

Chapter 2

Multi-factorial Studies: Populations and Linguistic Features

Monika S. Schmid

Abstract It was suggested in the introduction to this volume that cumulative evidence from studies investigating populations of learners from varying linguistic backgrounds, under different learning contexts, and with a range of experimental designs is necessary in order to gain further insight into fundamental questions of bilingual development—such as whether first and second language learning are qualitatively different from each other, or whether cognitive maturation specifically and independently affects language learning. Such a meta-approach, however, is easily compromised, as what may seem to be relatively minor differences and adjustments in participant selection, data acquisition, coding and analysis may eventually make it impossible to compare the findings from one study to that of another, or lead to conflicting findings. This chapter discusses the issue of what factors characterizing the populations being tested should and can be documented. We start with a discussion of how the lack of such documentation and differences in analysis have muddled the waters in the past.

Keywords Sociolinguistic and personal background factors • Second language acquisition • First language attrition

2.1 Cross-Study Variability and the Importance of Participant Documentation

“Even a supposedly clean study is prone to alternative interpretations of data” (Bialystok and Hakuta 1994, p. 69). This statement refers to the seminal study by Johnson and Newport (1989, henceforth J&N) which was long accepted as providing conclusive evidence for the existence of a maturational constraint on

Electronic supplementary material The online version of this article (doi:[10.1007/978-3-319-11529-0_2](https://doi.org/10.1007/978-3-319-11529-0_2)) contains supplementary material, which is available to authorized users.

second-language learning (the so-called ‘Critical Period’). J&N investigated a population of 46 Chinese and Korean learners of English with varying ages of acquisition (AoAs) by means of an extensive offline auditory grammaticality judgment task comprising 276 items. They found that the scores achieved on this task correlated strongly with AoA for the segment of their population who had acquired English before age 15 but not for those who had learned it after that age. This discontinuity was interpreted as evidence for a qualitative change of the learning mechanisms linked to biological development: “[L]anguage learning ability slowly declines as the human matures and plateaus at a low level after puberty” (p. 90).

Later investigations, however, are divided on whether or not to accept J&N’s conclusions. On the one hand, a re-analysis of the original data by Bialystok and Hakuta (1994) found that when the cutoff point between the younger and the older population was moved from 15 to 20, age of learning remained a significant predictor in both groups and the apparent discontinuity found in the original study disappeared. Bialystok and Hakuta further point out that the age-related decline observed across the sample is not necessarily the result of the deterioration of a *language-specific* acquisition ability, but may be due to other factors such as age at testing, different language learning circumstances in the different AoA segments and so on (p. 70f.). Similarly, Elman et al. (1996) propose different statistical ways of analyzing J&N’s data and suggest that ‘a simple linear dynamical alternative is available to account for the same nonlinear outcomes’ (p. 188).

An additional issue is that the population involved is inadequately described. It seems clear that other factors than AoA must be involved given the different conclusions that replications of the study have found: Birdsong and Molis (2001) find the exact opposite pattern of AoA effects among 61 native Spanish learners of English (no correlation between age of onset and score on the GJT for learners who were below 15 when they began the acquisition process but a strong correlation for learners beyond that age). Conversely, DeKeyser (2000), investigating 57 Hungarian learners of English, and DeKeyser et al. (2010), investigating 76 Russian learners of English and 64 Russian learners of Hebrew, found discontinuities in all populations which were similar to those discovered by J&N. Without detailed characterizations of the populations it is difficult to reconcile these results.

Given the large range of factors that impact on L2 development, it is not surprising that findings from different studies are often inconsistent or controversial. Inconsistencies such as the ones sketched above, however, severely limit our ability to come to hard conclusions about the language learning process and what constrains it. Lamentably, there is no overall consensus on how populations for linguistic investigations should be selected, and what factors it is vital to document for comparison purposes. This chapter will make an attempt to identify a number of these factors and suggest ways in which they may be assessed in order to ensure that populations and, eventually, studies, can be characterized more comprehensively. We hope that this may allow explanations of contradictory findings in terms of relevant group differences on the basis of cross-study comparisons.

The list of factors necessarily comprises biodata such as age, AoA, gender, education, learning context, SES and so forth, which are comparatively easy to elicit and record. Second, there are variable individual traits, potentials and limitations which, while not in themselves language-specific, may affect or impair performance on learning and on particular linguistic tasks and thus serve to distort the picture. This sort of variable comprises matters such as handedness; any potential impairments to sensory modalities: vision (including color blindness) or hearing; and cognitive factors like Working Memory (WM) and/or intelligence. In this context, we need to acknowledge the difficulties posed in measuring and comparing some of these factors across populations with different first languages. Lastly, the mere fact of knowing two languages will have an effect on knowledge and use of both languages, producing differences between monolinguals and bilinguals.

In some cases, information about the speakers should be collected beforehand in order to determine whether there are reasons to exclude them from participating in the study. For example, many executive control tasks such as the Stroop task (Stroop 1935), Go-No-Go tasks or language switching tasks use different colors as cues to which response should be given. Participants who have any kind of impairment to color perception will almost certainly provide different responses than participants with unimpaired perception. For tasks which rely on fast response times, participants above a certain age may be too slow, adding undue noise to the data being analyzed, and the researcher may therefore decide to establish an upper age limit for the population or exclude on an individual basis determined by the population distribution. How the criteria discussed here will be applied in any individual investigation will depend on the research design and research questions, so the procedures we outline here serve merely as suggestions.

Most of the factors discussed below, however, do not determine whether participants are included, but must be documented in order to be used in characterizing the populations. This relates to factors which may facilitate or constrain performance on certain tasks, such as level of education or language proficiency. For many of the more traditional statistical approaches, it is desirable that the different populations are as similar as possible with respect to these factors. Some more novel approaches, such as mixed effects and multiple regression models are better equipped for dealing with variability in the range of predictors and additionally do not force the use of artificial groupings within linear factors like AoA, but similar age ranges remain important for group comparisons.

In order to obtain the desired range of participants, balance criteria such as education level and gender, and apply exclusion criteria, it is a good idea to elicit or at least estimate some of these factors from potential participants before they are actually recruited for the study. This is comparatively straightforward for biographical criteria, such as age of learning, age at testing, length of exposure to the target language, and so on, which can relatively easily be collected by an online participant questionnaire. For other factors, such as handedness, the assessment is somewhat more complex, whereas measuring predictors such as language proficiency is invariably rather time-consuming. It is important that the pre-assessment

should strike the right balance here: If it becomes too lengthy, participants will be put off before they ever attend a testing session; if, on the other hand, it is too short and superficial, participants who do not fit the testing criteria (for example because their proficiency level is too low) will be included but their data may not be usable, and this may be expensive in terms of both laboratory costs and working time—not to mention unfair towards the participants.

2.2 Exclusion Criteria

Since we are usually keen to have populations as large as possible, excluding participants who have volunteered to take part in the study may not seem desirable. However, a number of participation criteria need to be established since otherwise results may be skewed, flawed or obscured. These exclusion criteria fall broadly into two categories: Factors that are connected with experience with additional languages not being tested, and physical and health factors that may affect the performance of the participant.

First, it is important to establish each individual's **language learning history**. For investigations that wish to study a particular process of bilingual acquisition it is often important that none of the participants has had substantial experience with languages other than those under investigation, in particular 1) outside the range of age of acquisition for which they are being tested and/or 2) containing a linguistic characteristic similar to the one under investigation (e.g. tone or gender), as this may facilitate its acquisition in the third language. This normally excludes speakers who grew up in bilingual families, have spent extended periods of time in third countries (these may include countries or regions where different varieties of the target language are spoken, for example Austria or Switzerland vs. Germany or Francophone regions in Canada or Africa vs. France) or otherwise achieved advanced proficiency in another language. We therefore suggest that in the course of the pre-screening, a short language learning questionnaire should be administered (see online supplementary material), and that it should be carefully considered to what extent any prior linguistic experiences may obscure or influence performance. Note that similar criteria should be applied to the control group, as well. One frequently finds investigations of, for example, native speakers of American English acquiring Spanish, where the reference population is comprised of speakers born in Latin American countries and residing in the US. While this is certainly a convenient and inexpensive way of recruiting a control group, such speakers may have experienced language dominance reversal and/or L1 attrition, particularly if they emigrated to the US early and were educated there, which may well distort the findings (Dussias 2004).

Second, it is advisable to establish whether participants are affected by a **medical condition** or **disorder** which might impact their performance. These include conditions that will impair the perception of or response to the stimuli, such as visual or hearing impairments (unless corrected), dyslexia, speech disorders (such as

stuttering) and color blindness. A number of neurological conditions (such as epilepsy, strokes or brain lesions) or neurodegenerative disorders may also result in linguistic difficulties, and the use of various medications, drugs or alcohol may interfere with task performance. It is wise not to include participants who are affected by any of these.

Lastly, **handedness** can be an important factor. For studies which measure response time by means of pressing a particular key on a computer keyboard or on a Serial Response Box, left-handed participants should be tested by means of a setup that is the reverse of that of right-handed ones. Although handedness does not strongly correlate with language lateralization, it does provide an indication and can muddy the waters when brain measures are used, since in left-handers there is a 20 % chance that the functional organization of the hemisphere is inverted (McKeever et al. 1995). For this reason, it is recommended that the population should comprise only right-handed individuals in studies employing neurological and neurolinguistic measures, such as EEG or MRI readings. Handedness, however, is not a dichotomous concept which neatly divides the population into left- and right-handed individuals, but rather a gradient phenomenon with more or less strongly pronounced levels. It has been established that the chance of right hemisphere dominance increases as the degree of left handedness increases (Knecht et al. 2000). For the purpose of our investigation we relied on an abbreviated version of the Edinburgh Handedness Inventory (Oldfield 1971; the version used in our study was part of the prescreening questionnaire which can be accessed in the online supplementary material, and excluded any participants who did not clearly emerge as right-handed.

2.3 Personal Variables

Many individual factors may have an impact on proficiency in both L1 and L2. Among these are biographical factors which are the outcome of lifetime events and impact indirectly on language development and processing. Others concern the cognitive skills that influence an individual's ability to learn and process language. Factors which depend on personal habits and/or attitudes are often directly linked to language proficiency. Not all studies will need to include all of these factors to the same level of detail, but in designing the study it is important to consider which of them should be assessed so that their impact on various populations can be examined. Even when a study is not directly concerned with some of these factors, it is recommended to include them to facilitate cross-study comparisons in the future. Furthermore, some factors may be used to establish the exclusion criteria discussed above.

2.3.1 *Biographical Data*

The factors treated in this section are characteristics of all participants, whether mono- or bilingual, which may have an impact on their performance on various tasks and should therefore be sampled similarly in the different segments of the population. First, and probably most straightforward, the **sex** of participants should be matched across populations. The debate on the impact of sex for language acquisition is too wide-ranging to be treated in detail here (as is the relatively recent debate on whether biological sex is a dichotomous factor; see for example Hall 2008). However, there is ample evidence to suggest that, both in instructed and in immersed language learning, females may have a degree of advantage over males (e.g. Pavlenko et al. 2001). This debate goes back to the age-old nature versus nurture controversy which asks to what extent such differences are due merely to socially constructed realities and expectations, or whether there is some kind of impact of differential (probably hormonally modulated) use of neurocognitive resources by the two genders that may make women more effective language learners (see the discussion by Ullman 2004). Given, however, that many studies of (first- and second-) language processing and production have found differences between male and female populations, it is wise to aim for equal representation of both genders. Should this not be possible, proportions of women and men should at least be approximately the same in all populations.

Second, the **age** of participants¹ at the time of testing is of importance. Performance on a variety of tasks fluctuates and changes throughout the lifespan, so that care should be taken to ensure that all populations are approximately age-matched. Furthermore, many recent studies suggest that age-related declines in cognitive skill may to some extent be attenuated in bilingual populations (e.g. Abutalebi et al. 2014; Bialystok et al. 2008; Gold et al. 2013). This means that if the populations comprise large numbers of individuals at higher age ranges, differences that are due to bilingual proficiency and the end state of long-term adult L2 learning (frequently referred to as ultimate attainment) may be confounded with differences contingent on the protective influence of long-term bilingualism on cognitive function. It is therefore strongly advised that a) not only the mean age but also the age range across populations be comparable, and b) a conservative upper age limit of 60 years be implemented, in particular for studies that use psycho- or neuro-linguistic tasks.

¹Note that for the purpose of the present text the discussion is limited to healthy adult volunteers. Age at testing may become a much more complex issue for studies wishing to include child or adolescent participants and compare their linguistic performance to that of adults, since it is often a great challenge to develop tasks that are suitable across age ranges from childhood to mature speakers. Furthermore, ethical approval is often more difficult to obtain for studies that seek to investigate younger participants; and the same is true for investigations of pathological conditions, such as aphasia. These restrictions and procedures should be kept in mind and addressed at an early stage of the research design in studies intending to target populations other than healthy adults.

Third, it is important to take into account the **education levels** of participants. Since the most frequently studied population in investigations of bilingualism consists of college or university students, this factor is often considered to be unproblematic. Investigations focusing on ultimate attainment in late learners, however, often do not have this option, as such studies rely on long-term immersed bilinguals whose exposure started after puberty, and student populations do not have a sufficient length of exposure. Measuring educational levels in immigrant populations can often be problematic, as Schmid (2011) has demonstrated: educational systems in different parts of the world may not be directly comparable. Furthermore, migrant biographies are often far less straightforward than those of individuals who have lived their entire life in the same country, since migration can be a disruptive event for a variety of reasons. That notwithstanding, education may affect performance on a wide range of tasks, in particular those with a written component or those that resemble test taking to any degree, making it important that it is carefully assessed and, as far as possible, comparable across populations.

2.3.2 Intelligence, Working Memory and Other Cognitive Factors

In addition to biographical factors such as aging or education, there are also long-term differences in cognitive capacities, including intelligence and working memory (WM), which may affect (aspects of) language learning and performance. The underlying assumption in much research on bilingual populations is that experimental designs should probe participants' ability to correctly identify or use particular linguistic structures. The task should ideally measure the mastery of this structure, and nothing else. However, a number of confounding factors can have an impact on the performance on any given task, facilitating the performance of some individuals while underestimating the actual level of others, irrespective of their underlying proficiency. As Bialystok and Hakuta (1994) point out, for example, the grammaticality judgment task used by Johnson and Newport (1989) was an extremely long one, comprising 276 items, and this might have disadvantaged the older participants who might not have been able to sustain full concentration across the entire test, while younger participants presumably had less trouble with fatigue (p. 70).²

There is ample evidence showing that individual capacities can enhance or constrain performance on many of the tasks frequently employed in linguistic research, such as acceptability judgments. For example, differences between high-

²Another potential age effect unrelated to the Critical Period is connected with education: the fact that most of the younger participants in the Johnson & Newport study were college students at the time of testing. Presumably, this population has recent experience with extended testing situations from which they would benefit.

and low-WM span individuals have been found with respect to interpretation (Miyake et al. 1995) and processing (Friederici et al. 1998) of complex sentences in the L1. Grammaticality judgments and comprehension tasks can also be affected by loading working memory with concurrent memory tasks (Blackwell and Bates 1995).

Despite the desirability of controlling for background variables associated with intelligence, working memory and other cognitive factors, the tests used to assess them present a number of challenges, first among them the considerable debate over what exactly they consist of, how stable they are among individuals, and particularly how to measure them even for monolingual populations. The challenges become even more formidable when it comes to comparing monolinguals and bilinguals. It is often debated which indicator of cognitive ability/potential is the best predictor of the level of performance on linguistic tasks. The question of which measure should be chosen is not an easy one for any linguistic investigation, since WM capacity and intelligence are not only strongly correlated (Conway et al. 2003) but it is also very controversial what subcomponents make up WM. Some studies find no evidence for such a specific WM capacity (e.g. Vos et al. 2001). Additionally WM and other cognitive measures seem to vary over time (the test-retest validity of the reading span test used by Caplan and Waters 1999 is as low as 0.41), meaning that their predictive power is concomitantly lower.

These controversies notwithstanding, given the impact of individual capacities on language processing in monolinguals, it is evidently important that this factor be measured and taken into account in linguistic experiments. However, where bilingual populations are concerned, this is anything but straightforward. Measures of intelligence have long been recognized to be problematic for bilingual populations. On the one hand, administering an IQ task in a bilingual's weaker language will inevitably lead to a depressed score. This fact resulted in the assumption, held for many decades of the 20th century, that bilingualism is detrimental to intelligence (for an overview of the debate see Pavlenko 2011). This problem cannot trivially be solved by providing bilinguals with a version of the test in their stronger language for two reasons: First, questions on verbal IQ tasks often have a cultural bias. For example, Baker (1988) points out that the question 'Who discovered America?', which forms part of the Wechsler Intelligence Scale for Children, has different acceptable answers for English and American children on the one hand and Welsh children (who are taught that it was discovered in the 12th century by Madoc, Prince of Gwynedd) on the other, while neither of these answers might be acceptable to Native American children (p. 12). And, of course, there are many parts of the world where Christopher Columbus is not the household name that it is for most Western cultures. It is virtually impossible to ensure that IQ tests across languages and cultures are identical in their difficulty and validity (Baker 1988: 12). The second problem with testing bilingual participants' IQ in their strongest language is that the question 'What is your strongest language?' may vary across domains. A speaker may, for example, have higher verbal skills in one language but prefer to do math in the other. All in all, these considerations suggest that IQ tasks and any other task that is administered verbally may not be suitable to establish a

valid range of individual cognitive capacity across monolingual and bilingual participants.

The same considerations are true for measurements of verbal WM, since bilingual individuals vary strongly in their performance on WM tasks administered in the L1 versus the L2 (McDonald 2006; Gass and Lee 2011). WM capacity is usually higher in the L1, but has been shown to positively correlate in the L2 with performance on sections of the TOEFL and L2 reading comprehension abilities (for a review, see Miyake and Friedman 1998), processing of gender and number agreement in the L2 (Sagarra and Herschensohn 2010), the ability to make use of interactional feedback in L2 classroom settings (Mackey et al. 2002) and general L2 proficiency (van den Noort et al. 2006). These findings point to some kind of trade-off effect: As processing in the L2 becomes more automatized with higher levels of proficiency, WM capacity increases, while the more controlled processing necessary for less proficient learners results in less storage space, in line with Baddeley's (2003) model (see Gass and Lee 2011). Where WM is measured with respect to the verbal component, proficiency and WM capacity may therefore be confounded for the L2 populations.

Given all of these considerations, it would seem that there is no ideal way of measuring individual cognitive capacities, even though these may constrain performance on linguistic tasks in studies which aim to compare monolingual and bilingual populations. Although the comments above rule out intelligence tests in general and many of the commonly used WM tasks of verbal recall, such as the serial-recall task (Daneman and Carpenter 1980) or the reading-span task (Conway et al. 2005), there is nevertheless one possible WM option, the *n*-back digit span task (Kirchner 1958). It should be admitted that this is probably not an optimal measure of WM, but that in the context of bilingualism research it very likely is the least-worst one.

It was for this reason that the *n*-back task was included in the project on which the present volume is based. In our version of the task, the participant was presented with a sequence of digits. Each digit was displayed on the computer screen for 500 ms, followed by a white screen for a period of maximally 1500 ms. During this interval, the participant had to indicate by means of pressing a button whether the current stimulus was a match for the one presented *n* steps earlier in the sequence. If this was the case, the target answer was yes (pressing a green button), while a mismatch was indicated by pressing a red button. The load factor *n* was adjusted to make the task more difficult: Participants completed two blocks of 2-back trials (a total of 104 trials) and two blocks of 3-back trials (162 trials). The script for this task is available in the online supplementary material.

2.3.3 *Attitude and Use*

The individual cognitive factors discussed in the previous section were acknowledged to be extremely problematic for bilingualism research. Not only do these

factors vary over time, they are also difficult to measure in a way that is valid and reliable for both monolingual and bilingual populations. These difficulties notwithstanding, in this area there exist established, widely used and standardized tasks which, at least in theory, allow them to be measured. The same cannot be said for another set of factors which impact strongly on the language development and proficiency of bilinguals, namely those predictors that are connected to patterns of language use and emotions.

The literature on the role of attitudes towards L1 and L2, as well as the frequency of language use in various domains, for the development of proficiency is vast (for an overview see Schmid 2011: Chaps. 7 and 8). One of the largest problems for linguistic research is that, unlike the factors mentioned above, these predictors cannot be independently measured. Researchers have to rely on self-assessments and self-reports, which are notoriously unreliable and often affected by how participants wish to be perceived, not by how they actually feel and behave. For example, a speaker who feels (for whatever reason) that her L2 proficiency should be better than it actually is might downplay the frequency of use of that language, and a participant who wants to be polite might not be entirely honest in reporting her attitudes towards the language and speech community with which the researcher is affiliated.

Furthermore, such reports can only provide a snapshot of attitudes and habits at a specific point in time and the reliability of reports of past experiences may decrease even further. Attitudes and the relative dominance of use of the various languages are extremely fluid and changeable over time. Instructed learners, for example, often develop their attitudes based on how much they like or dislike their teachers, so that a change in the instructor may lead to a reversal of the emotional stance. Migrants may arrive in their new country filled with enthusiasm, but become disenchanted through negative experiences. Use is possibly even more variable; language habits will change as people enter a speaker's social network, or disappear from the circle of friends, causing continual fluctuations in the actual amount of use of various languages. Technological and infrastructural changes over the past decades (internet, e-books, cheap telecommunication and Skype, cheap travel) further imply that the communicative possibilities open to most migrants are completely different today from what they were even two decades ago.

Lastly, one-dimensional measures of use of L1 and L2, which ask participants to estimate the proportion of use of both languages on a typical day, cannot capture the complex interplay of language use across a variety of domains—for example interactive versus receptive (reading, watching TV), written versus auditory, formal versus informal, and so on. Whether, for example, a participant routinely uses one of her languages in the workplace or at home with her partner may impact in very different ways on her acquisition of the L2 and the attrition of her L1 (see Schmid 2007, 2011).

All of these considerations imply that any attempt to comprehensively elicit a picture of both the current status of the participants' attitudes and language use and the development of these characteristics over time would necessitate a formidably complex instrument that would be extremely time-consuming to administer (and

might still not be entirely reliable). Again, a compromise is necessary, balancing available time versus the most desirable information. These considerations have lead to a questionnaire with some 100 questions, which is available in the online supplementary material. This questionnaire takes 45–60 min to administer and will allow you to obtain a comprehensive picture of attitude and use over time, contexts and modalities.

2.4 Language Proficiency

That language proficiency impacts strongly on performance on linguistic tasks and on language processing appears to be a trivial truth. However, it is an important consideration for assessments of L2 acquisition. There are two issues for which proficiency is particularly important. The first one concerns the comparison of groups with different L1s or L2s aiming to establish whether the developmental trajectory is similar for a specific aspect of language for various populations or whether path and outcome are affected by the L1 or L2 involved. In these cases, some more **general measure of proficiency** is essential in order to determine at what stage of language acquisition individuals are and to make sure that group differences are not confounded by proficiency. As well as providing valuable information for the single study, this general measure can act as a cross-study titration for meta-study analyses.

Second, there is the question of whether there are limits to the attainability of ultimate nativeness after a certain age, in particular where specific grammatical structures are assumed to be problematic. There are two important considerations here: First, L2 proficiency and age at acquisition are almost invariably negatively correlated in populations of learners. However, a substantial proportion of the decline in eventual proficiency among older learners is clearly due to changes in learning contexts, cognitive abilities and motivation, which co-occur with age and are associated with success in language learning (Bialystok 2001). Given these explanations for decline, identifying a limit to L2 development which is caused by independent maturational processes becomes difficult. It is therefore necessary to make sure that the L2 learners investigated are not only at the top of their cohort, but also at a level of **general proficiency** that is comparable to that of the reference population(s), i.e. that the population are near-natives.

Second, recent research has established that language processing and performance on certain tasks may also vary considerably within monolingual populations, and that one important contributing factor here is also language proficiency (2015). The general assumption of homogeneity within monolingual populations may be due to the fact that participants in linguistic experiments are so often university students—that is, groups of age-matched, highly educated and relatively young individuals who can be assumed to be comparatively highly proficient speakers of their L1. If such individuals are compared with monolinguals at lower levels of proficiency, very interesting differences emerge. For example, an ERP experiment

by Pakulak and Neville (2010) finds that high- and low-proficiency monolingual speakers of English differ substantially in their brain responses to sentences which violate English phrase structure rules. The differences between high- and low-proficiency native speakers found in this study are strikingly similar to those found between natives and second-language learners in many studies. Such findings indicate that the differences found between L2 populations of different ages of acquisition, such as the behavioral results reported by Johnson and Newport (1989) or the neurolinguistic findings presented by Weber-Fox and Neville (1996), may not solely be due to AoA but to the proficiency levels in the various age groups (which decrease with AoA in the Weber-Fox and Neville study, as was pointed out by Gillon Dowens et al. 2010; van Hell and Tokowicz 2010). Similarly, Dąbrowska (2012) suggests that individual differences between speakers, such as the level of education (and, by extension, the familiarity with formal, written registers) can lead to different underlying grammars even within monolingual populations.

For studies attempting to investigate whether there are any qualitative differences between (late) L2 learners and natives with respect to certain grammatical rules or structures, it is therefore of vital importance that the proficiency level of these populations be clearly described. Tremblay (2011) has treated this issue in detail and demonstrated that the common procedure of estimating proficiency based on hours of instruction is not reliable. Furthermore, this measure is obviously not useful in determining levels of knowledge of uninstructed learners, attriters or monolinguals.

We recommend a three-tiered approach, for practical reasons. First, recruitment should encourage participants to self-select on perceived level of proficiency, that is, the recruitment text should mention that the study will test advanced or very advanced speakers. Second, we highly recommend an online screening test before participants are invited to the experiment. For many languages, standardized instruments are available, for example the Goethe-Test for German (www.goethe.de/lrn/prj/pba/deindex.htm) or the DIALANG Placement Test for Danish, Dutch, English, Finnish, French, German, Greek, Icelandic, Irish-Gaelic, Italian, Norwegian, Portuguese, Spanish and Swedish (<http://www.lancaster.ac.uk/researchenterprise/dialang/about.html>). Using such a test will help ensure that the participants who are eventually selected are proficient enough to take part in the study.

In order to ensure homogenous proficiency levels across populations, the actual test battery should furthermore contain at least one further measure of overall proficiency. Again, previous studies vary in which task they prefer to this end. Some have used official standardized tests for L2 learners, such as the Test of Adolescent and Adult Language (TOAL-3; Hammill et al. 1994, used by Pakulak and Neville 2010) or the Oxford Quick Placement Test (2002), which is frequently used to assess L2 English proficiency. Such tests are, however, not available for all languages. We therefore recommend the widely-used C-Test, which is easy to construct and administer (as long as the language in question has a written register) and has been shown to be a reliable indicator of proficiency at advanced levels (Schmid 2011:183). The C-Test is constructed on the basis of real examples of short

texts, which may come from different genres (newspaper articles and columns, encyclopedia entries etc.). The first sentence of each text is left intact. Starting with the second sentence, the second half of every second word is removed and replaced by a gap (compounds, proper names and words that have been gapped before are skipped and in words with uneven numbers of letters, one more letter is removed than remains standing). Participants receive one point for each word which they are able to complete correctly; an exercise which requires them to make full use of the built-in redundancy of every text and integrate their linguistic knowledge from a number of levels.

Schmid (2014) used a C-Test comprising five short texts of ca. 20 gaps each, which she administered to very advanced L2 learners of German with English as their L1 and long-term L1 German migrants in Canada. A high level of reliability was found across these texts ($\alpha > 0.8$), and for the present study we therefore opted for a shortened version, using only two texts. The texts we used to assess proficiency in Dutch, English and German can be found in the online supplementary material.

To sum up we recommend that one or more measures of general proficiency be elicited by all investigations of bilingualism for purposes of cross-linguistic comparison. In addition we suggest that if there is a standardized test for the target language, this be preferred. Lastly, we recommend that the range of proficiency in the native group be established as well, since this can clearly also vary between studies and may explain inter-study variability as much as other factors.

2.5 Conclusion

The considerations offered in this chapter with respect to external and personal factors that may influence performance on linguistic tasks underscore the complex and multifactorial nature of linguistic research. Many predictors need to be taken into account, and most of them are not easily classifiable into a few neat categories (e.g. age before vs. after puberty). It should be kept in mind that some of these factors may interact with each other in complex ways. This suggests that more in-depth insight can only be gained on the basis of cumulative evidence from a range of studies.

In order to facilitate meta-investigations of such studies, it is necessary that all predictors be stringently controlled and documented. In addition, such a multi-level approach critically relies on comparable experimental procedure, and this is a substantial challenge, in particular where data collection at different sites, in different countries and in different laboratories is concerned. This is the challenge which the following chapter will address.

References

- Abutalebi, J., M. Canini, P.A. Della Rosa, L.P. Sheung, D.W. Green, and B.S. Weekes. 2014. Bilingualism protects anterior temporal lobe integrity in aging. *Neurobiology of Aging* 35(9): 2126–2133.
- Baddeley, A. 2003. Working memory and language: An overview. *Journal of Communication Disorders* 36(3): 189–208.
- Baker, Colin. 1988. *Key issues in bilingualism and bilingual education*. Clevedon: Multilingual Matters.
- Bialystok, E. 2001. *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press, Cambridge.
- Bialystok, E., and K. Hakuta. 1994. *In other words. The science and psychology of second-language acquisition*. New York: Basic Books.
- Bialystok, E., F. Craik, and G. Luk. 2008. Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 34 (4): 859.
- Birdsong, D., and M. Molis. 2001. On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language* 44(2): 235–249.
- Blackwell, A., and E. Bates. 1995. Inducing agrammatic profiles in normals: Evidence for the selective vulnerability of morphology under cognitive resource limitation. *Journal of Cognitive Neuroscience* 7(2): 228–257.
- Caplan, D., and G.S. Waters. 1999. Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences* 22(1): 77–94.
- Conway, A.R., M.J. Kane, and R.W. Engle. 2003. Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences* 7(12): 547–552.
- Conway, A.R., M.J. Kane, M.F. Bunting, D.Z. Hambrick, O. Wilhelm, and R.W. Engle. 2005. Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review* 12(5): 769–786.
- Dąbrowska, E. 2012. Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism* 2(3): 219–253.
- Daneman, M., and P.A. Carpenter. 1980. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19(4): 450–466.
- DeKeyser, R.M. 2000. The robustness of critical period effects in second language acquisition. *Studies in second language acquisition* 22(4): 499–533.
- DeKeyser, R., I. Alfi-Shabtay, and D. Ravid. 2010. Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics* 31(3): 413–438.
- Dowens, M.G., M. Vergara, H.A. Barber, and M. Carreiras. 2010. Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience* 22(8): 1870–1887.
- Dussias, P.E. 2004. Parsing a first language like a second: The erosion of L1 parsing strategies in Spanish-English Bilinguals. *International Journal of Bilingualism* 8(3): 355–371.
- Elman, J.L., E.A. Bates, M.H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett. 1996. *Rethinking innateness*. Cambridge: MIT Press.
- Friederici, A.D., K. Steinhauer, A. Mecklinger, and M. Meyer. 1998. Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology* 47(3): 193–221.
- Gass, S., and I. Lee. 2011. Working memory capacity, inhibitory control, and proficiency in a second language. In *Modeling Bilingualism: From Structure to Chaos*, ed. M.S. Schmid, and W.M. Lowie, 43–59. Amsterdam/Philadelphia: John Benjamins.
- Gold, B.T., N.F. Johnson, and D.K. Powell. 2013. Lifelong bilingualism contributes to cognitive reserve against white matter integrity declines in aging. *Neuropsychologia* 51(13): 2841–2846.
- Hall, K. 2008. 15 Exceptional speakers: contested and problematized gender identities. *The handbook of language and gender* 25: 353.

- Hammill, D.D., V.L. Brown, S.C. Larsen, and J.L. Wiederholt. 1994. *Test of adolescent and adult language: Assessing linguistic aspects of listening, speaking, reading, and writing*. Austin, TX: Pro-Ed.
- Johnson, J.S., and E.L. Newport. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21(1): 60–99.
- Kirchner, W.K. 1958. Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology* 55(4): 352.
- Knecht, S., B. Dräger, M. Deppe, L. Bobe, H. Lohmann, A. Flöel, and H. Henningsen. 2000. Handedness and hemispheric language dominance in healthy humans. *Brain* 123(12): 2512–2518.
- Mackey, A., J. Philip, T. Egi, A. Fujii, and T. Tatsumi. 2002. Individual differences in working memory, noticing interactional feedback and L2 development. In *Individual differences and instructed language learning*, ed. P. Robinson, 181–209. Amsterdam/Philadelphia: John Benjamins.
- McDonald, J.L. 2006. Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language* 55(3): 381–401.
- McKeever, W.F., K.S. Seitz, A.J. Krusch, and P.L. Van Eys. 1995. On language laterality in normal dextrals and sinistrals: results from the bilateral object naming latency task. *Neuropsychologia* 33(12): 1627–1635.
- Miyake, A., and N.P. Friedman. 1998. Individual differences in second language proficiency: Working memory as language aptitude. In *Foreign Language Learning: Psycholinguistic studies on training and retention*, ed. A.F. Healy, and L.E. Bourne, 339–364. Mahwah: Lawrence Erlbaum.
- Miyake, A., P.A. Carpenter, and M.A. Just. 1995. Reduced resources and specific impairments in normal and aphasic sentence comprehension. *Cognitive Neuropsychology* 12(6): 651–679.
- Oldfield, R.C. 1971. The assessment and analysis of handedness—The Edinburgh Inventory. *Neuropsychologia* 9(1): 97–113.
- Oxford Quick Placement Test. 2002. Oxford University Press, Oxford.
- Pakulak, E., and H.J. Neville. 2010. Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience* 22(12): 2728–2744.
- Pavlenko, A. (ed.). 2011. *Thinking and speaking in two languages*. Clevedon: Multilingual Matters.
- Pavlenko, A., A. Blackledge, I. Piller, and M. Teutsch-Dwyer (eds.). 2001. *Multilingualism, second language learning, and gender*. Berlin: Walter de Gruyter.
- Sagarra, N., and J. Herschensohn. 2010. The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua* 120(8): 2022–2039.
- Schmid, M.S. 2007. The role of L1 use for L1 attrition. In *Language attrition. Theoretical perspectives*, eds. Köpke B., M.S. Schmid, M. Keijzer and S. Dostert, 135–153. Amsterdam/Philadelphia: John Benjamins.
- Schmid, M.S. 2011. *Language attrition*. Cambridge University Press, Cambridge.
- Schmid, Monika S. 2014. The debate on maturational constraints in bilingual development: A perspective from first language attrition. *Language acquisition* 21(4): 386–410.
- Stroop, J.R. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18(6): 643.
- Tremblay, A. 2011. Proficiency assessment standards in second language acquisition research. *Studies in Second Language Acquisition* 33(3): 339–372.
- Ullman, M.T. 2004. Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92(1): 231–270.
- Van den Noort, M.W., P. Bosch, and K. Hugdahl. 2006. Foreign language proficiency and working memory capacity. *European Psychologist* 11(4): 289–296.

- Van Hell, J.G., and N. Tokowicz. 2010. Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research* 26(1): 43–74.
- Vos, S.H., T.C. Gunter, H.H. Kolk, and G. Mulder. 2001. Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology* 38(1): 41–63.
- Weber-Fox, C., and H. Neville. 1996. Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Cognitive Neuroscience* 8(3): 231–256.

Designing Research on Bilingual Development

Behavioral and Neurolinguistic Experiments

Schmid, M.; Berends, S.M.; Bergmann, C.; Brouwer, S.M.;

Meulman, N.; Seton, B.; Sprenger, S.; Stowe, L.

2016, VII, 107 p. 12 illus., Softcover

ISBN: 978-3-319-11528-3