n.s.
EngAtt			n.s.	n.s.
EngUse				n.s.

Figures

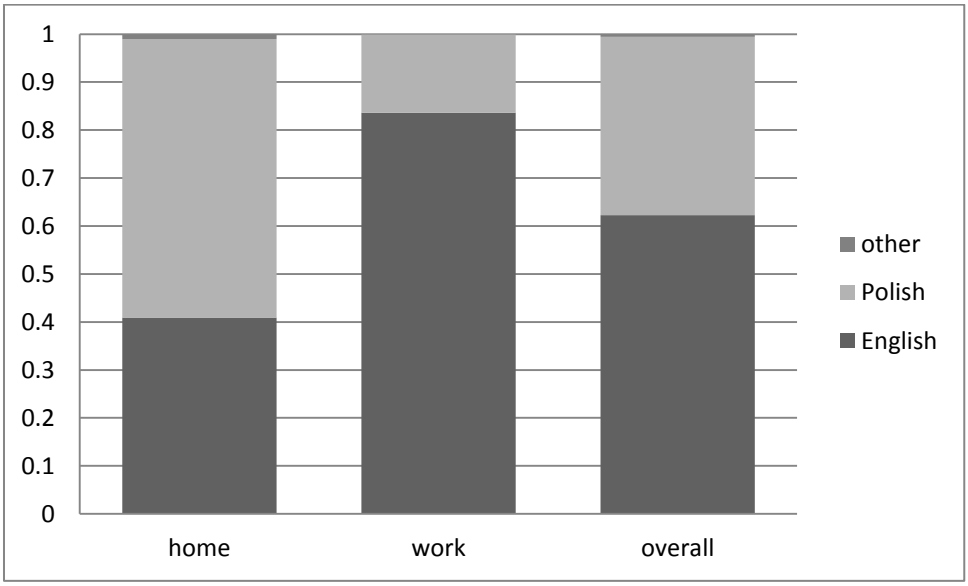

Figure 1. Frequency of usage of English, Polish and other languages at home and at work (self-report, n = 26).

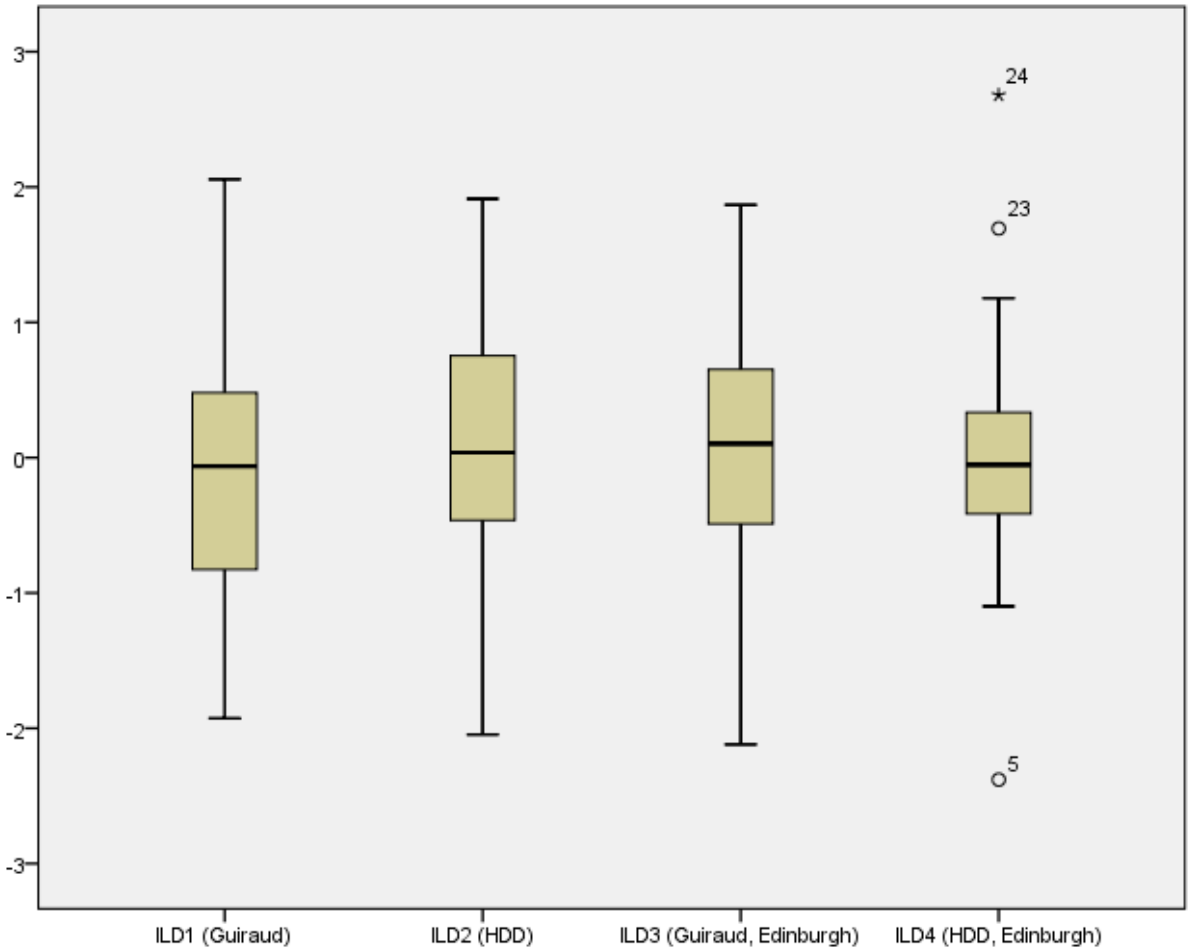

Figure 2 Indices of Language Dominance (z-scores), N = 22

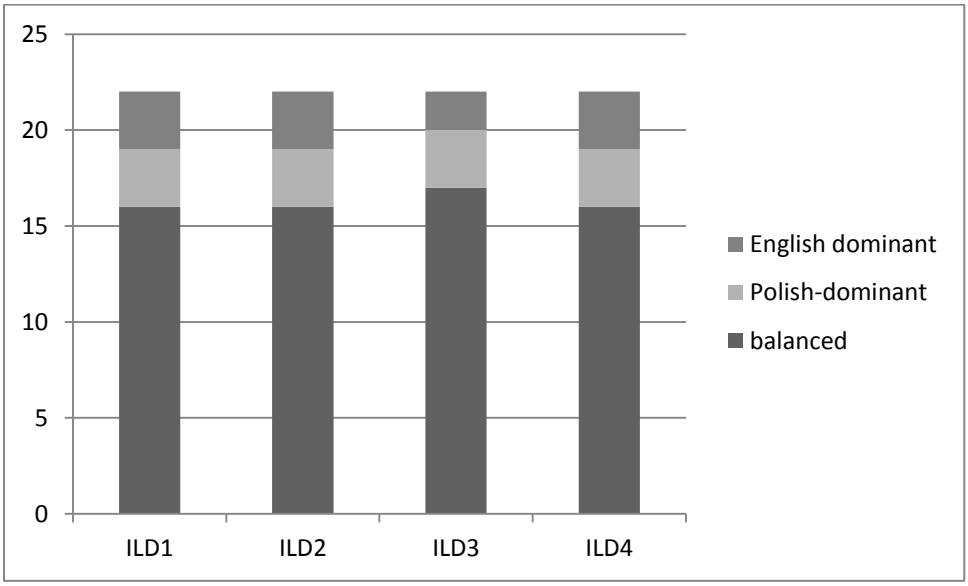

Figure 3. Number of balanced bilinguals, English-dominant and Polish-dominant bilinguals according to the four Indices of Language Dominance (n=22).

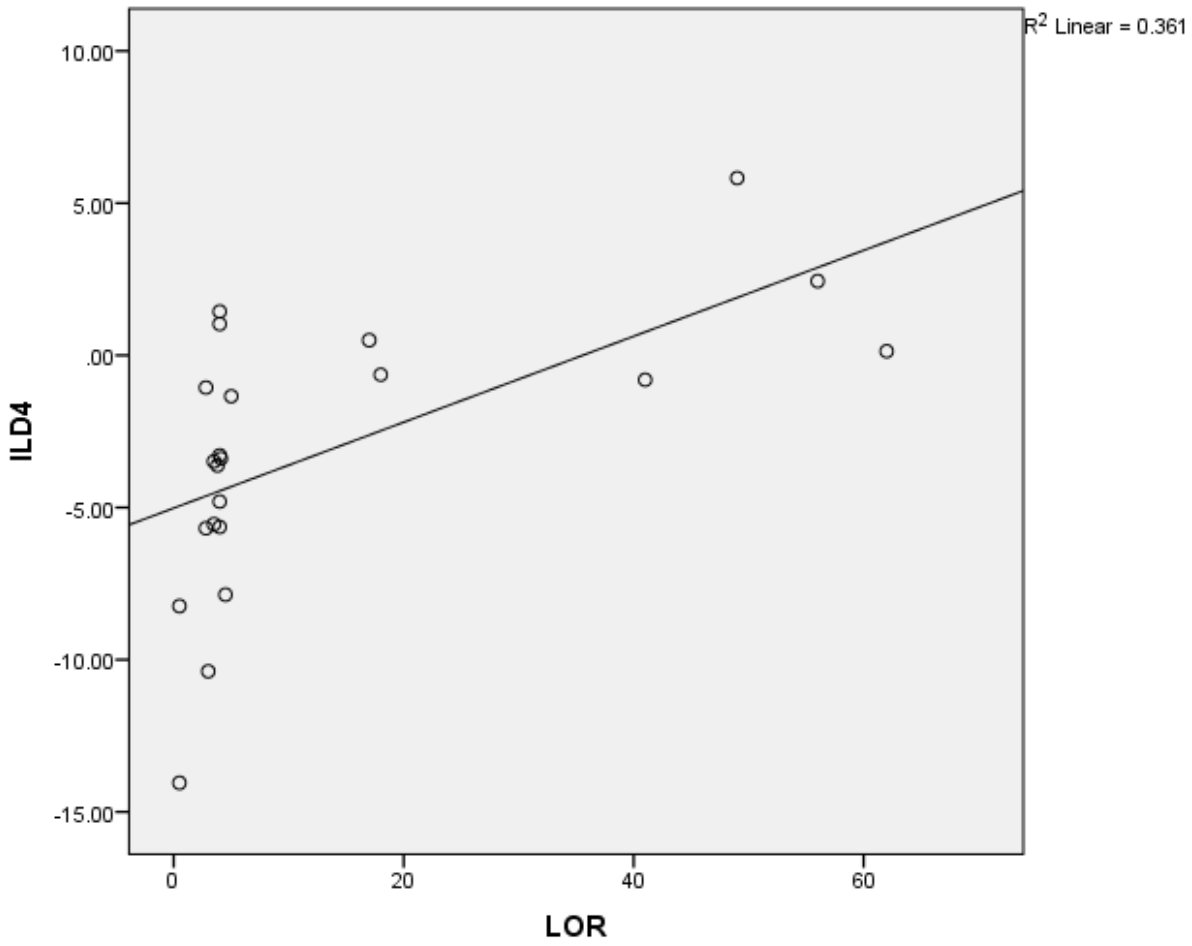

Figure 4. Correlations between length of residence and the ILD4

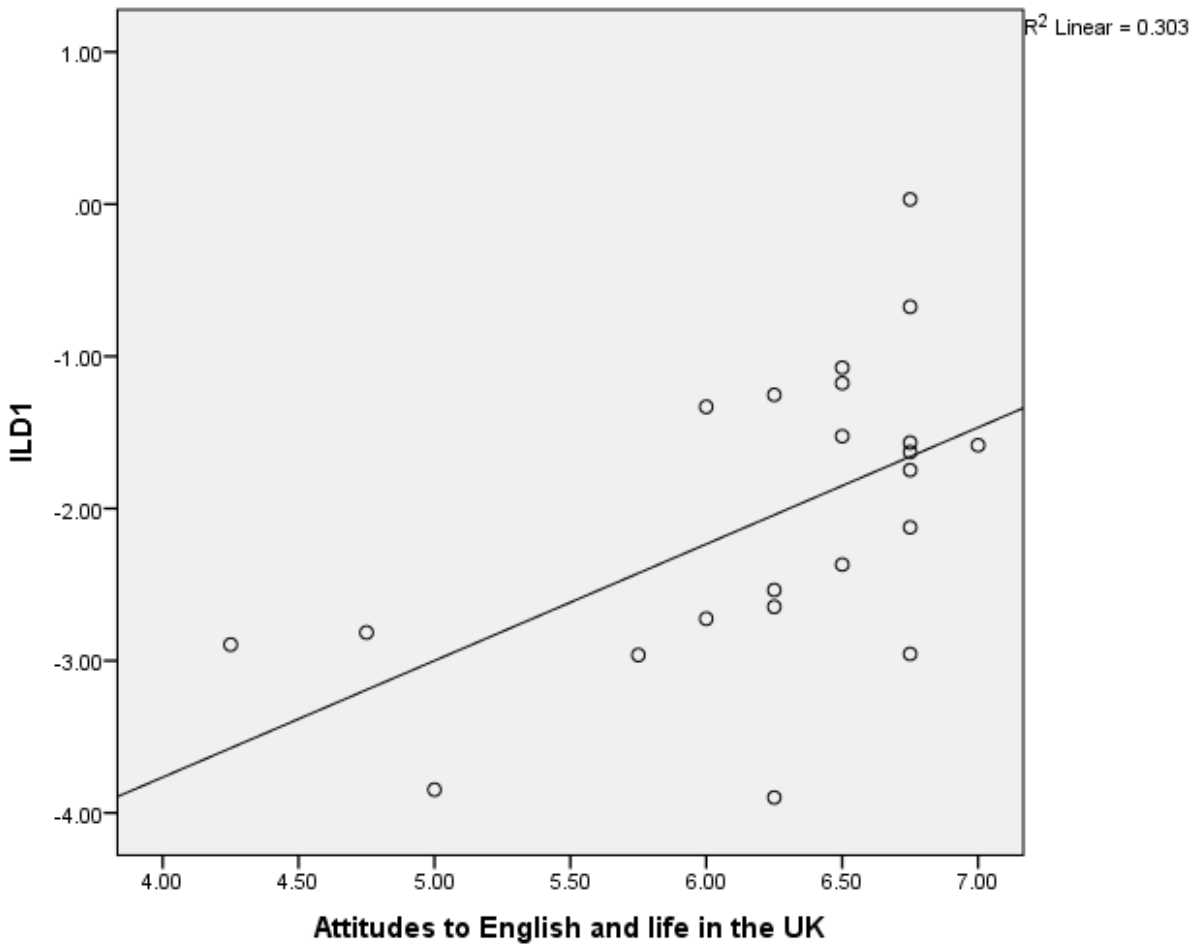

Figure 5. Correlations between ILD1 and EngAtt