Future Research Directions for Applied Linguistics

nninger

Future Research Directions for Applied Linguistics

Edited by
Simone E. Pfenninger and Judit Navracsics

$$G \ \frac{6}{232}$$

MULTILINGUAL MATTERS
Bristol • Blue Ridge Summit

Contents

Contributors

Editors

Simone E. Pfenninger, PhD, is Assistant Professor at the University of Salzburg. Her principal research areas are multilingualism, psycholinguistics and individual differences (e.g. the age factor) in SLA, especially in regard to quantitative approaches and statistical methods and techniques for language application in education. Simone is currently conducting research into early versus late learning of multiple foreign languages and the cognitive and psycholinguistic mechanisms that drive language change. Her work is published in several books, edited volumes, and in journals such as *Second Language Research* and the *International Journal of Bilingual Education and Bilingualism*. She has been involved in EFL for 10 years at different levels: secondary school, adult education, higher education, assessment of processes and outcomes in language education, and language policy. Simone is co-editor of the Second Language Acquisition book series for Multilingual Matters.

Judit Navracsics is Professor of Applied Linguistics and Head of the Institute for Hungarian and Applied Linguistics at the University of Pannonia, Chair of Pannon State Language Examination Centre and Editor-in-Chief of the *Hungarian Journal of Applied Linguistics*. Her research fields are bi- and multilingual development in early childhood, the bilingual mental lexicon and bilingual processing. Judit wrote her PhD thesis on early third language acquisition. Recently she has been working on bilingual word recognition and sentence comprehension with psychophysical tests carried out among bilinguals and Hungarian learners of English with different L2 proficiency levels. Based on her scientific findings, Judit with her colleagues has developed a smartphone application, 'Lystra', for learners of English preparing for the B2-level language examination. She is the author of five books, 14 edited books and more than 70 articles and book chapters. Judit is also the main organizer of the Summer School of Psycholinguistics held each year in Balatonalmádi, Hungary.

Authors

Szilvia Bátyi received her MA degrees in teaching English language and literature and history at the Ferenc Rákóczi II Transcarpathian Hungarian Institute (Ukraine), in 2007. In the same year she entered the Applied Psycholinguistics programme at the Linguistic Doctoral School, University of Pannonia (Veszprém), where she later received her PhD in 2011. In the period 2007–2011 she received several national and international scholarships such as the Ministry of Education Scholarship and the Hungarians Outside Hungary Scientific Scholarship. Her Erasmus professional practice was spent at CLCS, Trinity College, Dublin (tutor: David Singleton), and a research visit to Comenius University, Bratislava, was covered by the Visegrad Fund. In 2012 Szilvia joined the academic staff of the Department of Applied Linguistics at the University of Pannonia where she works as an Assistant Lecturer. Her research interests include bi- and multilingualism, linguistic landscape, and language attrition.

Lars Bülow is Lecturer and Assistant Professor of German Linguistics at Passau University (Germany). He is also postdoc at the University of Salzburg. His research and research interests range from language change theory, evolutionary linguistics, variation and language change, dynamic systems theory, sociolinguistics, genderlinguistics, textual linguistics, stylistics, forensic linguistics, semantics, pragmatics, philosophy of language, morphology, and language learning theory. Lars Bülow's doctoral thesis is about Language Dynamic in the Light of Evolutionary Theory – Towards an Integrative Model of Language Change. Lars Bülow was visiting scholar at the University of Groningen (NL) and Cambridge (UK).

Valéria Csépe is Professor of Cognitive Psychology and Neuroscience. She is member of the Hungarian Academy of Sciences and the Academia Europaea, and is head of the research group of Neurocognitive Development at the Brain Imaging Centre of the Research Centre of Natural Sciences of the Hungarian Academy of Sciences. Valéria's research focuses on topics such as higher order auditory processes (speech and music), and the dynamics of domain-specific (reading, language processing, visual cognition) and domain-general (executive functions) subsystems of cognition. Behavioural measures and methods of neuroscience (electrophysiology and neuroimaging) are used in combination to study these processes. Valéria has been acting for two terms (2008–2014) as Deputy Secretary General of the Hungarian Academy of Sciences, the first female elected to a high position in the Academy's history. She has been and is a member of numerous scientific societies and organizations as well as a member of the editorial boards of international and Hungarian scientific journals. Valéria has been decorated with many awards for outstanding scientific achievement. She is member of the International Scientific Council's (ICSU) Committee on Strategic Planning and Review. She has more than 200 publications (50% in English) with more than 2000 citations. Her most cited publications are on the neurocognitive correlates of the typical and atypical development of speech, language, reading and executive functions.

István Csernicskó was born in Chop, the westernmost town of the Soviet Union, in 1973. He received his university degree as a teacher of Hungarian language and literature at Uzhgorod State University in independent Ukraine in 1995, and received his PhD from Eötvös Loránd University in Budapest (2001). He habilitated in 2012 at Eötvös Loránd University in Budapest. István has taught at the Ferenc Rákóczi II Transcarpathian Hungarian Institute since 1997, and is Associate Professor at the University of Pannonia (Veszprém, Hungary, since 2014). Since 2001 he has been Head of the Antal Hodinka Linguistic Research Center in Beregovo, Ukraine.

Kees de Bot is Professor of Applied Linguistics at the University of Groningen and the University of Pannonia. His research interests are the psycholinguistics of multilingualism, the application of dynamic systems theory in the study of language development, and the cognitive consequences of multilingualism in different age groups. Kees has published in all the leading applied linguistics journals. His most recent book is *A History of Applied Linguistics 1980–2010*, which was published by Routledge.

Gyöngyi Fábián is a language teacher, lecturer and teacher trainer specializing in teaching English and Hungarian languages. She has worked as a materials writer and curriculum and materials developer with several national and international projects. She has a PhD degree in educational studies, and currently works as Associate Professor in the Institute of Hungarian and Applied Linguistics at the University of Pannonia, Hungary. Gyöngyi is also working as a materials developer for the State Language Examination Center Pannon. Besides teaching EFL, her teaching experience covers both general education courses and specialized courses for language teachers and school leaders in a variety of areas. Gyöngyi has written several course books. As a researcher, her special interests include general and language teachers' roles, behaviour and thinking in Hungary.

Rüdiger Harnisch holds the Chair of German Linguistics at the University of Passau (Germany). Before that he was Assistant Professor at the University of Bayreuth, Researcher at the Center of General Linguistics (Zentrum für Allgemeine Sprachwissenschaft), Berlin, and Professor at the University of Oldenburg (Germany). He is member of the Review Board for Linguistics at the German Research Funding Organization (Deutsche Forschungsgemeinschaft) and a member of two Commissions at the Bavarian Academy of Sciences

(Munich): those for Dialectology and for Territorial History. His research interests range from systemic to variational linguistics, from language typology to language (change) theory, from phonology and morphology to semantics and textual linguistics. Rüdiger's doctoral thesis was a Natural Morphology and Phonology of a South Thuringian dialect, and his habilitation thesis was a syn- and diachronic analysis of the typological parameter 'base form- vs. stem-principle' in German noun morphology. Rüdiger has published several anthologies, e.g. about strengthening processes (remotivation and resegmentation) and dialectological research. He has co-edited several collections of papers, e.g. about noun grammar or onomastics, and he has contributed articles to linguistic handbooks on morphology, sociolinguistics, dependency grammar, language minority research and the history of linguistics.

Ulrike Jessner is Professor at the University of Innsbruck and the University of Pannonia, Veszprem. She has published widely in the field of bilingualism and multilingualism with a special focus on the acquisition of English in multilingual contexts. She is the co-author of *A Dynamic Model of Multilingualism* (with Philip Herdina in 2002) and *Linguistic Awareness of Multilinguals: English as a Third Language* (2006). Ulrike is also Founding Editor of the *International Journal of Multilingualism* and the book series Trends in Applied Linguistics (with Claire Kramsch from Berkeley University). Ulrike has also been engaged in the development of the research area of third language acquisition/multilingualism, both as organizer of the biennial international L3-conferences and as founding member and President of the International Association of Multilingualism.

Kristina Cergol Kovačević is Assistant Professor at the Faculty of Teacher Education, University of Zagreb, Croatia, where she teaches courses in English linguistics and research methods in SLA studies. She learnt about research methodology in SLA while working on her master's thesis at the Research Centre for English and Applied Linguistics at the University of Cambridge and subsequently focused on bilingual language processing in her doctoral, postdoctoral as well as recent research. Accordingly, her research interests involve SLA studies and experimental psycholinguistics. Kristina is particularly interested in research methodology in the abovementioned fields and the mysteries of multilingual language processing mechanisms.

Wouter Penris initially worked as teacher and musician for some time and then he upgraded his BA in ESL teaching to an MA in Applied Linguistics and an MA in TEFL at the University of Groningen, the Netherlands. He currently works as teacher educator at the NHL University of Applied Sciences in Groningen, and as teacher of English Language and Culture and Music at a secondary school in Apeldoorn. Wouter is an eclectic researcher with interests in the fields of advanced SLA and statistics, but also in the

medical field, having co-authored two articles on the prevalence of feeding problems and cleft palates (one currently under review). In the coming years Wouter plans to start a PhD programme that will explore the relation between language measures and the Common European Framework of Reference in second language development.

Gyula M. Sáry, MD, PhD, DSc, is Head of Department of Physiology at the University of Szeged, Hungary, where he teaches medical physiology to medical, dental and pharmacy students. His main research field is the visual system, especially shape and object vision in monkeys and humans. His group conducts experiments in regard to the following areas: the neuronal coding of shapes in the visual areas; the binding of different modalities into a coherent percept; the effects of aging on the visual system; and the role of the magno- and parvocellular pathways in visual recognition. Gyula has recently become involved in linguistic research, where he is developing psychophysical software for studying bilingual written word recognition and sentence comprehension. He is the co-author of several articles and book chapters in this field.

David Singleton received his BA from Trinity College Dublin and his PhD from the University of Cambridge. He is a Fellow Emeritus of Trinity College Dublin, where until 2013 he was Professor of Applied Linguistics. He now holds the title of Professor in Hungary at the University of Pannonia. He has served as President of the Irish Association for Applied Linguistics, as Secretary General of the International Association of Applied Linguistics and as President of the European Second Language Association. David's publications number more than 180, his books and articles ranging across a wide spectrum of topics – focusing mainly on crosslinguistic influence, the second language lexicon, multilingualism, and the age factor in language acquisition. He is responsible for editing the Second Language Acquisition book series, published by Multilingual Matters.

Valentina Török is currently finishing her PhD thesis at the University of Innsbruck. Her research interests are predominantly based in the area of applied linguistics, including multilingualism, bilingualism, foreign language learning, third language acquisition, and attrition.

Vincent J. van Heuven obtained a BA (1971) and MA (1974) in English language and literature, an MA in general linguistics (1973) and his doctorate (1978) from Utrecht University. He took a postgraduate year at Edinburgh University and a postdoctoral year at UCLA. Vincent is Emeritus Professor of Experimental Linguistics and Phonetics at the Leiden University Centre for Linguistics (LUCL) and the Leiden Institute for Brain and Cognition, Honorary Professor at Nankai University (Tianjin, China) and at Groningen

University, and Professor at the University of Pannonia (Veszprém, Hungary). He was the Scientific Director of the Holland Institute of Linguistics and later of LUCL. Vincent was (Associate) Editor of *Netherlands Phonetic Archives, Speech Research, Journal of Phonetics* and *Phonetica*. Forty doctoral dissertations were defended under his (co-) supervision. He is a life member of the Royal Netherlands Academy of Arts and Sciences.

Marjolijn Verspoor received her PhD in 1991 from the University of Leiden, the Netherlands. She is a Professor at the University of Groningen, the Netherlands, and at the University of Pannonia, Veszprém, Hungary. Her primary research interest is second language development within a usage-based, emergentist perspective. She uses methods and techniques from dynamic systems theory to follow the development of different linguistic measures longitudinally and has published in several major journals on the topic. Another research interest is how insights from cognitive linguistics can help improve language teaching. Marjolijn has written several overview articles for encyclopaedias that show how cognitive linguistics may be used in teaching. Finally, Marjolijn is interested in transferring her own theoretical insights to her own students and has written and co-authored several textbooks. She co-authored a cognitive approach (construction grammar) to sentence analysis and a very popular, general introduction to cognitive linguistics, which has been translated (and re-authored) into several other languages: Afrikaans, Dutch, French, German, Greek, Italian, Korean, Polish and Spanish. Recently she co-authored an advanced resource book about second language acquisition, which gave the original impetus to applying dynamic systems theory to language development research.

1   Introduction

Simone E. Pfenninger and Judit Navracsics

As a genuine 'language-centered problem-solving enterprise' (Grabe, 2010: 34), applied linguistics has been at the busy intersection between various allied fields of general linguistics, such as psychology, education, sociology, public policy, information technology and others. Indeed, many scholars feel that linguistics represents only one among the many disciplines contributing to applied linguistics, rather than its core (see Cook, 2009). The call for papers to the 2011 AILA conference goes even further and lists 28 areas in applied linguistics (Schmitt & Celce-Murcia, 2010). It is therefore not surprising that this is a field that is notoriously difficult to define – almost 'elusive of definition' (Widdowson, 2005: 12) – embracing, inter alia, neuro-linguistic, psycholinguistic and sociolinguistic issues in the 'real world', language education and policy, and second language acquisition (SLA), foreign language (FL) teaching and learning. Thus, it does not represent one scholarly domain but rather a multi-, inter- and transdisciplinary science; what goes under the name of applied linguistics is acknowledged to be highly diverse (Hall *et al.*, 2011; Kaplan, 2010; Widdowson, 2005). The diversity of research issues addressed in applied linguistics, the steady growth of the field, the variety of (alternative) methods and procedures, the contextual diversity of acquisition/learning processes, the considerable increase in new statistical analyses, and the rate at which the field has been changing require those involved in this discipline to constantly reinvent themselves, to have a wide and thorough knowledge of new methods of data collection and analysis, and to make the transition from old-school to new-school approaches to language development. Furthermore, the greater attention currently being paid to this field has multiplied the number of points of view from which it can be analyzed.

To the extent that this collection, as its title implies, offers a snapshot of future implications for applied linguistics, our intention is to contribute to the progress and academic development of this broad discipline. This edited volume aims to be a timely overview of current thinking and directions for further research in applied linguistics by bringing together in a single volume a range of perspectives regarding original research agendas and innovative methodological approaches. The focus is thus not only on the challenges that applied linguistics research has been facing in recent years but also on

producing workable and productive research designs (possibly in uncharted territories) and on identifying ways as to how alternatives to conventional research methodologies can be operationalized. We are now at a point where it is becoming essential to initiate a reassessment of central themes and controversial issues in this branch of linguistics, to examine who is practising, where, and with what methods. Many applied linguists are beginning to do just that. There have been several excellent reviews of key theories, methodologies and the history of applied linguistics (e.g. de Bot, 2015; Gabryś-Barker, in prep.; Loewen & Plonsky, 2015; Paltridge & Phakiti, 2010), as well as collections that offer directions for further research with respect to specific areas of applied linguistics or a specific theoretical framework (e.g. Bruthiaux, 2005; Verspoor *et al.*, 2012). However, particularly for those new to these debates, it is necessary to stress a selection of salient points, such as: (1) new theoretical perspectives on research conundrums in applied linguistics, e.g. the so-called 'bilingual advantage' (BA) in psycho- and neurolinguistics or the optimal starting age debate in FL learning; (2) the growing interest among applied linguists in more nuanced and more complex (statistical) data analysis; and (3) the priority given to more descriptive and social approaches to linguistics rather than to theorizing. In what follows we examine these three main themes that emerge from the 12 chapters in this volume, starting with the contribution of applied linguistics to issues such as the BA debate.

New Theoretical Perspectives on Research Conundrums in Applied Linguistics

The volume opens with a piece by de Bot that reflects on the ongoing debate regarding the BA, and provides a conceptual framework for the more narrowly focused discussions that follow in Chapters 3 and 4. Analyzing some of the main language use activities that are assumed to be responsible for the BA, de Bot describes a new phase in bilingualism research, a phase that is characterized by finer grained analyses of task effects and a focus on the effects of a BA on different populations in different contexts and life circumstances. In lieu of definite answers, de Bot suggests that 'rather than trying to find "the" BA it is better to see what advantages bilingualism may bring for different populations and tasks' (p. 28). He also calls attention to the general tendency in academic publishing to favour positive outcomes over so-called null effects. In particular, he points out how results supporting a BA are still more likely to be published than results challenging such an advantage (de Bot, 2015; de Bruin *et al.*, 2015; but cf. Bialystok *et al.*, 2015). Csépe (Chapter 3) discusses insights into the BA that can be gained through the use of recently developed neuroscientific methods that improve the detectability of any advantage for bilinguals, as well as language use activities that are assumed to be responsible for the BA. She shares with de Bot the

belief that the debate on the benefits and drawbacks of growing up bilingual is still in its infancy – at least from a methodological point of view. Csépe does, however, seem less troubled by task effects and the low ecological validity of tests assessing the BA. Her main concern is the importance of neuroscientific research that focuses on questions of separability and domain specificity. In Chapter 4, Navracsics and Sáry raise the question of the quality of linguistic awareness as reflected by the age and manner of becoming bilingual. Having employed a psychophysical technique measuring reaction time and accuracy judgement, they see the development of semantic awareness independent of age or manner. At the same time, phonological awareness at the written level is much more dependent on age and manner of language acquisition, and they find second language (L2) learners faster but not less successful than natural bilinguals in phonological processing. In terms of the validity of the bilingual cognitive advantage, the data provide evidence that processing two languages is more time consuming, and it should be taken into account in all testing circumstances.

The potential and the benefits of research in neurolinguistics and the impact of language learning on brain processing is perhaps not a prominent immediate concern of applied linguistics (see Grabe, 2010: 39). However, significant advances in the relationships between brain functioning and language learning in recent decades suggest that research insights from neurolinguistics have become too important to ignore. This is also a main focus in Chapter 5 by van Heuven, who discusses alternative neurolinguistic methods such as functional magnetic resonance imaging (fMRI) and event-related potentials (ERP), which may supplement, or even replace, conventional paradigms to gain a better grip on the mental and cerebral representation of the phonology and phonetics of bilinguals (see also Sereno & Wang, 2007). Van Heuven – like Navracsics and Sáry in Chapter 4, and Pfenninger and Singleton in Chapter 6 – also addresses the thorny question of the impact of starting age on L2/FL attainment. Focusing on four groups of listeners in naturalistic settings – a monolingual Dutch group, a monolingual English group, late bilinguals and early bilinguals – he presents evidence for the hypothesis that even very gifted late learners of the target language still differ – subtly but measurably – from native speakers of the target language in terms of their perceptual representation of the sound system of the target language. Late bilinguals may acquire pronunciation skills that are indistinguishable from those of monolingual speakers of either target language, but are non-native in the details of their perceptual representation of either sound system. Taking as a starting point the classic myth that childhood is the best time to start to learn an FL, Pfenninger and Singleton (Chapter 6) review research that has a closer look at the 'earlier = better' claim. While findings concerning the age factor in naturalistic settings speak in favour of an early start to L2 learning in these settings – as exemplified in van Heuven's Chapter 5 – research in formal instructional L2 learning

settings has confirmed the finding relating to the initial faster rate of older starters (e.g. Álvarez, 2006; Cenoz, 2003; Mora, 2006), but has not confirmed the long-term benefits of an early start. Classroom studies in a range of countries have shown not only a rate advantage for late starters over early starters, but also very few linguistic advantages to beginning the study of an FL earlier in a minimal input situation (see, for example, Larson-Hall, 2008; Muñoz, 2006, 2011; Myles & Mitchell, 2012; Náves, 2009; Pfenninger, 2014a, 2014b; Unsworth et al., 2012).

It is subjects such as these that exemplify research at the heart of applied linguistics: studies of real-world issues, 'facts', beliefs and assumptions that constitute different ways of thinking and living. According to Widdowson (2005), the essential issue for applied linguistics is 'whether, how, and how far the ideas and findings that have been refined out of actual data by idealization and analysis can be referred back reflexively to the domains of folk [e.g. teacher/learner] experience whence they came and be made relevant in practice' (Widdowson, 2005: 20). As Widdowson (2005: 21) concedes, this is a difficult thing to do, 'and it is very tempting to simplify matters' such as by ignoring contextual factors, neglecting the diversity of local domains and making overgeneralizations. This leads us to the next theme of this volume.

New Methodological Approaches to the Complexity of Real-world Issues

Many authors in this volume endorse the development and use of new methodological approaches in applied linguistics. If we are to engage with real-world issues and take account of the complex variety of the 'real world', we need to: (1) take account of the interconnectedness of subsystems present; (2) develop research designs that base themselves on quantitative as well as qualitative findings such as mixed methodology; and (3) make more use of statistical models with built-in ecological validity, i.e. models that take account of both participant and item variability, allowing for the assessment of the impact of context-varying factors on age (over time).

It will be suggested throughout this book (e.g. by de Bot, Penris & Verspoor; Cergol Kovačević; and Bátyi) that there is a need to see language (acquisition) as comprising factors that are dynamic rather than static, complex (in the sense of patterns emerging from components interacting within the ecology in which they operate), interdependent, developing in a non-linear hence unpredictable fashion, highly dependent on initial conditions, and constantly changing and subject to perturbations – all of which are main points of the concept of complex dynamic systems theory (CDST) (for a fuller treatment of dynamic systems theory as it has been applied to language and SLA, the interested reader is referred to de Bot, 2008; de Bot et al., 2007; Dörnyei et al., 2014; Jessner, 2008; Larsen-Freeman, 2015;

Larsen-Freeman & Cameron, 2008). Penris and Verspoor's study (Chapter 11) makes a convincing case that CDST has far-reaching consequences, beyond what one might normally expect with a new theory. They set out to identify measures that work for writing at advanced levels, especially of academic writing, where nominalizations and other non-finite constructions are used. To this end, the texts of one student were rated on academic style by seven independent judges, and a number of measures were correlated with the median of the ratings. In line with the main premises of CDST, Penris and Verspoor conclude that 'linguistic development in writing is a rather erratic process when examined up close'. Given the fact that throughout the life span of a multilingual person many factors involved in multilingual development are subject to constant change, Jessner and Török (Chapter 10) argue that multilingualism lends itself to being researched from a CDST perspective. The aim of their chapter is to provide further insight into multilingual processing by shedding light on one of the key elements of the so-called M(ultilingualism)-factor, namely the implementation of strategies when trying to decode an unknown language. Research on strategies conventionally takes a rather reductionist approach, whereas the methodology discussed in this chapter takes into consideration the dynamics of the linguistic and other systems of the multilingual learners involved, thus proposing an alternative in the methodological approach to research on multilingual benefit.

The multidisciplinary approach of applied linguistics is also very strongly reflected in the greater focus on more sophisticated and more appropriate quantitative analyses from natural sciences methodology rather than the general and often inadequate approaches of the past. Attention to new statistical methods in applied linguistics has become particularly salient in view of two main developments in this field. One is the paradigm shift from qualitative to quantitative research designs that took place in applied linguistics in the 1990s and early 2000s. Lazaraton (2005: 214) sees three main problems with this paradigm shift: (1) she fears that general linear models (GLM) such as ANOVA and t-tests are used 'in violation of at least some of the assumptions of the procedure' (see also de Bot, 2012: 13–22), such as the inclusion of correlated data in linear models; (2) a great deal of the research becomes obscure for all but the most statistically literate; and (3) using high-powered parametric procedures may tempt one to overgeneralize results to other contexts or to other language users, when, in fact, many research designs do not use random selection from a population or random assignment to groups, but rather employ intact groups of a very limited demographic profile. Pfenninger and Singleton (Chapter 6) depart from the conventional approach to numeric data in applied linguistics with GLM as they address methodological and assessment issues that have been raised by the upsurge of interest in research on the age factor in FL settings. They suggest that multilevel modelling (MLM) approaches are ideal for a potentially generalizable study of age effects, as these analyses encourage us to shift

from a myopic focus on a single factor such as the age factor to examining multiple relationships among a number of variables, including contextual variables – or in Brown's (2011) words: 'you are more likely to consider all parts of the picture at the same time, and might therefore see relationships between and among variables (all at once) that you might otherwise have missed or failed to understand' (Brown, 2011: 11–12). Another pressing issue in current quantitative methods in SLA that is also addressed in Chapter 6 is to find a method that enables us to maximize the generalizability of the findings without sacrificing too much variability. Age-related research in this past decade has demonstrated an increased sensitivity to constraints on generalizability (or transferability) of results as well as features of the research design which support the causal connection between treatment and outcome that is inferred from the results (e.g. randomization). This is a reality for many areas in applied linguistics; as Gass (2006: 216) says, '[as] Applied Linguists, we do live within the real world and are subjected to real-world constraints on data'.

Besides pushing the agenda for more statistical advances (see also Plonsky, 2013, 2014), the community of applied linguistics researchers has become increasingly receptive again to qualitative research in the last decade. Previous reviews have identified a trend of the rising presence of qualitative studies published in leading applied linguistics and language learning journals (e.g. Benson et al., 2009; Lazaraton, 2005). In particular, mixed methods research – the meaningful merging of qualitative and quantitative approaches – offers a radically different new strand of research methodology which allows investigators to obtain data about both the individual and the broader societal context and brings out the best of the qualitative and the quantitative paradigm while also compensating for their weaknesses (Dörnyei, 2009; Dörnyei & Ushioda, 2011; Singleton & Pfenninger, 2015). According to the first issue of the Journal of Mixed Methods Research, this new research approach is defined as 'research in which the investigator collects and analyzes data, integrates the findings, and draws inferences using both qualitative and quantitative approaches or methods in a single study or program inquiry' (Tashakkori & Creswell, 2007: 4; see also Teddlie & Tashakkori, 2009). Thus, qualitative and quantitative research are combined at the data collection level and/or at the analysis level. Promoting the use of a mixed methods approach in future attrition research, Bátyi's paper (Chapter 13) reports on the findings of a study carried out with former learners of Russian, which was a mandatory school subject for many years in Hungary. The result is a paper that exemplifies the mutually beneficial interchange that can occur between theoretically grounded qualitative research at one end of the spectrum and data-based quantitative approaches at the other: while attrition researchers generally insist on using uniform research tools to study the phenomenon, Bátyi reminds us to bear in mind that language learning and attrition are dynamic processes that may differ considerably between individuals. To understand these processes, both qualitative and quantitative data are needed in the future.

Future Implications for Language Policy and Education

Several chapters in this volume offer a sample of the richness of two academic disciplines within applied linguistics that are directly concerned with bridging theory and practice and are thus of direct relevance to the practices of a range of professionals, including language educators and language policy makers: language policy and FL teaching. Language policy – comprising language practice and language beliefs or ideology – fits into applied linguistics, as it constitutes arguably the most extensive branch of 'language management', that is, the effort by individuals, groups, or institutions with (or claiming to have) authority to modify the language practices or beliefs of other speakers. Language management, in turn, is said to be 'at the very core of applied linguistics' (Spolsky, 2005: 30). In Chapter 7, on language policy in Ukraine, Csernicskó creates a strong case for examining language policies that involve macro-level general national language policy. He first describes the language situation in Ukraine, paying particular attention to the changes in FL learning and teaching since the demise of the Soviet Union. The major portion of his chapter is devoted to a description of how the political elite in Ukraine strived to settle the language issue by introducing language laws that were supposed to regulate the use of Ukrainian and Russian. By comparing and analyzing five language acts, he illustrates the priorities of the power elite in passing these laws, and the implications for the future directions of Ukrainian language policy. His work exemplifies the view that language planning is a multidimensional activity that requires academic and community input. Also in the realm of language policy, Bülow and Harnisch (Chapter 9) look into gender-sensitive language use and its widespread implications for the future expression of gender equality. Their study elaborates a concern that real-world language policy making ought to adopt systematic and rational approaches informed by sociolinguistic research. Promoting a bottom-up rather than top-down approach, they demythologize the assumption that the generic masculine discriminates against women – a powerful argument in political and academic discourse, as the notion of 'genus' (a grammatical category) is easily conflated with the concept of 'gender', which is a social and biological category. Their findings support their view that 'we need a non-ideological discourse in the future which takes account of the context and is based on valid empirical research and not on personal attitudes'.

Five of the 12 papers in this volume present theoretical and empirical evidence in the realm of language teaching and learning, which has been a domain that has often been considered the principal concern of applied

linguistics (see, for example, Pennycook, 2005). Far ranging and comprehensive, Fábián's Chapter 9 draws readers' attention to the need to investigate micro language planning activities. She anchors her discussion firmly in the world of the classroom as she provides a four-component model in order to extend the existing theories of the critical thinking concept into a wider conceptual framework to allow a more comprehensive approach to the study of critical thinking practices in classroom environments worldwide. While a lot of attention is paid to developing and improving methods of teaching critical thinking in the classroom, little evidence of teachers' critical thinking practices has been gathered. Her chapter considers some of the dilemmas involved in researching classroom critical thinking and makes an attempt to identify further areas of investigation in the subject. As mentioned above, Penris and Verspoor (Chapter 11) take a CDST approach to investigating the academic writing development of a young adult as 'a long, complex, dynamic process' over the course of 13 years. To trace the linguistic development of different subsystems from one level to the next, they zoom in on the dynamic relations between 13 variables measuring lexical and syntactic complexity across 49 written samples. The declared goal of their study is to show the high degrees of variability within the investigated variables and demonstrate how the latter grow in relation to each other. Making reference to introspective data from third language learners, the aim of Jessner and Török's chapter (Chapter 10) is to take a holistic approach to strategy research in order to define the systemic interaction between the multilingual learner's linguistic systems and the cognitive processing within the multilingual's mind – in line with the dynamic model of multilingualism proposed by Herdina and Jessner (2002). The chapter perfectly exemplifies the idea that, in order to be able to offer a real-world response to the problems they face, applied linguists would be well advised to conduct their studies in close collaboration with the people who are experiencing the problem or whose needs are to be met (see also Hall *et al.*, 2011: 18). Cergol Kovačević's empirical study in Chapter 12 provides a deeper insight into the processing of identical and semi-cognates by comparing the visual and auditory processing of (semi-)cognates by Croatian learners of English. Studies of bilingual cognate processing have long been used as a means of uncovering details about the organization of the bilingual mental lexicon and lexical access. Since Croatian is an orthographically shallow language, it is visually processed differently from English with its deep orthography, which is why results from both visual and auditory tasks need to be integrated into the interpretation of the lexical processing of such language combinations. Fittingly in the context of this volume, these chapters span theory and practice, but also add specificity to the view that applied linguistics 'should be concerned with bringing theory to bear on language-related problems while allowing practitioners to test their theories against systematically collected data' (Bruthiaux, 2005: 9).

Due to the fragmented and interdisciplinary nature of applied linguistics studies, this book is divided into three discipline-specific parts (Part 1: 'Future Implications for Bilingualism'; Part 2: 'Future Implications for Second Language Acquisition and Language Policy: Theoretical Considerations'; Part 3: 'Future Implications for Instructed Second Language Acquisition: Empirical Evidence'). Having said that, we accept that this division is in some cases problematic. For example, the chapters on the future of the BA (Chapter 2), the multilingual brain (Chapter 5), and academic writing development (Chapter 10), as well as others, could just as easily have been placed in Part 2 due to the interrelatedness of theory and practice in applied linguistics, which we have tried to establish in this Introduction chapter. Furthermore, the scope of the included studies is relatively broad, yet it represents the colourful tapestry of the state of the art and the interdisciplinarity of the fields in question, and it reflects the wider development of applied linguistics in the last 40 or 50 years. We hope that viewing these perspectives collectively in one volume will lead to a deeper understanding and appreciation of the merits of the theoretical and methodological diversity in applied linguistics.

Acknowledgements

References

Álvarez, E. (2006) Rate and route of acquisition in EFL narrative development at different ages. In C. Muñoz (ed.) *Age and the Rate of Foreign Language Learning* (pp. 127–155). Clevedon: Multilingual Matters.

Benson, P., Chik, A., Gao, X., Huang, J. and Wang, W. (2009) Qualitative research in language teaching and learning journals, 1997–2006. *Modern Language Journal* 93, 79–90.

Bialystok, E., Kroll, J.F., Green, D.W., MacWhinney, B. and Craik, F.I.M. (2015) Publication bias and the validity of evidence: What's the connection? *Psychological Science* 26, 944–946.

Brown, J.D. (2011) Quantitative research in second language studies. In E. Hinkel (ed.) *Handbook of Research in Second Language Teaching and Learning* (Vol. 2.) (pp. 190–206). New York: Routledge.

Verspoor, M., de Bot, K. and Lowie, W. (2012) *A Dynamic Approach to Second Language Development.* Amsterdam: John Benjamins.
Widdowson, H.G. (2005) Applied linguistics, interdisciplinarity, and disparate realities. In P. Bruthiaux, D. Atkinson, W.G. Eggington, W. Grabe and V. Ramanathan (eds) *Directions in Applied Linguistics* (pp. 12–25). Clevedon: Multilingual Matters.

Part 1

Future Implications for Bilingualism

2   The Future of the Bilingual Advantage

Kees de Bot

Introduction

One of the hot topics in research on multilingualism in recent years is what has been labelled 'the bilingual advantage' (henceforth, BA). Inspired by the seminal work by Ellen Bialystok and her colleagues, there is now a large amount of research on this phenomenon. In this contribution we discuss a number of aspects related to the BA and critical research on it in recent years.

In lay terms, the BA has been defined in terms of skills areas in which bilinguals outperform monolinguals. Bilinguals are assumed to be better with regard to:

- creative thinking;
- metalinguistic awareness;
- logical thinking;
- flexibility in thinking;
- the enhanced ability to learn additional languages.

Not all the links between these abilities and bilingualism have been supported by relevant research findings, and some of them are hardly testable. Operationalizing complex and often vague concepts like 'creativity', 'logical thinking' and 'flexibility' makes it difficult – if not impossible – to provide the empirical work to support the claims that these abilities are somehow enhanced by bilingualism. For a scientific approach, more refined and testable aspects need to be used. Most scientific research on the BA has focused on three components of cognitive processing:

- *Updating*. Keeping and refreshing information in memory: this is measured using tests like the so-called 'number recall test'.
- *Inhibitory control*. The ability to ignore irrelevant information: this is typically measured using the Stroop test and the Flanker test.

- *Task switching.* Swift switching between tasks, traditionally measured with tasks like the dimensional card sorting task. In this task, cards have to be categorized first on the basis of the colour of pictures and then on the basis of the shape of objects on the pictures. The switching between tasks typically leads to a slowing down of reaction times after the switch.

A yet to be solved issue is whether these three cognitive functions should be treated as separate modules in the cognitive system. It seems to be the assumption behind the idea that domain-specific training leads to domain-general applicability: if the functions only exist with the cognitive content they are working on, then it is hard to see how they can exist independently of that content. Think of intelligence: it doesn't exist by itself, but only becomes visible when certain tasks have to be carried out. Hence the question arises as to whether there is a module that supports task switching or whether task switching is an integral part of other activities, such as picking up the phone while cooking.

How is the Bilingual Advantage Assessed?

In current research, the existence of a BA is assumed to be demonstrated when, on certain tasks, there is a significant difference between bilinguals and a matched monolingual control group (see also Csépe, this volume, for a neurolinguistic perspective of the phenomenon). A major problem for this particular line of research is the question as to what counts as 'matching'. As we will see later, there are many factors that could cause a difference between treatment group and control group. This is especially true for the research on the advantages of bilingualism in ageing. In the literature there is a substantial debate about the degree to which an advantage in older age is in fact a BA (e.g. Bialystok *et al.*, 2004, 2006; Kousaiue & Philips, 2012; Salvatierra & Rosselli, 2010).

Experimental tasks

The three executive functions mentioned above have been studied extensively, using a number of tasks, in particular the anti-saccade task, the Flanker task, the Simon task and the Stroop task, as well as various formats of go/no-go tasks. In what follows we present a brief description of these different tasks.

Anti-saccade task

In this task, the informant has to focus on an asterisk in the centre of the screen. Another asterisk is projected on the left or the right of the centre. Depending on a cue (e.g. a colour) the participant has to look either in the direction of the new asterisk (congruent items) or in the opposite direction (incongruent items). The average difference between congruent and incongruent is the score used.

Flanker task

In this task, the informant sees a series of arrows and has to indicate the direction of the middle one. For example, for the series -> -> <- <- <-, <- is the right choice (the arrow keys of a computer keyboard are typically used to indicate the direction). Again there may be congruent items (-> -> -> -> ->) and incongruent items (-> -> -> <- <-), and the difference between these two is the score used to measure the effect.

Simon task

In this task, the informant sits in front of a screen with two buttons, one left and one right. These buttons are labelled for colour, e.g. red for right and green for left. On the screen, stimuli are presented in either red or green, appearing in different positions (left/right/middle). Again, there are congruent and incongruent items: when a stimulus is projected on the side that matches with the colour of the item, it is congruent; when it is projected at the side of the other colour, it is incongruent.

Stroop task

There are various versions of the Stroop task. The most widely used one is with words in different colours. The task for the informant is to name the colour of the print of the word rather than read the actual word. The stimuli can be colour names or other words. There are again congruent and incongruent items. Congruent items are words denoting a colour that are printed in that colour. Incongruent items are words denoting a colour that are printed in another colour (e.g. 'blue' printed in red ink). These items are mixed with neutral items: non-colour words printed in any colour.

Go/no-go tasks

Again there are a variety of tasks. A widely used format is when informants hear words or numbers and have to tap with a finger for each item apart from one or more specified ones. For example, they hear a series of numbers (3519849506439203) and have to tap after every number they hear apart from 5. The error rate is used as the score of the task.

These tasks have different ways of tapping into the three executive functions of updating, inhibition and task switching (Miyake *et al.*, 2000) – all of which are supposed to be sensitive to age-related decline (see below). There is substantial debate about the transfer of language-related versus more general abilities. To give an example: does a bilingual's inhibiting of a language over many years lead to their ability to inhibit information more generally, for instance when driving a car in a busy street?

The idea is that these abilities are domain general, i.e. that they are relevant to different cognitive domains. The other assumption is that training in one domain will lead to changes in the brain that also have an impact on other domains. Thus, being active for instance with music as a performer implies the use of inhibition, updating and task switching, so musical activity contributes

to each of these executive functions. By implication, speaking another language should contribute to musical skills, but there is no research on this as yet. The most obvious example of potential transfer from linguistic exercise and skills to more general abilities is code switching and task switching. Frequent code switching is assumed to enhance task switching as an executive function. Consequently, people who frequently switch codes are expected to perform better in other switching tasks. We will come back to this issue shortly.

What Causes a Bilingual Advantage? Language Use as Exercise Assumption

The general idea is that using multiple languages leads to the activation of skills that also affect non-linguistic skills. That may be the result of using the languages regularly over someone's lifespan. Bialystok suggests that an early start by itself may lead to cognitive advantages: 'If the boost given by childhood bilingualism is sufficiently strong, bilingualism may continue to influence certain control processes throughout the life span' (Bialystok *et al.*, 2004: 301). Pelham and Abrams (2014: 313) describe 'the BA' as 'the result of proficient, habitual use of 2 languages and not of developmental changes associated with becoming bilingual during childhood'.

Language use is a complex social and cognitive activity:

Speaking any language appears to require substantial amounts of monitoring, switching, and inhibitory control. To provide just a few examples, conversational participants must monitor the environment for signals regarding turn-taking, misunderstandings, possible use of sarcasm, changes of topic, and/or changes in register contingent upon who enters or leaves the conversation. These lead to switches from speaker to listener, switches from one knowledge domain to another, and so forth. (Paap & Greenberg, 2013: 256)

The task for bilinguals is even heavier:

Fluent bilinguals have extensive experience in language switching that involves monitoring the situation to select the appropriate language, activating the selected language, and inhibiting the other language. (Paap & Greenberg, 2013: 232)

The idea behind this is that language use acts as an exercise. Through exercise the brain develops special traits that are transferred from the language domain into the general skills domain.

One relevant question in this context is whether bilinguals have that much more to do than monolinguals. Quoting Paap and Greenberg (2013) again:

Fluent bilinguals have additional needs for monitoring, switching, and inhibitory control, but these unique requirements may not be substantial enough to generate group differences in cognitive control. (Paap & Greenberg, 2013: 256)

Whether or not this is true, there is simply not enough evidence to show how much activity and contact is needed to have the extra effect of bilingual use. In addition to frequency and amount of use, the kinds of linguistic tasks done may be of relevance. Is reading Thomas Mann's *Der Zauberberg* in German as a foreign language more effective than reading German comic strips? What makes a task in L2 more difficult, and is there any relevance to the level of difficulty ('if it doesn't hurt, it doesn't work')? Defining the difficulty of tasks is actually one of the main challenges in research on task-based learning (see Ellis, 2005, for an extensive treatment of this issue).

Different Bilingual Advantages for Different Populations?

Although in the literature (Bialystok *et al.*, 2004, 2006; Valian, 2015) authors refer mostly to 'the BA', it could be argued that there are different (types of) BAs. Research on the BA for young bilingual children in terms of cognitive processing focuses on different issues from research on the impact of bilingualism on ageing and dementia. The following groups can be distinguished:

- young early bilinguals (age of onset 0–5);
- young late bilinguals (age of onset 6+);
- young adults (typically university students);
- healthy elderly ($M = 65/M = 80.5$);
- elderly with dementia.

The first group typically involves children from bilingual families who are bilingual from birth. A large part of the research on the BA in children has focused on this group and has successfully found a BA. The second group typically involves subjects in educational settings, such as bilingual kindergarten and primary schools. Within this group a distinction can be made between fully fledged bilingual schools in which half of the teaching is actually in a second language and schools with early foreign language teaching where the children receive a couple of hours of English teaching a week. The third group is often the control and/or experimental group in psychological research. In studies of healthy elderly adults, the range of ages is rather large. Many studies report a mean age in the early sixties, while others tested adults of 80 and older, e.g. Kavé *et al.*, 2008 ($M = 83$), Schreuder and Marian, 2012 ($M = 81$). There is no clear cut-off point for this. The last group can show

more or less severe forms of dementia, caused by a range of neuro-degenerative diseases such as Alzheimer's, Parkinson's or Korsakow, but also transient ischemic incidents (TIAs) and other minor strokes. What they share is a decline of white and grey matter in the brain, which affects areas of the brain that are essential for cognitive and physical functioning. This group has received substantial attention because some of the research on bilingualism and ageing suggests that bilingualism delays the onset of dementia within a couple of years (Bialystok et al., 2007). The form of the BA will be different for each of these groups. The elderly group may benefit from less impact of age-related cognitive changes, while for dementia patients, the delay of onset of crippling conditions would be the most important.

Continuous Versus Discontinuous Bilingualism

If we accept the idea of language use as an exercise, then this use can take various forms. As indicated for the different age groups above, there is a difference between bilingualism from birth and bilingualism at a later age. We will not go into the issue of the critical age in language learning (see van Heuven, Navracsics & Sáry, and Pfenninger & Singleton, all this volume), and refer to all participants who learned a second language after age three as late acquirers, although it should be noted that it does make a difference whether a child learns English in primary school or whether an adolescent learns a foreign language at the age of 12 or 30. But from the perspective of the BA we can simply lump them together. Then there is the issue of continued and continuous use versus discontinued use. A child may grow up in a bilingual household, but for various reasons stops using one of the two languages, more or less returning to a monolingual state. Another child from the same family may continue using multiple languages. Along similar lines, someone who learnt French as a second language in school may continue using it or not. So we have a 2 × 2 matrix with age of onset and continuity of use as the two variables (Table 2.1).

By continued use we also mean using multiple languages now, that is, at the time of testing. The expectation is that the BA will be largest for Group 1, then Groups 2/3 and then Group 4. We have no evidence that suggests a larger effect for early/late onset versus continued/discontinued bilingualism. In the literature on the BA, information on age of onset is often mentioned

Table 2.1 Age of onset and continuity of bilingual language use

	Continued	
Age of onset	+−	
From birth	1	2
Later	3	4

(e.g. Salvatierra & Rosselli, 2010: 11 years and older; Fernandes et al., 2007: two languages from birth), but information on continued or discontinued use is sparse. In future research, these factors should definitely be included.

The Role of Language Proficiency

Defining bilingualism is notoriously difficult. Level of proficiency in the languages involved and frequency of use are mentioned most often to define someone's bilingual state (see, for example, Aronin & Singleton, 2012 for an extensive overview of this debate). Needless to say, very low levels of proficiency are unlikely to lead to a BA, but on the other side of the continuum it may be worth asking to what extent using multiple languages leads to cognitive load when the speaker is very fluent and uses their languages on an everyday basis. Maybe the largest gains can be achieved at intermediate levels, since there the effort needed to understand and use the other language may be higher than at very low and very high levels of proficiency.

A related issue is the linguistic distance between the languages of a bilingual speaker. It could be argued that, for a Dutch/German bilingual, keeping the two languages apart may be difficult because there is so much overlap between the languages, but the effort of using both languages will be lower, as they support each other through this overlap. For Dutch/Hungarian bilinguals the situation is the reverse: keeping the languages apart is easy, but knowing one of them provides little help in learning and using the other.

Bilingualism as a Process Rather Than a State

In the literature on the BA, bilingualism is typically treated as a state, that is, something that is fixed. People 'are' monolingual or bilingual. It could be argued that such stability is unlikely: there is probably an interaction between the advantages that come from being bilingual and knowledge of multiple languages. Better updating of information, such as retaining words in working memory long enough for them to be transferred to long-term memory will lead to higher proficiency and vice versa: added knowledge of the language will generate more effective updating. From a dynamic systems perspective[1] bilingualism should be seen as something that is dynamic and constantly changing (see also Penris & Verspoor; Jessner & Török; Cergol Kovačević, and Bátyi, all this volume).

What Evidence is There for the Bilingual Advantage?

A series of publications by Bialystok and colleagues (Bialystok et al., 2004, 2007) have provided evidence for a BA in children, adolescents and elderly

more or less severe forms of dementia, caused by a range of neuro-degenerative diseases such as Alzheimer's, Parkinson's or Korsakow, but also transient ischemic incidents (TIAs) and other minor strokes. What they share is a decline of white and grey matter in the brain, which affects areas of the brain that are essential for cognitive and physical functioning. This group has received substantial attention because some of the research on bilingualism and ageing suggests that bilingualism delays the onset of dementia within a couple of years (Bialystok *et al.*, 2007). The form of the BA will be different for each of these groups. The elderly group may benefit from less impact of age-related cognitive changes, while for dementia patients, the delay of onset of crippling conditions would be the most important.

Continuous Versus Discontinuous Bilingualism

If we accept the idea of language use as an exercise, then this use can take various forms. As indicated for the different age groups above, there is a difference between bilingualism from birth and bilingualism at a later age. We will not go into the issue of the critical age in language learning (see van Heuven, Navracsics & Sáry, and Pfenninger & Singleton, all this volume), and refer to all participants who learned a second language after age three as late acquirers, although it should be noted that it does make a difference whether a child learns English in primary school or whether an adolescent learns a foreign language at the age of 12 or 30. But from the perspective of the BA we can simply lump them together. Then there is the issue of continued and continuous use versus discontinued use. A child may grow up in a bilingual household, but for various reasons stops using one of the two languages, more or less returning to a monolingual state. Another child from the same family may continue using multiple languages. Along similar lines, someone who learnt French as a second language in school may continue using it or not. So we have a 2 × 2 matrix with age of onset and continuity of use as the two variables (Table 2.1).

By continued use we also mean using multiple languages now, that is, at the time of testing. The expectation is that the BA will be largest for Group 1, then Groups 2/3 and then Group 4. We have no evidence that suggests a larger effect for early/late onset versus continued/discontinued bilingualism. In the literature on the BA, information on age of onset is often mentioned

Table 2.1 Age of onset and continuity of bilingual language use

		Continued	
		+	−
Age of onset	From birth	1	2
	Later	3	4

(e.g. Salvatierra & Rosselli, 2010: 11 years and older; Fernandes *et al.*, 2007: two languages from birth), but information on continued or discontinued use is sparse. In future research, these factors should definitely be included.

The Role of Language Proficiency

Defining bilingualism is notoriously difficult. Level of proficiency in the languages involved and frequency of use are mentioned most often to define someone's bilingual state (see, for example, Aronin & Singleton, 2012 for an extensive overview of this debate). Needless to say, very low levels of proficiency are unlikely to lead to a BA, but on the other side of the continuum it may be worth asking to what extent using multiple languages leads to cognitive load when the speaker is very fluent and uses their languages on an everyday basis. Maybe the largest gains can be achieved at intermediate levels, since there the effort needed to understand and use the other language may be higher than at very low and very high levels of proficiency.

A related issue is the linguistic distance between the languages of a bilingual speaker. It could be argued that, for a Dutch/German bilingual, keeping the two languages apart may be difficult because there is so much overlap between the languages, but the effort of using both languages will be lower, as they support each other through this overlap. For Dutch/Hungarian bilinguals the situation is the reverse: keeping the languages apart is easy, but knowing one of them provides little help in learning and using the other.

Bilingualism as a Process Rather Than a State

In the literature on the BA, bilingualism is typically treated as a state, that is, something that is fixed. People 'are' monolingual or bilingual. It could be argued that such stability is unlikely: there is probably an interaction between the advantages that come from being bilingual and knowledge of multiple languages. Better updating of information, such as retaining words in working memory long enough for them to be transferred to long-term memory will lead to higher proficiency and vice versa: added knowledge of the language will generate more effective updating. From a dynamic systems perspective[1] bilingualism should be seen as something that is dynamic and constantly changing (see also Penris & Verspoor; Jessner & Török; Cergol Kovačević, and Bátyi, all this volume).

What Evidence is There for the Bilingual Advantage?

A series of publications by Bialystok and colleagues (Bialystok *et al.*, 2004, 2007) have provided evidence for a BA in children, adolescents and elderly

adults. The large formal meta-analysis on the BA by Adesope *et al.* (2010) presents data to show that there is a BA for the younger groups. Bilinguals were found to outperform monolinguals on combined measures of metalinguistic and metacognitive awareness and on measures of abstract and symbolic representation, attentional control and problem solving. There was, however, significant variability in these effect sizes:

These results indicate that the process of acquiring two languages and of simultaneously managing those languages – or inhibiting one so the second can be used without interference – allows bilinguals to develop skills that extend into other domains. These skills appear to give bilingual speakers insight into the abstract features of language and into their own learning processes. They also appear to give bilingual speakers an enhanced capacity to appropriately control and distribute their attentional resources, to develop abstract and symbolic representations, and to solve problems. (...) The evidence reviewed in the current analysis suggests that earlier, rather than later, acquisition of a second language is also more likely to be associated with greater metalinguistic and metacognitive awareness. Although bilingual speakers of all ages demonstrated significant advantages with respect to representation and attention, only the youngest bilinguals (who, by definition, must have acquired their second language early in life) showed significant advantages with respect to metalinguistic or metacognitive awareness. (Adesope *et al.*, 2010: 228)

With respect to elderly adults, the overview presented by de Bot *et al.* (in prep.) suggests that various types of BA are found for both cognitive functions in healthy elderly and delay of symptoms in dementia.

There is some – albeit inconsistent – evidence of a BA in children: 'Together, these findings indicate that while exposure to a second language in a classroom setting may not be sufficient to engender changes in cognitive control, it can facilitate verbal memory and verbal learning' (Kaushanskaya & Marian, 2014: 564). For elderly adults with dementia there are a number of articles pointing to a BA. The first study showing a delay in the onset of symptoms of Alzheimer's disease was the seminal study by Bialystok *et al.* (2007). Similar observations were documented in Canada and India by Freedman *et al.* (2014). Alladi *et al.* (2013) reported on a large-scale study on dementia including 648 patients with dementia, 391 of whom were bilingual and 36% of whom were illiterate. They found a 4.5-year delay in onset of dementia for bilinguals. The BA for illiterates was significantly higher than for literates. More than two languages had no additional effects.

There is a large set of studies showing mixed effects, with positive findings on some tasks and no or negative findings on others. These will not be

elaborated on further here (see the overviews in Adesope *et al.*, 2010; Hilchey & Klein, 2013; Valian, 2015).

Opposing Voices

After the first excited reactions to the findings on cognitive advantages in children and, in particular, elderly people, there is now a stream of research that casts doubt on the existence or at any rate detectability of the BA (Gathercole *et al.*, 2014; Hilchey & Klein, 2011; Paap & Greenberg, 2013). Some researchers (Valian, 2015; Gathercole *et al.*, 2013) have argued that there are many factors that have been shown to enhance cognitive functioning in younger and older people, including exercise, playing an instrument, playing computer games, intellectually challenging work and social engagement. Bilingualism is another factor that might be added to this list, but it is virtually impossible to isolate bilingualism as a factor. As a matter of fact, it has yet to be determined what exactly the mechanism is that turns bilingual processing into changes in the brain that have a lasting effect on cognitive processing, in particular with respect to executive functions that are responsible for information updating, switching and inhibition of irrelevant information. Interestingly, in a large-scale study, Gathercole *et al.* (2014: 236) found no BA: 'Card sorting, Simon and metalinguistic judgment task (650, 557 and 354 participants, respectively) reveal little support for a bilingual advantage'. On the basis of an extensive analysis of the literature, Hilchey and Klein (2011: 629) conclude that 'the research findings testing for bilingual advantages in executive functions do not provide coherent and compelling support for the hypothesis that the bilingual experience causes improved executive functions'.

In her review of the literature on bilingualism and cognition, Valian (2015) carefully analyzes the different components of the BA. She starts by pointing out two possible positions with respect to the bilingual benefit:

(1) There *is* a benefit of bilingualism for executive function, but that benefit competes with other benefits that both mono- and bilinguals have to varying degrees. Depending on the composition of each group in any given experiment, the other benefits may be more plentiful in the monolingual than bilingual group (or sufficiently plentiful in both groups), so that the benefits of bilingualism are invisible. This is the possibility that I favour.

(2) There is *no* cognitive benefit of bilingualism. In experiments that have found a benefit, the effect is either due to the accidentally larger number of other positive factors, such as high SES [socio-economic status], that bilinguals have in that particular sample, or due to the correlation of bilingualism with some other active property that is difficult to separate from bilingualism.

Valian's final conclusion seems to be that there may be a BA, but that it is virtually impossible to prove its existence. Her carefully worded conclusions are somewhat different in tone from those expressed by Morton (2014):

> For several years now, headlines around the world have been trumpeting the cognitive control advantages enjoyed by bilinguals. As the story goes, a lifetime of experience selecting between two competing languages leads to improvements in control that generalize beyond the domain of language. It even protects the brain from the untoward effects of ageing. Despite the wide adulation of this view, critical readers like me find the whole story to be an insufferable mixture of excessive claims and weak evidence. (Morton, 2014: 929)

As for the conditions or activities that were used in the various studies Valian discusses, there is an impressive list. To name a few: high levels of education; high SES; a stimulating environment; aerobic training; engaging in challenging activities; playing a musical instrument; playing video games; and even cooking and visiting relatives and friends. And of course: speaking and using more than one language. But the latter is not necessarily more effective than the other factors and, depending on our research setup and statistical analysis, we may give priority to certain activities and let them take as much of the variance as possible. The pure effect of all these potential contributing factors cannot be measured, since all factors are related to each other. For instance, both SES and education appear to have an impact on cognitive functioning, but they are also correlated. A high SES and level of education will both impact on cognition, which in turn is related to many other variables, such as international orientation, holidays abroad, and interest in culture and arts. Similarly, extensive physical training will be related to motivational factors, diet and general health.

A Publication Bias for the Bilingual Advantage

There is a tendency in academic publishing to favour positive outcomes over null effects: in any branch of science, articles reporting positive results have a higher chance of being published than articles reporting no effect (Easterbrook et al., 1991). As Dickersin et al. (1987) point out, this has nothing to do with the quality of the research, since studies with significant results do not appear to be superior to studies with a null result with respect to quality of design or the way the experiment was carried out. Gathercole et al. (2013) point to the likelihood of a publication bias for the BA and, in an interesting setup to test this, de Bruin et al. (2015) looked at presentations at conferences on the BA and later assessed which of these presentations were ultimately published in academic journals. They found a strong bias for

studies reporting a positive effect. The question as to why the studies showing no BA were published less frequently remains to be answered: the authors may have decided that there was no point in submitting it, given the seemingly 'disappointing' results, but it is also possible that the reviewers and/or journal editors may have favoured the positive findings over the negative ones. Such a bias leads to a false picture of the state of play for a specific issue, because it suggests that 'the right experiment' should lead to a positive result, as that is what is generally reported in the literature; therefore the researcher's own study, which did not show such a positive result, must be wrong in one way or another.

In addition to the publication bias, there are also indications that there may be a citation bias. De Bot (2015) selected a large set of publications on the BA and looked at the number of citations generated by publications showing a BA, no BA, or mixed findings. It turned out that the number of citations was larger for the studies with a positive result than for the studies in the two other categories.

It is becoming increasingly clear that rather than one all-purpose BA, there may be different advantages – but also disadvantages – for specific groups (early/late bilinguals, migrants, high/low education) and specific cognitive functions (working memory, attention, task switching).

Switching Costs

A recent development is the interest in what has become known as 'switching costs'. Ever since the publication of Meuter and Allport's (1999) study on switching costs, a substantial number of experimental studies have been carried out on this topic. In the Meuter and Allport study, bilinguals had to name numerals as quickly as possible either in their first language or their second language, depending on a cue. The data show that response latencies on switch trials, i.e. items following a switch, were longer than items not preceded by a switch. Language-switching costs were larger when switching to the dominant L1 than when switching to the L2. The data generally reveal an effect of code switching (henceforth CS) in terms of timing of speech. Such switching costs are seen as reflecting the additional effort a bilingual speaker must invest to switch between languages as opposed to a continuation of speech in the current language. For a switch, the current language has to be switched off, and the other language has to be activated. Prior and Gollan (2011) show that the size of the BA is related to the reported amount of CS in bilinguals. Similar results were found by Verreyt et al. (in press), who reported that the degree of CS correlated with BA on other tasks. How reliable these findings are is unclear, however. Self-reports on CS are notoriously unreliable (see, for example, Bourhis & Harvey, 2010; Dewaele & Li, 2014).

To what extent such experimental findings reflect what happens with CS 'in the wild', that is, in non-experimental everyday language use, is also unclear, and it could be argued that such experimental findings tell us basically nothing about normal language use. The literature on switching costs is based on a model of language in which different languages are seen as separate entities that can be switched on and off. This type of model has recently come under attack, along with a larger array of psycholinguistic tasks that have created a world of experimental findings which has become a field of research itself, with an increasingly weaker link with real language use. In the experimentation on switching costs, the timing of cued switches in tasks like picture naming is compared with that of non-switched words. The underlying assumption here is that being as fast as possible is the goal in CS. For various reasons, this assumption is problematic. In conversations, switches are anticipated by changes in topic or conversational settings and sometimes 'softened' by trigger words. The notion of languages that need to be switched on and off is problematic if we view language as essentially conversational/interactional in nature, that is, shared, embedded and embodied.

What can CS corpora tell us about switching costs? For CS there are hardly any corpus data that can inform us about the switching costs. In a small case study, we looked at CS data from a 68-year-old Dutch immigrant in Australia who has developed a speech style with heavy CS. Analyzing pauses before switch and non-switch words, we found more pauses before a switch to English, which was hypothesized to be the weaker language, as she was brought up in the Netherlands and had acquired only basic skills in English. At first glance, these findings would then suggest a cognitive cost. But can we really deduce that from an analysis of length of pauses? There may be many reasons for the lengthening of pauses, even a signalling of an upcoming switch. It is also possible that particular words are harder to access than others, but these are not necessarily from a specific language. Many factors, such as frequency, saliency, topic, and recent use in a prior conversation define the accessibility of words. Hence, interpreting pause length as reflecting switching costs is difficult. Maybe there are no costs at all in normal CS conversations, which makes the experimental data on switching costs interesting within the realm of that type of metalinguistic activity but essentially irrelevant for everyday CS behaviour among multilinguals.

Speakers in a CS setting have to be constantly aware of changes in the conversational setting that may call for the use of a specific language. In particular, the inhibition of the non-selected language is supposed to have an impact on various types of non-linguistic skills. The executive function tasks mentioned are typically based on the need to ignore confounding information. For instance, in the Flanker task, the informant has to indicate the direction of a pointer. As mentioned earlier, there are two conditions – a congruent one and an incongruent one. So the task is to ignore the incongruent information. This activity is assumed to be similar to what a bilingual does. But if we think about this a little further, we may wonder to what extent that argumentation holds. When is a bilingual in a situation in which a specific language has to be inhibited? Let us take a situation in which a group of speakers of the same language are talking together, when suddenly someone else joins in who does not speak that language. The speakers have to switch languages if they want to be polite. But does that mean that all of a sudden they have to constantly inhibit the language they were speaking before the intruder joined them? The moment they noticed his presence they switched languages, but at what cost? There was no need to react in the other language as quickly as possible. At that conversational level, a 300-msec switching effect is irrelevant – if it exists at all. What is more, the situation is not unique for bilingual settings. When, at a company party, collaborators of the same department are talking together, typically about their departmental issues, and someone from another department joins in, they have to adjust their language and choice of topic to that new setting. So the experience of having to switch codes is a frequently occurring one and certainly not specific to bilinguals. If there are no costs, or training of inhibition, then what creates the BA?

Another fundamental difference between CS experiments and natural CS is that in the former speakers are forced to switch based on specific cues, e.g. 'use English when the target word is preceded by a red dot' and 'use French when the word is preceded by a green dot'. In normal CS conversation speakers are never forced to switch from one item to the next. They switch because that suits them; words and constructions are triggered by what happens in the conversation. There is priming on the lexical, syntactic and conceptual level and some words or expressions just come more easily in that language than in the other. Maybe even the expression 'in that language' is inadequate to describe what is happening. A word or larger unit is used because it presents itself, so to speak, and from what language it stems is basically irrelevant. At the metalinguistic level we can assess that the speaker used this language and then switched to another language, but that may not be what happens in the speaker's mind: she is just using what is convenient and cost-effective, and that can be the use of entities that belong to one language or another, but that is a metalinguistic statement, not a reflection of what may be somewhat misleadingly called CS.

However, even if we were able to measure the unique contribution of language-related factors, what would be the point? In an ecological and dynamic approach to bilingualism, the use of multiple languages is not something that can be singled out or put between brackets without consequences for other aspects of life. Life in multilingual settings involves much more than the use of one code or another. There may be cultural aspects that impact on the communication, and there are norms for the use of specific languages and CS as a communicative activity.

Bilingual Advantages and Disadvantages

Why is a large part of the multilingualism community so excited about this whole debate about the BA? If we have a closer look, then the 'advantage' is measured by a limited number of tests that suffer from extremely low ecological validity. Why are we excited about a 120-msec effect in the Simon or Flanker task or in some small but significant differences on the ANT, go-no go or task-switching task? Maybe we should worry more about the bilingual disadvantage in various tasks such as decision experiments, where monolinguals outperform bilinguals and trilinguals (Ransdell & Fischler, 1989). Bilinguals are typically slower than trilinguals, and trilinguals slower than bilinguals in a range of lexical tasks (Mägiste, 1985), including semantic fluency (Gollan et al., 2002). Bilingual adults are often slower at lexical access (Ivanova & Costa, 2008) and show more tip-of-the tongue problems (cf. Gollan & Acenas, 2004). Navracsics et al. (2014) compared monolinguals and bilinguals on phonological and semantic rhyming and found no advantage for the bilinguals (see also van Heuven, this volume).

It is not clear how we should weigh these findings if we want to translate the experimental findings into real-world tasks. In addition, it is not clear how we should interpret longer latencies. Longer latencies can be seen as a sign of effort or as compensatory strategies to reduce effort. What indications do we have that the findings on such tasks have in any sense an impact on bilinguals' wellbeing or professional performance? Are bilinguals better drivers because they are better at ignoring irrelevant information, which is crucial in heavy traffic? Do children in bilingual programmes perform better on mathematics at a later age? Am I a better bridge player because I am multilingual?

Different components of executive function may be positively correlated, uncorrelated or negatively correlated with one another, depending on the task. As Paap and Greenberg (2013) show, even tasks that appear superficially similar, such as the Simon task and the Flanker task, tend not to correlate, which demonstrates the need for a fine-grained task analysis.

The Future of the Bilingual Advantage

The state of affairs with respect to the BA is far from clear at the moment, and it looks as though it would be wise not to see the BA as the Holy Grail that is there or not. It is becoming clearer now that rather than trying to find 'the' BA it is better to see what advantages bilingualism may bring for different populations and tasks. An example could be a study of inhibition in elderly adults without dementia with an early onset and continued use of multiple languages at C1 level. Earlier on we pointed out that there are different age groups with different language skills and language use patterns and that each of them requires a specific approach in order to have a BA.

At the same time, it has to be acknowledged that research on the BA has societal repercussions and that the research on the BA by Bialystok and her colleagues has put bilingualism on the map. The issue is that bilingualism should not be seen as a panacea for all cognitive ills. While cognitive activity is beyond doubt beneficial for many neural degenerative diseases, its impact is yet to be established. An example could be the impact of cognitive activity on Parkinson's disease. The benefits of mental and physical activity for a reduction of symptoms is well established (see Tanaka et al., 2009, for an overview) but whether the learning of a difficult language like Hungarian as a foreign language has benefits for Parkinson's is unclear. Here a distinction can be made between the cognitive spin-off of having the brain work on such a difficult task and the satisfaction generated by successful learning.

The Bilingual Advantage in Additional Language Learning

Another aspect of the BA that deserves to be explored further is the assumed impact of learning the next language in bilinguals. While this has been mentioned repeatedly (see the special issue on L3 learning of the journal Bilingualism: Language and Cognition, 2015), there is remarkably little research on this. Is it easier to learn the next language because the executive functions that appear to benefit from bilingual use are enhanced, or have bilinguals developed more effective learning strategies through the experience of learning multiple languages?

The whole discussion on the BA in executive functioning can be repeated here. How and to what extent do age of onset, level of proficiency and use of languages influence the successful learning of yet another language? Is there a cumulative effect in the sense that learning the fourth language is easier than the third, which in turn is easier than the second? What about cognateness of languages? The situation of learning Spanish by German/French bilinguals is probably different from that of Dutch/English bilinguals learning Hungarian. But how different these situations are is yet to be established.

Concluding Remarks

Research on the BA is now entering a third and new phase: after the initial sensational findings there has been a countermovement that problematized some of the issues involved, such as the compatibility and validity of tasks, the selection of informants and the designs of the original studies. The state of play is that there may be a BA for different populations, but it

may be difficult to single that out among many other factors that may be beneficial over someone's life span. This third phase will entail the integration of the two previous stages, in which different types of advantages for different populations will be the object of research. The fact that being a bilingual is beneficial both socially and psychologically is beyond doubt, but the specifics are waiting to be explored.

Note

(1) We will not elaborate on this here. The interested reader is referred to various publications (e.g. de Bot, 2008; de Bot & Larsen-Freeman, 2011; Verspoor et al., 2011).

References

Adesope, O.O., Lavin, T., Thompson, T. and Ungerleider, C. (2010) A systematic review and meta-analysis of the cognitive correlates of bilingualism. Review of Educational Research 80 (2), 207–245.

Alladi, S., Bak, T.H., Duggirala, V., Surampudi, B., Shailaja, M., Shukla, A.K., Chaudhuri, J.R. and Kaul, S. (2013) Bilingualism delays age at onset of dementia, independent of education and immigration status. Neurology 81 (22), 1938–1944. doi: 10.1212/01.wnl.0000436620.33155.a4.

Aronin, L. and Singleton, D. (2012) Multilingualsim. Amsterdam: John Benjamins.

Bialystok, E., Craik, F.I., Klein, R. and Viswanathan, M. (2004) Bilingualism, aging, and cognitive control: Evidence from the Simon task. Psychology and Aging 19 (2), 290–305.

Bialystok, E., Craik, F.I. and Ryan, J. (2006) Executive control in a modified antisaccade task: Effects of aging and bilingualism. Journal of Experimental Psychology: Learning, Memory, and Cognition 32 (6), 1341.

Bialystok, E., Craik, F.I. and Freedman, M. (2007) Bilingualism as a protection against the onset of symptoms of dementia. Neuropsychologia 45 (2), 459–464.

Bourhis, R.Y. and Harvey, S.P. (2010) Intergroup allocation bias. In J.M. Levine and M.A. Hogg (eds) Encyclopedia of Group Processes and Intergroup Relations (Vol. 1, pp. 445–448). Thousand Oaks, CA: Sage.

de Bot, K. (2008) Second language development as a dynamic process. The Modern Language Journal 92 (2), 166–178.

de Bot, C. (2015) Circadian rhythms and second language development. International Journal of Bilingualism 19 (2), 142–156. doi: 10.1177/1367006913489201.

de Bot, K. and Larsen-Freeman, D. (2011) Researching second language development from a dynamic systems theory perspective. In M. Verspoor, K. de Bot and W. Lowie (eds) A Dynamic Approach to Second Language Development. Methods and Techniques (pp. 5–23). Amsterdam: John Benjamins Publishers.

de Bot, K., Houtzager, N. and Kochynek, A. (in prep.) A review of the bilingual advantage in aging.

de Bruin, A., Treccani, B. and Della Sala, S. (2015) Cognitive advantage in bilingualism: An example of publication bias? Psychological Science 26 (1), 99–107.

Dewaele, J.M. and Li, W. (2014) Intra- and inter-individual variation in self-reported code-switching patterns of adult multilinguals. International Journal of Multilingualism 11 (2), 225–246.

Dickersin, K., Chan, S. and Chalmers, T.C. (1987) Publication bias and clinical trials. Controlled Clinical Trials 8 (4), 343–353.

Easterbrook, P.J., Berlin, J.A., Gopalan, R. and Matthews, D.R. (1991) Publication bias in clinical research. Lancet 337 (8746), 867–872.

Ellis, R. (2005) Task Based Learning. Oxford: Oxford University Press.

Fernandes, M.A., Craik, F., Bialystok, E. and Kreuger, S. (2007) Effects of bilingualism, aging, and semantic relatedness on memory under divided attention. Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale 61 (2), 128.

Freedman, M., Alladi, S., Chertkow, H., et al. (2014) Delaying onset of dementia: Are two languages enough? Behavioural Neurology 2014, 290–304.

Gathercole, V.C.M., Thomas, E.M., Roberts, E.J., Hughes, C. and Hughes, E.K. (2013) Why assessment needs to take exposure into account: Vocabulary and grammatical abilities in bilingual children. Issues in the Assessment of Bilinguals 20–55.

Gathercole, V.C.M., Thomas, E.M., Kennedy, I., Prys, C., Young, N., Guasch, N.V., Roberts, E., Hughes, E.K. and Jones, L. (2014) Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. Frontiers in Psychology 5, 11.

Gollan, T.H. and Acenas, L.R. (2004) What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish-English and Tagalog-English bilinguals. Journal of Experimental Psychology: Learning, Memory, and Cognition 30 (1), 246.

Gollan, T.H., Montoya, R.I. and Werner, G.A. (2002) Semantic and letter fluency in Spanish-English bilinguals. Neuropsychology 16 (4), 562.

Hilchey, M.D. and Klein, R.M. (2011) Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. Psychonomic Bulletin & Review 18 (4), 625–658.

Ivanova, I. and Costa, A. (2008) Does bilingualism hamper lexical access in speech production? Acta Psychologica 127 (2), 277–288.

Kaushanskaya, M. and Marian, V. (2009) Bilingualism reduces native language interference during novel-word learning. Journal of Experimental Psychology: Learning, Memory and Cognition 35 (3), 829–835.

Kavé, G., Eyal, N., Shorek, A. and Cohen-Mansfield, J. (2008) Multilingualism and cognitive state in the oldest old. Psychology and Aging 23 (1), 70–78.

Kousaie, S. and Phillips, N.A. (2012) Ageing and bilingualism: Absence of a 'bilingual advantage' in stroop interference in a nonimmigrant sample. The Quarterly Journal of Experimental Psychology 65 (2), 356–369.

Mägiste, E. (1985) Development of intra- and interlingual interference in bilinguals. Journal of Psycholinguistic Research 14 (2), 137–152.

Meuter, R. and Allport, A. (1999) Bilingual language switching in naming: Asymmetrical costs of language selection. Journal of Memory and Language 40 (1), 25–40.

Miyake, A., Friedman, P., Emerson, M., Witzki, A. and Howerter, A. (2000) The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: A latent variable analysis. Cognitive Psychology 41, 49–100.

Morton, J. (2014) A sunny review casts a foreboding shadow over status quo bilingual advantage research. Applied Psycholinguistics 35 (5), 929–931.

Navracsics, J., Sáry, G., Bátyi, S. and Varga, C. (2014) Testing linguistic awareness among learners of Hungarian. In M. Pawlak and L. Aronin (eds) Essential Topics in Applied Linguistics and Multilingualism (pp. 111–129). New York: Springer.

Paap, K.R. and Greenberg, Z. (2013) There is no coherent evidence for a bilingual advantage in executive processing. Cognitive Psychology 66 (2), 232–258. doi:http://dx.doi.org/10.1016/j.cogpsych.2012.12.002.

Pelham, S.D. and Abrams, L. (2014) Cognitive advantages and disadvantages in early and late bilinguals. Journal of Experimental Psychology: Learning, Memory, and Cognition 40 (2), 313–325.

Prior, A. and Gollan, T.H. (2011) Good language-switchers are good task-switchers: Evidence from Spanish English and Mandarin-English bilinguals. Journal of the International Neuropsychological Society 17 (4), 682–691.

Ransdell, S. and Fischler, I. (1987) Memory in a monolingual mode: When are bilinguals at a disadvantage? *Journal of Memory and Language* 26, 392–405.

Salvatierra, J.L. and Rosselli, M. (2010) The effect of bilingualism and age on inhibitory control. *International Journal of Bilingualism* 15 (1), 26–37.

Tanaka, K, de Quadros, A., Santos, R., Stella, F., Gobbi, L. and Gobbi, S. (2009) Benefits of physical exercise on executive functions in older people with Parkinson's disease. *Brain & Cognition* 69, 2, 435–441.

Valian, V. (2015) Bilingualism and cognition. *Bilingualism: Language and Cognition* 18 (1), 3–24.

Verspoor, M., Lowie, W. and de Bot, K. (2011) *A Dynamic Approach to Second Language Development: Methods and Techniques.* Amsterdam: John Benjamins.

3   The Multilingual Brain: Implications for the Future

Valéria Csépe

Introduction

A large part of our current knowledge about cognitive brain functions results from studies of the modular approach which classifies brain areas as independent units responsible for complex cognitive functions. However, brain imaging studies of the last decade have shed light on the limitations of this approach and especially on the general conclusions based on modular explanations, showing that a re-evaluation of the role and function of brain areas including those seen as pinnacles of modularity (e.g. primary sensory areas) is needed. Therefore, the era of simplistic views about mapping cognitive constructs onto individual brain areas is over. New insights into cognitive functions including language are provided by studying the conjoint functions of brain areas working within the concerted action of large-scale networks. A great step towards a high-complexity approach results from recent investigations into sophisticated methods.

The concept of brain networks that has been rediscovered by recent neuroscience has emerged from research data on the structural connectivity and functional interdependence of specialized networks formed by the anatomical linkage of the corresponding neurons. There are various neuronal populations with internal circuitry configurations throughout the brain, and some of these act as network nodes. Network nodes are defined as neuronal assemblies of local structural organization with large-scale structural connectivity and local functional activity different from those of their neighbours. The functional interdependence of brain network nodes refers to joint activity in different brain structures that are co-dependent according to their functional and/or behavioural parameters. Nodes of large-scale functional networks are described by functional neuroimaging studies using the functional magnetic resonance imaging (fMRI) blood oxygen level dependent (BOLD) signal, by relating the joint activation of brain areas to

different cognitive functions. More than two decades ago, Mesulam (1990) described the brain network concept, as well as providing a thorough analysis of anatomically distinct large-scale brain networks responsible for different cognitive functions. However, it is still an open question as to how cognitive functions and especially languages are served by large-scale brain networks. Perhaps an even more pertinent question is whether neuroscientific investigations should contribute to debates about particular questions of multilingualism (e.g. the bilingual advantage (BA), see below and Chapters 2 and 4) or if they might better use highly sensitive and holistic approaches in order to help to understand the real nature of multilingual performance. The majority of neuroscientific studies on multilingual performance rely on measuring a fraction of the functions, predominantly focusing on two of the major core networks, i.e. (1) the language network anchored in Wernicke's and Broca's areas, and (2) the working memory–executive functions network anchored in the prefrontal and inferior parietal cortices. However, the neural activity associated with different stages of linguistic processing is characterized by spatial and temporal overlap as well as by high complexity and variability due to different combinations of the task-efficient networks recruited.

This chapter therefore focuses on questions about multilingualism that are answered by brain research methods, which have typically produced more questions than answers in recent decades. One of the crucial issues addressed here is how the neural networks recruited especially by language show (domain-specific) change in respect of those who speak two or more languages and whether these changes are more structural or more functional in nature. Moreover, as has been assumed by several researchers, the linguistic processes required by two or more languages rely heavily on a domain-general network known as the executive function network, the changes in which may contribute to the improvement of the various types of these functions required by cognitive tasks. This issue will be discussed in detail, not only because the so-called bilingual advantage is one of the most frequently debated issues, but particularly because of the general tendency to underestimate the limitations of interpreting and generalizing neuroscientific data. Finally, the linguistic processes that contribute to the proficient use of two or more languages and that are seemingly as important as code switching will be discussed, in order to highlight the third aspect of the challenges cognitive neuroscience faces in aiming to understand the multilingual brain. For this a thorough – although not exhaustive – review of the relevant benchmark, as well as recent, corresponding, contradictory and critical studies on bilingualism will be discussed. One of this review's objectives is to shed light on the interpretational differences in neuroscientific data acquired by different methods and measured on different populations that are often neither balanced for languages nor controlled for contributory cognitive factors.

Altered Brain Networks for Language

Despite all the efforts of psycholinguistic and neuroscientific research into how more than one language is acquired, represented and controlled, the true nature of bilingual language use is still an issue for debate. Two of the most studied topics are language representation and language control, as part of the process of resolving lexical competition and the selection of the appropriate language of communication.

In the following, relevant studies that address naming as one of the performance measures associated with proficiency will be briefly reviewed. Using a whole-brain mapping technique (MRI and diffusion tensor imaging (DTI)), Lee and colleagues (2007) showed that vocabulary knowledge in monolingual adolescents is positively correlated with grey matter 'density' of the bilateral posterior supramarginal gyri. The effect was specific to the number of words learned, regardless of verbal fluency or other cognitive abilities. In a prior investigation performed by Andrea Mechelli et al. (2004), seen as a benchmark study, increased grey matter density was found in Italian-English bilinguals when compared to monolingual English speakers. The study tested grey and white matter density difference between: 25 monolinguals with little or no exposure to a second language; 25 early bilinguals acquiring a second European language before the age of five years and practising it regularly ever since; and 33 late bilinguals learning a second European language between 10 and 15 years of age and practising it regularly for at least five years. Twenty-two Italians who learned English as second language between the ages of two and 34 years were also recruited for further testing. Second-language proficiency (reading, writing, speech comprehension) was estimated for each subject with a battery of standardized neuropsychological tests using principal component analysis, and the grey and white matter voxel-based morphometry was compared as a function of age at acquisition. The authors, reporting their findings in a brief communication to *Nature*, demonstrated that 'learning a second language increases the density of grey matter in the left inferior parietal cortex and that the degree of structural reorganization in this region is modulated by the proficiency attained and the age at acquisition'. In a replica study by Richardson et al. (2010), participants of a very large age range (7–75 years) were tested. Their findings were similar to those of Mechelli et al. (2004) as well as to those of Lee et al. (2007), in that a positive association between grey matter probability in the posterior supramarginal gyri and vocabulary knowledge was found. Moreover, the Richardson et al. data shed light on more dynamic structural changes related to bilingualism than was expected by the previous studies.

While the cortical regions identified as the main areas of proficiency-related dynamics showed significant changes in adolescents only, the grey matter density of two left posterior temporal regions (supratemporal sulcus

(STS) and temporo-parietal junction) showed a positive correlation with vocabulary knowledge across the lifespan. The authors suggested that changes found in the adolescents investigated arose from typical modes of learning in formal education, whereas changes across the lifespan might arise from linking semantic and syntactic information to prior knowledge due to day-to-day exposure to language (Richardson *et al.*, 2010). Although this general interpretation is not supported by any study available in the literature, recent research by Bellander *et al.* (2016) showed that 10 weeks of vocabulary learning led to an increase in grey matter volume in the right hippocampus. However, this association was not significant when baseline memory measures were introduced in the model, revealing that short-term memory performance was the key component that predicted the change found.

In their early study on brain structural changes, Mechelli *et al.* (2004) also found a higher probability of more grey matter in the left inferior parietal cortex of bilinguals compared to monolinguals, subsequently interpreted as the consequence of greater experience or expertise with a second language leading to structural reorganization in this brain region. This region is in close vicinity to the region found to predict vocabulary knowledge in the two studies described above (Lee *et al.*, 2007; Richardson *et al.*, 2010). However, Stein and colleagues (2012) did not find any changes in this brain region when examining the structural plasticity of native English speakers who had moved to Switzerland to learn German. In order to follow any changes in the targeted brain areas, all participants were MRI scanned upon their arrival in the country and five months later. The grey matter volume difference between the first and second measures was significant in the left inferior frontal gyrus (IFG) and the left anterior temporal lobe. This change showed positive correlation with the participants' second language proficiency, although with marked individual variation. However, no change could be found in the left inferior parietal cortex and this was explained as being the effect of the short duration of language exposure. However, individual differences as well as the time factor in second language experience rarely feature as experimental variables in brain studies. Moreover, studies on the effects of formal education and lifespan changes in vocabulary have not been investigated systematically in multilingualism brain research, although ageing as a modulating factor should be taken into account for the latter. It is important to note that lifespan changes in bilingual vocabulary are still not focused on in bilingualism studies. Instead, the neural regions presumed to account for the bilingual benefit have been studied, recently in ageing populations.

Abutalebi *et al.* (2015) aimed to discover the neural background of the postponed cognitive decline in bilinguals shown by previous behavioural studies (Alladi *et al.*, 2013; Craik *et al.*, 2010; Schweitzer *et al.*, 2012). They scanned aging bilinguals and matched monolinguals from Hong Kong and found increased grey matter in the inferior parietal lobules of the bilinguals alone. The authors interpreted this finding as being the neuroprotective effect of bilingualism. However, unlike in younger adults, age of L2 acquisition did not predict the grey matter volumes of the elderly. Furthermore, as grey matter changes may occur in those subcortical structures that play a very important role in a number of language functions, including the monitoring of speech production and language control, the neural dynamics attributed to bilingualism need further investigation of the structural and functional changes of these regions as well. A new study by Burgaleta *et al.* (2016) revealed significantly expanded subcortical structures for bilinguals when compared to monolinguals. The subcortical grey matter changes were found in the bilateral putamen and thalamus, as well as in the left globus pallidus and right caudate nucleus. As the authors suggested, a 'more complex phonological system in bilinguals might lead to greater development of the subcortical brain network involved in monitoring articulatory processes' (Burgaleta *et al.*, 2016: 437).

In general, more complex and therefore more reliable conclusions about brain and behavioural studies could be drawn from systematic investigations if multidisciplinary approaches and a broad repertoire of methods were applied. Moreover, as the effect of proficiency is one of the main factors investigated, a thorough and deep analysis of the contribution of proficiency-related factors (age, method and intensity of second language acquisition (SLA), years and frequency of daily use, etc.) to performance (very often picture naming only) is also needed. To our best knowledge there is no neuroscientific investigation into the effect of differences of languages spoken by bi- and multilinguals in terms of grammar, e.g. the surface structure of languages acquired and used. Although systematic investigations of these factors are important for the behavioural studies as well, a standard and reliable set of measures of proficiency is even more crucial when structural brain changes are investigated in order to understand how the function network of language changes with multilingual use.

Traditional mapping studies show that changes found in bilinguals involve many regions of the core network, including one of the subsystems responsible for vocabulary operations. It has been commonly shown that structural changes occur in the left and right posterior supramarginal gyri, in the posterior STS and in the temporo-parietal cortex. Although these areas show differences between bilinguals and monolinguals in most of the studies, this mapping-type approach does not provide sufficient information on the core network assumed to be different in bi- and multilinguals as compared to monolinguals. A recent study by García-Pentón *et al.* (2014) compared Spanish monolinguals and Spanish-Basque early bilinguals, revealing two structural subnetworks connected by white matter (WM) tracts to a larger extent in bilinguals than in monolinguals. One of the subnetworks found comprised the left frontal and the parietal-temporal areas, and one was composed of the left occipital, the left parietal-temporal regions and the

right superior frontal gyrus. Although all these regions have already been described by previous studies (for example, Golestani & Pallier, 2007; Luk *et al.*, 2011; Mohades *et al.*, 2012) as being related to language processing and monitoring, the authors interpreted their findings in relation to its complexity and not just in terms of the changes in a number of areas. They proposed that the efficient bilingual communication of the participants investigated might be served by specialized language subnetworks. The authors' conclusion and proposal for subnetworks developed in bilinguals was based on a complex network analysis (a graph theoretical approach to examine global and local graph network efficiencies), which revealed a higher capability of the subnetworks for between-nodes information transfer in bilinguals than in monolinguals.

The functional network approach used by García-Pentón and her colleagues (2014) led to a more complex interpretation than the one provided by simple mapping of the language performance into brain regions. From their point of view, an early acquisition of two languages might be a prerequisite for developing specialized structural subnetworks within the core network characterized by increased connection density and graph-efficient information flow. An intriguing finding of the García-Pentón group was that the assumed emergence of these subnetworks was associated with less efficient between-nodes parallel information transfer in the whole brain. Therefore, one may ask whether the tradeoff is a less optimal configuration of the general network and, if the answer is yes, what kind of disadvantages bilinguals face and how we interpret the phenomenon known as the 'bilingual advantage' (BA).

The BA is a general term relating to the performance changes found in bilinguals in the domain-general system of executive functions, a highly debated concept that has given a large impetus to many neuroscientific studies of the last decade (see Chapters 2 and 4). The concept is especially interesting because the domain-specific system, e.g. the language core system, consists of the subnetworks involved in resolving phonological, syntactic and semantic interference between languages as well as in contributing to word recognition, reading and semantic processing. Although the language core network changes through the use of more than one language with sufficient proficiency and may modulate some functions of the domain-general system of executive functions, the extent of these changes and the neural mechanism underlying them are not very well studied, i.e. are highly controversial. Here we would like to stress again that language performance is not exceptional in terms of the contribution of executive functions, since all cognitive task performances require planning, controlling, monitoring and error correction, just to mention some of the most important executive functions. Therefore, as shown by empirical data, subnetworks of the language core functional network involve – either through overlaps with or links to – frontal regions playing an important role in the bilingual performance of increased language demands.

Extended Network of Executive Functions

One of the most intriguing findings in the area of bilingualism is the observation that bilinguals outperform monolinguals across a number of tasks that measure non-linguistic executive functions. The linguistic awareness as well as the metalinguistic development reported in bilinguals (Bialystok, 1987, 1988) has led to the concept of the BA mentioned above. The better performance of bilinguals with respect to several executive functions has been attributed to the cognitive demands and attentional control required by speaking two languages, and was originally proposed by Ellen Bialystok and Kenji Hakuta in 1994. Their book about the developmental aspects of the 'bilingual mind' – as well as Bialystok's subsequent publication in 1999 – first evoked enthusiasm followed by intensive debates, and gave a big impetus to empirical studies using behavioural and/or neuroscientific methods. The most parsimonious explanation behind the concept was that acquisition as well as performance in more than one language not only changes the neural network of language, but general executive functions will also be affected. The neuroscientific studies searching for the brain correlates of the BA found in the behavioural experiments produced a large amount of data with a very similar interpretation. Changes assumed to occur in the executive function system via bilingual practice were presumed to have a general effect on performance in non-linguistic tasks as well (for a review, see Bialystok *et al.*, 2012).

The last decade has seen an increasing number of publications questioning the general advantage effect. Stocco *et al.* (2012), for example, preferred the idea of cognitive advantage, yet they proposed a slightly different explanation. According to their interpretation, a common pathway is shared by all kinds of task switching, i.e. the neural pathway is common and the executive functions' system per se does not undergo significant changes in bilinguals. According to their suggestion, the possible area to reach via conditional routing for linguistic signals is the frontal cortex reached through the basal ganglia. This concept is based on a thorough review of the brain research data in several studies (for details, see Stocco *et al.*, 2012); the main conclusion drawn is based on brain research data revealing a rich connectivity of the frontal regions and basal ganglia as well as the fronto-striatal loops shared by executive functions in non-linguistic (e.g. set-shifting) and linguistic (e.g. language switching) tasks in bilinguals.

However, we have to bear in mind that most of the language switching studies, including those reviewed by Stocco *et al.* (2012), use different methods compared to those of the mainstream executive function studies. The most frequently used experimental paradigms of language switching are picture or digit naming tasks, where participants name the items in two consecutive trials in the same language (non-switch trials) or in different

languages (switch trials). The difference in naming latencies between switch and non-switch trials is referred to as the *switch cost*. Meuter and Allport (1999) are among the first researchers to conduct switching experiments in bilinguals performing digit naming tasks where colour serves as a cue in either the participants' L1 or L2. The authors, as expected, found longer naming latencies in switch trials than in non-switch trials. The effect found was asymmetrical, as switching to the L1 required more time than switching to the L2. It has to be mentioned, however, that all participants in this study were unbalanced bilinguals so that L1–L2 balance – an important factor of naming latency – could not be taken into account. The authors interpreted the larger L1 switch cost as the consequence of overcoming the L1, an effect seen by many authors as the empirical evidence of strong inhibition control.

The significant role that inhibition control plays in language switching (see below) is further supported by event-related brain potential (ERP) studies. Jackson et al. (2001), for example, found an asymmetrical behavioural effect, i.e. switching to the L1 was slower than switching to the L2, accompanied by a significant amplitude increase of the N2 ERP component. Such a change is usually elicited by stimuli requiring no-go response as compared to those of go-response in go/no-go tasks. The N2, an ERP wave observable as negative shift over the fronto-central sites with 250–350 msec latency after stimulus onset, has been related either to response inhibition (e.g. Jodo & Kayama, 1992; Thorpe et al., 1996) or to response conflict monitoring, but not to inhibition (e.g. Donkers & van Boxtel, 2004; Nieuwenhuis et al., 2003). The ERP changes found were interpreted as the result of response conflict monitoring influenced by switch-related modulation, i.e. an access and not a response inhibition. Interestingly, the N2 changes found by Jackson et al. (2001) were significant only when the participant switched to their L2 (requiring more inhibition) and not when switching to their L1 (requiring less inhibition). However, the general view of the possible cortical generators of the N2 is that its changes reflect the inhibition intensity, so that the ERP findings are not in contradiction to the reaction time measured as overt behaviour.

One of the most influential models, the inhibitory control (IC) model, assumes inhibition as a crucial mechanism of non-specific selection. The IC model proposed by Green (1998) suggests that language task schemas external to the bilingual lexico-semantic system compete to control the output, so that all active lemmas corresponding to the target language are inhibited. Therefore, when the weaker language (i.e. the L2) is in use, the dominant language's (L1) task schema is suppressed, and the L1 lemmas undergo active inhibition in the bilingual lexico-semantic system. There are, however, several authors who argue against the inhibition concept and question it as a source of the switching cost measured as the response latency of naming produced by bi- and multilinguals. Costa and Santesteban (2004), for

example, could replicate the findings of several previous studies, showing that asymmetrical switch costs can be measured in unbalanced Catalan-Spanish bilinguals. However, in balanced trilinguals with equal proficiency in Spanish and Catalan, the switch cost was symmetrical for the L1 and L2. Therefore, Costa and Santesteban (2004) interpreted their data in terms of the L1–L2 proficiency balance and suggested that inhibition might play a role only in unbalanced bilinguals. This would mean that inhibition is not used in balanced bilinguals, which would explain the lack of switch costs. Moreover, the inhibition concept was not only questioned in the study of unbalanced bilinguals by Gollan and Ferreira (2009), who pointed to an important problem when showing that symmetrical switch costs could be measured in bilinguals only when switched voluntarily between languages, also pointed out in the study by Verhoef et al. (2009). The latter found both symmetrical and asymmetrical switch costs in unbalanced bilinguals whose response depended on the stimulus preparation time manipulated in the experiment. Changes of the N2 ERP component elicited by pictures correlated well with the switch costs as well as with the naming latency of the same pictures measured in separate sessions. While a short inter-stimulus interval between the pictures and their linguistic cues resulted in asymmetrical switch costs, longer intervals produced symmetrical costs and larger N2 amplitude.

One may raise the question of whether inhibition is the main mechanism accounting for switch costs and how well it explains the observed switch cost symmetry differences. In an ERP study by Christoffels et al. (2007), behavioural and ERP correlates of switch costs were measured in a bilingual picture naming task applied in blocked- and mixed-language contexts. Moreover, the authors manipulated form similarity of the translation equivalents (cognate status), and in the blocked context all pictures were named in German or in Dutch. Although naming latencies were faster for the L1 than for the L2, the mixed language context resulted in different naming performance. In the L1 two ERP components' amplitude (peaking at around 400 and 500 msec) was found to be significantly larger, and the response latency was also longer. Furthermore, the cognate facilitation effect found in both conditions was explained as the effect of activation of the non-response language. Although the ERP effect associated with the cognate status was small, the non-response language activation was assumed to modulate the phonological level. Therefore, it is logical to ask how bilinguals prevent intrusions from the non-response language and whether the lexical selection is language specific. As Christoffels and colleagues suggested (Christoffels et al., 2007; see also Acheson et al., 2012), the sustained and transient components of language control might be related to different brain processes. It is thus crucial to distinguish between them in order to understand the real nature of bilingual performance. Moreover, it seems that the reactive inhibition indicated by asymmetry in switch is not easily

replicated, and the language switch costs found in bilingual naming depend on many factors. This also means that the switching costs theories should be revisited, especially because several experimental studies available do not favour or do not directly support the active inhibition theory. As mentioned above, Costa and Santesteban (2004) proposed a different explanation for not finding asymmetric switching costs for highly proficient bilinguals by pointing to the development of *language-specific selection mechanisms*. Costa *et al.* (2006) extended their proposal on this mechanism, successfully demonstrating symmetric switching costs for highly proficient bilinguals in older age and in trilinguals for L2 and L3. Again, these data did not square with the IC model.

However, it is not only the brain mechanisms assumed to contribute to switching that are problematic. There are several behavioural as well as neuroscientific language switching studies that include methodical problems as well as misinterpretations and misapplications. Therefore, it is a timely question to ask: (1) how proficiency is judged in different studies, (2) how bilingualism is defined; (3) how broad the repertoire of measures used is; and especially (4) how adequate the neuroscientific methods used are to draw general conclusions on bilingual performance and representation. Moreover, there are alternative concepts, such as, for example, the so-called language strength.

Although language strength is a frequently used term in behavioural studies, it is rarely defined or described as a combined factor of spoken and written proficiency as well as of lexical robustness. Schwieter and Sundemann (2008, 2009) defined lexical robustness as estimated global proficiency that incorporates vocabulary knowledge and fluency and refers to the relationship between words and the concepts they represent. These authors calculated the total lexical robustness scores by adding the responses given by bilingual participants in semantic and phonological fluency tasks. Their regression analysis results were evaluated in combination with the picture-naming data. Schwieter and Sunderman (2011) then further addressed the role of language strength in language switching by investigating the same speakers of two non-native languages of different strengths. The L1 switch costs were expressed by longer reaction time latency in comparison with L2 or L3, while the response latency was longer for L2 than for L3. However, these findings were in contradiction to the results of Costa *et al.* (2006), among others, who reported mixed findings. While most of the behavioural studies are interested in the organization of the bilingual lexicon, a few of them focus on the cognitive mechanisms of lexical selection. Moreover, the vast majority of neuroscientific studies aim to investigate the neural correlates of lexical selection, especially because it can be better operationalized for the brain measures available.

It seems that some, yet by far not all studies have found symmetrical switch costs, which means that there is no clear evidence supporting the

hypothesis of active inhibition. On the contrary, many behavioural studies obtained asymmetrical switch costs, although bilingual balance, context, preparation time, L2 exposure, age of acquisition and many other factors varied to a great extent. Moreover, some studies have challenged the involvement of inhibitory mechanisms underlying language switching and questioned the strong role of inhibition (cf. Runnqvist *et al.*, 2012).

In sum, the behavioural and electrophysiological studies discussed above do not provide unequivocal evidence for the IC model. However, that is not to say that active inhibition does not play any role at all in language switching. In the following, we consider whether the fMRI studies on bi- and multilingualism help our understanding by investigating the functional networks active during language switching. We will first discuss the literature on the brain areas involved in inhibition in general and then provide an overview of fMRI studies on language switching in particular.

As mentioned above, inhibition was assumed to play a significant role in code switching as a domain-general (i.e. not domain-specific in modular terms) mechanism in the control of linguistic performance and to overcome the unwanted responses. However, inhibition is only one of the domain-general executive function mechanisms contributing to measurable differences in motor responses. Jahfari and collaborators (2010, 2011) found a strong involvement of the cortical – IFG, supplementary motor area (pre-SMA) and subcortical (subthalamic nucleus (STN)) – regions during response inhibition. Jahfari *et al.* (2011) investigated the effective connectivity patterns recorded in a combined Simon task (faster response to stimuli occurring at the required response's location) and stop-signal (performance difference between executing and withholding the planned response) task and described two important pathways: the hyper-direct (fronto-subthalamic) and the indirect (fronto-striatal-pallidal) pathways, connecting the frontal regions with the basal ganglia. These pathways are the same as the ones that were described by Stocco *et al.* (2012) as contributing to conditional routing for linguistic signals in bilinguals. The common mechanism assumed to participate in linguistic and non-linguistic performance is response inhibition; the naming response is inhibited in the first and the simple motor response in the second. Moreover, Jahfari *et al.* (2011) MRI-scanned their participants, who responded to coloured shapes by pressing either a left or a right button matching or mismatching the spatial location of the stimulus, e.g. during the performance of a Simon task. An acoustic stop-signal to inhibit the response was given in one-third of the trials. The activity changes elicited by the stop signals could best be explained by the contribution of a right-lateralized network including the hyper-direct pathway of inhibition. However, it is rather unclear which aspects of the IC model were required by the Simon task and whether other mechanisms could also interact with the performance.

All these results, however, do not necessarily mean that language switching per se shapes the executive function network. Although this cannot be

excluded, a reliable answer needs further fMRI studies, especially those that directly address the question of whether switching is controlled by the brain regions related to domain-general inhibition. Therefore, one may ask why the majority of language switching studies focused for a long time on the executive functions only and did not pay attention to other domain-general functions or to the activity of a broader cortical network. Although Abutalebi and Green (2007) proposed that language switching required the activity of a core functional network including the prefrontal cortex, anterior cingulate cortex (ACC), caudate nucleus, and the supramarginal gyrus, we have to bear in mind that this network is not specifically allocated to language switching. The ACC, for example, is involved in error detection, conflict monitoring and conflict resolution, while the basal ganglia is activated in motor control and planning as well as in various tasks requiring executive control. Therefore, the latest study by the Abutalebi group (see, for example, Branzi *et al.*, 2016) addressed the role of cognitive control in bilinguals performing different tasks expecting the contribution of distinct mechanisms. The authors' main question was whether the core network of executive functions was differently engaged in non-linguistic and linguistic tasks. Again, a switching task with a blocked design was carried out. Bilingual participants were investigated by using event-related fMRI. A thorough analysis of the neuroimaging data gave a promising answer to the question as to whether prefrontal cortical activity was similar in the linguistic and non-linguistic tasks. While the left prefrontal cortex was active both in linguistic and non-linguistic tasks, suggesting its domain-general role in response selection, the dorsal anterior cingulate cortex (dACC) and pre-supplementary motor area (pre-SMA) were recruited by monitoring demands of the second language. Moreover, the prefrontal region, the inferior parietal areas as well as the caudate were active in response selection. These data suggest a functional differentiation of the bilinguals' language control network composed by a response selection system of bilateral frontal areas and a monitoring control system residing in the d ACC and the pre-SMA (part of the supervisory attentional system).

The authors proposed that language control is hierarchically organized with the dorsal subsystem (ACC/pre-SMA) *supervisory attentional system* recruited for the increased monitoring demanded by the second language. Moreover, the prefrontal and inferior parietal areas as well as the caudate, supposed to be part of the *response selection system*, are active during linguistic tasks (among others). Although this proposal is in agreement with the functional network approach, it further elaborates on the previous assumptions that language switching does not primarily engage the language system. While the Branzi *et al.* study (2016) further contributed to our understanding of the role of domain-general executive control functions (Abutalebi & Green, 2008), the recent hot debate about the effect of bilingualism on executive functions was not addressed.

In a meta-analysis of fMRI studies on language switching, Luk *et al.* (2012) reported eight areas related to the executive control of language processing. These were the left IFG, left middle temporal gyrus (MTG), left middle frontal gyrus (MFG), right precentral gyrus, right superior temporal gyrus (STG), pre-SMA and bilateral caudate nuclei. Moreover, the dorsolateral prefrontal cortex (DLPFC) showed increased activation in picture naming tasks of mixed and blocked design performed by early Spanish-English bilinguals (Hernandez *et al.*, 2000). Along similar lines, Wang *et al.* (2007) and Hosoda *et al.* (2013) reported that switching between two languages required different activation of the functional network; switching to the L2 was associated with higher activation of the right DLPFC, the left STG, the ACC, the left IFG and the left caudate nucleus as compared with switching from L2 to L1. Abutalebi *et al.* (2013) also found activation differences in the left caudate nucleus, the pre-SMA and the ACC in trilingual language switching as compared to non-switching. However, only the caudate nucleus activation correlated with language proficiency.

While most of the language studies on switching costs used mixed contexts, Guo *et al.* (2011) as well as Branzi *et al.* (2016) measured brain activity in local versus global inhibition contexts. Local inhibition meant controlling a restricted set of memory items, e.g. specific lexical items as tested in blocked- and mixed-language naming; global inhibition was examined by comparing the order (L1 after L2 versus L2 after L1) of languages within blocked naming. Although the two studies found the same network to be active, Guo *et al.* (2011) explained the differences found exclusively by inhibition mechanisms without referring to the executive-attentional system. However, neither of these studies addressed directly how the networks proposed interacted or whether distinct subnetworks in concerted action contributed to changes that might explain the assumed BA.

Bilingual Advantage Revisited

Before going into discussion of the ongoing debate on the BA we want to stress that speaking more than one language has several advantages. This means that being bilingual is an advantage, although not in the original meaning of the term 'bilingual advantage' proposed by Bialystok in 1999. We instead suggest developing a new model of bilingual performance including the delicate network of executive functions revealed by recent behavioural and neuroscientific studies. In this we follow the theoretical framework by Miyake and Friedman (2012), who defined three components of the executive functions, i.e. updating, shifting and inhibiting. Executive functions here mean a set of general-purpose control processes that regulate one's thoughts and behaviours. However, any target executive function is embedded within

a specific context, so that all measures include systematic variance attributable to non-executive processes. This is especially valid for linguistic tasks where the systematic variance included is attributable the different non-executive processes. This, the so called task-impurity problem, is solved in the cognitive studies by using the latent variable approach. Here multiple exemplar tasks sharing little systemic variance of the executive functions are chosen and the latent variant can be statistically verified by different methods (confirmatory factor analysis, structural equation modelling). This approach was used in our behavioural and ERP studies on different aspects of the executive functions in non-linguistic tasks (Kóbor et al., 2014, 2015). Unfortunately, the linguistic studies executed in order to show how bilingualism contributes to enhanced executive functions rarely use this approach. This means that the BA studies suffer from significantly more methodical problems, including the statistical methods used, than are mentioned and discussed in the critical paper by Paap et al. (2015).

The widely accepted view that bilinguals are better than monolinguals at non-linguistic tasks goes back to Bialystok and Hakuta (1999). Along these lines, Bialystok (2011) also suggested that coordination of the executive functions develops better in bilingual children because language switching enhances these mechanisms. Recently, however, more and more authors have challenged the interpretation of language group differences and argue that many findings are indicative of a bilingual disadvantage rather than advantage. Paap et al. (2015) question the BA in executive functions and heavily criticize the widely used behavioural tests in studying the effect of the BA on general-purpose executive functions. One of their strong arguments is that the proportion of positive results is relatively small, and the significant differences found in performance by many authors may be due to questionable statistics. Moreover, as was discussed above in this chapter, brain imaging studies have not so far been able to provide any strong support for the BA hypothesis. Although it could be demonstrated that the neural correlates of task performance in executive function tasks differed between bilinguals and monolinguals, the interpretation of these results as neural reorganization of the executive function networks is contradictory. One of the frequent problems discussed by Paap et al. (2015) is that the results of many behavioural studies could not be replicated. Moreover, there is no compelling evidence provided by the neuroscientific investigations. The alignment problem of behavioural and neural data is present in a large number of studies reporting on Simon, flanker and switching tasks. Paap et al. (2015) drew the attention of researchers working with ERPs to the misinterpretation of changed component amplitudes (see above). As the authors stated, a BA in executive functions probably did not exist, so that the neuroscientific studies better study conflict monitoring and turn to a more complex executive function model like that of Miyake and Friedman (2012).

Beyond Code Switching

Code switching is measured by using various tasks requiring language production. However, as several studies show, the L1's acoustic attributes and their representation contributes to robust changes in speech perception and influences word recognition (for a review, see Sebastián-Gallés, 2005). This means the 'L1-based processing' can be characterized by interference or transfer from L1 to L2 during non-native, second or third language processing. Therefore, successful L2 speech processing and word recognition should rely on the representation qualities of the L2 phonological system. Unfortunately, only a few proficiency studies focus on the L1–L2 interaction of phonology. According to the proposal of Munson et al. (2005), phonological knowledge emerges with lexical growth and improvement of phonology affects further acquisition of the vocabulary. Converging data, however, are available only within and not across languages. Empirical data (e.g. Bundgaard-Nielsen et al., 2011; Majerus et al., 2008) on the phonology–vocabulary relationship speak for more efficient derivations of phonological regularities in individuals of larger L2 vocabulary.

Several studies examining the behavioural and brain correlates of phonetic perception revealed large individual differences in detecting or discriminating non-native phonetic contrasts (see, for example, Golestani, 2014). Moreover, as was shown by previous structural and functional brain imaging studies (see, for example, Golestani et al., 2007), learned non-native contrasts are processed by the bilateral auditory cortices and the left IFG. However, a better discrimination performance of the non-native contrasts was associated with lower activation of the left IFG (Golestani & Zatorre, 2004), a region belonging to the domain-general working memory–executive functions core network. As was shown by further neuroscientific investigations (dual-site transcranial magnetic stimulation (TMS)), another portion (posterior) of the IFG, the Brodmann's area (BA) 44 as well as the left supramarginal gyrus were always active during phonological processing and subvocal rehearsal (Hartwigsen et al., 2010). Furthermore, structural changes associated with faster learning of the non-native contrast showed differences in the Heschl gyrus and the parietal cortex as compared to slower learners (Golestani et al., 2007).

While several studies were performed to investigate adult performance as well as developmental changes in processing segmental speech cues, only a few studies focused on the prosodic perception in bilinguals as compared to monolinguals. From a theoretical and practical point of view it is very important to know whether a better stress pattern discrimination develops in bilinguals and how this contributes to dominant stress contrasts. A significant body of behavioural studies attempted to understand how a bilingual environment affected the perceptual processing of phonological (segmental

and prosodic) information in newborns and young infants. While several ERP studies were performed in order to follow the development of segmental processing, less attention was paid to the suprasegmental. Bijeljac-Babic and colleagues (2012) investigated French-learning 10 month olds raised in bilingual as opposed to monolingual environments. The authors used the classic version of the Headturn Preference Procedure (HPP) and found an increased sensitivity in the bilingual infants to stimuli that were different from those heard in their bilingual environment. The number of ERP studies using the difference-sensitive mismatch response (MMR) to word stress is very low in general, and particularly low in infants. In all word stress MMR studies, monolingual infants were investigated and familiar patterns with non-familiar (Friedrich et al., 2004, 2009) or legal ones with illegal (Ragó et al., 2014) were contrasted. The term legal and illegal refers to the single rule, e.g. fixed stress on the initial syllable, of the language investigated (Hungarian). The only study, available as a poster presentation (Neophytou et al., 2015), showed that bilinguals exposed to similar but not identical stress patterns, e.g. Welsh and English, showed an unclear MMR pattern. The results suggested that although both monolingual and bilingual infants developed segmentation abilities at a similar time, the underlying processes might differ.

Future Challenges

It is clear that bilingualism affects the brain regions that mediate language abilities. Our understanding would depend on the theoretical frame of the identification of domain-specific (e.g. language-related) functions and domain-general (e.g. executive) functions. Effects of the bilingual experience on the brain networks should be taken into account both for perception and production. The combination of behavioural and neuroscientific methods would play the lead role in discovering the interactions assumed and the contribution of the executive functions proposed. For this approach, better selected tasks and experimental and statistical methods as well as studies aiming to investigate the replicability of the results published are needed.

Acknowledgements

References

Abutalebi, J. and Green, D. (2007) Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics* 20, 242–275.

Abutalebi, J. and Green, D. (2008) Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes* 23, 557–582

Abutalebi, J., Della Rosa, P., Castro Gonzaga, A., Keim, R., Costa, A. and Perani, D. (2013) The role of the left putamen in multilingual language production. *Brain and Language* 125, 307–315.

Abutalebi, J., Guidi, L., Borsa, V., Canini, M., DellaRosa, P.A., Parris, B.A. and Weekes, B.S. (2015) Bilingualism provides a neural reserve for aging populations. *Neuropsychologia* 69, 201–210.

Acheson, D., Ganushchak, L., Christoffels, I. and Hagoort, P. (2012) Conflict monitoring in speech production: Physiological evidence from bilingual picture naming. *Brain and Language* 123 (2), 131–136.

Alladi, S., Bak, T.H., Duggirala, V., et al. (2013) Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology* 81 (22), 1938–1944.

Bellander, M., Berggren, R., Mårtensson, J., et al. (2016) Behavioral correlates of changes in hippocampal gray matter structure during acquisition of foreign vocabulary. *Neuroimage* 131, 205–213.

Bialystok, E. (1987) Influences of bilingualism on metalinguistic development. Second Language Research 3, 154–166.

Bialystok, E. (1988) Levels of bilingualism and levels of linguistic awareness. Developmental Psychology 24, 560–567.

Bialystok, E. (1999) Cognitive complexity and attentional control in the bilingual mind. Child Development 70 (3), 636–644.

Bialystok, E. (2011) Coordination of executive functions in monolingual and bilingual children. Journal of Experimental Child Psychology 110, 461–468.

Bialystok, E. and Hakuta, K. (1994) *In Other Words: The Science and Psychology of Second Language Acquisition*. New York: Basic Books.

Bialystok, E. and Hakuta, K. (1999) Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (ed.) *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah, NJ: Lawrence Erlbaum.

Bialystok, E., Craik, F.I.M. and Luk, G. (2012) Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences* 16 (4), 240–250.

Bijeljac-Babic, R., Serres, J., Höhle, B. and Nazzi, T. (2012) Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. PLoS ONE 7 (2), e30843. doi:10.1371/journal.pone.0030843.

Branzi, F., Della Rosa, P., Canini, M., Costa, A. and Abutalebi, J. (2016) Language control in bilinguals: Monitoring and response selection. *Cerebral Cortex* 26 (6), 2367–2380.

Bundgaard-Nielsen, R., Best, C. and Tyler, M. (2011) Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Journal of Applied Psycholinguistics* 32 (1), 51–67.

Burgaleta, M., Sanjuán, A., Ventura-Campos, N., Sebastian-Galles, N. and Ávila, C. (2016) Bilingualism at the core of the brain. Structural differences between bilinguals and monolinguals revealed by subcortical shape analysis. *Neuroimage* 125, 437–445.

Christoffels, I., Firk, C. and Schiller, N.O. (2007) Bilingual language control: An event-related brain potential study. *Brain Research* 1147, 192–208.

Costa, A. and Santesteban, M. (2004) Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language* 50, 491–511.

Costa, A., Santesteban, M. and Ivanova, I. (2006) How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory and Cognition* 32, 1057–1074.

Craik, F.I., Bialystok, E. and Freedman, M. (2010) Delaying the onset of Alzheimer disease: Bilingualism as a form of cognitive reserve. *Neurology* 75 (19), 1726–1729.

Schwieter, J. and Sunderman, G. (2009) Concept selection and developmental effects in bilingual speech production. *Language Learning* 59 (4), 897–927.

Schwieter, J. and Sunderman, G. (2011) Inhibitory control processes and lexical access in trilingual speech production. *Linguistic Approaches to Bilingualism* 1 (4), 391–412.

Schweitzer, T.A., Ware, J., Fischer, C.E., Craik, F.I.M. and Bialystok, E. (2012) Bilingualism as a contributor to cognitive reserve: Evidence from brain atrophy in Alzheimer's disease. *Cortex* 48 (8), 991–996.

Sebastian-Gallés, N. (2005) Cross-language speech perception. In D. Pisoni and R. Remez (eds) *The Handbook of Speech Perception* (pp. 546–566). Malden, MA: Blackwell.

Stein, M., Federspiel, A., Koenig, T., Wirth, M., Strik, W., Wiest, R. and Dierks, T. (2012) Structural plasticity in the language system related to increased second language proficiency. *Cortex* 48, 458–465.

Stocco, A., Lebiere, C., O'Reilly, R. and Anderson, J. (2012) Distinct contributions of the caudate nucleus, rostral prefrontal cortex, and parietal cortex to the execution of instructed tasks. *Cognitive, Affective and Behavioral Neuroscience* 12 (4), 611–628.

Thorpe, S., Fize, D. and Marlot, C. (1996) Speed of processing in the human visual system. *Nature* 381, 520–522.

Verhoef, K., Roelofs, A. and Chwilla, D. (2009) Electrophysiological evidence for endogenous control of attention in switching between languages in overt picture naming. *Journal of Cognitive Neuroscience* 22, 1832–1843.

Wang, Y., Xue, G., Xue, F. and Dong, Q. (2007) Neural bases of a symmetric language switching in second-language learners: An ER-fMRI study. *Neuroimage* 35, 862–870.

4   Phonological and Semantic Awareness of Bilinguals and Second Language Learners: Potential Implications for Second Language Instruction

Judit Navracsics and Gyula Sáry

Introduction

There are two major sources of information for the study of phonological and semantic processing in bilinguals: neuro-imaging data and experimental behavioural data. The aim of this study is to find out to what extent the two sources converge. While there are a number of studies on regional fMRI activation, topography and lateralization during the performance of phonological and semantic tasks (Khateb *et al.*, 2000; Perrin & Garcia-Larrea, 2003, among others), there are hardly any findings available when it comes to the processing of two languages at a time (cf. Pillai *et al.*, 2003).

In this chapter, we will analyze the phonological and semantic awareness of early and late bilinguals, employing a psychophysical technique measuring reaction time (RT) and accuracy of judgement.

Recent Neurolinguistic Findings on Processing Two Languages

As Csépe (this volume) points out, the existing neuroimaging literature on bilingual lexico-semantic representation is contradictory concerning how multiple languages are represented in the brain. Neurophysiological and neuroimaging studies support a shared cerebral representation of L1 and L2 lexicons in both early and late bilinguals (Chee *et al.*, 1999; Fabbro, 2001; Illes *et al.*, 1999)

and claim that cortical activation for the L1 and L2 is located in identical regions of the left hemisphere. At the same time, other fMRI and PET studies have found distinct neural representations for the L1 and L2 within the classical left hemisphere language regions (Kim *et al.*, 1997; Perani *et al.*, 1998). There is also evidence that although the languages may be represented in different portions of the cortex in the multilingual brain (e.g. Hervais-Adelman *et al.*, 2011), this may be a function of L2 proficiency or age of L2 acquisition (see also van Heuven, and Pfenninger & Singleton, this volume). For instance, studies that examine lexico-semantic processes in bilinguals in the less proficient language (e.g. Leonard *et al.*, 2010) show that greater activity is modulated by L2 proficiency. They also demonstrate that the L1 and L2 are not completely isolated from one another, and they both mutually interfere with and support each other (Leonard *et al.*, 2010). Kovelman *et al.* (2008) suggest that the neural processing of a bilingual person differs across the two languages, as they found differential behavioural and neural patterns in studying the sentence comprehension skills of English monolingual and English-Spanish bilingual participants. The basic difference observed was that bilinguals had a significantly greater increase in the blood oxygenation level when they processed English as compared to the English monolinguals. This led the authors to hypothesize that there might be a 'neural signature' of bilingualism, as differential activation sheds light on different components of language processing for bilinguals from those for monolinguals. Pillai *et al.* (2004) investigated the activation topography with semantic and phonological tasks and found that it was different when the tasks were performed in the second language (English) from that in the first language (Spanish).

In event-related potential (ERP) tests, interference effects can be observed in bilingual tasks. Recent evidence (e.g. Huster *et al.*, 2010) suggests that an early component may reflect response selection, while a later component may reflect inhibitory cognitive components. The later ERP component seems to be systematically greater in amplitude in bilingual than in monolingual participants (Moreno *et al.*, 2008). Moreno and her colleagues suggest that this component reflects enhanced cognitive control mechanisms related to the everyday demands of a bilingual.

With respect to lexical access, frequency of use has been reported to be a major influential factor. The more frequently a language is used, the faster the words in that language are identified as members of the language; in turn, the more frequently used language causes greater interference if it is not the target language (Ng & Wicha, 2013). In their fMRI study, Paulesu *et al.* (2000) show that in addition to frequency, the regularity and familiarity of the word, as well as the orthographic pattern of the language that the word belongs to affect brain activation. They found selective activation for reading words in English, which has a deep orthography, and for Italian, which is much more on the shallow side of the continuum of orthographic consistency; in other words, the predominant process while reading is the

letter-to-sound conversion or the grapheme–phoneme correspondence. This will be discussed in more detail in the next section.

Processing Reading

Reading involves the joint activation of orthography, phonology and semantics. The question is whether these processes are independent or co-dependent, whether they are performed sequentially or in parallel, and whether they are automatic or strategic (Rastle, 2007). There are two presumed pathways to lexical access: (1) a direct one from orthography to semantics; and (2) an indirect one which, according to the 'phonological mediation hypothesis', includes phonology (see Tan & Perfetti, 1999). Price *et al.* (1996) and Price (2000) suggest that reading frequent, regular words does not require precise phonological recoding. In contrast, phonological forms are accessed directly and automatically. Braun *et al.* (2009) carried out an ERP study to analyze the time course of visual word recognition as well as the role of phonological processing. RT data, ERPs and low-resolution brain electromagnetic tomography (LORETA) results showed phonological activation in silent reading as well, which serves as evidence for phonological processes being involved in visual word recognition.

Van Heuven and Dijkstra (2010) reviewed the existing EEG and fMRI evidence for various psycholinguistic models of bilingual word recognition (see also van Heuven, this volume). They found support for their bilingual interactive activation model (BIA+), according to which the bilingual lexicon is integrated and words are accessed in a language non-selective manner. The question arises, however, as to whether there is a difference between phonological and semantic processing.

The Aims of our Study

Models of visual word recognition help us understand how we process language. There is neuroimaging and behavioural evidence for models, but our aim is to see how the models work in test situations, and whether the theoretical presumptions are valid among the language users. In other words, we want to show how the theory behind practice may help instructors with compiling teaching materials so that they are in harmony with the natural language processing scheme.

In this study, we intend to provide possible guidance for language teachers in relation to what is easier to process for language learners: semantics or phonology. The main question at stake is how we process bilingual visual information. Do we access the meaning based on the orthographic display of the word or do we also embed phonology in the process? Do we activate the

phonological and phonetic levels together when we recognize a word? Is there a need for training the ability to make semantic categorization in the foreign language classroom? Based on our findings, we wish to suggest some implications for future SLA classroom activities in terms of teaching and learning vocabulary, always bearing in mind that there is likely to be a gap between the laboratory type of data presented here and the strategies and activities in real language teaching classes.

To be more specific, we shall investigate phonological and semantic awareness of Hungarian L1 and English L2 users. Our goal is to see whether there is a difference in the processing time and in the accuracy of written word comprehension at the phonological and semantic levels in the second language. By awareness we mean the ability to successfully apply information gained by implicit and explicit learning during language development. To this end, we constructed two tests: a 'phonological rhyming test' and a 'semantic rhyming test'. With this procedure we follow Khateb *et al.* (2000), who used a monolingual phonological rhyming test in order to analyze how their monolingual subjects recognized the rhymes in pairs of orthographically dissimilar words. They also investigated how category relations in semantically related and unrelated pairs of words were recognized.

The experiments discussed in this chapter represent a follow-up of our previous lexical decision and language decision tests, where we found that in lexical decision the participants relied on their orthographic and phonological awareness, i.e. pseudo-words were considered to be real words belonging to either of the languages, depending on their orthographic and phonological structure (see Navracsics *et al.*, 2014). Semantics seemed to play no role in word identification. This being a real surprise, we then decided to investigate the temporal and performance features of phonological and semantic awareness. To this end, we designed two rhyming tests, one with phonology and another with semantics in focus, in order to see the time and success rate differences between the processing of the two linguistic levels. As the tests are presented in two languages (Hungarian and English), the results may have special implications for the development of psycholinguistic models of language processing, in particular, reading. The findings may also prove useful for the study of bilingual visual word recognition when the two languages composing bilingualism are different with regard to the degree of orthographic or phonological consistency.

Methodology

Participants

Forty-eight healthy, right-handed adults (mean age 23 years) participated in this study (see Table 4.1). Participants reported no history of psychological

Table 4.1 Language history of the groups in the study

Group	Number of participants	Age of L2 acquisition	Manner of L2 acquisition
L1	31	Late (9 years)	School instruction
HL	17	Early (from birth)	Natural + school instruction

or neurological impairment, and all had completed or were completing tertiary education. Language history and proficiency in both languages were assessed by a detailed questionnaire that asked participants to rate learning sources and degrees of exposure to both languages, and their reading, writing and speaking abilities in both languages (Marian *et al.*, 2007). Participants gave informed, written consent to participate in the experiments.

Thirty-one of the participants are native Hungarian L1 speakers and sequential L2 English learners. They are students of an English major programme at the University of Pannonia (henceforth the L1 Group). They began learning English when they were in third grade (mean age of onset of acquisition: nine years). They will be referred to as 'late bilinguals' as opposed to the other 'early bilingual' group. Their language proficiency in both languages was at C1 level or higher as they had to prove their C1 level proficiency in English when applying for university. Another 17 participants (henceforth the HL Group) were early bilinguals, who acquired both languages before puberty.

Participants in the L1 Group, born into monolingual Hungarian families in Hungary, learned English through school instruction. Participants in the HL Group acquired both English and Hungarian in a natural way in their bilingual homes as the families they came from used their immigrant language at home. Even the ones that came from mixed marriages had acquired Hungarian from birth. However, as English became the dominant language for them, they were not happy with their Hungarian language command. In order to improve their home language they started to attend Hungarian classes at Balassi Institute in Budapest, which offers Hungarian courses to Hungarians living abroad. The institute provides scholarship programmes for expatriates, and their offspring can learn about Hungarian history and culture as well as the language.

Material

The test material was constructed for Hungarian-English bilinguals. We constructed word pairs from both languages, Hungarian and English, in order to test how bilingual visual word recognition was going on and whether there were any differences in the processing of bilingual and monolingual word pairs. Hungarian is an agglutinative language and Finno-Ugric in origin, and English is typologically analytic and of Indo-European origin. For our experiment, it is essential to stress that the two languages differ in

terms of orthographic and phonological consistency. English has a deep writing system with many different varieties of letter–sound (or rather grapheme–phoneme correspondence) combinations, whereas Hungarian is rather shallow with some elements of deep writing to serve morphological procedures, such as suffixation, conjugation, etc.

Phonological rhyming test

Forty mixed (English-Hungarian) pairs of words served as the test material, and there were 40 pairs of same language, monolingual words (20 per rial, and there were 40 pairs of same language, monolingual words (20 per language) serving as controls. The task was to decide whether the word pairs phonologically rhymed or not. In the test condition, there were eight interlingual homophones, i.e. phonologically and phonetically rhyming mixed word pairs in English and Hungarian (e.g. test pair: SIGH–SZÁJ, LOW–LÓ, SHIRT–SÖRT). The orthographic overlap between stimulus pairs was strictly controlled for their phonetic similarity. For instance, the words 'high' and 'báj' are entirely different orthographically, but very similar in pronunciation; they are minimal pairs differing only in the initial consonant 'h' and 'b'. Pairs shared:

(a) low orthographic and high phonological similarity (e.g. test pair: HIGH–BÁJ);
(b) low orthographic and low phonological similarity (e.g. WAY–VÁJ);
(c) high orthographic and low phonological similarity (e.g. VAIN–VAN);
(d) high orthographic and high phonological similarity (e.g. CAR–KÁR).

In other words, there were homophones, pseudohomophones (PsH) or phonologically non-rhyming pairs. Homophones had identical phonological/phonetic forms in both languages; PsH were the neighbours that differed in only one feature in their phonological components (e.g. SHOES–HÚZ, SHUT–CSAT); and the non-rhyming word pairs were dissimilar both in orthography and phonology (e.g. DOVE–ÓV). Following this pattern, 40 same-language controls were added to the test. Twenty of them were English and 20 were Hungarian pairs. They were constructed the same way as the test word pairs were: there were phonologically rhyming but orthographically dissimilar words in English, such as SQUEEZE–SNEEZE, TAIL–TALE. Such PsH could not be constructed in Hungarian because of its higher degree of orthographic consistency. Still, there were phonologically rhyming neighbours (e.g. PÉK–FÉK, VILLA–TRILLA) and non-rhyming pairs, such as VIASZ–PAPÍR, KÁRTYA–SZŐNYEG.

Semantic rhyming test

Forty Hungarian-English mixed word pairs served as the test material. The name of the test is metaphorical, as the aim of the test is to discover whether the participants are aware of the relationship between the meanings of the words that appear at the same time on the screen. Twenty of the word pairs were semantically related and 20 were unrelated. For the control word

pairs, 20 Hungarian and 20 English word pairs were used. Ten in each language were semantically related and 10 were non-related. All the words were checked for frequency and familiarity, and the task was to decide whether the word pairs were related or not, i.e. whether they rhymed or not.

In order to check the frequency of English words we used the Corpus of Contemporary American English (COCA), which is a 400 million word database, from which the 5000–60,000 word lists are based on the only large, genre-balanced, up-to-date corpus of American English (http://corpus.byu. edu/coca/). The 100,000-word list supplements the COCA with detailed frequency data from Corpus of Historical American English (COHA), the British National Corpus (BNC) and the Corpus of American Soap Operas (for very informal language). In COCA the dispersion of the words is given, i.e. how evenly the specific word is distributed in the corpus. All the words selected for the test are from the MRC Psycholinguistic Database, which contains 150,837 words (http://www.psych.rl.ac.uk). We chose our test material from the 2500 words of this database for which psychological measures are recorded.

For the frequency of Hungarian words we used the Hungarian National Corpus, which currently contains 187.6 million words (http://corpus.nytud. hu/mnsz/index_eng.html). In this database, rank, lemma, word category, absolute and relative occurrence, and genre are given, but there is no dispersion or familiarity list available.

The English words used in the test were included in the 5000–60,000 frequency list with dispersion between 0.92 and 0.98. Their average familiarity was 567 (minimum 393, maximum 643) on the 100–700 scale. The average frequency of words was 0.0000732. Another 40 words made up mixed pairs with Hungarian words with average familiarity of 550 (minimum 379, maximum 644). Their average frequency was 0.0000508.

The average frequency occurrence of the 40 Hungarian words (i.e. 20 Hungarian word pairs) was 0.00002. The frequency occurrence of the other 40 Hungarian words that were used for the mixed pairs with English words was 0.00001. The word pairs were made up with or without their semantic relations (a detailed description of the different kinds of meaning relations is provided in the Appendix). In the analyses of both test results, we will refer to the word pair conditions using the same abbreviations: TRC (test, i.e. mixed language rhyming correct decision), TNRC (test, i.e. mixed language non-rhyming correct decision), CRC (control, i.e. monolingual rhyming correct decision) and CNRC (control, i.e. monolingual non-rhyming correct decision). These abbreviations will be used in the figures as well.

Procedure

A custom-made software program (MATLAB, MatLab Inc.) was used for the experiments. Word pairs were presented on a white background, using

black characters (Arial, font size 14) in the middle of the screen. The viewing distance was set to be the appropriate normal viewing distance of a computer screen (~50 cm).

Participants received written instructions at the beginning of the experimental session. This ensured that every subject received the same instruction. In the phonological test, the task was to decide whether the words just seen were phonologically rhyming or not. In the semantic test, the instruction for the participants was as follows:

In the middle of the computer screen you will see a fixation point. Keep your eyes on it. After 2 seconds, a word pair will appear at the place of the fixation point. Your task is to decide by pushing the right or left arrow of the keyboard whether the two words on the screen are semantically related or not. If you think that the two words are related, push the right arrow, if not, please use the left arrow. You will see the word pairs for 5 seconds while you have to make a decision. When the time is over, the computer does not accept decisions, and the fixation point appears again for 2 seconds before the next word pair comes up.

Trials started with the onset of a fixation spot in the middle of the screen, which was followed by a word pair chosen from the pool. The inter-trial interval was set for 2 sec, the word pairs stayed on the screen for 5 sec (exposure time). During this time participants were required to hit the right arrow key if they thought the word pair on the screen rhymed, and the left arrow key if they thought the words did not rhyme. If no response key was selected during the exposure time, the program did not record anything and the next trial started (fixation onset for 2 sec, etc.). The task was machine paced to ensure a constant level of attention of the participants.

In an initial training phase, the participants were shown five pairs in order to become familiar with the procedure. After a short break, the 80 word pairs of the phonological rhyming test and then the 80 word pairs of the semantic rhyming test were presented in a semi-random fashion (test phase). The program recorded correct and incorrect hits as well as response latency times.

In the data analyses we only included the correct decisions. Data were analyzed with Statistica software (StatSoft, Inc.), running Wilcoxon tests, Mann–Whitney U-tests and ANOVA. The significance level was set at 5%.

Results

Phonological rhyming test: Reaction time results

Table 4.2 presents the descriptive statistics for the RT results of the phonological rhyming test. Latencies on rhymes and non-rhymes in the mixed

Table 4.2 Means and standard deviations of phonological RT results

	Mean		SD	
	L1	HL	L1	HL
TRC	1.560129	1.933855	0.361366	0.576657
TNRC	1.840369	2.003375	0.468644	0.641128
CRC	1.586367	1.669170	0.412403	0.509289
CNRC	1.662660	1.852456	0.415629	0.661776

language pairs (TRC versus TNRC) were significantly different ($Z = 4.27$, $p < 0.01$) in the L1 Group, but no such difference was observed in the HL Group. Latencies were shorter in the rhyming conditions (TRC: 1.56 sec) than in the non-rhyming ones (TNRC: 1.84 sec) in the English-Hungarian mixed word pairs.

In the HL Group, we found differences between the RT results in the monolingual rhyming and non-rhyming word pairs ($Z = 1.96, p < 0.05$) (CRC versus CNRC). Again, the average amount of time spent on the decisions about the rhyming word pairs (CRC) was shorter (1.67 versus 1.85 sec) than the decisions about the non-rhyming pairs.

In the comparison of the latencies of decisions of mixed and monolingual rhyming word pairs (TRC versus CRC), only the HL Group produced statistically different results ($Z = 2.53, p < 0.02$). For them the decision was faster in the case of monolingual pairs (1.67 versus 1.93 sec). While there was no difference in the HL Group between the non-rhyming mixed and monolingual word pairs' RT, the L1 Group produced a significant difference in this comparison ($Z = 3.43, p < 0.01$), with the monolingual word pairs' RT being faster (1.66 versus 1.84 sec).

Phonological rhyming test: Judgement results

Table 4.3 presents the descriptive statistics for the judgement results of the phonological rhyming test. There was a significant difference in the L1 Group's accuracy results of mixed rhyming conditions (TRC versus TNRC,

Table 4.3 Means and standard deviations of the phonological judgement results

	Mean		SD	
	L1	HL	L1	HL
TRC	80.80645	71.17647	10.73132	22.11684
TNRC	57.41935	61.76471	25.68628	23.44816
CRC	78.06452	65.58824	10.92988	18.86484
CNRC	88.54839	75.88235	7.87333	21.59538

$Z = 3.63, p < 0.01$) and monolingual conditions (CRC versus CNRC, $Z = 3.47$, $p < 0.01$), with 80% average accuracy in the case of the mixed rhyming and 57% in the case of the mixed non-rhyming pairs. While in the mixed word pairs (TRC), rhymes were more accurately recognized than non-rhymes, in the monolingual pairs non-rhyming conditions (CNRC) resulted in better performance (88% versus 78%). However, the statistical analyses did not show any difference between the groups.

In the HL Group, the only difference found was concerned with the decisions about the monolingual word pairs ($Z = 2.01, p < 0.05$). Non-rhyming pairs were recognized significantly better (76%) than rhyming ones (65%).

Group comparison of phonological rhyming test

Phonological rhymes in the mixed word pairs were recognized significantly faster ($U = 157.0, p < 0.03$) by the L1 Group (1.56 sec) than by the HL Group (1.93 sec). However, the accuracy of judgement was the same for the two groups. In the decision about the monolingual word pairs, the L1 Group proved to be slightly better at both rhyming (78% versus 65%) and non-rhyming (88% versus 76%) conditions, but there was no difference between the two groups.

Semantic rhyming test: Reaction time results

Table 4.4 presents the descriptive statistics for the RT results of the semantic rhyming test. In both the L1 and HL groups there were significant differences in the latencies of rhyming and non-rhyming conditions in the mixed word pairs (TRC and TNRC, $Z = 4.08, p < 0.01$ and $Z = 2.53, p < 0.02$, respectively) as well as in the monolingual word pairs (CRC and CNRC, $Z = 4.68, p < 0.01$ and $Z = 3.41, p < 0.01$). Whereas in both groups there was a time-related difference (L1 Group: $Z = 2.48, p < 0.02$, HL Group: $Z = 2.22$, $p < 0.03$) between the recognition of mixed and monolingual rhyming pairs of words, there was no such difference in either group between the mixed and monolingual non-rhyming conditions.

Table 4.4 Means and standard deviations of the semantic RT results

	Mean		SD	
	L1	HL	L1	HL
TRC	1.351649	1.622993	0.255597	0.504779
TNRC	1.589970	1.832711	0.343973	0.496566
CRC	1.303384	1.472933	0.269870	0.399378
CNRC	1.594079	1.857462	0.328224	0.529097

Table 4.5 Means and standard deviations of the semantic judgement results

	Mean		SD	
	L1	HL	L1	HL
TRC	88.06452	78.43750	10.54178	21.42574
TNRC	83.87097	75.00000	13.21046	23.52304
CRC	90.48387	73.43750	10.82858	21.65785
CNRC	88.06452	72.18750	8.62916	22.80122

Semantic rhyming test: Judgement results

Table 4.5 presents the descriptive statistics for the judgement results of the semantic rhyming test. The only difference that we observed in the accuracy of judgements in any of the conditions was the average percentage of the correct recognition of mixed (84%) and monolingual (88%) non-rhyming word pairs in the L1 Group ($Z = 2.06, p < 0.04$). In the HL Group, no difference was found in any of the conditions.

Group comparison of the semantic rhyming test

No time-related differences were observed between the L1 and the HL Groups for the semantic rhyming test. However, the L1 Group performed significantly better in the monolingual rhyming and non-rhyming conditions. In the rhyming conditions, the mean accuracy of decisions in the L1 Group was 90% while that of the HL Group was 73%. In the non-rhyming conditions, the accuracy of decisions of the L1 Group reached 88% as opposed to that of the HL Group: 72% (CRC: $U = 95, p < 0.01$; CNRC: $U = 100, p < 0.01$).

Processing monolingual word pairs

Decisions were significantly faster in the semantic than in the phonological rhyming test ($F = 8.874, p < 0.01$; see Figure 4.1). No difference was observed in the given conditions between the groups. Non-rhyming conditions, whether phonological or semantic, took more time to process. The recognition of phonologically non-rhyming word pairs was the slowest for both groups.

Participants performed less successfully in the phonological test ($F = 5.21$, $p < 0.04$) as compared to their performance in the semantic test. There was also a significant difference between the two groups' accuracy rates ($F = 7.86$, $p < 0.02$; see Figure 4.2) in the phonological rhyming condition.

Processing bilingual word pairs

Figure 4.3 shows the RT of both groups in the bilingual context. Phonological processing took significantly longer ($F = 6.952, p < 0.02$) than
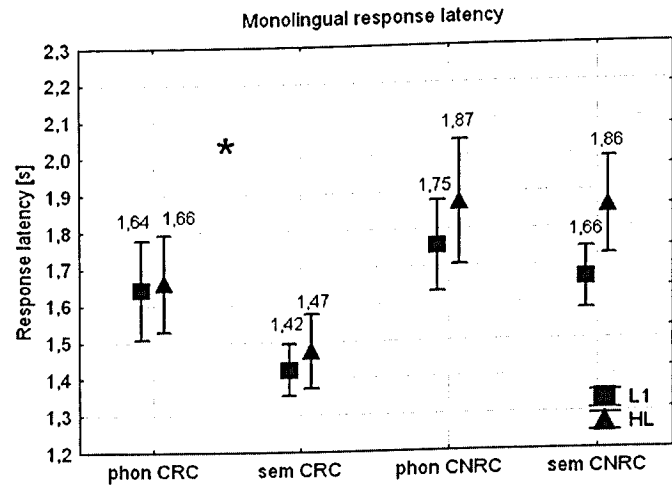
Figure 4.1 A comparison of processing times in monolingual phonological and semantic rhyme conditions

Figure 4.3 A comparison of processing times in bilingual phonological and semantic rhyme conditions

semantic in the bilingual condition as well. While the RT differences between the groups were minor in the monolingual condition, here we found somewhat greater differences in all the conditions except for the semantically non-related one. In all cases the bilingual group was slower; however, the statistical analysis did not show any significance. The greatest latency

difference between the two groups could be observed in the phonological rhyming condition.

Figure 4.4 illustrates the performance results. Our data support that processing bilingual data is harder as the decisions are less successful than those of the monolingual data. The difference between the two groups is smaller.

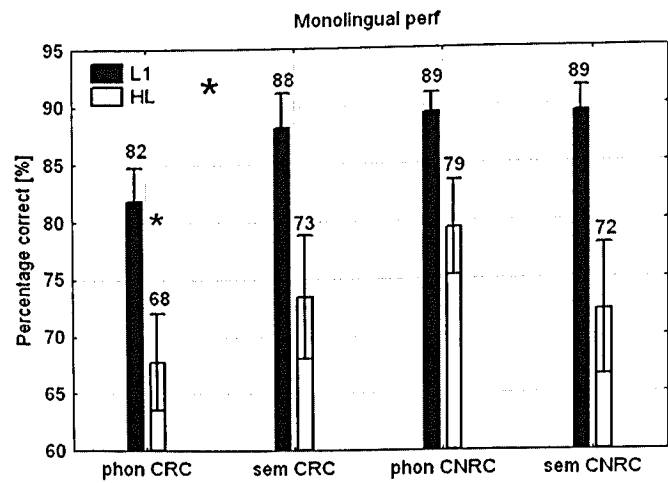**Figure 4.2** A comparison of rhyming judgements in monolingual phonological and semantic rhyme conditions

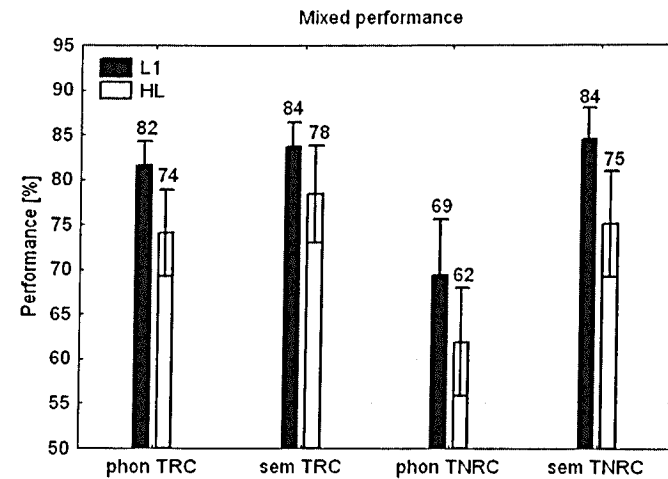**Figure 4.4** A comparison of rhyming judgements in bilingual phonological and semantic rhyme conditions

Discussion

The main goal of our study was to investigate the question of whether the processing of two languages is similar to the processing of only one language. To this end we looked at two typologically unrelated languages, namely English and Hungarian. These languages differ not only according to their phonological typology, but also in terms of their orthographies. The fact that the two languages are different in their phonological typology raises the question of whether and how bilingual word recognition takes place at the sub-lexical and lexico-semantic levels in bilinguals who speak these two languages. In this way, when we study written language processing, we must take orthographic, phonological and semantic awareness into consideration.

In a review of neuroscientific data about the bilingual brain, Buchweitz and Prat (2013) state that the cognitive processing of bilinguals must be reflected in features in the bilingual brain, such as magnitude and expansion of activation and synchronization of activity in different brain areas. Marian *et al.* (2003) suggest parallel activation and shared cortical structures in the early stages of language processing with a special emphasis on phonetic processing. They propose that the two languages may be using separate structures at later stages of processing, such as lexical processing. When the two systems are active at the same time, the elements of both languages have to be taken into account, which involves the activation of different neural connections as well (cf. Hervais-Adelman *et al.*, 2011; Kovelman *et al.*, 2008). This makes the process more difficult and complex and this is the reason for interferences.

A further goal was to see what temporal and performance results could be observed in the phonological and semantic processing by our two groups, i.e. our intention was to examine whether these processes were identical or if they showed differences when compared. We hypothesized that the processing of languages very much depended on the age and manner of acquisition. We recruited participants who differed in these two variables: early bilinguals acquired their L2 in a natural setting, and late bilinguals through school instruction. In order to check how bilingual visual processing occurs, we constructed two types of tests: one at the sub-lexical, and one at the lexical level. In this way we intended to check Marian and her colleagues' (2003) claim concerning shared and separate processing at different linguistic levels.

Following Khateb *et al.* (2000), who carried out similar tests among monolinguals, we made up a test for the investigation of the phonological and another for the semantic awareness of bilinguals. In our test material, we included mixed word pairs from both languages as well as word pairs from the same language in the hope of seeing the difference between processing mixed-language information and same-language information.

Orthographic and phonological consciousness shows how the two languages can be kept apart. If monolingual pairs of words are processed faster and better, and bilingual pairs take a longer time to judge, it indicates that the cognitive burden is much bigger in the bilingual context.

What we examined was the reading quality of words belonging to, on the one hand, a shallow and on the other hand a deep writing system. We tested the processing of words in the same and in mixed contexts with words from both languages. With the phonological rhyming task our intention was an immanent investigation of phonological awareness, the recognition and activation of sounds corresponding to letters or letter combinations. We aimed to see how accurately rhymes might be recognized in mixed and monolingual word pairs, i.e. crosslinguistically and in a single language context.

As reading requires the activation of orthography as well as phonology and semantics, we were interested in analyzing to what extent school instruction was important in the phonological and semantic awareness of bilinguals. The results concerning the latencies and the accuracy of the recognition of rhymes in the two conditions showed that phonological awareness was more developed in the L1 Group, who were late bilinguals and acquired their L2 by instruction. As opposed to Khateb *et al.* (2000), who did not find significant RT differences in the phonologically rhyming and non-rhyming pairs, in our test rhymes were always recognized faster and more accurately than non-rhymes. Still, a difference could be observed in the processing times of the two groups. Those who learnt their L2 through school instruction had shorter recognition times. The early bilingual participants (HL Group), who acquired their L2 naturally in the family, were slower in general. At the same time, in the comparison of the word pair conditions, their processing time of monolingual word pairs was shorter than that of mixed, bilingual word pairs. Even though we proposed a shared representation of languages for early bilinguals that would result in longer RT supposing that both their lexicons were activated to the same degree, we could see that there was a difference between the processing times of monolingual and mixed word pairs. Interestingly, when considering the accuracy of the decisions, being faster did not necessarily mean being better as well. Bilinguals took a longer time to make decisions but their accuracy results were similar to those of the L2 learners.

What we might infer from our results is that school instruction must have a crucial role in the development of phonological awareness since those who went through instructed learning were faster in the phonological decoding. The fact that monolingual words are processed faster suggests that for the brain it is easier to get on within just one language. If the language teacher puts enough emphasis on the role of phonology, not only will the pronunciation of the learner be better but also the reading processes will be faster and more automatized.

We found evidence that words from a more shallow writing system (Hungarian) were easier to read and judge on their phonological similarity than words from a language with a deep phonological consistency. This proves that phonological processing is language selective. Language instructors have to bear this in mind, and give time and special training to their students for phonological development.

Once the learners are good readers, they do not need to spend a considerable amount of time on the grapheme–phoneme correspondence rules to make the written material meaningful for them. If the readers are more experienced, the letter–sound combinations may be more automatized for them, and semantics may be linked to the frequently read word without processing it at the sub-lexical level.

For semantic awareness, school instruction seemed to play a less crucial role than for phonological awareness. Based on our results, we can claim that semantic awareness is more based on the general categorization faculty of the human mind. Semantic awareness relies more on implicit learning and seems to be less language dependent.

In our tests we found that for both groups the time spent on making decisions about meaning relations was shorter in the rhyming than in the non-rhyming conditions. This is in harmony with the monolingual results of Khateb et al. (2000), where the responses were also faster if the word pairs were related than when they were unrelated. In addition, we observed the same in mixed word pairs as well, which makes us assume that meaning relations are language independent.

Our results showed that late and early bilinguals took the same amount of time for semantic processing, but for the late bilingual participants (L1 Group) the processing of monolingual information was more successful. For early bilinguals (HL Group), the processing time was prolonged, especially in the mixed language non-rhyming conditions.

While there was no difference between the groups in the judgement results of mixed word pairs, they differed in the judgement accuracy of monolingual word pairs: the L1 Group outperformed the HL Group. This means for us that, although the meaning relations are there for both languages of the bilingual, the age of becoming bilingual (early or late) influences the distance between the elements of the languages (see also van Heuven, this volume). Relations are tighter within one language if the person becomes bilingual at a later age, and so it is easier for the languages to be kept apart for late bilinguals.

Semantic processing proved to be faster than phonological. This implies that semantic categorization must have had a parallel activation in both languages, or rather that semantics is shared between the languages, at least in the categories that were part of the test. This finding supports the idea of the bilingual mental lexicon's organization based rather on meaning relations than on languages (Navracsics, 2007), and that representation also depends

on the category of the word (de Groot, 1993, 2011). Our results support the hypothesis that the bilingual mental lexicon is to a large extent common for both languages. We also provide evidence that the mental lexicon is built on meaning relations as is reflected in the literature (Fox, 1996; Franceschini et al., 2003). However, we also claim that languages can be kept apart in the lexicon (see Cook, 2003; Singleton, 2007), depending on the orthographic, phonological, morphological and lexico-semantic organization of words and on the age of L2 acquisition (but cf. de Bot, this volume).

Based on our observation, data analyses and results, processing semantics seems to be easier than processing phonology, even in bilingual conditions. This means for our study that word meanings that are composed of the same semantic features across the languages must be processed similarly, and the same brain activity must underlie the processing of semantically related words, irrespective of the language they belong to. This finding seems to be supported by neuroimaging data as well. Buchweitz and Prat (2013) also propose a similar semantic representation at the neural level in bilinguals.

Referring to the literature (Leonard et al., 2010) that proposes that the L1 and L2 are not completely isolated from one another and that they both mutually interfere and support each other, and based on our findings about the difference between phonological and semantic processing, we suggest that during phonological processing languages interfere while during semantic processing they support one another. The phonological structure is more language dependent and the semantic structure is more shared between the languages. Instructed language learning may improve phonological awareness but it has no effect on semantic awareness.

Conclusions

In our chapter we intended to analyze how the neuroimaging data and experimental behavioural data converged in our psychophysical test on bilingual phonological and semantic awareness. As is stated in the literature, when the two language systems are active at the same time, the elements of both languages have to be taken into account, and processing is more difficult and complex, which causes interferences. In our study it became clear that interferences have different impacts on processing languages at different linguistic levels.

The investigation into phonological and semantic awareness showed that, for bilinguals, phonological processing is a more difficult cognitive task than processing semantics if we take latencies as a measure. This difference may be due to the fact that languages of different phonological typology are processed differently in the bilingual brain, whereas lexical semantics and sense relations do not differentiate languages in the cerebral representation, and the semantic representation is common for all the languages.

When considering the validity of the bilingual cognitive advantage (de Bot, 2015, this volume), we need to keep in mind that bilinguals need more time to process their languages than monolinguals, but this delay does not imply that they are any weaker in performance. Language learners, i.e. late bilinguals who have help while learning their L2 through school instruction, may be faster at recognizing the phonological structure of the word than early bilinguals, but they are not better at solving the linguistic task. This is of great importance to consider in bilingual visual word recognition and other types of reading tasks. Through school instruction, language learners may get to be faster in recognizing the grapheme–phoneme correspondence. At the phonological level, the manner of L2 acquisition makes a difference between natural and school-instructed bilinguals in terms of recognition time. Based on our results, school instruction has a great role in phonological processing, and with conscious learning, with a special attention to phonology, language learners may reach a higher level of phonological awareness than through natural language acquisition. In this respect, early bilinguals do not show any advantages in phonological processing. On the other hand, the fact that bilinguals' processing time is longer compared to that of L2 learners should not imply the interpretation that this delay is a disadvantage of bilingualism as opposed to monolingualism, since there is no difference in the quality of the processing in relation to the correctness of similarity judgements. The time-related difference may also be due to the greater brain area involved in language processing, which must be the result of the different phonological representations of the bilinguals' languages.

Finally, frequently used words are recognized faster and their meanings are accessed automatically without the need for processing the phonological structure of the word completely. This might be the reason why semantic processing is faster than phonological in our study.

There are two main implications for the future from this study.

(1)   While our findings are to some extent relevant to instructors, we want to be careful not to take this too far. Language learning and teaching are complex processes in which many factors interact at different levels: at the individual level, at the group level, at the school level and in the larger society. There is an increasing awareness that all these factors potentially play a role, which makes issues like the ones raised in this contribution relevant for theoretical development, which ultimately should enhance our understanding of the processes of learning and teaching. However, we should be careful when it comes to direct implications. Semantic and phonological awareness are important aspects of language learning, but the findings cannot be translated directly into classroom practices. They should be part of teacher training programmes, since teachers have to be aware of such issues. It is certainly possible to develop course material aimed at the raising of awareness at

different levels and evaluate such an intervention, but that is the type of research that follows from a more fundamental research than that which is presented here.

(2)   In our test material we incorporated frequently used, supposedly familiar words having a concrete reference, whose semantic features must be shared across the languages. However, we had difficulties in finding the right words and matching them for frequency and familiarity due to the differences of depth and extension in which the available corpus linguistic databases provide their data in the relevant languages. In our case, the available databases differed in size and in comparability as well. We thus agree with de Bot (this volume) that there is a need for unified, comparable corpora-based databases in order for researchers to be able to make up valid test materials in different languages.

Acknowledgements

References

Braun, M., Hutzler, F., Ziegler, J., Dambacher, M. and Jacobs, A. (2009) Pseudohomophone effects provide evidence of early lexico-phonological processing in visual word recognition. *Human Brain Mapping* 30, 1977–1989.

Buchweitz, A. and Prat, C. (2013) The bilingual brain: Flexibility and control in the human cortex. *Physics of Life Reviews* 10, 428–443.

Chee, M., Tan, E. and Thiel, T. (1999) Mandarin and English single word processing studied with functional magnetic resonance imaging. *Journal of Neuroscience* 19, 3050–3056.

Cook, V. (2003) Introduction: The changing L1 in the L2 user's mind. In V. Cook (ed.) *Effects of the Second Language on the First* (pp. 1–18). Clevedon: Multilingual Matters.

de Bot, K. (2015) A bias in citation patterns in applied linguistics: The case of the bilingual advantage. In Sz. Bátyi and J. Navracsics (eds) *First and Second Language: Interdisciplinary Approaches*. Studies in Psycholinguistics 6 (pp. 22–32). Budapest: Tinta Könyvkiadó.

de Groot, A. (1993) Word-type effects in bilingual processing tasks: Support for a mixed-representational system. In R. Schreuder and B. Weltens (eds) *The Bilingual Lexicon* (pp. 27–53). Amsterdam and Philadelphia, PA: John Benjamins.

de Groot, A. (2011) *Language and Cognition in Bilinguals and Multilinguals*. New York and Hove: Psychology Press.

Fabbro, F. (2001) The bilingual brain: Cerebral representation of languages. *Brain and Language* 79 (2), 211–222.

Fox, E. (1996) Cross-language priming from ignored words: Evidence for a common representational system in bilinguals. *Journal of Memory and Language* 35, 353–370.

Franceschini, R., Zappatore, D. and Nitsch, C. (2003) Lexicon in the brain: What neurobiology has to say about languages. In J. Cenoz, U. Jessner and B. Hufeisen (eds) *The Multilingual Lexicon* (pp. 153–166). Dordrecht: Kluwer.

5   Perception of Checked Vowels by Early and Late Dutch/ English Bilinguals – Towards a New Measure of Language Dominance[1]

Vincent J. van Heuven

Introduction

There has been considerable debate on the question of to what extent someone can acquire a native pronunciation in a second or foreign language. It is generally observed that adults who immigrate to a foreign country do not normally acquire the language of the new country without an accent that is reminiscent of their mother tongue (e.g. Flege *et al.*, 2006; Piske *et al.*, 2001). The adult second or foreign language (henceforth L2) learners acquire the new language after their first language (L1) is well established. However, when young children immigrate, they usually learn to speak the new language without a trace of a foreign accent.

When one has learnt a first language, speech sounds are typically perceived in terms of the (phoneme) categories of the native language. These categories were shaped during the first 12 months after birth (e.g. Kuhl & Iverson, 1995). According to the motor theory (MT) of speech perception (e.g. Liberman *et al.*, 1967), sounds that belong to the same phoneme category are not readily discriminated by native listeners; however, even a small phonetic difference between two sounds is easily perceived if these sounds belong to different phonemes, i.e. are separated by a phoneme boundary. A more recent theoretical development holds that a physically equal difference between two sounds is perceived as smaller the closer the sounds lie to a category prototype (Native Language Magnet theory (NLM) of speech perception; Kuhl & Iverson, 1995). Together, the mechanisms explained by MT

and NLM conspire to make it difficult for adults to fully acquire a new sound system. The difference in success of the acquisition of the L2 phonology between young and older learners is not fully understood and is subject to ongoing debate. Loss of neural plasticity with age has been advanced as one cause that prevents learners from modifying the perceptual categories that were developed in the early stages of life (e.g. Lenneberg, 1967; see also Navracsics & Sáry, and Pfenninger & Singleton, this volume). However, L2 acquisition at a young age (in a natural, non-supervised learning environment) was also found to be strongly correlated with intensity of exposure, which by itself could explain the greater success of young learners (e.g. DeKeyser, 2013; Flege *et al.*, 2006).

There are indications that at least some adults have been able to learn to pronounce a foreign language in a way that cannot be distinguished from that of born-and-bred native speakers, despite the fact that the learning process did not involve early L2 exposure. For instance, Bongaerts *et al.* (1997) recorded a number of intermediate and highly advanced Dutch L2 speakers of English. Each L2 speaker produced five English test sentences, which were subsequently judged by native English listeners and rated on a scale for nativeness. The recordings randomly alternated with control utterances spoken by native English speakers. Seven of the 11 highly advanced L2 speakers (typically professors or lecturers in English at the level of university or teacher training college) obtained a pronunciation score that was equal to, or even better than, that of some of the native English speakers (see Figure 5.1).
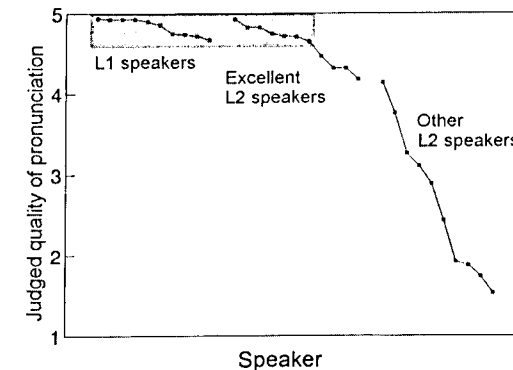
Figure 5.1 Quality of British-English pronunciation rated by native English listeners (5 'No accent at all, definitely native'–1 'Very strong foreign accent, definitely nonnative') of three samples of speakers: native speakers of British English (L1); excellent L2 speakers with Dutch as the mother tongue; and an intermediate group of L2 speakers (including some students of English)
Notes: The grey area marks the overlap between native L1 and excellent L2 speakers.
After Bongaerts *et al.* (1997).

Production and perception may not be in perfect harmony, however. It might be the case, therefore, that such extremely advanced L2 speakers still have a mental representation of the L2 sound categories that deviates from that of L1 speakers. The first aim of the present paper is to establish the extent to which this might be the case.

It is generally recognized that young children who grow up in a bilingual setting acquire a native pronunciation in both languages they are exposed to (see also Navracsics & Sáry, and Pfenninger & Singleton, this volume). Bilingual settings may occur, for instance, when one parent is a native speaker of language A (e.g. English) and the other of language B (e.g. Dutch). When each parent communicates with the child in his/her native language, the child will end up with a (virtually) equal command of, and native pronunciation in, both languages. Alternatively, the parents may have immigrated to a new country but rear the child in the native language. When the child goes to school (typically between the age of four to six) she is regularly and massively exposed to the language of the host country, which it will acquire at the native level – without losing her proficiency in the parents' language. Such bilingually raised children are referred to as 'natural' or 'early' bilinguals. The term 'early bilingualism' is used in contrast with 'late bilingualism', which is the situation referred to above, when someone acquires the L2 after the complete acquisition of the L1. It is the second aim of this paper to establish if there is any difference in the perceptual representation of the sound systems of two languages that are spoken by early and late bilinguals.

Within the category of early bilinguals a further subdivision has been proposed depending on language dominance. It is hardly ever the case that even an early bilingual has a perfectly balanced command of both languages. Rather, it will be the case that one language is more often heard or produced than the other, so that tasks in this language can be carried out faster and with fewer errors. For instance, Sebastián-Gallés and Soto-Faraco (1999) had Spanish-dominant and Catalan-dominant early bilingual Spanish-Catalan listeners identify members of binary vowel and consonant contrasts in Catalan (/e-ɛ/, /o-ɔ/, /ʃ-ʒ/, /s-z/). The former three contrasts are unique to Catalan and are not part of the Spanish sound system. The fourth contrast, /s-z/, however, occurs in both languages. The results showed that the Catalan-dominant early bilinguals differentiated the Catalan contrasts more accurately and earlier in the time course of the stimulus than the Spanish-dominant counterparts. Crucially, no difference between the two groups was found for the shared /s-z/ contrast. The Spanish-dominant bilinguals had been monolingual in Spanish during the first 24 months of their lives. Likewise, the Catalan-dominant bilinguals did not hear any Spanish before the age of three. These results indicate that even in near-perfect bilinguals early exposure to one language over the other may have a lasting influence on the sharpness with which the sound categories of the two languages are defined perceptually.

It is well-known that late bilingualism comes at a price. When one learns the sounds of an unfamiliar language, several possibilities may present themselves. Flege's (1995) Speech Learning Model (SLM) makes a distinction between three scenarios.

(i) A sound category (phoneme) in L2 may be indistinguishable from its counterpart in the learner's L1. In a phonetic transcription the L1 and L2 sounds would be written with exactly the same base symbol and diacritics. In this case of 'identical sounds' the learner will use the L1 category in the L2 without introducing even a hint of foreignness. The nasal consonants /m, n, ŋ/ in English and Dutch would be examples of identical sounds.

(ii) In certain other cases the L2 has a sound category that has no match in the learner's L1. The L2 category may be between two L1 categories or it may find itself somewhere in the phonological space where the L1 has no categories at all. Typically, the International Phonetic Alphabet (IPA) has distinct base symbols to represent the L1 and L2 sounds. Here the learner will – sooner or later – become aware of the difference between the 'new sound' in the L2 and his native L1 categories. SLM predicts that the learner will ultimately set up new categories for these sounds, which will be faithful approximations of the L2 sound categories. When these new sounds are in place, they will no longer contribute to the foreign accent on the part of the learner. Moreover, since the new sounds are added to the set of categories that exist in the learner's L1, the latter will not be affected in any way by the inclusion of the additional categories, so that the pronunciation of the L1 will not change. English /æ/ as in *bad* would be an example of a new sound for Dutch learners of English; conversely, the Dutch /œy/ as in *huis* 'house' would be a new sound for an English learner of Dutch (Collins & Mees, 1984).

(iii) The third scenario applies when the L2 has sound categories that are similar but not identical to their nearest equivalent in the learner's L1. The learner is not aware of a subtle difference between the L2 sound and its L1 equivalent but native listeners of the target language perceive the deviant sound as foreign. Similar sounds would be designated by the same IPA base symbol but differ in their diacritics. English /s/, for instance, has its spectral centre of gravity at a higher frequency (and therefore has a sharper timbre) than its Dutch counterpart. Dutch learners of English consistently fail to notice this difference and substitute their own, duller, Dutch /s/ when they speak English, and as a result sound foreign to the native English listener (Jongman & Wade, 2007). Since the learner is not aware of the difference between the L1 and the L2 sound, a more widely defined sound category is formed which includes the observed variability of both the L1 and the L2 sounds. The result of this fusion is a compromise category which is noticeably

incorrect both in the L2 and in the L1 of the learner: the learner will not only sound foreign in the L2 but also in the L1.

We may predict that the perceptual representation of the sound categories in the L1 and L2 of late bilinguals will be different from that of early bilinguals. In the case of early bilinguals we expect the sound categories to be similar in nature within one language. That is to say, all the phoneme categories in L1 will have the same sharpness of definition. The same applies to the categories in the other language. This hypothesis does not rule out the possibility that the sound categories of one language may be more sharply defined than those of the other language. Very likely, one of the two languages that are spoken by an early bilingual will be more prominent (or dominant) than the other; this difference in dominance may come to light in the well-definedness of the sound categories. We suspect that the categories in the dominant language will be more sharply defined, i.e. have smaller boundary widths than those in the other language – but within each of the two languages the categories have similar boundary widths.

In the case of late bilinguals, we hypothesize that the categories in the L1, which were formed during childhood, will be more sharply delineated than those in the L2. Also, in the case of similar sounds, in terms of the SLM, the ideal (prototypical) location of the sound in the phonetic space will be in between the preferred locations found for the L1 sound and its equivalent in the L2.

It is our ultimate prediction that the perceptual representation of the sounds in the L1 and the L2 will be different – in terms of the location of the prototypes and in the sharpness of the category boundaries – for monolingual speakers of a language as compared to bilingual speakers, whether early or late, and that there will also be differences between the early and late acquirers. More generally, we predict that even excellent L2 learners, whose pronunciation can no longer be distinguished from native L1 speakers, can still be shown to have perceptual representations of the L2 sound system that deviate from those of monolingual speakers of the target language.

The first question I address in the present paper is whether such excellent L2 speakers as are exemplified in Bongaerts *et al.* (1997; see also Figure 5.1) have a mental representation of the English vowel sound system that is equal to that of native L1 speakers of British English. I would hypothesize that if the excellent L2 speakers cannot be differentiated from native speakers by their speech production, their perceptual representation of the sound categories should also be similar.

Secondly, we would like to know to what extent the late bilingual listeners differ in their mental representation of the English and Dutch vowel systems from monolingual listeners, not only in terms of the location of the prototypes in the vowel space but also in terms of the well-definedness of the vowel categories.

Thirdly, we want to know if language dominance is reflected in the relative strengths of the vowel categories in the early bilingual listeners. Answering this question requires the availability of instruments to determine (the degree of) language dominance in bilinguals. This chapter aims to contribute to the discipline by proposing – and provisionally testing – a new method for estimating language dominance in the phonology of bilinguals. In the next section we will introduce a short-cut method to establish the well-definedness (in lieu of establishing boundary widths between categories) of the sound categories in the phonological systems of bilinguals and show how the measure can be transformed into an estimate of the relative dominance of one language over the other in the mental representation of the vowel systems in a bilingual listener.

Approach

It follows from the introduction that we needed four different groups of subjects in order to find an answer to the research questions formulated. These are (i) a monolingual group of native English speakers, and (ii) a similar group of monolingual Dutch native speakers.[2] These two groups served as control conditions that establish the baseline against which the performance of the bilingual speakers was to be gauged. Two bilingual groups of listeners took part in the study, i.e. (iii) an early bilingual group, whose members all had learnt English and Dutch from childhood onwards in a typically bilingual setting (details see below), and (iv) a group of late bilinguals, which was comprised of the same type of learners as was targeted by Bongaerts *et al.* (1997).

In order to keep the experiment within reasonable bounds, we limited the study to the perceptual representation of only the short vowel phonemes in two related languages, i.e. (British) English and (Netherlandic) Dutch. Apart from six long vowels, three true diphthongs and – depending on the depth of the phonological analysis – a number of centring diphthongs, English has six short vowels (also termed lax vowels), which – phonotactically – cannot occur at the end of a word.[3] These are /ɪ, e, æ, ʌ, ɒ and ʊ/, i.e. the vowels that occur in the words *hid, head, had, hud, hod,* and *hood.* In addition to four long vowels and three diphthongs, Dutch has eight short vowels, five of which are phonologically lax (and cannot occur at the end of a word), i.e. /ɪ, ɛ, ɑ, œ, ɔ/ as in *bid* 'pray', *bed* 'id.', *bad* 'bath', *put* 'well', and *bod* 'offer', and three more which are tense but remain phonetically short (unless followed by /r/ in the first syllable of a trochaic foot, cf. Kooij & van Oostendorp, 2003), i.e. /i, y and u/ as found in *bied* 'offer', *buut* 'terminal point' and *boet* 'make repairs', respectively.

We decided to generate an artificial vowel space that included all the short vowels in Dutch and English, and to sample vowels at regular intervals from this space. This approach is reminiscent of Schouten (1975) with the

exception that we used (perceptually) uniform sampling, whereas Schouten sampled his vowels by interpolating in small steps between the prototypes of all the vowels of English (whether long or short). We reasoned that such non-uniform sampling is inefficient and compromises the comparability between the vowels in the system. By limiting our study to only the short vowels, we could synthesize the vowels with realistic durations, while Schouten (1975) synthesized his vowels at a duration that was halfway between that of short and long vowels. As a result Schouten's vowels were always non-ideal, or non-prototypical, exemplars of their category. Another point of difference between Schouten's stimulus materials and ours is that, where Schouten used isolated vowels produced out of any spoken context, we synthesized the vowels in the context of a carrier phrase, embedded between an initial and a final consonant.

The location of the prototypes of the various vowel categories can be found by asking listeners to indicate, first of all, which of the six English or eight Dutch vowels they associate most readily with a given stimulus vowel, and secondly, how good they think the token is as an exemplar of the category chosen. In this way, we may map out the perceptual representation of the vowel space of the English or Dutch listener in terms of the location of the prototypes, i.e. the most preferred vowel tokens in the set of artificial vowel sounds. Although it is also possible to define the boundary widths from these judgements, I decided to establish the boundary widths by a shortcut method proposed by van Heuven and van Houten (1989). The shortcut is to present each vowel token for identification twice (in different random orders) and to compute the consistency with which the listeners label the two tokens of each vowel type. The poorer the definition of the category boundary, the poorer the labelling consistency. Language dominance can then be established by determining the ratio between the mean consistency observed in sound categories of the two competing languages.

Methods

Stimulus materials

The English vowels /ɪ, e, æ, ʌ, ɒ and ʊ/ were placed in the context *I say m_f* and recorded in a soundproofed recording studio. The set was repeated three times at conversational speed by a late bilingual speaker of English and Dutch, who was an experienced phonetician with Dutch as his native language. This procedure was repeated with /ɪ, ɛ, ɑ, œ, ɔ, ɪ, y and u/ in the Dutch context *Ik zei m_f*. The mean duration of all the Dutch vowel tokens (127 msec) and of all the English vowel tokens (133 msec) was measured. On the basis of this result, it was decided to adopt a single vowel duration of 130 msec for all the vowels in the stimulus set to be generated.

The same speaker then produced extreme versions of the Dutch point vowels /i, u, a, ɑ/ (as approximations of cardinal vowels 1, 4, 5 and 8) in the context of the Dutch carrier *Ik zei m_f* 'I said ...'. The recordings were analogue-to-digital converted (12 bit, 10 KHz) and submitted to a so-called robust analysis of formant frequencies and bandwidths, using the split-Levinson algorithm for Linear Predictive Coding (Willems, 1987) with a 25-msec window and a 10-msec frame shift. Fundamental frequency (vocal pitch) was computed by the method of subharmonic summation described in Hermes (1988). The same algorithm was used to decide for each analysis frame whether it was voiced or voiceless. Formants F_1 to F_5 and the associated bandwidths B_1 to B_5 were computed. Table 5.1 lists the values for the lowest two formants measured at the temporal midpoint in each of the four cardinal vowels.

A two-dimensional grid was defined as a vowel space which was spanned between the extreme point vowels in Table 5.1. The vowel height dimension was sampled with nine steps at 0.7 Bark apart along the first formant, interpolating between 308 and 1050 Hz (from 3.2 to 8.8 Bark).[4] The combined vowel backness and rounding dimension was sampled with 10 steps of 0.7 Bark along the second formant between 1033 and 2732 Hz (from 8.7 to 15.0 Bark). This defines a 9 (F_1) × 10 (F_2) grid (see Table 5.2), from which cardinal 5 and the two types closest to it were later eliminated on the grounds that these tokens sounded highly unnatural. The F_1 and F_2 values chosen were the targets to be attained at the temporal midpoints of the vowels. The target values of the higher formants were kept constant at the values found for the token containing cardinal vowel 4.

The resulting 87 vowels were embedded in either the Dutch or the English carrier phrase using LPC resynthesis in order to ensure smooth and realistic formant transitions between the initial /m/ and the final /f/ sound of the target word. The target vowel consisted of a steady-state segment of six frames with a six-frame transition (i.e. lasting 60 msec) on either side, the entire phrase being 105 frames for both language conditions. The stimulus sentence was resynthesized with constant segment durations and a fixed F_0 contour with an accent-lending rise–fall pitch movement on the target syllable.

Table 5.1 Formant centre frequencies and bandwidths measured for the four cardinal vowels which formed the basic framework for the stimulus space in the experiment

	[i]	[a]	[ɑ]	[u]
F_1	300	980	790	360
F_2	2640	1620	1210	1157
B_1	30	98	79	36
B_1	264	162	121	116

Table 5.2 The vowel stimulus space, defined by 10 F_2 (front–back) by 9 F_1 (close–open) values

Second formant F2 (front–back)

Step #			1	2	3	4	5	6	7	8	9	10
Hertz			2732	2464	2221	2001	1800	1617	1451	1299	1160	1033
Bark			15.0	14.3	13.6	12.9	12.2	11.5	10.8	10.1	9.4	8.7
First formant F_1 (height)	1	308	3.2	1,1	2,1	3,1	4,1	5,1	6,1	7,1	8,1	9,1	10,1
	2	381	3.9	1,2	2,2	3,2	4,2	5,2	6,2	7,2	8,2	9,2	10,2
	3	459	4.6	1,3	2,3	3,3	4,3	5,3	6,3	7,3	8,3	9,3	10,3
	4	541	5.3	1,4	2,4	3,4	4,4	5,4	6,4	7,4	8,4	9,4	10,4
	5	628	6.0	1,5	2,5	3,5	4,5	5,5	6,5	7,5	8,5	9,5	10,5
	6	722	6.7	1,6	2,6	3,6	4,6	5,6	6,6	7,6	8,6	9,6	10,6
	7	822	7.4	1,7	2,7	3,7	4,7	5,7	6,7	7,7	8,7	9,7	10,7
	8	932	8.1	1,8	2,8	3,8	4,8	5,8	6,8	7,8	8,8	9,8	
	9	1050	8.8	1,9	2,9	3,9	4,9	5,9	6,9	7,9	8,9		
#	Hz	Bk										

Note: Step number, hertz and Bark values are indicated. Three combinations were omitted from the stimulus set.

For each language, the 87 stimulus utterances were converted back and recorded on analogue audiotape in one random order, preceded by three practice items, and repeated in reversed order, preceded by the same three practice items. The interstimulus interval was 3 sec (offset to onset); a short beep was recorded after every 10th item.

Subjects

Fifteen monolingual English speakers (seven males) were recruited at Edinburgh University with a mean age of 21;2 years and a range from 19 to 25. All but one were raised in the south of England and spoke RP English. The 15th subject was born in Portugal but raised by RP-speaking parents and later attended public school in the UK. None of these subjects had any working knowledge of Dutch.

Fifteen Dutch-speaking subjects (four males) were recruited at Leiden University, the Netherlands. All were selected on the basis of their ABN (Algemeen Beschaafd Nederlands) variety, the Dutch equivalent of RP. The mean age for the monolingual Dutch subjects was 26;4 years with a range between 19 and 37. None of these subjects had a special affiliation with either the English language or English-speaking countries.

Fifteen bilingual speakers were found, eight of whom were early bilinguals. The remaining seven were native speakers of Dutch who had acquired

an exceptional command of English after the age of puberty (late bilinguals). Inclusion proceeded on the basis of three main criteria: (i) they were assessed as having native competence by native speakers of each language; (ii) they made regular use of both languages; and (iii) they subjectively perceived themselves as bilingual. In distinguishing between the natural (early) and artificial (late) bilinguals, a further requirement was that the former had acquired their languages, preferably simultaneously, before the age of 10.

An attempt was made to post hoc categorize early bilinguals in terms of English versus Dutch language dominance. This was done on the basis of a written questionnaire, in which the participants provided details with respect to their language acquisition, use and preference.

Procedure

Stimuli were presented to listeners over good-quality headphones (Sennheiser HD414) in individual sessions. Subjects indicated for each of the utterances they heard which vowel they recognized, with forced choice from either the six English short vowels or from the eight Dutch short vowels – depending on the presentation mode. Listeners were issued response sheets that listed the six or eight vowels in a quasi-phonetic spelling, which was explained and exemplified with unambiguous sample words in the instruction text. The English response categories were *i* (as in *pit*), *e* (as in *pet*), *a* (as in *pat*), *o* (as in *pot*), *U* (as in *putt*) and *u* (as in *put*). The Dutch categories were *i* (as in *rit* /rɪt/ 'ride'), *e* (as in *red* /rɛt/ 'save'), *a* (as in *rat* /rɑt/ 'id.'), *o* (as in *rot* /rɔt/ 'rotten'), *u* (as in *Ruth* /rœt/ 'id.'), *ie* (as in *riet* /rit/ 'reed'), *uu* (as in *Ruud* /ryt/ 'Rudy') and *oe* (as in *roet* /rut/ 'soot').[5]

The monolingual English listeners heard only the 2 × 87 items in their own language and responded with forced choice from the eight English short vowel categories. The bilingual listeners and the Dutch monolinguals took the test twice, the first time listening to the Dutch materials (and responding in the Dutch mode), the second time listening to the English version (and responding in the English mode).

Results

Preliminary screening of the responses revealed that the categories *ie* and *uu* (the high front unrounded and rounded vowel type, respectively) were severely underrepresented, with no more than 120 responses to either of these categories (against up to 864 responses for the other categories). Possibly, the frequency of the F_1 was not low enough to elicit a sufficient number of responses to these categories. Consequently, we excluded these two vowels as possible response categories from further data analysis – so that both the Dutch and the English response sets are limited to six possibilities.

We will now present the results in three successive stages. First, we will consider the location of the prototypical realizations of each of the six short vowel types in Dutch and in English, as determined from the responses by the four listener groups. Secondly, we will look at the size of the categories as can be derived, rather loosely, from the magnitude of the dispersion ellipses that can be drawn around the prototypes (or centroids) and, thirdly, we will quantify the sharpness of the perceptual representation of the various vowel types in terms of the response consistency measure defined by van Heuven and van Houten (1989).

Location of prototypes

Figure 5.2 shows the location of the vowel prototypes in the F_1-by-F_2 space, which was determined by the mean F_1 and the mean F_2 value of all the vowel types identified by a group of listeners as an instance of that category weighted by the number of responses. The three graphs in the left column of Figure 5.2 present the centroids of the short vowels of English as perceived by early bilinguals (top row), by late bilinguals (middle row), and by Dutch learners of English as a foreign language (bottom panel). The vowels perceived by monolingual English listeners are repeatedly shown in the three graphs as the corners of the shaded polygons.

The early bilinguals have virtually the same locations of the six short English vowels as the monolingual English reference listeners (top panel). Also, both early and late bilinguals, as well as the Dutch learners of English, show only minor discrepancies in the locations for the vowels /ɪ, e, ɒ/. However, discrepancies can be observed for the remaining three short vowels /æ, ʌ, ʊ/. These are vowels that would be classified as new sounds in terms of Flege's (1995) SLM. The Dutch learners of English ('monolingual Dutch') deviate considerably from the RP-English targets. The discrepancy is moderate for /æ/ (about 0.5 Bark) in the Euclidean vowel space, and larger for /ʌ, ʊ/ (roughly 1.0 Bark). Predictably, the centroids for /æ, ʌ/ are raised (closer vowel qualities) since the nearest vowel in Dutch would be /ɛ/ and /æ/, respectively. The location of /ʊ/ in the perceptual representation of the monolingual Dutch listeners is further back than for the RP listeners. Again, this follows directly from the circumstance that the nearest Dutch vowel is a very back /u/. Interestingly, these findings correspond closely to the production data of Dutch learners of English and those of native English speakers as reported by Wang and van Heuven (2006).

Crucially, the late bilinguals have their perceived ideal location of /ʌ, ʊ/ in positions that are intermediate between those of the native and early bilingual listeners, on the one hand, and those of the Dutch monolinguals on the other. It would seem, therefore, that the late bilinguals' perceptual representation of the English vowels is affected by the presence of an interfering vowel in the competing language, i.e. Dutch. A rather unexpected
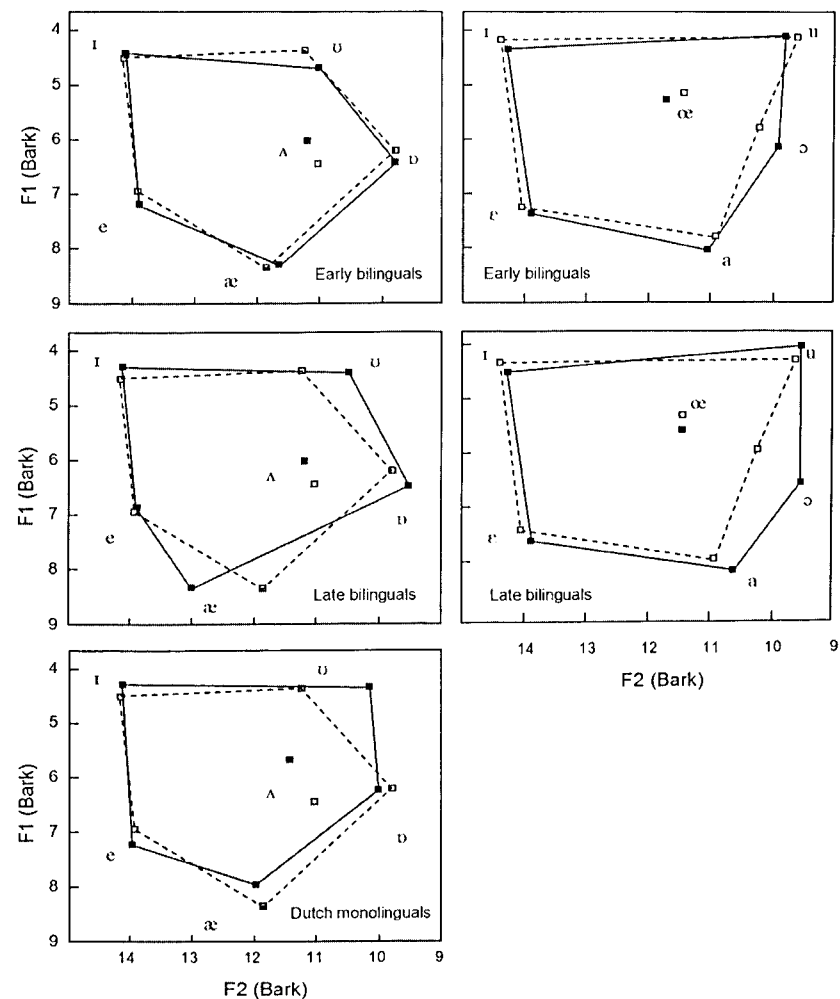
Figure 5.2 Vowel centroids in the F_1-by-F_2-space (in Bark) by early bilinguals (top row), late bilinguals (middle row) and Dutch monolinguals, as perceived in the English (left column) and Dutch (right column) response modes
Note: Monolingual English listeners (left column) and Dutch (right column) listeners are repeatedly indicated by the shaded polygons for reference purposes.

phenomenon is seen in the ideal location of /æ/ as perceived by the late bilinguals, which is as low as that of the native RP listeners but considerably more fronted. This might be seen as a tendency on the part of the excellent learners of English as a foreign language to overcompensate (or exaggerate) the difference between the ash-vowel and its competitors. Finally, the late

bilinguals' results show a negative effect of their excellent command of English on the perceptual representation of the L1. This is clearly seen in the lowered and backed location of the vowel /ɔ/, which is identical to these listeners' perceptual representation of English /ɒ/.

Dispersion of perceived vowels

Figure 5.3 displays not only the centroids of the perceived vowels but also the dispersion of the vowels associated with a particular category. The dispersion ellipses were drawn at ±1 standard deviation around the vowel centroids along the two principal components of the scatter cloud of each vowel type in the F_1-by-F_2 plane. The classification of each vowel type by the majority of the listeners is indicated by linking the vowel type to the centroid of the preferred category. The left and right top panels show the dispersion ellipses for the English (left) and Dutch (right) short vowels as perceived by monolingual English and Dutch listeners, respectively.

There is not much difference in terms of dispersion between the early and late bilinguals' perceptual representation of the English vowels. The degree of overlap between adjacent categories is roughly comparable, with one exception. For both the late bilinguals and the Dutch learners of English the dispersion ellipses for the vowels /e/ and /æ/ overlap considerably. The origin of this might lie in the fact that Dutch has just one low-mid front vowel whereas English has two. This type of native-language interference can be expected from learners of English as a foreign language, but it is somewhat surprising that the interference should still be found with the late bilinguals.

A second observation would be that the bilingual listeners seem to have narrower vowel categories than the monolingual listeners. Possibly, then, the vowel space of bilingual speakers, whether early or late, is divided into a larger number of categories (i.e. the union of the vowel inventories of both languages), which would leave less space for each vowel in the combined inventory.

Labelling consistency

A first exploration

Each listener responded to each of the 87 stimulus vowel tokens twice. A consistency index was computed by dividing the number of vowel repetitions responded to in a like manner by the total number of repetitions (=87). The means are illustrated in Table 5.3.

First, Table 5.3 shows that the consistency in the Dutch listening mode is substantially better than in the English mode. We have no explanation for this difference, although we observe that the large spread of the centralized back vowel /u/ in English does lead to an unusual overlap (and therefore
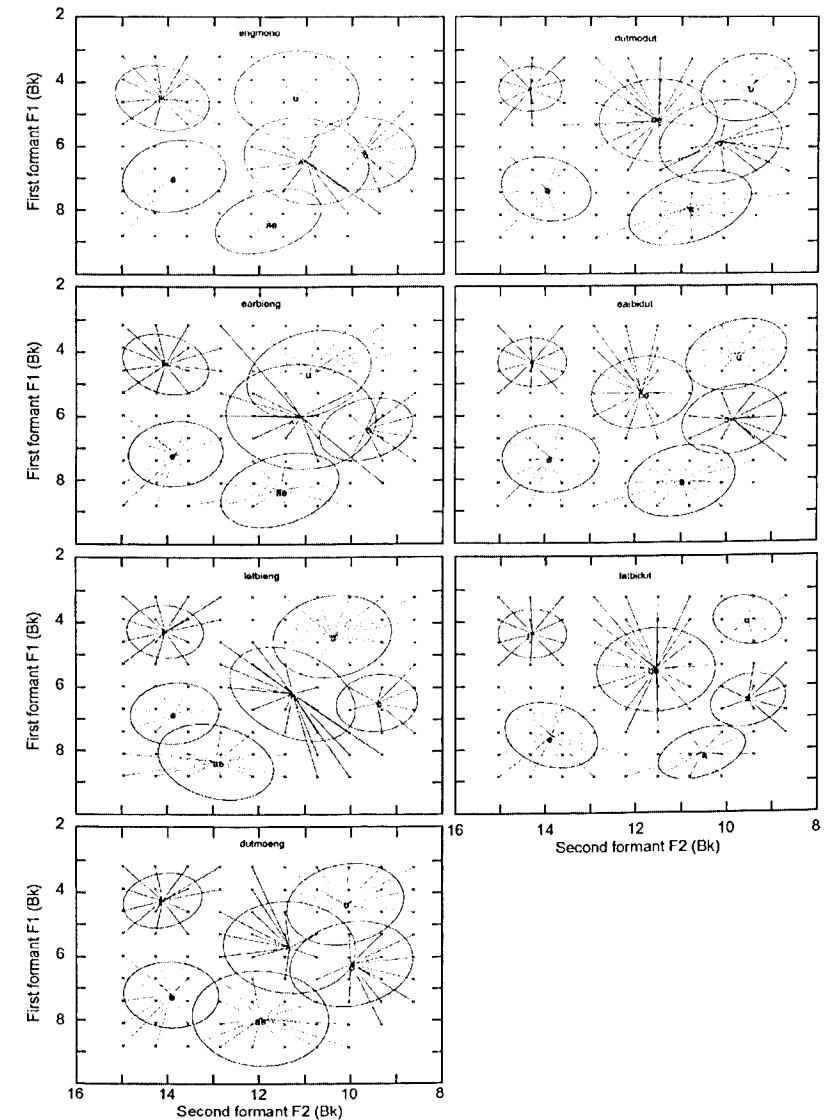
Figure 5.3 Centroids and dispersion ellipses (±1 standard deviation) in the F_1-by-F_2 space (Bark) for six perceived short vowels in the English listening mode (left column) and in the Dutch listening mode (right column) by monolingual native listeners (top row), early bilinguals (second row), late bilinguals (third row) and Dutch L2 listeners of English (bottom)

Table 5.3 Consistency index per language (listening mode) and subject group

	Listening mode	
Subject group	English	Dutch
Monolingual	0.70	
Bilingual	0.71	0.79
Early	0.68	0.75
Late	0.75	0.82
Dutch	0.66	0.80

poorly defined category boundaries), which is avoided in Dutch, with its extremely back-articulated /u/ that has to be kept distinct from its front rounded neighbour /y/. The /y/, of course, is not a phonemic category in English, which is the reason why /u/ may have such a wide dispersion in RP. The results further confirm that non-native listeners categorize the L2 vowel system less consistently (0.66) than native listeners do (0.70). Also, although bilingual listeners, on average, have the same consistency in vowel identification as monolingual L1 listeners, the data show a difference when we split up the group of bilingual listeners. The early bilinguals tend to be less consistent in either language than the monolinguals. Moreover, it would seem that the late bilingual listeners are the most consistent vowel identifiers, irrespective of the stimulus language. One reason why the late bilinguals are such consistent labellers might be that these subjects – in contrast with the other listener groups – are professional linguists/phoneticians and pronunciation instructors. They are conscious of the differences between the Dutch and the English vowels in terms of their location in the vowel space, and may therefore be unusually intent on keeping the competing vowel categories separate – hence the smaller dispersion ellipses in Figure 5.3 and the greater consistency in Table 5.3.

The consistency data were submitted to a repeated measures analysis of variance (RM-ANOVA) with stimulus language (Dutch, English) as a within-subject factor and with type of bilingualism (early, late, poor) as a between-subjects factor. Since the RM-ANOVA requires correlated data, the English monolingual group could not be included (since no Dutch stimuli were presented to these listeners). The RM-ANOVA indicates that the effect of stimulus language is significant, $F(1, 27) = 14.2$ ($p = 0.001$, $p\eta^2 = 0.344$). However, the effect of listener group fails to reach significance, $F(2, 27) = 1.1$ ($p = 0.363$, $p\eta^2 = 0.072$) as does the language × group interaction, $F(2, 27) = 1.1$ ($p = 0.362$, $p\eta^2 = 0.073$). A one-way ANOVA on the consistency indexes obtained in the English presentation mode by all four listener groups, i.e. including the English monolinguals, again fails to show significance for listener group, $F(3, 41) < 1$.

Consistency as a correlate of language dominance

As a last exercise we may see if there is any value in the consistency index as a tool to differentiate bilinguals in terms of language dominance. Given that the consistency index is significantly higher for Dutch than for English we defined for each listener a difference value Δ as the index for Dutch minus the index for English. Table 5.4 lists, among many other variables, the index scores for English, Dutch and the Δ for each of the eight early bilinguals.

We have seen that the consistency scores were better overall when the stimuli were presented and responded to in the Dutch mode. For a fair comparison of the consistency scores in the two modes by the early bilinguals, some form of normalisation would therefore be needed. Moreover, the proper comparison should be between the early bilinguals in each response mode in separate groups combined with monolingual listeners responding in the same language mode. Accordingly, z-transformation was applied to the English consistency scores for the early bilinguals combined with the group of monolingual English respondents. Similarly, z-scores were computed for the Dutch consistency indexes obtained by the members of the combined early bilingual plus monolingual Dutch listeners. In both arrays of z-scores the early bilinguals have lower scores on average than their monolingual comparison groups, but the difference is not significant (or else the effect of listener group should have been significant in the above analyses of the raw scores).

Language dominance should correlate with the difference between the z-scores obtained by the early bilinguals in the Dutch versus English listening mode. Figure 5.4 displays the consistency z-scores obtained in the English response mode plotted against the z-scores in the Dutch mode for each of the eight early bilinguals, who are identified in the scatterplot by arbitrary two-digit numbers. Five out of eight early bilinguals have roughly the same consistency in their vowel labelling in both language modes. These would then be balanced bilinguals. Two individuals have clearly better consistency in English than in Dutch. They appear below the 45° reference line and are marked by filled squares. These would be clear cases of English-dominant bilinguals. It should be observed that these two individuals (Listeners 24 and 25) do not have exceptionally good consistency within their peer group. In fact, they belong to the poorer performers. The point is that within themselves they do better in English than in Dutch. Listener 20 (marked by a filled circle) represents the other extreme of the dominance continuum. This person has by far the poorest consistency in English of all eight early bilinguals but is in the upper half of the peer group in the Dutch response mode.

Now let us see if there is anything in the language acquisition history of the eight early bilinguals that might explain the differences in language dominance observed in the consistency indexes. The top part of Table 5.4 summarizes the information gathered on the early bilinguals by means of a questionnaire. The bottom part of the table presents the individual

Table 5.4 Consistency scores (raw and z-transformed, see text) and variables relating to language acquisition for eight early (natural) bilingual listeners

ID number	16	17	19	20	21	22	24	25
(1) Gender	F	F	F	F	F	F	M	M
(2) Age	22	22	23	24	23	26	21	24
(3) Father	E	E	E	E	D	D	D	E
(4) Mother	E	E	D	D	D	D	D	E
(5) Home child	D/E	D/E	D/E	E	D/E	D/E	D/E	D/E
(6) Prim school	D/E	D	D	D	D	D/E	D/E	D/E
(7) Sec school	D	E	D	E	D	E	E	E
(8) Home now	E	D/E	D/E	E	D/E	D/E	D/E	D/E
(9) Work now	E	D	E	D	D	E	E	E
(10) Preference	E	D/E	E	E	D/E	E	E	E
(11) Age D	5-10	1.5-12	1-12	5-12	4-8	0-2, 8-12	9-12	6-12
(12) Age E	0-5, 10-10	0-1.5	≥12	0-5	0-4, 8-12	2-8	0-9	0-6
(13) Years D	8	20	23	19	23	26	21	24
(14) Years E	22	22	23	24	19	25	21	24
(15) Consist D	82.56	80.46	77.91	82.76	72.41	89.66	52.87	63.22
(16) Consist E	77.01	73.26	65.06	44.83	72.41	83.91	65.43	64.71
(17) ΔConsist	5.55	7.20	12.85	37.93	0.00	5.75	-12.56	-1.49
(18) Consist D(z)	0.4453	0.2313	-0.0290	0.4657	-0.5891	1.1689	-2.5815	-1.5267
(19) Consist E(z)	0.5649	0.2784	-0.3470	-1.8908	0.2141	1.0911	-0.3186	-0.3740
(20) ΔConsist(z)	-0.1196	-0.0470	0.3180	2.3565	-0.8032	0.0778	-0.2629	-1.1527

Notes: (3) L1 of father; (4) L1 of mother; (5) Language used at home when child; (6) Language used at home when child; (7) Language of instruction in primary school; (7) Language of instruction in secondary school; (8) Language used at home at the time of the experiment; (9) Language used in public domain at the time of the experiment; (10) Preferred language in present daily life; (11) Age when exposed mainly to Dutch; (12) Age when exposed mainly to English; (13) Number of years using mainly Dutch; (14) Number of years using mainly English; (15) Consistency index in English; (16) Consistency index in Dutch; (17) Difference between consistency in Dutch and English. (18)–(20) are the z-transformed versions of (15)–(17).
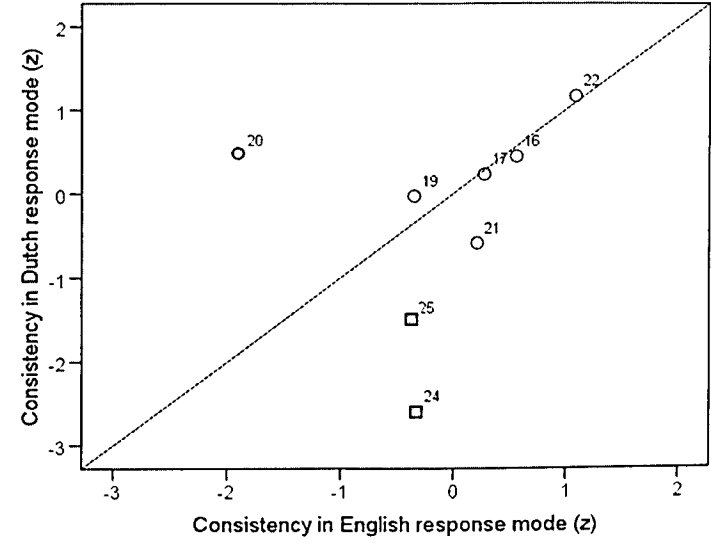
Figure 5.4 Consistency index (z-scores) in English (horizontal axis) and Dutch (vertical axis) response mode for eight early bilinguals
Note: The 45° reference line represents perfect balance of English and Dutch consistency (for further details see text).

consistency scores in English and Dutch, the difference between them as well as the z-transformed values.

The most striking difference that distinguishes Listeners 24 and 25 from the other early bilinguals is the fact that they spent the first nine and six years, respectively, of their lives in England, and only then crossed the North Sea to live in the Netherlands. These two listeners also turn out to be the only males in the sample, but it is hard to imagine that this would be causally related to their English language dominance. It is probably not accidental that the English dominance of the male who spent the first nine years in England is considerably stronger than that of the person who left England at the age of six. It suggests that spending time with English-speaking peers in an English school environment is of decisive importance here.

Listener 20 turned out to be a bit of a problem case. She spent the first five years of her life in England, and yet she obtained by far the lowest consistency score for English, which makes her the most Dutch-dominant listener in the sample. Reconstructing her language past from the table, it would appear that her mother spoke Dutch to her (or English with a Dutch accent), while the language the parents used to communicate with one another was English – which is understandable since the family lived in England at the time. Right at the time when the child was about to go to school, the family moved to the Netherlands. This child was probably never

exposed to a rich English-speaking school environment such as the one Listeners 24 and 25 experienced. During primary school (until the age of 12) she grew up with Dutch-speaking peers. The family continued to live in the Netherlands (given that the young woman speaks Dutch outside the home). Most likely, the language she used at home continued to be English (with a Dutch-accented mother), while she attended an international school in the Netherlands, where the language of instruction was English but spoken by a great many non-natives. This, to my mind, leads to poorly defined perceptual norms for the English vowels – a hypothesis which is supported by the low consistency score in English.

It would seem, in sum, that the (z-transformed) difference in labelling consistency for vowel categories would be a useful tool in determining language dominance in a bilingual speaker. It may well be, however, that a more sensitive measure of language dominance could be defined if we were to take specific vowel categories into account. Some vowels may be more prone to perceptual confusion than others. By zooming in on the poorly defined vowel categories we may enhance the sensitivity of the instrument.

Consistency in the perception of poorly defined vowel categories

Table 5.5 provides a full survey of the vowels provided by the listeners at the first and second presentation of the same stimulus. For ease of presentation, each of the four listener groups is listed separately, broken down further by the language mode of the stimulus presentation (Dutch versus English). Note that the English monolingual listeners were never asked to classify the vowels in the Dutch mode. Dutch monolinguals are not really monolinguals but always have a basic knowledge of English as a result of six to eight years of English lessons at school; they could therefore be asked to respond in the English as well in the Dutch mode. The numbers in Table 5.5 are row percentages, and add up to 100. The actual number of responses is specified for each row under N.

The table shows, again, that the high tense vowels were unpopular response categories in Dutch. Front /i/ was the least frequent response vowel in this group, followed by /y/. The response frequency of /u/ is better, although it still falls short of the non-tense vowel types. When comparing the performance of the groups, it would seem better to disregard the results for the high tense front vowels of Dutch. In that case there are no striking differences in the performance of the three Dutch groups (i.e. early bilingual, late bilingual and monolingual).

The most striking difference is seen in the English response mode, viz. between the consistency of the early bilinguals and that of the monolinguals. Although the mean difference is insignificant, as shown above, we now see that the English monolinguals have more consistent responses than the early bilinguals for the vowels /æ, ʌ, ʊ/, while the reverse is seen for the vowels /ʌ, ʊ/. In particular, the inconsistency concerning the pair /ʌ, ʊ/ seems

Table 5.5 Relative frequency (% of row total) of vowel response on first versus second presentation of the same stimulus, broken down by language group by response mode

Vowel category responded (first presentation)

Language group	Stimulus	Mode: English							Mode: Dutch								
		i	e	æ	ʌ	ɒ	ʊ	N	i	ɛ	ɒ	ɔ	æ	ɪ	y	u	N
Monolingual English (N = 15) Second presentation	i	**69**	18		2	4	7	141									
	e	3	**80**	7	5	2	4	271									
	æ	1	7	**77**	11	3	2	131									
	ʌ	1	9	10	**51**	8	20	178									
	ɒ	1	1	7	15	**62**	14	163									
	ʊ	4	3	1	12	7	**74**	382									
Bilingual early (N = 8) Second presentation	i	**80**	10	2	1	1	6	104									
	e	4	**82**	3	2	6	3	140									
	æ	2	18	**45**	20	12	3	60									
	ʌ	3	4	11	**40**	5	37	114									
	ɒ		2		7	**84**	7	90									
	ʊ	4	3		18	6	**69**	175									
	i								**72**	10	1	1	9	2	4	1	93
	ɛ								1	**85**	6	2	4	1	1	1	142
	ɒ									4	**84**	6	2	1	1	1	96
	ɔ								1	5	10	**71**	6		2	5	108
	æ									5	1	3	**81**		5	5	149
	ɪ								15	3	15	8	8	**15**	31	8	13
	y										2	7	29	4	**45**	16	38
	u								2	2	2	7	9	9	9	**67**	55

(Continued)

Table 5.5 (*Continued*) Relative frequency (% of row total) of vowel response on first versus second presentation of the same stimulus, broken down by language group by response mode.

Vowel category responded (first presentation)

Language group	Second presentation	Mode: English — I	e	æ	ʌ	ɒ	ʊ	Mode: Dutch — I	ɛ	ɒ	ɔ	æ	i	y	u	N
Bilingual late (N=7)	e	80	8	18	8	7	2	92		74	15	1	7	2	1	85
	æ	3	70	74	5	5	1	96	2	92		59	5	1		129
	ʌ	4	16	10	75	4	2	61		15		6	7			71
	ɒ		1		14	84		159	1		2	78	16			63
	ʊ	3		1	20	5		58	50	50	2	1	91	2	1	179
								143					50			4
	y								5			30		55	5	20
	u											9	3	86		58
Monolingual Dutch (N=15)	e	76	9	8	8	4	1	193	77	14	1	8	1			155
	æ	5	68	18	17	2		194	1	92	3	4				256
	ʌ	1	21	54	70	6		179			85	6	5	1		209
	ɒ	3	2	5	22	56		424		3	15	18	62	1	2	136
	ʊ		3	7	28	2		164	2	2	3	85		1	3	415
		2				68		144				38	50	13		8
	y								5	5	10	67	2	10	5	21
	u									2	1	20		7	63	105

Note: Consistent responses are on the main diagonal (bold type in shaded cells). Total number of valid responses in listed under *N*.

characteristic for the early bilinguals, with confusions (a change of response category between first and second presentation from one to the other) of 37% and 18%. This suggests that the perceptual categories for these two vowels are less well defined in the minds of the early bilinguals than for the monolinguals. Similarly, the vowel /æ/ is often confused with its neighbours in the vowel space, i.e. /e/ and /ʌ/. The triplet /æ, ʌ, ʊ/ is known to present pronunciation problems for Dutch learners of English (see, for example, Wang & van Heuven, 2006). The vowel with the weakest definition is the central vowel /ʌ/. This is also the vowel with the largest dispersion ellipse in the corresponding panel in Figure 5.3. The competition between neighbouring vowel categories would appear to be limited to the mid-range of F_2 (front–back) values. As a result, there is vacillation or uncertainty between the lower vowel /æ/ and the higher vowel /ʊ/. This perceptual uncertainly does not arise in the case of the more peripheral vowels /I, e, ɒ/, which have more extreme F2 values, i.e. higher in the case of /I, e/ and lower in the case of /ɒ/. Moreover, the English vowels /I, e, ɒ/ would be rather good and readily identifiable exemplars of Dutch /I, ɛ and ɔ/, respectively. It would appear, then, that early bilingualism comes at a (small) price in that the perceptual representation of some vowels is poorer than is seen for monolingual speakers.

In conclusion to this section it should be observed that the inconsistency patterns seen in Table 5.5, i.e. the vacillation in the responses to the repeated presentation of the same stimulus vowel sound, are a way to qualify the degree of overlap between the dispersion ellipses in Figure 5.3. Obviously, the more two ellipses overlap in Figure 5.3, the more strongly the vacillation between the two vowel categories involved.

Conclusions

In this chapter we asked whether excellent speakers of a second or foreign language (L2), whose pronunciation sounds as native as that of native (L1) speakers, might still be different when it comes to their perceptual representation of the sound categories in the L2 (and possibly even of their L1). Our results show that the excellent adult Dutch learners of English have a perceptual conception of at least some of the vowels of RP English that differs from that of monolingual English listeners. Specifically the location of the 'new sound' /æ/, although more open than that in the pronunciation of less advanced Dutch learners of English, is more fronted than is the case in the perceptual representation of this vowel by monolingual English listeners (and of early bilinguals). We also noticed that the perceptual representation of the mid back vowel for the late bilinguals is the same in English and in Dutch – which it should not be since English /ɒ/ differs from Dutch /ɔ/; the late bilinguals' representation of this vowel is correct for English but wrong

for Dutch. This shows that excellent learners of a foreign language, whose pronunciation sounds perfectly native as judged by phonetically trained native listeners (Bongaerts *et al.*, 1997), may still have an imperfect perceptual representation of the target-language sound system.

The second question we asked is whether the perceptual representation of the vowel systems of early bilinguals would differ from those of late bilinguals. The results indicate that this is indeed the case. Whereas we found only small differences between the perceptual representation of the vowel systems of monolingual listeners and those of early bilinguals, the late bilinguals differed in several respects. Not only did the location of at least one vowel differ between early and late bilinguals (and monolinguals) in both English (L2) and Dutch (L1), see above, but we also found that the late bilinguals' vowel categories, in both L1 and L2 tend to be more narrowly defined, i.e. with less allowed deviation from the category prototype (smaller spreading ellipses) and with sharper boundaries (better labelling consistency). We conclude, then, that early (or natural) bilinguals and late (artificial) bilinguals have different mental representations of the vowel systems of both languages they command.

Early bilinguals do show subtle differences in their mental representation of the vowels of English (more than of Dutch) in that the boundaries between /æ, ʌ, ʊ/ are less sharply defined than those of /ɪ, ɛ, ɔ/. We may note that this observation runs counter to the hypothesis we formulated in the introduction that the well-definedness of the vowels within the same language system would be uniform. This aspect of vowel representation in the mind of early bilinguals deserves more future research. The consistency analysis that we advocated in this study may turn out to be a useful tool to analyse the stability and well-definedness of sound categories in monolingual and bilingual language users. We found evidence that the consistency index may serve as a diagnostic of language dominance in (early) bilinguals. The relative difference in labelling consistency between the two languages of the listener seemed to correlate well with the individual's language acquisition history. The sample of early bilinguals, however, was small. More and more systematic research is needed in this area in order to develop a better understanding of the potential of the labelling consistency index as a correlate of language dominance.

Notes

(1) This chapter appeared in abridged form as van Heuven *et al.* (2011). It is based on an MA thesis written by Nicole Broerse under the supervision of the present author. I thank Ing. Jos Pacilly for his invaluable help in generating the stimulus materials used in the experiment. I am also grateful to Professor Antonella Sorace for her part in the supervision of the work done at Edinburgh University.

(2) Since Dutch children take compulsory lessons in English from the age of 10 onwards, a purely monolingual adult speaker of Dutch is hard to find. It is generally accepted, however, that Dutch speakers of English have no clear idea of the sound categories of English – with the exception of students of English either at the university or at teacher training colleges, who are explicitly trained to pronounce English without an accent. The latter type of subjects were not included in the monolingual speaker group.

(3) Schwa was not included in the sets of short vowels, neither in English nor in Dutch. We used vowels that could occur in stressed monosyllabic words, which requirement rules out the inclusion of schwa. Also in terms of its phonotactics, schwa is not on a par with the regular short vowels, since it can occur at the end of words.

(4) The Bark transformation is an empirical formula that adequately maps the differences in Hertz-values onto the perceptual vowel quality (or timbre) domain. A difference of 1 Bark, in whatever direction, is a perceptually equal difference in vowel quality, irrespective of its location in the F_1-by-F_2 space. We used the Bark formula proposed by Traunmüller (1990): Bark = $[(26.81 \times F)/(1960 + F)] - 0.53$, where F represents the measured formant frequency in Hertz.

(5) The response vowels were spelled as is usual in English and Dutch orthography in closed syllables. Capital versus lower case *u* were used to differentiate the English vowels in *putt* and *put*, respectively. Although this is somewhat contrived, the use of examples turned out to be adequate to resolve any remaining ambiguity.

References

Bongaerts, Th., van Summeren, C., Planken, B. and Schils, E. (1997) Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition* 19, 447–465.

Collins, B. and Mees, I. (1984) *The Sounds of English and Dutch*. Leiden: E.J. Brill/Leiden University Press.

Flege, J. (1995) Second language speech learning: Theory, findings, and problems. In W. Strange and J.J. Jenkins (eds) *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 233–277). Timonium, MD: York Press.

Flege, J., Birdsong, D., Bialystok, E., Mack, M., Sunge, H. and Tsukada, K. (2006) Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics* 34, 153–175.

Hermes, D. (1988) Measurement of pitch by subharmonic summation. *Journal of the Acoustical Society of America* 83, 257–264.

Jongman, A. and Wade, T. (2007) Perception of native and foreign-accented English by native and non-native listeners. In O.-S. Bohn and M. Munro (eds) *Language Experience in Second Language Speech Learning* (pp. 135–150). Amsterdam and Philadelphia, PA: John Benjamins.

Kooij, J. and van Oostendorp, M. (2003) *Fonologie. Uitnodiging tot de klankleer van het Nederlands [Phonology. Invitation to the Sound Structure of Dutch]*. Amsterdam: Amsterdam University Press.

Kuhl, P. and Iverson, P. (1995) Linguistic experience and the 'perceptual magnet effect'. In W. Strange and J. Jenkins (eds) *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 121–154). Timonium, MD: York Press.

Lenneberg, E. (1967) *Biological Foundations of Language*. New York: Wiley.

Liberman, A., Cooper, F., Shankweiler, D. and Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychological Review* 74, 431–461.

Piske, T., MacKay, I. and Flege, J. (2001) Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29, 191–215.

Schouten, M. (1975) Native-language interference in the perception of second-language vowels. Doctoral dissertation, Utrecht University.

Sebastián-Gallés, N. and Soto-Faraco, S. (1999) Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition* 72, 111–123.

Traunmüller, H. (1990) Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America* 88, 97–100.

van Heuven, V. and van Houten, J. (1989) Vowel labeling consistency as a measure of familiarity with the phonetic code of a language or dialect. In M. Schouten and P. van Reenen (eds) *New Methods in Dialectology* (pp. 121–130). Dordrecht: Foris.

van Heuven, V., Broerse, N. and Pacilly, J. (2011) Perception of checked vowels by early and late Dutch/English bilinguals. In W. Zonneveld, H. Quené and W. Heeren (eds) *Sound and Sounds. Studies Presented to M.E.H. (Bert) Schouten on the Occasion of his 65th Birthday* (pp. 103–116). Utrecht: Utrecht Institute of Linguistics OTS.

Wang, H. and van Heuven, V. (2006) Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. In J. van de Weijer and B. Los (eds) *Linguistics in the Netherlands 2006* (pp. 237–248). Amsterdam and Philadelphia, PA: John Benjamins.

Willems, L. (1987) Robust formant analysis for speech synthesis applications. In J. Laver and M. Jacks (eds) *European Conference on Speech Technology* (pp. 250–253). Edinburgh: CEP Consultants.

Part 2

Future Implications for Second Language Acquisition and Language Policy: Theoretical Considerations

6 Recent Advances in Quantitative Methods in Age-related Research

Simone E. Pfenninger and David Singleton

Introduction

The upsurge of interest in research on the age factor in foreign language settings in recent decades has raised new methodological and assessment issues. Although much research has been devoted to identifying age effects and to their interaction with social-psychological, personal and affective variables, the specific impact and contribution of different quantitative approaches has more often than not been disregarded. This is surprising, considering that methods in age research have evolved significantly over the past couple of decades, and that it would therefore not have been unrealistic to expect certain methodological innovations to have entered this domain. Multilevel modelling (MLM) – a subgroup of linear mixed-effects regression modelling – has for some time been finally finding its way into certain SLA subfields. Research on the age factor, however, has only recently – and tentatively – begun to adopt these kinds of statistical models.

This chapter discusses the benefits that MLM can furnish to any SLA research that involves the sampling of populations, within educational establishments or naturalistic settings, and that has a particular focus on chronological age and the age of onset of acquisition in their roles as continuous and categorical predictors. Since the emphasis here is on conceptual issues and practical recommendations, technical details are deliberately kept to a minimum and mathematical details of the methods in question will be avoided.

First, we explore some central issues in the age factor discussion, following up on van Heuven's discussion in Chapter 5. Secondly, we review methods that have developed in respect of linguistic approaches to the age factor in SLA and go on to outline the benefits and advantages of mixed-effects models, in particular regarding perceived gaps in age-related research. It is hoped that this discussion will contribute to ensuring the consistent choice

of the most insightful analysis in relation to any given dataset. Unfortunately, given the constraints of space, it will not be possible to discuss the (very important) complementary role of qualitative work in this area (see Pfenninger & Singleton, 2016; Singleton & Pfenninger, 2015).

The Age Factor: Portrait of a Complex Variable

Many researchers still blithely talk about 'the age factor' as if it were a simple, single, solitary factor. This is despite the fact that for many years it has been authoritatively pointed out that such a view is almost laughably simplistic and deeply unsatisfactory. Moreover, the notion of the age factor being a rather more complex phenomenon than how it has customarily been portrayed is not linked to any particular theoretical stance on, for example, the critical period.

Thus, Montrul, who broadly favours the notion of the existence of a critical period, sees age of acquisition as a macrovariable that subsumes other interrelated factors, such as 'maturational state, biological age, cognitive development, degree of first and second language proficiency, amount of first and second language use, among others' (Montrul, 2008: 1). For Flege (2009), who is generally seen as a critical period sceptic, age of onset (AO) is a proxy for a multitude of variables, including state of neurological development, state of cognitive development, state of L1 phonetic category development, levels of L1 proficiency, language dominance, frequency of L2/L1 use and kind of L2 input (native speaker versus foreign accented).

The significance of initial age of learning may be difficult to determine precisely because of the fact that it cannot be disentangled from other variables. Adopting this approach, Jia and Aaronson (2003) argue that AO is a confounded indicator of neurobiological maturation because it co-varies with environmental factors. Moyer, for her part, has recently had the following to say on this matter:

> ... a host of interrelated variables is at play, having to do with learner orientation and experience. ... One valuable contribution of sociolinguistic work in SLA has been to call attention to social, cultural, and psychological circumstances relevant to individual L2 users – a reminder to take a more nuanced look at what underlies age effects in SLA. (Moyer, 2013: 1)

There is also the question of whether age-related differences should really be regarded as individual differences. The usual line is to place them alongside individual variables like gender, aptitude, richness of environment, motivation, learning styles, learning strategies and personality (see, for example, DeKeyser, 2012; Paradis, 2011; Zafar & Meenakshi, 2012). R. Ellis (2006), however, excludes it from his inventory of individual differences. The

reasons he gives are interesting. He states that age does not belong to any of his four categories of individual differences: '"abilities" (i.e., cognitive capabilities for language learning), "propensities" (i.e., cognitive and affective qualities involving preparedness or orientation to language learning), "learner cognitions about L2 learning" (i.e., conceptions and beliefs about L2 learning), ... "learner actions" (i.e., learning strategies)' (R. Ellis, 2006: 529). He takes the view that age transcends these categories and potentially impacts on all four. He also touches on the different views that have been advanced in relation to age – and their controversial nature. He concludes: '[t]he question of the role played by age in L2 acquisition warrants an entirely separate treatment' (R. Ellis, 2006: 530).

Ellis's uncharacteristic wariness in relation to the age factor and his stated reasons for such wariness speak volumes about the complexity of this variable, confirming the kinds of arguments he earlier expressed very eloquently. The inference must be that researching the age question demands both a very comprehensive and a very delicate (in the Hallidayan sense of 'fine-grained') perspective. Our view is that it necessitates both qualitative and quantitative methodologies, and that the quantitative approach used, which is what we focus on here, needs to go well beyond the kinds of the general linear model (a family of statistical models that assumes a normal distribution among other features, e.g. t-tests, ANOVA or multiple regression models; see Cohen, 1968; Plonsky, 2013) that have been employed in this area in the past.

Quantitative Research on the Age Factor I: Where We Are Now

In age-related research, as in many other areas of SLA, quantitative research is currently perceived as more prestigious than qualitative research, at least in so far as it dominates empirical research in many of the most prestigious journals (Benson et al., 2009; Richards, 2009). This trend is reflected in the steady increase of published studies using sophisticated statistical tests as well as in the multiplication of the range of tests used (see, for example, Lazaraton, 2005; Plonsky, 2013, 2014; Plonsky & Gass, 2011).

Age-related research has followed this trend. From the 1990s, a trend became apparent for researchers to move beyond a focus on the influence of the age factor on L2 attainment as a stand-alone variable, and for them to begin to explore its interaction with other variables. This period was characterized by a marked increase in the use of statistical methods: inferential statistics such as t-tests (e.g. Jia & Fuse, 2007; Johnson & Newport, 1989; Mora, 2006) or (multivariate) analyses of (co)variance (e.g. Flege et al., 1999; Larson-Hall, 2008; Llanes, 2012; Llanes & Muñoz, 2013; McDonald, 2006, 2008; Muñoz, 2006; Torras et al., 2006) or multiple regression analyses (e.g. Muñoz, 2003, 2014) or a factor analytic approach (e.g. Csizér & Kormos,

2009; Moyer, 2004), as well as correlations (e.g. DeKeyser *et al.*, 2010; García Lecumberri & Gallardo, 2003; Kinsella & Singleton, 2014; Miralpeix, 2006). Recently, we have been able to observe the emergence of the citing of effect sizes and confidence intervals, which can be attributed to the requirements of a number of applied linguistics journals (see Brown, 2011; Vacha-Haase & Thompson, 2004) – for example, *Language Learning*, which has tried to steer writers away from relying too much on significant *p*-values by asking them to 'always present effect sizes and their confidence intervals for primary outcomes' (N. Ellis, 2000: xii).

On the other hand, quantitative methods have also been critically evaluated, and numerous limitations of empirical efforts in SLA have been documented (see Lazaraton, 2005; Norris & Ortega, 2000; Oswald & Plonsky, 2010; Plonsky, 2011, 2013, 2014; Plonsky & Gass, 2011). In what follows, we address some of the main points that have featured in this critique of established quantitative procedures in the context of SLA research and having particular reference to age-related studies.

Generalizability

One of the main differences between qualitative research and quantitative research is that, in the latter tradition, scholars usually define their scope more broadly and seek to make generalizations about large numbers of cases. (Note, however, that in dynamic systems theory and other process-oriented research agendas, scholars advise against making universal generalizations and instead focus on 'particular generalizations' (Gaddis, 2002: 62, quoted in de Bot & Larsen-Freeman, 2012: 19) without implying that they are applicable beyond our own research site and data.) For example, when comparing differences between qualitative and quantitative research in contemporary political science, Mahoney and Goertz (2006: 238) state that 'in quantitative research, where adequate explanation does not require getting the explanation right for each case, analysts can omit minor variables to say something more general about the broader population'. The generalizability issue has long been a controversial one in debates about quantitative research methods in SLA. In Boulton (2011), we read:

> Quantitative research ... may be more generalizable as it irons out some individual differences; but that is also its disadvantage as it can result in 'over-simplicity, [making] it a blunt and meaningless instrument' (Leakey 2011, 251). [...] The methodology [in quantitative research] is limited and constraining. (Boulton, 2011: 5)

Flynn and Foley (2009: 30) comment critically that '[a] commonly noted limitation to this general approach [in quantitative works] is that the narrow focus risks missing important contextual information or other variables'. In

other words, recent SLA research has become increasingly aware that the variation between individuals is crucial and not just 'noise' (N. Ellis & Larsen-Freeman, 2006: 564) and that 'learners are more than bunches of variables' (Dewaele, 2009: 637).

This also affects age factor research. As mentioned above, age interacts with social-psychological, personal and affective variables that have been found to be under the influence of situation. In both naturalistic and institutional environments, age effects need to be considered in light of macrocultural and microcultural phenomena that can have a bearing on interpersonal relations which influence, shape, increase or decrease the impact of variables that interact with age, such as motivation. Furthermore, recent thinking on age suggests that external factors also need to be addressed as environmental influences that interact with age effects and possibly mediate them. It would thus be a gross error of omission to filter out or fail to address such influences.

Randomization

Related to generalizability in classroom-based research is the problem of randomization. As Vanhove (2015: 135) points out, intervention studies and controlled experiments in which participants are randomly assigned to the treatment or control group are often considered 'the gold standard for establishing the effectiveness of language learning methods'. However, randomization comes with a variety of problems in age-related classroom research. Besides the problem that random assignment to experimental conditions has often not been implemented in classroom research, randomization has been recently questioned in quantitative research, since (1) it is 'the process of de-individualization, that is, the uniqueness of each person is ignored' (Navidinia & Eghtesadi, 2009: 59), and (2) it is frequently neglected or not dealt with appropriately in statistical models. In discussing the inappropriateness of ignoring the effects of assigning whole groups of participants to the experimental conditions, Vanhove (2015) discusses the various traditional ways of dealing with background variables in randomized controlled interventions.

One common way is to group participants according to variables that are deemed important *before* randomization, e.g. by assigning half of the boys to the treatment group and half to the control group (see Oehlert, 2010, Chapter 13, quoted in Vanhove, 2015). Despite the validity of this procedure, grouping according to background variables is rather difficult in a classroom setting, where the participant samples are defined at the onset of the data collection (see discussion above). A more practical solution has been to first run so-called balance tests (e.g. *t*-tests or ANOVAs, or χ^2-tests; see Vanhove, 2015, for a discussion of balance tests) to ensure that the different groups are comparable in all relevant respects save for the independent variable (e.g. AO), on the basis of the belief that randomization is a mechanism for creating samples that are balanced with regard to potential confound variables (Vanhove, 2015); and,

secondly, to equate subjects on basic pretest and prior ability measures (and then run an analysis of covariance with the selected background variables as the covariates), as well as to equate the treatment practices as much as possible in terms of task demands (see Chaudron, 2001: 67). However, many authors today (e.g. Mutz & Pemantle, 2013) deem balance tests 'superfluous', mainly because statistical tests already take account of fluke findings due to randomization (see also Oehlert, 2010, Chapter 2, quoted in Vanhove, 2015), and p-values already take chance findings due to randomization into consideration, which make it unnecessary to use balance tests to establish whether a sample is indeed balanced with respect to the background variables measured. Furthermore, covariates that are not actually related to the outcome 'decrease statistical precision since they fit noise in the data at the cost of degrees of freedom' (Vanhove, 2015: 139), which is why researchers have to limit themselves to a small number of background variables.

There are also practical problems that come with randomization in instructed settings. In his review of nine decades of classroom-based research in *The Modern Language Journal*, Chaudron (2001) laments the fact that most school contexts do not allow for the random sampling of subjects, or even random assignment into classes or groups; thus, 'intact groups are the norm' (Chaudron, 2001: 66–67). Use of intact classes (so-called group- or cluster-randomized interventions) – whether for convenience or to preserve ecological validity – impedes random group assignment. Thus, if randomization occurs, it often occurs not at the individual level but at a higher level in classroom research, which has dramatic consequences for the outcome: 'ignoring the fact that randomization took place at the group level drastically affects the insights gained from the study' (Vanhove, 2015: 142).

General linear models such as ANOVA cannot take account of the various unmeasured aspects of the upper level units (e.g. schools or classrooms) that affect all of the lower level measurements (e.g. measurements within subjects or students within classrooms) similarly for a given unit. Accordingly, a t-test (or, equivalently, an ANOVA) may well yield a statistically significant result when there is, in fact, no effect. This has to do with the fact that there are a variety of possible upper-level variance-covariance structures relevant to the relationships among the lower level units, e.g. the relationship between students within a classroom. This leads us to our next topic.

The notion of context in language learning

The classroom is a notoriously complex context. It is difficult to document and quantify classroom processes and classroom effects; however, it is indispensable to include reference to such processes and effects if differences in learner outcomes are to be adequately explained (Nunan, 2005: 232). Under classroom effects we understand a complex interplay between effects of individual characteristics including self-confidence, personality, emotion,

motivation, degrees of learners' control over their learning, perceived opportunity to communicate and willingness to communicate, and classroom environmental conditions such as topic, task, interlocutor, receptivity to the teacher and pedagogical approach, classroom dynamics and group size (see, for example, Borg, 2006; Cao, 2011; Dewaele, 2009; Kozaki & Ross, 2011; Walls *et al.*, 2002; Wen & Clément, 2003). Kumaravadivelu (2001) states:

> ... all pedagogy, like all politics, is local. To ignore local exigencies is to ignore lived experiences. ... [and that] ... language pedagogy, to be relevant, must be sensitive to a particular group of teachers teaching a particular group of learners pursuing a particular set of goals within a particular institutional context embedded in a particular sociocultural milieu. (Kumaravadivelu, 2001: 539)

According to Seltman (2009: 375), it thus seems likely that students within a classroom will be more similar to each other than to students in other classrooms due to whatever school level characteristics are measured (so-called cohort effects). In MacIntyre and Mercer's (2014) words, 'contexts in which language learning occurs are diverse, nuanced, and *they matter*' (MacIntyre & Mercer, 2014: 165, our emphasis).

The question now, of course, is how to operationalize such an ecological perspective on the age factor in foreign language classrooms, e.g. the interrelationship between variables interacting with starting age in class. For a variety of reasons, the general linear model cannot capture the complexity of contextual effects on individual learning. For instance, Chaudron (2010: 68) laments the 'inadequate attention to the unit of analysis (whether students, class groups, teachers, or schools) when the statistical inferences [in classroom studies between 1916 and 2000] have typically been made on the assumption that the individual subjects were the unit for error rates'. This is a serious problem, since 'ignoring even small degrees of interrelatedness within clusters can invalidate the analysis' (Vanhove, 2015: 142).

Considering the shortcomings of traditional quantitative methods outlined in the last three sections, the main task in quantitative age research is now to find a method that takes enough variability in the data into account in order to be able to maximize the generalizability of the findings in age-related research.

Centrality of time in research on the age factor

The final and perhaps most serious problem of traditional quantitative analysis currently practised in age-related research is the centrality of time in research on the age factor. Ortega and Iberri-Shea (2005: 26) suggested that many, if not all, fundamental issues concerning L2 learning that SLA researchers investigate are in part issues relating to 'time', and that any

claims about 'learning' (or development, progress, improvement, change, gains, and so on) can be most meaningfully interpreted only within a fully longitudinal perspective. Usually language researchers will not only want to assess whether the influence of the field effect generalizes beyond the participants sampled to the wider population, while taking into account any random variation observed, but also want to test if results generalize both to the wider population of people and the wider population of linguistic materials (see Cunnings & Finlayson, 2015). However, as Flynn and Foley (2009: 31) point out, longitudinal studies often have the characteristics of qualitative work, whereas studies with a more quantitative approach often use cross-sectional sampling. It is also important to add that SLA research has not been exactly to the fore in employing sophisticated procedures to analyse truly longitudinal data (Piniel & Csizér, 2014: 165). This is a serious limitation, particularly for age-related research, where, like in no other SLA domain, many questions are fundamentally questions of time and timing. For example, what do we know about the pace and pattern of L2 development throughout mandatory school time (in an instructional setting) or throughout the lifetime of L2 learners (in a naturalistic setting)? What critical transition points in L2/FL development need to be taken into account when planning educational policy for early versus late learners?

Given, then, the centrality of time in research on the age factor, more attention to longitudinal research practices is desirable and also to findings gleaned from longitudinal studies (see also Ortega & Iberri-Shea, 2005: 28). However, longitudinal data in age research are often analyzed by recourse to the same inferential statistics that are employed in cross-sectional research (t-tests, multivariate analysis, etc.). While ANOVA methods can provide a reasonable basis for a longitudinal analysis in cases where the study design is very simple, they have many shortcomings that have limited their usefulness in applications (see Fitzmaurice et al., 2009; Maxwell & Tiberio, 2007). For instance, in many longitudinal studies there is considerable variation among individuals in both the number and timing of measurements. As mentioned above, ANOVA cannot take account of such unbalanced data. Given this, Ortega and Iberri-Shea (2005: 41) caution that if 'more large-size longitudinal quantitative studies are conducted in SLA, it will be important to train ourselves in the use of statistical analytical options that are available specifically for use with longitudinal designs and data'.

Quantitative Research on the Age Factor II: Quo Vadis?

These findings, along with other suggestions for reform, point to the presence of weaknesses in quantitative research on the age factor. The good news is that even though general linear models – such as ANOVA, t-tests

or multiple regression models – are still widely used in second language research in general (see, for example, Cunnings, 2012; Cunnings & Finlayson, 2015; Plonsky, 2013, 2014), there is some evidence of an increase in statistical sophistication in terms of the types of analyses performed in age-related studies. For instance, the class of statistical models known as multilevel modelling (MLM) – a subclass of linear mixed-effects regression modelling (e.g. Baayen et al., 2008; Jaeger, 2008; Quené & van den Bergh, 2008) – appears to be increasing in this body of research. To illustrate the advantages of these models and the problematic nature of traditional analyses, let us compare a traditional multivariate analysis of variance described in Pfenninger (2014) with the multilevel data analysis in Pfenninger and Singleton (2017), using the same dataset. The following summarizes the main research question of these studies: what is the strength of the association between L3 English performance with starting age, on the one hand, and with type of instruction, on the other, in learners with a long learning experience (more than 10 years)? A total of 200 Swiss participants (89 males and 111 females; mean age 18;9) were recruited at the end of mandatory school time from 12 different classes in five different schools. In other words, the sampled students were nested in a hierarchical fashion within classes within schools. They were divided into four groups of 50 participants each according to AO and learning constellation in primary and secondary school. Among other tasks, each participant filled in 20 gaps in a listening comprehension task, which were later rated as correct or incorrect.

In Pfenninger (2014) these data were analyzed using a general linear model, i.e. two-tailed t-tests and multivariate analysis of variance (MANOVA), which means that the data were initially aggregated, averaging first over participants, i.e. the four groups, and secondly over the 20 items. That is, all of the measurements for a given age group category were assumed to have uncorrelated errors. The results of a two-tailed t-test for independent means and ANOVA revealed that there were significant differences between the listening skills of (a) the four groups ($F = 46.39$, $df = 3$, $p < 0.001$) and (b) the 100 early starters versus the 100 late starters ($t = -2.75$, $p = 0.006$). With respect to the impact of age, MANOVA indicated that listening comprehension reached statistical significance, with a small effect size ($\eta^2 = 0.038$), an earlier start emerging as advantageous. There was also a significant interaction between AO and type of instruction ($F = 7.89$, $df = 1$, $p = 0.005$, $\eta^2 = 0.024$).

However, assuming that measurements for a given age group category have uncorrelated errors is somewhat problematic, as it could be that performance correlates between students within the same class (and school) in a way that is not observed between different classes (and schools), and it would be beneficial to take such variance and covariance into account statistically in order not to maximize age effects (see discussion above). While correlated data are explicitly forbidden by the assumptions of standard

(between-subjects) (M)AN(C)OVA and regression models (see, for example, Seltman, 2009: 357), mixed-effects models were developed to shed light on precisely such situations (Goldstein, 1987, 1995; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). Rather than the data being averaged over the 50 participants per group and the 20 items, multilevel analyses require no prior aggregation and are run on unaveraged data. This takes into account: (1) compositional effects, which are known to mediate the trajectories of age differences in (growth of) proficiency; and (2) the fact that some participants may generally have higher scores than others in a particular task (and some participants might do well on all the items in a given task), and that some items may generally yield lower scores than others. Accordingly, in Pfenninger and Singleton (2017), a multilevel analysis of the same dataset was used, in which the independent variable of interest, 'age of onset', was taken to be a fixed effect (meaning that it was assumed that the effect *did not* vary randomly within the population of classes), while the participants sampled from a larger population of L2 learners and the language stimuli sampled from a much larger population of linguistic materials were the random factors. Furthermore, there were also significant random school and class effects for all dependent variables in this study. (Remember that the hierarchical structure of the data on all skills tested consisted of two levels: class (level 1), and school (level 2).) This made a significant difference in regard to results yielded by the dataset in question. When we subjected it to a multilevel analysis, which achieves adequate estimates of variances and therefore correct standard errors, correct inferences and correct (likelihood-based) *p*-values, there was no longer any sign of age effects for listening comprehension ($\beta = -2.20$, $SE = 0.80$, $t = -2.75$, $p = 0.130$). This illustrates how drastically clustering effects of streamed classes can minimize age effects.

Finally, note that the model above, which contains random intercepts, allows mean values for each participant and each item to vary. However, in theory, we also need to include random slopes, which take account of the fact that different classes and/or different items may vary with regard to how sensitive they are to the manipulation at hand. For instance, it could be that age effects are restricted to certain items or certain tasks (or certain classes), as stipulated by the idea that the age factor represents an individual difference variable (see discussion above). Whatever the effect of AO is, is it the same for all subjects, items, classes and schools? Furthermore, whereas AO might vary between classes and schools, in a longitudinal research design, the continuous predictor 'time' varies *within* them, as each student and each class and each school are tested at multiple points in time. As such, students and classes and schools may not only differ in overall average proficiency, but also in their sensitivity to the change in proficiency over time. Random slopes are required to model this type of variance (see Cunnings & Finlayson, 2015).

While we could not test for AO varying across classes in the study mentioned above due to the fact that early and late starters were not integrated in the same classes, MLM enabled us to investigate if AO worked similarly across settings (i.e. schools) and items in the task or whether it was influenced by characteristics of the setting and/or the items – and, if yes, whether there were school variables that could help us understand why those outcomes are different. In our case, likelihood ratio tests showed that school-specific, item-specific slopes for the fixed effect AO were not necessary for any dependent measure in this specific case (which is why we contented ourselves with random intercept models, see Pfenninger, forthcoming). This supports R. Ellis's (2006) idea of excluding age from his grouping of individual differences, as the effect of age was not different for different subjects or items (but cf. Pfenninger & Singleton, 2016, 2017). It also illustrates nicely that it is only through MLM that we can actually get a reliable estimate of classroom and school effects like that.

Another relatively recent study by Admiraal *et al.* (2006) also demonstrated the use of MLM with AO as one of the fixed effects. They analyzed the effects of the use of English as the language of instruction in the first four years of secondary education in the Netherlands on the students' language proficiency in English and Dutch, and on achievement in subject matter taught through English. The study involved 584 students participating in bilingual education and 721 students following a regular programme, who belong to one of four cohorts in one of five schools, three of which offered bilingual education. The participants were tested on two different occasions. Thus, the hierarchical structure for, as an example, receptive word knowledge, included four levels: occasion (i.e. the data of the dependent variables at Time 1 and Time 2), students, cohort and school. Multilevel analyses were performed using a multilevel repeated measures design. The vocabulary test scores were the dependent variables, school programme (bilingual education or regular education) and time (in terms of the number of months attending the school programme) were the independent variables, and student characteristics were the covariates. The analyses concerning the covariates were conducted separately for students' gender, their entry ability level, and language background information (home language, language contact, and motivation to learn English), respectively. By contrast, the hierarchical structure of the data on reading comprehension consisted only of three levels (student, cohort and school), since the reading comprehension test was administered only once, which meant there was no growth curve involved. Instruction effects for oral proficiency and reading comprehension were found, with bilingual education leading to better results, but there were no effects for receptive word knowledge. It is important to mention that MLM is also ideal in longitudinal designs that use shorter inter-measurement intervals than the studies mentioned in this chapter, i.e. studies in which change is expected to be ongoing or repeated rather than permanent or unidirectional.

Other researchers have used mixed-effects data analysis to focus on bio-logical age (rather than starting age) as a fixed effect. For instance, Haenni Hoti and Heinzmann (2012) used a multilevel model to compare the French listening and reading skills of two groups of Swiss learners (with previous English instruction, $n = 542$, and without previous English instruction, $n = 351$) in Grades 5 and 6, when students were approximately 11 and 12 years old. They controlled for a large number of other variables which might influence the scores on the achievement tests in French: biological age, gender, cantonal affiliation, nationality, length of residency in Switzerland, number of family languages, L1 spoken at home, literacy of the household, type of study plan (regular or special curriculum), metacognitive, cognitive and social learning strategies, motivation, self-concept as a learner of French, feelings of being overburdened and fear of making mistakes, attitudes towards French speakers and countries, parental assistance with learning French, and German reading skills. The study showed that the biological age of the learners played a role. With respect to listening skills, older learners' scores were significantly lower in the French listening test than the scores of learners who were younger at both measurement times. This study also showed that the educational background of the household in which the children live is important. Children of families with ample literacy resources as measured by the number of books at home (more than 100 books) dem-onstrated significantly higher listening skills after one year of French instruction than children of families with limited educational resources (less than 51 books) (Haenni Hoti & Heinzmann, 2012: 198). Thus, in Haenni Hoti and Heinzmann's dataset, the families and classes were not nested hierarchically (as in Pfenninger & Singleton, 2016) but are instead crossed at the same level of sampling, as the children came from different families. MLM can model such crossed random effects as well (Raudenbush, 1993), which is particularly important in naturalistic settings (as described in van Heuven, this volume).

It is also important to note that the fixed effects component of a multi-level model can not only feature age as a categorical factor (e.g. early AO versus late AO), a continuous predictor (e.g. chronological age or proficiency, if measured on a continuous scale), or a mixture of the two, but age can also function as a control variable if it is not of primary interest and we are mainly interested in assessing something else. One of the benefits of multi-level models is that properties of both the participants (such as chronological age or AO) and/or the items tested can be included in the analysis. Under traditional methods, the inclusion of such control predictors would involve various additional analyses (see Cunnings, 2012: 375) – for example, ANCOVA with age as a covariate – but these linear models would not take age effects on certain items or subjects into account. Finally, MLA can handle unbalanced data, where not everyone is necessarily measured at the exact same times, whereas the ANOVA design requires that all assessments at the

second data collection time be obtained at the same time for each individual (not some classes measured xx months after Time 1, and others measured after yy months).

In sum, MLM thus turns out to be a convenient method for obtaining ecologically valid tests of age effects. It reduces arbitrariness because it more closely reflects the power of situations as they are encountered in the stu-dents' daily lives: if the classes in our two AO groups differ in any important ways, then those between-class differences are going to contaminate our standard errors and thus our inferences (p-values). The bottom line is that we need to collect and correctly analyze data at multiple levels to try to sepa-rate out these kinds of effects – otherwise any differences found on the dependent measures may be due to uncontrolled differences among the par-ticipating groups rather than the main independent variable, e.g. AO or bio-logical age. MLM also capitalizes on the strength of each of the traditional methodological approaches to motivation, in its ability to incorporate indi-vidual difference effects, experimental group effects, and the study of behav-ioural variability and change through repeated measurement across situations and time (see Brown & Ryan, 2007).

Limitations of Quantitative Research

Despite the obvious advantages of recent methodological improve-ments in quantitative approaches to the age factor we have to bear in mind that by their nature all models are flawed (because they are idealized sim-plifications of Nature), but that some are clearly more useful than others (Seltman, 2009: 372). Other problems of even the most sophisticated models involve the interconnectedness of age with other variables (see, for example, de Bot, 2008) or the causal direction of relationships (see, for example, Dewaele, 2009; R. Ellis, 2006), both of which we will discuss in detail in Pfenninger and Singleton (2017). What is more, even multilevel models do not necessarily produce entirely accurate p-values (see Vanhove, 2015 for a discussion of this). Also, even though multilevel models may show that a student's performance in a task such as the English listening task described above is dependent on which class he or she is in, what exactly leads to these class differences (group dynamics, teacher personal-ity, quality of instruction, school environment, etc.) can often not be clarified with quantitative methods due to limitations of the research design. In this kind of context qualitative and quantitative research can complement each other. For example, when statistical results concerning the effects of causes are reported, the qualitative analysis is helpful for understanding e.g. the direction of causation *in a given specific case*. This enlightening complementarity is one reason why MLM research recom-mends itself.

Conclusions and Directions for Future Research

This chapter constitutes a preliminary step towards more fully understanding the use of multilevel models in the age factor research area. In the examples we have discussed in this chapter, we have attempted to show that multilevel models can provide a number of benefits to the age factor researcher, particularly in a classroom setting, where cluster-level randomization is the norm rather than the exception. To conclude, let us briefly sum up the six advantages of these models that we consider most important for our discussion here:

(1) multilevel models are ideal for a potentially generalizable study of age effects, since they permit multiple random factors (i.e. they can model different types of random effects structures that arise during random population sampling) and thus take account of both participant and item variability, allowing for 'the simultaneous generalization of the results on new items and new participants' (Gagné & Spalding, 2009: 25);
(2) they can be used for assessing the impact of context-varying factors on age;
(3) they can cope with measurements within and between sampled students that are nested in a hierarchical fashion within classes within cohorts within schools, as they flexibly give correct estimates of treatment and other fixed effects (e.g. age effects) in the presence of the correlated errors that arise from a data hierarchy (Seltman, 2009: 378);
(4) they are robust against missing data and imbalanced designs, obviating the need to replace missing values using debatable imputation techniques (Quené & van den Bergh, 2004, 2008), which is advantageous in longitudinal studies as well as experimental studies, in which L2 learners may exhibit high numbers of missing responses (Cunnings, 2012);
(5) they can model time effects (Goldstein, 1987, 1995), e.g. change and growth of EFL of different age groups over time and the success of different AO groups in longitudinal studies;
(6) they can handle interval-scale measures (e.g. age or proficiency, if measured on a continuous scale) through linear mixed-effects models (see Baayen et al., 2008) as well as categorical measures (e.g. early AO versus late AO) using mixed logit models (see Cunnings, 2012; Jaeger, 2008).

These models thus encourage us to shift from a myopic focus on a single factor such as the age factor to examining multiple relationships among a number of variables, including contextual variables; or in Brown's (2011) words: 'you are more likely to consider all parts of the picture at the same time, and might therefore see relationships between and among variables (all at once) that you might otherwise have missed or failed to understand' (Brown, 2011: 191–192).

We do not deny that the general linear model has several obvious attractions. However, the fact is that we often have hunches and theoretical grounds for believing there is more to a particular picture than these models reveal. For instance, more often than not, classroom and teacher effects have been neglected in quantitative research for practical reasons. Thus, in quantitative analysis as currently practised, there are problems not only for the age researcher but for anyone conducting classroom research on language learning. Looking forward, research on the age factor will best be served by a broadening and an enrichment of its methodological practices.

Acknowledgements

References

Admiraal, W., Westhoff, G. and de Bot, K. (2006) Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English. *Educational Research and Evaluation* 12 (1), 75–93.

Baayen H., Davidson D. and Bates, D. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 390–412.

Benson, P., Chik, A., Gao, X., Huang, J. and Wang, W. (2009) Qualitative research in language teaching and learning journals, 1997–2006. *Modern Language Journal* 93, 79–90.

Borg, S. (2006) The distinctive characteristics of foreign language teachers. *Language Teaching Research* 10, 3–32.

Boulton, A. (2011) Blending research methods: Qualitative and quantitative approaches to researching computer corpora for language learning. *Proceedings of KAMALL 2011: New Directions for Blended Learning in EFL* (pp. 1–9). Daejeon, South Korea: Pai Chai University.

Brown, J.D. (2011) Quantitative research in second language studies. In E. Hinkel (ed.) *Handbook of Research in Second Language Teaching and Learning* (Vol. 2.) (pp. 190–206). New York: Routledge.

Brown, K.W. and Ryan, R.M. (2007) Multilevel modeling of motivation: A self determination theory analysis of basic psychological needs. *Oxford Handbook of Methods in Positive Psychology* (pp. 530–541). Oxford: Oxford University Press.

Cao, Y. (2011) Investigating situational willingness to communicate within second language classrooms from an ecological perspective. *System* 39, 468–479.

Chaudron, C. (2001) Progress in language classroom research: Evidence from The Modern Language Journal, 1916–2000. *Modern Language Journal* 85, 57–76.

Cohen, J. (1968) Multiple regression as a general data-analytic system. *Psychological Bulletin* 70, 426–443.

Csizér, K. and Kormos, J. (2009) An investigation into the relationship of L2 motivation and cross-cultural contact among elementary school students. In M. Nikolov (ed.) *Language Learning Research: Young Learners* (pp. 62–75). Bristol: Multilingual Matters.

7   Language Policy in Ukraine: The Burdens of the Past and the Possibilities of the Future

István Csernicskó

Introduction

Several analyses have summarized the linguistic situation of Ukraine, highlighting various aspects of the problematic issues of Ukraine's language policy (Besters-Dilger, 2009; Maiboroda *et al.*, 2008 among others). The fundamental problems of the linguistic situation of Ukraine are the lack of consensus regarding the issue of what role the Ukrainian language has in constructing the new post-Soviet identity and in nation-building, and what status the Russian language should be given in Ukraine (Bowring, 2014; Korostelina, 2013; Kulyk, 2014; Polese, 2011; Zhurzhenko, 2014). Both researchers (Lozyns'kyi, 2008: 436; Pavlenko, 2008: 275; Stepanenko, 2003: 121; Ulasiuk, 2012: 47) and specialists of international organizations (e.g. HCNM, 2010: 2; Opinion, 2011: 7; UN, 2014) have repeatedly pointed out that the question of languages is heavily politicized in Ukraine, and the fact that it is not clearly settled can lead to the emergence of language ideologies as well as conflicts between ethnic groups and languages.

The aim of the present chapter is first to show how politicians in power have delayed solving the language issue in order to retain their power (see the second section). The events in Eastern Ukraine were not, naturally, caused by the language issue. However, in the third section I will demonstrate how the language issue has served as a pretext in the development of the conflict. In the following section, through the comparative analysis of four linguistic rights documents, I will show how the Ukrainian political elite attempted to maintain, between 1989 and 2014, the social equilibrium through introducing legislation aimed to regulate language use – unsuccessfully, as has been borne out by the events. In the final section I will outline a language policy model through which the language situation in Ukraine could potentially be unravelled.

A Periodization of Ukrainian Language Policy

Ukraine has had a longstanding tradition of political parties not delivering on their campaign promises after their election victories, particularly with regard to language issues, whatever those promises were; instead, attempts have been made to create an equilibrium between the regions of the country, divided into two language areas (Kulyk, 2006). Constantly aiming to balance the situation is one of the characteristic features of Ukrainian language policy (Stepanenko, 2003: 129).

Language policy aimed at creating Balance (1989–2004)

Ukraine's first President, Leonid Kravchuk (1991–1994), followed the strategy of careful balance in language policy. During his time in office he did not force Ukrainization, but he did pass many positions over to the nationally dedicated elite, which gained significant victories in Ukrainizing administration and education. He was often criticized for not taking any steps against the Russian domination of the written media and cultural life. While Kravchuk's slogan was 'there is no nation without a language', he often stressed that excessive Ukrainization would be a mistake (Besters-Dilger, 2011: 355–356). The time of his presidency was a period of direction seeking and legitimation in Ukrainian politics. The main goals of the time (right after gaining sovereignty without any struggle for independence) were to avoid denominational, ethnic and language-based conflicts, as well as to avoid becoming a 'cleft country' (Huntington, 1993).

Leonid Kuchma (1994–2004), President for two terms, was a master of political balancing. He was able to pass himself off as a protector of speakers of Ukrainian at some times and of speakers of Russian at others, depending on what his interests required (Kulyk, 2007: 309; Stepanenko, 2003: 129). In the 1994 presidential campaign against Kavchuk, who used heavy national rhetoric, he gained the support of the voters by promising to strengthen relations with Russia and making Russian an official language. At the time of his swearing in, he did not even speak Ukrainian – he learned it during his time in office. Then, in the 1991 presidential campaign, when his opponent was the communist Petro Symonenko, he proclaimed that Ukraine could have only one official language – Ukrainian. In one of his 1999 speeches he also stated that 'the Russian language cannot be foreign in Ukraine' and that 'we only have one official language, Ukrainian' (Kulyk, 2007: 308–309).

Another good example of Kuchma's shuttlecock policy is the Ukrainian Constitution passed in 1996.[1] Its Article 10 states that: '[t]he state language of Ukraine is the Ukrainian language. The State ensures the comprehensive development and functioning of the Ukrainian language in all spheres of social life throughout the entire territory of Ukraine.' However, according to the following paragraph, '[i]n Ukraine, the free development, use and

protection of Russian and of other languages of national minorities of Ukraine is guaranteed'. According to Bilaniuk (2010: 108): '[t]his wording was a compromise between the opposing camps of Ukrainophones and Russophones, but its ambiguity was used strategically by "centrist" political groups to de-emphasize the language issue, avoid a commitment to a clear course of action, and generally uphold the status quo.'

Article 92 of the Constitution places the settlement of the language issue under the jurisdiction of other laws. However, Kuchma's 10 years as President were not enough to modify the 1989 language law. The controversial situation of Ukraine's ratification of the European Charter for Regional or Minority Languages was also the result of political haggling (Kresina & Yavir, 2008: 190–196). Zhurzhenko (2014: 253) accuses Kuchma of establishing a laissez-faire language policy.

Orange language policy (2004–2010): Balance upset

In the years following the 2004 Orange Revolution, the most important goal of Ukraine's language policy became the practical implementation of Ukrainian as a state language. The political goal was to defuse the tension between the *de jure* situation (according to which Ukraine is a monolingual country) and the de facto situation (the reality where most of the population is multilingual). As President Viktor Yushchenko wrote in 2010: 'It can be stated that the survival of the Ukrainian state depends on the introduction of the Ukrainian language in all the spheres of the state and social life. Under the current circumstances, language is the guarantee of national security, territorial integrity, national consciousness, and the people's historical memory.'

In his campaigns, Yushchenko was unable to avoid policy aimed at achieving a balance as far as the language issue was concerned either. After the election, Yushchenko acted as a fighter for the Ukrainian language and national identity, even though during his presidential campaign he had talked about 'a historical compromise':

> The historical compromise lies in us, Ukrainophones admitting that the Russian language constitutes more for our society than just being the language of a national minority or of a neighboring state. Russophones, however, have to agree that the Ukrainian language, which was persecuted for centuries, has the right to be positively discriminated. (Shumlianskyi, 2006: 98)

The Orange forces led by Yushchenko and Yulia Tymoshenko were of the opinion that 'Ukraine inherited a deformed language situation' which the state language policy 'has to straighten out', regardless of whether there is a societal consensus backing this (Edict, 2010). That is, the situation believed

to be ideal had to be reached where Russian is only one of the many minority languages, and Ukrainian is omnipotent for all formal functions (Besters-Dilger, 2009: 9). Many politicians and intellectuals of national leanings talked about two Ukraines, regarding half of the country as being Ukrainian in its language and ways, and the other half being 'creole' (Riabchouk, 2003). People in the latter were perceived as Russified Ukrainians who 'had to be reconverted' to the Ukrainian language and nation (Bilaniuk, 2010: 116–117). Those who thought along these lines often categorized the population of the country into moral groups on an ethno-linguistic basis (Masenko, 2007: 57). Many considered people of Ukrainian ethnicity but of Russian mother tongue to be traitors or 'janissary' (Kulyk, 2001: 211; Osnach, 2015; Pavlenko, 2011: 48–49). Due to the 'deformed' language situation (Masenko, 2007: 7), part of the political and social elite regarded the strongest and widest propagation of the Ukrainian language as the primary language policy task of the state (Pavlenko, 2011: 50).

The central idea of this attitude was that the shared Ukrainian language was a special symbol of the newly formed and unified political nation and 'a means of strengthening the state's unity' (Edict, 2010). Thus, according to nationally inclined politicians, all those arguing for two state languages or believing that minority languages would be given official status were acting against the idea of a new Ukrainian state and a unified Ukrainian nation and state (Maiboroda & Panchuk, 2008: 207–209; Masenko, 2007: 11). They considered it all too natural that the Ukrainian people who gained state independence wanted to have Ukrainian monolingualism after a historically long period of forced and asymmetrical bilingualism (Shemshuchenko & Horbatenko, 2008: 168).

Among Ukraine's presidents, however, it was the winner of the Orange Revolution of 2004, Yushchenko, who turned out to be least effective in Ukrainization. Proof of his failure was his inability to pass a new language law that would have strengthened the position of the Ukrainian language. It was around this time that the pro-bilingualism Party of Regions became strong (Besters-Dilger, 2011: 353–361).

The Yanukovych era (2010–2014)

The Orange elite was defeated in the 2010 presidential elections; it was winner Viktor Yanukovych who, in his presidential campaign, promised to settle the status of the Russian language in Ukraine.

According to the campaign promises of his party, the Party of Regions, language policy would be based on the real linguistic situation, and the de facto bilingualism of the country would be codified through elevating Russian to the status of a state language along with Ukrainian. An important element of their argument was that forceful Ukrainization endangered the linguistic and nationality rights of the Russian-speaking population and

pushed Russian culture and language into the background (Bowring, 2014). But proponents of this view demanded the right of the free use of Russian not only in the southern and eastern regions but in the entire country, including regions where the proportion of the Russian speaking population is minimal.

The fears of those worried for the Ukrainian language were exacerbated also by the fact that, similarly to Kuchma before him, Yanukovych also started to learn Ukrainian only when he became a highly placed politician.[2]

Language Policy Becomes (One of) the Pretext(s) of the Conflict

Before the war

Following the tight election results of the 2010 presidential election (Yanukovych received 48.95% of the votes in the second round while Tymoshenko received 45.47%), the tense political language situation was difficult to ease, partly because maintaining the tension was in the interest of both sides of the political arena. Mobilizing voters for the local elections on 31 October 2010 was important for both.

In line with Ukrainian political traditions (Stepanenko, 2003; Zaremba & Rymarenko, 2008), the language issue was an important topic of the campaign at this time as well. Its activity rose especially when the 2012 parliamentary elections were approaching. In 2010 only about 2.43% of all political actions were related to the language issue; in 2011 1.80% were, whereas in 2012 this proportion rose to 10.45% (Ishchenko, 2013: 34).

A good indication of how motivated politicians were to keep the issue of language on the agenda was that in 2011 (a year without political elections in Ukraine) political parties and other political actors took part in organizing 46% of political actions that touched on the issue of language, while in the years of elections (2010 and 2012), the corresponding proportions were 64% and 66%, respectively (Ischenko, 2013: 36). In 2012 92% of the 380 events of social activity touching on language were in support of the Ukrainian language and against the elevation of the status of Russian (Ishchenko, 2013: 34).

Following its ascent to power through their parliamentary election victories in 2006, 2007 and 2012 as well as the presidential election victory of 2010, which made Yanukovych President, the Party of Regions got down to modifying the Constitution and the language law despite widespread protests. They lacked the political power to modify the Constitution, but managed to modify the Soviet era language law left over from 1989 in 2012. However, as became evident by the autumn of 2013, the new language law failed to settle the language issue in a satisfactory manner.

On 21 November 2013 it became apparent that the President of Ukraine – a country heading towards state bankruptcy at the time – was not going to sign the free trade agreement or the association agreement with the European Union in Vilnius. The Ukrainian government opted for the very favourable Russian loan rather than the IMF loan, which would have brought unpopular and strict austerity measures, threatening its power.

On 23 November protests for Ukraine's European integration started in Kyiv (Euromaidan). The protests organized in the city's main square were peaceful for a while and were known as the Revolution of Dignity. On 30 November an unreasonably brutal use of force by the police propelled the lukewarm protests into a national movement. The parliamentary majority backing the President modified several laws on 16 January 2014, in order to limit people's right of assembly. This triggered the protests to escalate to uncontrolled violence, resulting in many casualties as well.[3] The Parliament repealed the laws of 16 January on 28 January, and Prime Minister Mykola Azarov resigned. On 22 February Yanukovych fled the country.

A quick realignment occurred in the Parliament. Representing different parties from before, the same Members of Parliament formed a parliamentary majority which repealed the language law on 23 February 2014, on the initiation of Member of Parliament Vyacheslav Kyrylenko of the Fatherland party.[4] Russia immediately announced that it would defend the Russian-speaking minority of Ukraine and protect it from Ukrainian nationalism. In the territory of Crimea, which was transferred in 1954 as the Crimean Autonomous Republic from the Russian Federation to the Ukrainian Soviet Socialist Republic, military personnel from the Russian army bearing no insignia of their affiliation (called *little green men* in popular discourse)[5] appeared on the same day.

The steps taken to repeal the language law were not met with the disapproval of Russia alone. The Special Rapporteur of the UN stated: 'Steps to abolish the 2012 Law on the Principles of the State Language Policy, although vetoed, created anxiety amongst some communities, including ethnic Russians, who fear that that minority language rights will be eroded' (UN, 2014).

Intellectuals from the western Ukrainian city of Lviv addressed the Kyiv government in an open letter on 25 February as follows:

> We demand that the Highest Council, the newly appointed members of the government, and the temporary head of state lead a balanced cultural and language policy. [...] We must respect the cultural and linguistic needs of the population of southern and eastern Ukraine so they do not feel like aliens in the country.[6]

On the same day, President of the International Association for Ukrainian Studies, Michael Moser, also addressed an open letter to Ukraine's

politicians: 'Ukraine is a multi-ethnic and multilingual country, and that constitutes its richness. The Ukrainian language has to be a state language, but every language spoken in Ukraine has to be given as much protection as possible.'[7]

On 26 February the western Ukrainian city of Lviv and the southern Ukrainian city of Odessa announced, in bilingual Ukrainian-Russian language posters, a day of solidarity with the other half of the country for the next day. Lviv asked its population to speak Russian at home, at work, on public transport and everywhere else on this day, whereas Odessa asked its population to use Ukrainian all day.[8]

Temporarily filling the positions of both President and Speaker of the Parliament, Oleksandr Turchynov assessed the situation and decided, on 27 February, not to sign the document that would have repealed the language law of 2012, which thus remained in force.

A need for a new language law?

Instead of repealing the language law, Turchynov proposed that a new language law should be drafted,[9] and the Parliament decided to form a committee to draft such a law on 4 March.[10] Representatives of all parliamentary parties joined the 11-member committee, headed by Ruslan Koshulynskyi, one of the deputy heads of the Parliament and representative of the far-right Freedom party. The formation of the committee and the appointment of Koshulynskyi as its head were calculated steps. By forming the committee intended to draft the language law, Turchynov sent a message to both Ukrainian and Russian speakers. Those striving for the exclusive use of Ukrainian were reassured that the new political leadership did not wish to keep the language law associated with Yanukovych. By appointing the right-wing Koshulynskyi to head the committee, the political leadership also signalled that the strengthening of the position of Russian would not be the main aim of the new language law. At the same time, the Russian-speaking population was given reason for hope: the language law of 2012 remained in force, and the drafting of the new law was begun within a parliamentary framework.

The committee planned to meet a total of five times. Three of these meetings, however, were cancelled due to insufficiently low attendance to pass resolutions.

In its first meeting on 4 March, the committee decided to invite experts from all over Ukraine (including linguists, literary people and artists) recommended by the committee members.[11] Using Article 10 of the Ukrainian Constitution, it was also decided that the following documents would be used as the point of departure during the work: (a) the language law that was in effect between 1989 and 2012; (b) the European Charter for Regional or Minority Languages; (c) Ukraine's law ratifying the Charter; (d) the language law in force; (e) the text of a draft language law compiled by three

opposition parties on 10 January 2013[12]; and (f) a draft prepared by a working group headed by Leonid Kravchuk (but not submitted to the Parliament).[13]

In a meeting a week later the list of documents used as a basis for the new draft was extended with '[a]n outline of the state's language policy' (Edict, 2010), and the statements of the Venice Commission[14] and the OSCE's commissioner on minority matters[15] regarding the drafts of the 2012 language law.

At this meeting the committee members also decided, by majority vote, that they would not draft a new proposal for a law but would take the proposal drafted by Kravchuk's working group as a starting point for theirs.[16] This working group was formed at Yanukovych's request in late 2012 'to perfect' the new language law.[17] The reason for the formation of the working group was the series of protests that spread to much of the country in 2012 in opposition to the language law.[18]

The attempt to repeal the controversial 2012 language law became one of the pretexts of the Ukrainian crisis (Drozda, 2014; Osnach, 2015). This was, naturally, felt in the Ukrainian Parliament as well, and its members voted to release a statement, the 'Memorandum of understanding and peace' on 20 May 2014, which states the following regarding the status of languages:

> In parallel with the constitutional standing of the Ukrainian language as a state language, the Supreme Council of Ukraine guarantees ensuring the status of the Russian language. The state will likewise guarantee the support of the languages of national minorities in the territories compactly inhabited by the minorities.[19]

This memorandum, however, was too little too late: in April 2014 further internal political developments in Ukraine pushed the matter of the new language law into the background. An armed conflict erupted in the eastern Ukrainian Donetsk and Luhansk counties with the support of the Russian army. The war, euphemistically called 'the Ukrainian crisis' in international media, further exacerbated the extremely grave economic situation and unbalanced internal political relations of the country.

Wartime language policy

The eastern edge of Ukraine has been consumed by war since the spring of 2014. On 7 April the Donetsk People's Republic, then on 27 April the Luhansk People's Republic declared their independence. As a reaction to the manifestations of separatism and the armed incidents accompanying them, the Kyiv government started the war known as the 'Antiterrorist operation' ('Антитерористична операція', ATO) against the separatist rebels of eastern Ukraine and the unmarked Russian soldiers supporting them. On 11

May the separatist republics affirmed their sovereignty in referendums. Language policy, however, is assigned symbolic importance in Ukraine even amidst a war – as the events of February 2014 had already demonstrated. In this game, Ukraine's President Petro Poroshenko, voted in in the 25 May 2014, presidential election, was forced to follow some delicate balancing acts when following in the footsteps of his predecessors.

In the 26 October 2014 parliamentary elections, Poshenko's party, the Petro Poroshenko Bloc, finished second to the People's Front led by the then interim prime minister, Arseniy Yatsenyuk. As a result, the President does not have a secure backing even though his party has the largest faction in Parliament. Poroshenko tries to project the image of a self-confident President who stands up for the unity of the country and does not give in to efforts aimed to harm the national interests (and territory) of the country, even while he has to demonstrate to the mostly Russian-populated eastern and southern regions a readiness for compromise.

The President made gestures of goodwill to the Russian-speaking citizens of Ukraine in both his victory speech and his New Year's greeting, switching to Russian at certain points during both and stating that Ukraine is loved in Russian just as much as in Ukrainian.[20] According to Poroshenko, the decision of Parliament that deprived Russian of its status as an official language was wrong, because it made the language issue into a problem threatening national unity.[21] He supported the movement that tried to ease tensions in the crisis-hit country by displaying, on billboards, posters and leaflets, in the corner of the screen of national TV channels and in video clips,[22] the bilingual Ukrainian-Russian slogan 'Єдина країна – Единая страна' ('Unified country'), which even has a page of its own on one of the most popular social networking sites.[23]

This gives a reason for nationalist intellectual circles to keep attacking the President (Masenko & Horobets, 2015; Osnach, 2015). One of the best known Kyiv-based linguists, Larysa Masenko, stated the following on an internet portal where language-related issues are discussed regularly:

> The currently popular slogan 'Єдина країна – Единая страна' is faulty: it solidifies bilingualism on a country level, that is, it strengthens Russian as a second state language. In other words, it does not unify the country but divides it. By saying Единая страна in Russian we turn to those living in the east and assert that Ukrainian, which is a special symbol of the unification of the nation, is not obligatory, even for the President, and with this we betray Ukrainians, primarily those living in the east, who are loyal to their language despite the constant pressure from the Russian speaking environment. (Masenko & Orel, 2014)

At the same time as making goodwill gestures to Russian speakers, Poroshenko made several statements to reassure those living in the western

part of the country, firmly saying that the country has and will continue to have only one state language, Ukrainian.[24] It is quite symbolic that he repeated this statement on 11 February 2015, in the meeting of the extended cabinet,[25] just a few hours before leaving for Minsk, Belarus, to attend a decisive meeting with Chancellor Angela Merkel and Presidents François Hollande and Vladimir Putin about the chances of ceasefire and peace in Ukraine. The message was clear: the Ukrainian President may have to accept compromises in Minsk, but he will not compromise on issues of language.

Language Laws in Ukraine (1989–2015)

In this section I will compare Ukraine's four language laws:

- Law on Languages in the Ukrainian Soviet Socialist Republic (1989–2012).[26]
- Law on Ratification of European Charter for Regional or Minority Languages (2003–).[27]
- Law on the Principles of the State Language Policy (2012–).[28]
- Law on the Development and Use of Languages in Ukraine (draft – 2012/2014).[29]

By comparing these it will be possible to identify the main directions of Ukrainian language policy in the period between 1989 and 2014 and see why all parties – the speakers of Ukrainian, Russian, and of other languages of Ukraine – are dissatisfied with legislation regarding language.

The 1989 Law on Languages (LL1989)

The Law on Languages of 1989, passed before independence, constituted a compromise between Ukrainization and the preservation of the status quo (Besters-Dilger, 2011: 354). According to Kulyk (2006: 310), the law equally assisted Ukrainian nation building and the continuing presence of the Russian language in many domains of life. Ulasiuk (2012: 33–34) interpreted it thus: '[w]hile the law did not give Russian the status of the second state language, its position as a "language [of inter-national communication] of the peoples of the USSR" was confirmed.' However, Bilaniuk's (2003: 50) interpretation is that the Law on Languages took the 'first legal steps towards de-Sovietization and independence of the country in 1991'.

The Law of Ukraine 'On Ratification of European Charter for Regional or Minority Languages' (ECRML/UA)

Ukraine ratified the European Charter for Regional or Minority Languages for the first time in 1999.[30] However, the document did not come

into force. The law of ratification was repealed by the Constitutional Court of Ukraine (CCU) on 12 July 2000 (Ulasiuk, 2012: 36–37).[31] After that, several draft versions of the ratification law were developed (Kresina & Yavir, 2008: 196). However, the next ratification of the Charter happened only on 15 May 2003. The ratification document was deposited with the President of the European Council, and the Charter came into force in Ukraine as of 1 January 2006.

This law was heavily criticized (Bowring, 2014: 74), among other reasons, for the list of languages it protects and especially for including the Russian language in that list (Masenko, 2007: 38–40). Some pointed out that the Charter was ratified using a bad Ukrainian translation (Kresina & Yavir, 2008: 197), which misled MPs, who voted on a law protecting minority languages rather than endangered languages, even though in Ukraine it is not Russian but Ukrainian that requires protection. In 2004 46 Members of Parliament requested that the Law on the Ratification of the Charter be declared unconstitutional. According to them, the ratification of the Charter put unreasonable financial burdens on Ukraine, and this was not taken into account at the time of the ratification. The Constitutional Court, however, refused to discuss this request[32] (Kresina & Yavir, 2008: 200–201). Another opinion was that there should be a moratorium on the implementation of the Charter in Ukraine, and a new law on ratification should be drafted so that the list of languages to be protected could be re-examined and modified (Shemshuchenko & Horbatenko, 2008: 162).

According to the statement of Ukraine's Ministry of Justice issued on 10 May 2006, the faulty translation of the Charter caused political, legal and social problems in Ukraine.[33] The European Council's expert report[34] also mentions (on its page 4) that Ukraine wants a new translation of the Charter made. This, however, has not happened to date.

The Law on the Principles of the State Language Policy (LL2012)

The law on the Principles of the State Language Policy was passed after several attempts, with several political deals in the background and amidst scandalous circumstances, on 3 July 2012, and another long road led to its being signed by then Speaker of Parliament Volodymyr Lytvyn and President Yanukovych. It was finally published on 10 August 2012, in the official publication Голос України (Voice of Ukraine), thus coming into effect.[35]

Several attempts were made to have this law declared unconstitutional. In 2012 51 MPs requested that the Constitutional Court declare it unconstitutional. According to them, the articles of the law 'modify the framework in which the Ukrainian language can be used in Ukraine, [...] [and] contribute to regional languages gaining dominance over the state language'. According to the statement (No. 10-y/2013) of the Constitutional Court issued on 27 March 2013, the MPs' request 'voices presuppositions' and

'presuppositions cannot be considered to be appropriate arguments in support of declaring a legal document or its regulations as unconstitutional'.[36]

On 7 July 2014, 57 MPs again addressed the Constitutional Court. They argued[37] that: this law narrows the range in which the state language can be used; it allows for the state language to be bypassed in certain public situations; it protects only 18 languages and leaves out another 12 minority languages in a discriminative fashion; and the law was passed in a way that conflicted with the standing orders of the Parliament. Since the Constitutional Court did not discuss the MPs' petition, social activists protested in front of the Constitutional Court building on 18 June 2015, demanding that the law be declared unconstitutional.[38]

The Kravchuk's Draft (KD)

The parliamentary committee headed by Ruslan Koshulynskyi did not, in the end, produce a draft of a language law. In the 11 April session of Parliament Koshulynskyi officially announced that the committee would submit to Parliament the draft law proposed by Kravchuk in 2012. Members of the committee and several experts asked to present their opinions added critical commentary to the draft (Hodinka, 2014).[39] This prompted Koshulynskyi to say that the text of the draft would be sent to the Venice Commission, OSCE, the European Council and the representatives of all Ukraine's minorities for their comments.[40]

At the same time, the report of the Advisory Committee on the Framework Convention for the Protection of National Minorities made public the following report regarding the Ukrainian situation on 2 April 2014:

> With the present language legislation remaining in force, there is no immediate necessity to adopt amendments. Moreover, doing so could create considerable further tension in the current context. The Advisory Committee urges the authorities to refrain from moving too hastily in this field and to engage in a comprehensive and effective consultation process with representatives of all minorities before taking any further steps.[41]

A similar opinion was voiced in the statement by the OSCE High Commissioner on National Minorities, Astrid Thors, after her visit to Ukraine, on 7 April 2014:

> There can be no doubt that the 2012 Law on the 'Principles of the State Language Policy' remains in force, and, legally, nothing has changed regarding the status of languages. [...] Therefore, there is no need for the Ukrainian Parliament to rush to make any new decisions. I hope that all political forces can adhere to this opinion.[42]

A comparison of the laws

LL1989 [3] and ECRML/UA [2] regulate the use of Ukraine's minority languages, while LL2012 [7.2] and KD [5.1] define the rights of the speakers of languages regulated by the law.[43] The difference between the two approaches is crucial, since there are considerable differences between groups of the population by ethnicity, and in censuses a greater proportion of the population profess themselves to be of Ukrainian ethnicity than of Ukrainian mother tongue (see Table 7.1).

LL1989 protects the language of all national minorities in Ukraine – close to 130 languages.[44] ECMRL/UA deals with the languages of 13 national minorities. LL2012 codifies the rights of the speakers of 18 regional or

Table 7.1 The population of Ukraine by ethnicity[a] and mother tongue[b] (based on data from the 2001 census)

	By ethnicity	By mother tongue
Ukrainian	37,531,510	32,570,743
Russian	8,334,141	14,273,670
Belarusian	275 763	56,249
Moldavian	258,619	185,032
Crimean Tatar	248,193	231,382
Bulgarian	204,574	134,396
Hungarian	156,566	161,618
Romanian	150,989	142,671
Polish	144,130	19,195
Jewish (Yiddish)	103,591	3,307
Armenian	99,894	51,847
Greek	91,548	6,029
Gypsy	47,587	22,603
German	33,302	4,206
Gagauz	31,923	23,765
Slovak	6,397	2,768
Karaim	1,196	96
Krymchak	406	21
Rusyn	10,183	6,725
Other	510,390	143,142
No response	–	201,437
Total		**48,240,902**

Source: [a]http://2001.ukrcensus.gov.ua/results/general/nationality/; [b]http://ukrcensus.gov.ua/eng/notice/news.php?type=2&id1=21 (both accessed 14 June 2015).

Table 7.2 Languages covered by the four laws

LL1989	ECRML/UA	LL2012	KD
All languages of national minorities in Ukraine, i.e. approximately 130 languages (22.18% of the population).	**13** languages of national minorities: Russians, Belarusians, Moldavians, Crimean Tatars, Bulgarians, Hungarians, Romanians, Poles, Jews, Greeks, Germans, Gagauz, Slovaks (20.81% of the population).	Mother tongue speakers of **18** regional or minority languages: Russian, Belarusian, Bulgarian, Armenian, Gagauz, Yiddish, Crimean Tatar, Moldavian, German, Greek, Polish, Romani, Romanian, Slovak, Hungarian, Rusyn, Karaim and Krymchak (31.77% of the population).	Mother tongue speakers of **17** regional or minority languages: Russian, Belarusian, Bulgarian, Armenian, Gagauz, Yiddish, Crimean Tatar, Moldavian, German, Greek, Polish, Romani, Romanian, Slovak, Hungarian, Karaim and Krymchak (31.75% of the population).

minority languages in Ukraine. KD does the same with 17, leaving out Rusyn (cf. Table 7.2).

Whether a law protects the languages of national minorities or the rights of mother tongue speakers of specific languages makes a great difference. At the time of the 2001 census 5.5 million people professed to be of Ukrainian ethnicity but of Russian mother tongue (cf. Table 7.3). Thus, the proportion of the Russian minority in the population was 17.28%, whereas that of people of Russian mother tongue was much higher. The 5.5 million Russian mother tongue speaker Ukrainians are not part of the national minority, but based on native language they are part of the minority language speaker community.

LL1989 [2], LL2012 [6] and KD [4] alike name Ukrainian as the only state language (while ECRML/UA does not name one at all). LL1989 assigns to the Russian language the status of 'the language which is used for communication between the peoples of the Soviet Union' [4]. The special status of Russian is also emphasized by the 1996 Constitution (Stepanenko, 2003: 117). The other three documents do not assign a special status to Russian, but treat it as one of the minority languages spoken in Ukraine, even though historically, from the point of view of its functions in the country, and based on the number of its speakers, Russian stands out significantly among Ukraine's minority languages, most of which have a relatively small number of speakers. The ECRML's Committee of Experts also pointed out the special status of Russian in Ukraine (Ulasiuk, 2012: 38).

As Table 7.4 demonstrates, 22% of the country's population belong to one of the ethnic minorities, while 33% claim to speak a minority language. However, the table also reveals that the vast majority (78%) of the members

Table 7.3 The population of Ukraine according to mother tongue and ethnicity

Ethnicity and mother tongue	No. people	%
Ukrainians (by ethnicity) whose mother tongue is Ukrainian	31,970,728	66.27
Russians whose mother tongue is Ukrainian	328,152	0.68
National minorities whose mother tongue is Ukrainian	278,588	0.58
Total no. of those whose mother tongue is Ukrainian	**32,577,468**	**67.53**
Russians whose mother tongue is Russian	7,993,832	16.57
Ukrainians whose mother tongue is Russian	5,544,729	11.49
National minorities whose native language is Russian	735,109	1.52
Total no. of those whose mother tongue is Russian	**14,273,670**	**29.59**
National minorities whose ethnicity and mother tongue are the same	1,129,397	2.34
National minorities who speak the mother tongue of another minority group as their native language	260,367	0.54
Total no. of those who speak minority languages	**1,389,764**	**2.88**
TOTAL NUMBER OF SPEAKERS IN UKRAINE	**48,240,902**	**100**

Source: Based on data from the 2001 national census (from Csernicskó & Ferenc, 2010: 330).

Table 7.4 Minority citizens by ethnicity and native language in Ukraine

Minorities by	People (%)	Of which:	People	Ratio in total population (%)	Ratio in total minority population (%)
Ethnicity	10,699,209 (22.18%)	Russian	8,334,141	17.28	77.89
		Other minorities	2,365,068	4.90	22.11
Mother tongue	15,663,434 (32.47%)	Russian	14,273,670	29.59	91.13
		Other minorities	1,389,764	2.88	8.87

Source: Based on 2001 census data.

of these minorities are ethnic Russians. The ratio of Russian speakers is more dominant at 91%. After all, it is not surprising that the minority issue in Ukraine is almost synonymous with the issue of the Russian community; other ethnic and linguistic groups are rather insignificant in comparison.

Within the population of Ukraine, people of Ukrainian and Russian ethnicity together constitute 95.1%, whereas mother tongue speakers of Ukrainian and Russian together total 97.1%.

According to the Constitutional Court of Ukraine's official interpretation, No. 10-рн/99 dated 14 December 1999 of Article 10 of the Constitution of Ukraine, the only state language of Ukraine is Ukrainian, and the state language is also an official language.[45] At the same time, according to Para. 2, Point 1 of this interpretation, the fact that Ukraine has only one state language does not mean that only the Ukrainian language can be used as a public means of communication in official language functions: 'In addition to the state language, the Russian language and other minority languages can also be used in the work of the local administrative organs of the executive branch, the organs of the Crimean Autonomous Republic, and the local self-governments the Russian language and other minority languages can also be used as regulated by the effective laws of Ukraine.' Widening the range of the use of regional or minority languages does not in any way diminish the range of the use of the state language. As a state and official language of Ukraine, the Ukrainian language is used all over Ukraine in the work of organs of the legislative, executive and jurisdictional branches of administration, in international treaties, and in all educational institutions of all levels. The law also allows for the presence of minority languages in the public sphere under certain circumstances.

LL1989 allows for the use of the languages of national minorities in state offices in case the given national minority amounts to an absolute majority (i.e. higher than 50%) within an administrative unit [3] – which is a difficult demographic benchmark to reach for a minority (Bowring, 2009: 85). The use of the minority language is optional rather than obligatory even in this case.

According to LL2012, regional or minority languages can be used in state offices and local government offices if their members constitute at least 10% of the population of the administrative unit [7.3]. In these cases the law makes it obligatory to use the minority language in oral and written dealings between office workers and citizens. Local governments are required to publish their resolutions in the minority language as well as in Ukrainian [7.7, 10–11].

KD wished to raise the 10% benchmark of LL2012 to 30% [5.2], without the automatism involved in the application of the law. If the proportion of the speakers of a minority language in an administrative unit is over 30%, a group of initiators have to collect the signatures of 30% of the population to request that the law be applied. The collected signatures have to be sent to the county government, which in turn presents the request of the local government to the Parliament. The minority language is allowed to be used instead of the state language in local state offices only if the Parliament approves each request individually.

ECRML/UA does not define a demographic benchmark. Instead, it allows the use of regional or minority languages in the work of local or regional authorities where the number of regional or minority language speakers justifies this [4c].

ECRML/UA also does not define the local or regional authorities that can use the regional or minority languages. LL1989 [3] and LL2012 [1] precisely determine the levels of administrative units (county, town and village)

where minority languages can be used. No matter that the proportion of Russian mother tongue speakers in Ukraine is 29.59%, according to LL2012, on the national level the only state and official language is Ukrainian. KD would limit the range of use of the minority languages: it would only allow their use on the local level (but not on the level of county or district), and only if certain conditions are met [5.2].

In order for minority languages to be used in the work of local governments, however, administration officials need to have proficiency in these languages. LL1989 requires that the administration officials of the state have proficiency in Ukrainian and Russian. In addition, in those regions where national minorities constitute a majority, they have to know the language of the minority to the extent that they are able to carry out their work [6]. In ECRML/UA Ukraine undertook appointing, in regions where regional or minority languages are spoken, administration officials who had proficiency in these, as far as possible [4c]. LL2012 [11.3] and KD [9.3] do not unequivocally require administration officials to have proficiency in minority languages, but only to use them (under certain circumstances).

The 1999 resolution of the Constitutional Court requires the obligatory use of Ukrainian in the work of state administration and local governments in the entire territory of Ukraine. After ECRML/UA and LL2012 were passed, a frequently used argument was that regional and minority languages were displacing the state language from administrative offices. However, all the analyzed laws emphasize in several articles that the use of the state language is compulsory in the work of jurisdiction, the executive branch, local governments and education.

According to LL1989 [25] and LL2012 [20], the choice of language in education is an inalienable right; the latter guarantees education in the mother tongue at all levels of education from preschool to higher education. According to ECRML/UA, only those minority students are provided the right of education in their mother tongue whose family requests this and only if there is a sufficient number of students to form a class [4a]. According to KD, non-state educational institutions can decide about their language of instruction themselves [17.3]. Educational institutions run by the state can use a language other than Ukrainian as the language of instruction only in places where the proportion of mother tongue speakers of that minority language is over 30% and where the use of the minority language has been approved by the Parliament [17.1]. According to KD, then, no state-run school using Russian as a language of instruction can exist in the capital Kyiv, since the proportion of Russian speakers there was 25.27% at the time of the 2001 census.

What languages can be used for personal identification documents is an issue that ECRML/UA does not touch upon at all. According to LL1989, the ID card of Ukrainian citizens is bilingual: entries are made in the state language and Russian [14]. LL2012 requires that personal documents should have entries made in Ukrainian; in addition to that, at the special request of the individual, data can also be entered in one of the 18 regional or minority languages recognized by the law, whereas the document is in Ukrainian if the individual does not request otherwise [13.1]. According to KD, the language of personal documents is Ukrainian or another language prescribed by law [10.1].

Unreasoned Ukrainian language policy

LL1989 had already created relatively favourable conditions for the use of minority languages before independence. The next law, ECMRL/UA, brought about a considerable backward step in the field of language rights (Bowring, 2009: 69). LL2012, again, significantly increased the rights of the use of minority languages. However, KD greatly narrowed the language rights of the minorities not only compared to LL2012, but even compared to LL1989.

This tendency very much fits in with the traditions of Ukrainian language policy and language planning. With some simplification, Goodman (2009: 20) characterizes Ukrainian language policy as follows: 'At the level of national language planning, Ukrainian history can be characterized by waves of "Russification" or "Ukrainization" [...]. [O]fficial language policy in Ukraine has generally focused on promoting the public use of either Russian or Ukrainian to the exclusion or denigration of the other.'

However, Ukraine is not a state based on justice and integrity in the Western sense, and, similarly to other laws, laws regulating language use are not applied consistently either (Goodman, 2009: 22–23; Polese, 2011: 47; Shemshuchenko & Horbatenko, 2008: 169; Stepanenko, 2003: 118). The following example is a case in point. Pressured by nationally minded political forces, President Kuchma signed modifications of the law on television and radio broadcasting in 2006, which made conditions of language use stricter and raised the proportion of obligatory Ukrainian language broadcasting (Bilaniuk, 2010: 114–115). When pro-Russian political parties and the general public protested against this, the President announced that these passages were more like recommendations and did not have to be understood literally (Besters-Dilger, 2009: 281).

There are aspects lacking in the implementation of LL2012 as well. Even though it requires [10.1] that laws be passed in Ukrainian and translated into Russian and other regional or minority languages in Ukraine, even the official translations of the law itself into minority languages have not been made.

According to Stepanenko (2003: 129): '[t]he contradictory state language policy is formally oriented to the "one state, one language" model but, at the same time, the authorities actually conduct a policy of "language balance" between Ukraine's two main languages.'

In Polese's view (2011: 43), the inconsistent enforcement of language policy serves to avoid conflicts: 'a partial lack of enforcement of laws or rules seems to prompt an easier acceptance of a Ukrainian identity even by those who might be unable to acknowledge all the identity markers proposed by the state.'

The inconsistent enforcement of laws and the use of political strategizing have not brought about social peace but, instead, increased dissatisfaction with the language issue. The members of the two largest language communities, Ukrainian and Russian, see the situation of the two languages in cardinally opposite ways (Bowring, 2014; Stepanenko, 2003). According to Taranenko (2007: 138), '[t]he treatment of the present-day language situation and the language policy by pro-Ukrainian and pro-Russian intellectual and ideological elites is incompatible'. Efforts made to strengthen Ukrainian identity and the status of the titular language activated a mobilization response on behalf of the minorities. This resulted in a paradoxical situation: minorities are displeased with the guaranteed language rights, while the elite of the titular nation fears for the current and future status of the Ukrainian language (Zaremba & Rymarenko, 2008: 276). Korostelina (2013: 313) sees the struggle of identities and languages as a tension-generating zero sum game which makes reaching a compromise impossible.

It is in this complex situation that a viable model of language policy needs to be found in order to arrive at a peaceful solution.

What Kind of Language Policy Can Be Successful in Ukraine?

The language issue has become so politicized during the years of independence that it is impossible to carry out dispassionate dialogue about it. In 2014 the language issue became one of the pretexts that sparked off the Ukrainian crisis. What perspectives can language policy have in this situation?

Chances of codifying bilingualism

In this politically taut situation, there has been a centrist position between the two extremes, Ukrainization and Russification, present the whole time (Stepanenko, 2003: 121–122). In order to avoid linguistic and ethnic conflicts, its representatives believed that for Ukraine it is best to preserve the situation where Ukrainian is fulfilling the functions of a state language only partially, and where Russian is used practically as an official language in much of the country (Kulyk, 2007: 308–315; Shumlianskyi, 2006). However, since February 2014 this solution has not been a realistic and viable alternative.

Ukraine did not follow in the footsteps of Belarus, which codified bilingualism, or those of the three Baltic states, which also became independent after the dissolution of the Soviet Union (Polese, 2011: 41; Zhurzhenko, 2014: 253). According to Masenko, Belarus said 'yes' to Russian and was fully Russified, and Latvia expressed a clear 'no' and began developing its national language. Ukraine keeps saying 'don't know' to the challenge, giving way to the manipulation of the language issue (Masenko & Horobets, 2015).

Finland could serve as an obvious example for Ukraine, with its former state language, Swedish, being an official language along with Finnish at present. Despite the fact that Swedish as a mother tongue speakers constitute only 6% of the population, Finnish intellectuals could not take it upon themselves to treat Swedish as a minority language. Those who are afraid for the Ukrainian language, however, tend to keep the example of Ireland in mind (Bowring, 2014: 57–58; Yushchenko, 2010). For these reasons, the resolution of the language situation along the lines of Canada, Belgium, Switzerland or Spain is not a viable option for Ukraine.[46]

It is clear, however, that 'the reality in most of Ukraine is of bilingualism' (Bowring, 2014: 70). 'Almost everyone in Ukraine is bilingual, to varying degrees' (Bilaniuk, 2010: 109). According to data from the 2001 census, 80% of the adult population of the country speak (at least) one other language fluently in addition to their mother tongue; 87.84% of the population spoke Ukrainian, and 67.71% spoke Russian (Lozyns'kyi, 2008: 199–254). In spite of this, due to negative historical experiences, bilingualism is stigmatized in Ukraine (Bilaniuk & Melnyk, 2008; Goodman, 2009; Masenko, 2007; Pavlenko, 2011), and that makes codification of bilingualism impossible at the state level (Masenko & Horobets, 2015). 'Having one single state language is the condition to the existence of a developed European state. The introduction of two state languages in Ukraine does not serve consolidation. Consolidation is only possible along one thing', wrote Yushchenko (2010).

In the years before the passing of LL2012, research results demonstrated that there was a realistic chance of reaching a compromise according to which Ukrainian would remain the single state language while certain languages could be given the status of regional official languages in certain parts of Ukraine (Besters-Dilger, 2009; Maiboroda et al., 2008). LL2012 codified exactly this solution. However, the majority of the Ukrainian political and intellectual elite reject bilingualism even on a regional level (Masenko, 2007; Masenko & Orel, 2014).

A Utopian model of language policy

The events surrounding LL2012 demonstrate that codifying the bilingualism of certain regions did not result in consolidation. However, the model of the ethnocentric and monolingual nation state did not turn out to be viable either. The use of Western models cannot be successful, as there is no standard 'Western' model of language policy (Pavlenko, 2011). Several attempts have been made (Besters-Dilger, 2009; Maiboroda et al., 2008; Ulasiuk, 2012, among others) to find an alternative, but no matter which model is chosen, the issue of what roles should be assigned to the Ukrainian and Russian languages cannot be avoided.

It is essential to try and break with the view that approaches the issue through the majority versus minority dichotomy. The solution could lie in

regarding the rights of mother tongue use as rights that constitute basic human rights, accorded to the majority and the minorities alike, rather than just extra rights accorded to minorities (Andrássy, 2012). However, this is not something that countries passing language laws and international organizations with many member states make explicit. Emphasizing, usually in legal documents mentioning minorities (such as language laws), that minorities have the right to use and be educated in their own languages, implicitly suggests that these are extra rights which are accorded to minorities due to the positive discrimination given to minorities by the nurturing and democratic state. Through only emphasizing these rights with regard to minorities, however, states run by majority elites suggest that these are not rights that should also be shared by the members of the majority. But it is easy to see that, in order for all the citizens of a country to have the rights of the free use of their mother tongues, it is enough to state this in terms of a general right, without any reference to national minorities (Andrássy, 2012; Kontra, 1999). The practical issues of the implementation of this right have to be worked out in a detailed set of instructions.

The Ukrainian nation, naturally, has the right to use its own language in the entire territory of the country. However, instead of applying victimization due to historical grievances, it would be important for the rights of the other communities living in the country to be recognized as well. The successful operation of the suggested 'utopian' model only requires that Article 10 of the Ukrainian Constitution be consistently applied and a balance found between the support of the state language and the protection of the minority languages (including Russian). And if the focus of the debate is shifted to general linguistic human rights, it should be possible to avoid politicians polemicizing solely about the interrelationship of Ukrainian and Russian.

Such an interpretation of linguistic rights certainly seems utopian in Ukraine (Pavlenko, 2011: 45). But a century ago people could only hope that racism would one day cease to be the accepted ideology. Ukraine on the brink of breaking out of the crisis has the opportunity to spearhead the codification of general linguistic human rights and becoming a positive example rather than a negative one.

Conclusions

In this paper I have discussed Ukraine being a bilingual country in practice, where the Ukrainian and Russian languages are both used widely. In spite of this, the country's political elite regards assuring the dominance of the state language at the expense of Russian to be the basis of societal consolidation and of the new national identity, and considers the codification of *de jure* monolingualism to be the right direction for the country. Many are against according minority languages (e.g. Hungarian, Romanian, Gagauz,

Bulgarian or Russian) an official status, even in those regions where the speakers of these languages live in great numbers. This language policy necessarily results in conflicts.

Four language laws of the country from four different periods have been compared. It has been concluded that LL1989 and LL2012 brought about positive changes in codifying minority language rights. Both laws codified only one state language, Ukrainian, and required the obligatory use of Ukrainian in all formal domains. A part of the Ukrainian political and intellectual elite, however, considered the support for Ukrainian to be too weak, and that of the minority languages (especially of Russian) to be too strong. Because of this, ECRML/UA and KD define narrower rights for the speakers of minority languages than the previous laws. Taking away rights that were previously granted to minorities met with opposition on the part of Ukraine's minority communities, but at the same time also decreased the number of conflicts resulting from an inconsistent application of laws regulating minority language use. This inconsistent application law, which is unacceptable in a state based on the rule of the law, made it possible for the language issue to become a political stumbling block for all the election campaigns. This, in turn, has politicized the language issue, provided it with various symbolic meanings, and made it a pretext for the eruption of the crisis in late 2013 and early 2014 in Ukraine of political instability and on the verge of economic collapse.

It took the country a long time to get to this point. Following the Russification-centred policies of Czarist Russia and the Soviet Union at the time of independence in 1991, Ukraine opted for monolingualism in Ukrainian through adopting LL1989 and the new 1996 Constitution. However, Ukrainization was mild and inconsistent in the first years of independence. The Ukrainian political elite was satisfied with according the status of state language to Ukrainian but did little to actually support this state language. Ukraine attempted to move towards a balanced 'nationalisation of the state' (Brubaker, 1996), which was made more difficult by the lack of and inherent permeability of a clear boundary between Ukrainian and Russian culture, language and identity.

It seems, however, that the country's political elite considered ceasing and maintaining power to be their most important goal rather than the building of the monolingual state. Leaving the language issue open would have been necessary for maintaining the societal balance. Thus, gradual Ukrainization was most successful in the state administration and also in education (Bilaniuk & Melnyk, 2008; Polese, 2011), whereas the position of the Russian language remained strong in the media, the economy and culture (Bilaniuk, 2010).

The language policy following the 2004 Orange Revolution took a new direction, especially as far as education was concerned, which caused conflicts and triggered the opposition of not only the Russian minority but also of the other minorities (Csernicskó & Ferenc, 2010).

After the policy of Ukrainization, Yanukovych and his team strove to find a consensus of sorts by replacing the outdated LL1989 and passing LL2012, which effectively meant that they renounced the idea of making Russian a second state language and an official language in the whole of Ukraine. However, LL2012 met with considerable opposition from most of society. Due to the political embeddedness of the language issue and the symbolic meanings that it has acquired, this law could not become the basis of consolidation.

After Russia's annexation of the Crimean peninsula and amidst the military conflict raging in Eastern Ukraine, hopes for a compromise fade by the day. Division over the language issue increases further as the 2004 Orange Revolution and the 2014 change of power (both resulting in removing Yanukovych from power) have been regarded by the United States and the European Union as well as by about half of the population of Ukraine as democratic revolutions and as *coups d'état* resulting in overthrowing lawful political power by the other half (Darden, 2014).

A more or less satisfactory closure could only be possible if the political leadership gave up its goals of centralization and homogenization and found the desired unity in diversity. Whether the path leading to it is called decentralization, federalization or regionalization, Kyiv has to hand over some of its power – especially over education, language rights and the development of the economy – to the regions. Finding the constitutional framework necessary for the administrative reform of the country could be a significant step towards the rule of the law, democracy, societal consolidation, political stability, economic prosperity and, last but not least, peace. Regardless of which political party will lead Ukraine and how it divides the country administratively, its political elite will have to face the fact that Ukraine's population is ethnically and linguistically heterogeneous. The linguistic rights situation of the country will have to be normalized accordingly – and this is in the common interest of Ukrainians, Russians, and the other minorities.

The Uncertain Future

The European Union has *de jure* no official language, i.e. it has 24 equally official languages in 2016. Nevertheless, the economic and political capital behind English makes it the de facto dominant language in the EU. Phillipson (2016: 142) points out that as in practice, the language policy of the EU is controlled by market forces and background negotiations, the supremacy of English is inevitable. He also suggests that this *laissez-faire* language policy escalates the effect of market forces, which means that it reinforces the advantageous position of languages that are already leading.

In Ukraine, 25 years after the declaration of independence, in the majority of public language domains, it is still Russian that prevails, especially in the media and the economy; however, the use of Ukrainian is not exclusive even in public administration (Csernicskó, 2016a). In May 2016, the state of war enhanced national cohesion and the loyalty to Ukraine. Just like in the euphoria following the Orange Revolution, there is again a chance for a new Ukrainian political state to be born. Besides the 25-year independence of the Ukrainian state and the Orange Revolution in 2004 – which set patriotic feelings on fire – the Revolution of Dignity ending in the forced retreat of Yanukovych's regime as well as the war going on since the spring of 2004 in some parts of the country have been strengthening the position of the Ukrainian language. The higher proportions of Ukrainian native speakers and thus the expansion of Ukrainian language use are enhanced by the annexation of the mainly Russian-speaking Crimea and the fact that Kyiv has lost control over the primarily Russian-speaking counties of Donetsk and Luhansk. Tens of thousands of refugees have left the Crimea and the eastern regions of Ukraine inflicted by fights, several of whom have gone abroad. With consideration of the war, Poland, the Czech Republic, Greece and Hungary have been assisting their linguistic minorities in relocating (Csernicskó, 2016b: 120–123). Ukraine is therefore becoming increasingly more homogeneous from an ethnic and linguistic point of view.

At this time, it is still unknown how the above-mentioned processes will affect the ethnic and linguistic composition of the population, their collective identity, linguistic ideologies and the language policy (Zhurzhenko, 2014: 260–261), but one thing is certain: after peace has been established, the political elite will have to abandon its former ways in language policy. Instead of the controversial yet still *laissez-faire* language policy, they will have to provide real governmental support to the Ukrainian language in public administration and public life. Simultaneously, the use of regional and minority languages will have to be ensured in all situations where there is a need for it. The imbalance in favour of Russian – caused by political and economic power – must be counterweighted by a consistent language policy that regards both the Ukrainian language and linguistic diversity as values. Some kind of language policy is needed in which the support of the state language does not go hand in hand with the exclusion of minority languages.

According to Grin (2003: 30) language policy is a systematic, rational and scientific activity which aims at changing the linguistic environment in order to increase wellbeing. We believe that in Ukraine, which is now struggling with economic, political and military crisis, a science-based rational language policy should be implemented. In our paper we point out that language policy in Ukraine has not really focused on the creation of language balance throughout its short history. The main goal of language policy in the country is to serve the interests of opposing political powers. If we consider linguistic rights as general human rights, we can develop a language policy model based on the theoretical and practical aspects and findings of applied linguistics (see also Bülow & Harnisch, this volume), which (if implemented) may contribute to the settling of the language question in Ukraine. The

settling of the language problem, which intensifies emotions inside and outside the country, could be a small but crucial step in the direction of ending the armed conflict which transforms the ethnic and linguistic composition of Ukraine, threatens the security of Europe and pushes back the world's economic development. In this way, applied linguistics can contribute to the solution of a serious problem. The short- and long-term social effects and benefits are unquestionable.

Notes

(1) Конституція України. See http://zakon4.rada.gov.ua/laws/show/254%D0%BA/96-%D0%B2%D1%80 (accessed 21 June 2015).
(2) See http://tabloid.pravda.com.ua/focus/471e03a6885e3/ (accessed 25 June 2015).
(3) The casualties are remembered as 'Heaven's Hundred Heroes'.
(4) Ukraine abolishes law on languages of minorities, including Russian (23 February 2014). See http://rbth.co.uk/news/2014/02/23/ukraine_abolishes_law_on_languages_of_minorities_including_russian_34486.html. The text of law: http://w1.c1.rada.gov.ua/pls/zweb2/webproc4_1?pf3511=45291 (accessed 10 June 2015).
(5) On 16 March 2014, the population of the Crimean Peninsula expressed their wish in a referendum organized with Russian help to be joined to Russia, which annexed the peninsula in a treaty signed in Moscow on 18 March.
(6) See http://obozrevatel.com/politics/79534-intelligentsiya-lvova-vstupilas-za-rossijskij-yazyik.htm (accessed 16 June 2015).
(7) See http://www.mau-nau.org.ua/_private/novyny/novyny39.htm (accessed 16 June 2015).
(8) See http://www.pravda.com.ua/news/2014/02/26/7016309/ (accessed 16 June 2015). According to the 2001 census, 89.6% and 34.2% of the populations of Lviv and Odessa were Ukrainian speakers, respectively, whereas the proportions of Russian speakers were 9.7% and 64.8%, respectively.
(9) See http://iportal.rada.gov.ua/news/Novyny/Povidomlennya/88685.html (accessed 25 June 2015).
(10) See http://zakon4.rada.gov.ua/laws/show/851-18 (accessed 5 February 2015).
(11) See http://iportal.rada.gov.ua/news/Novyny/Povidomlennya/89010.html (accessed 9 February 2015).
(12) See http://w1.c1.rada.gov.ua/pls/zweb2/webproc4_1?pf3511=45410 (accessed 12 June 2015).
(13) See http://iportal.rada.gov.ua/news/Novyny/Povidomlennya/89010.html (accessed 9 February 2015).
(14) See, for instance, http://www.venice.coe.int/webforms/documents/?pdf=CDL(2011)010-e and http://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-AD(2011)047-e (accessed 12 June 2015).
(15) HCNM (2010).
(16) See http://iportal.rada.gov.ua/news/Novyny/Povidomlennya/89249.html (accessed 25 June 2015).
(17) See http://tyzhden.ua/News/57359 (accessed 25 June 2015).
(18) See, for instance, http://www.unian.ua/politics/669714-aktsiji-proti-movnogo-zakonu-vidbuvayutsya-v-bagatoh-mistah-ukrajini.html (accessed 25 June 2015).
(19) Постанова Верховної Ради України Про меморандум порозуміння й миру. See http://zakon3.rada.gov.ua/laws/show/1280-18 (accessed 25 June 2015).
(20) See http://www.pravda.com.ua/news/2014/08/23/7035607/ (accessed 11 February 2015).

(21) See http://korrespondent.net/ukraine/politics/3385352-reshenye-o-lyshenyy-russkoho-yazyka-statusa-rehyonalnoho-bylo-oshybkoi-poroshenko (accessed 25 June 2015).
(22) See https://www.youtube.com/watch?v=O_FfRvzZSv8 (accessed 25 June 2015).
(23) See http://vk.com/ukrone (accessed 25 June 2015).
(24) See http://www.president.gov.ua/en/news/status-ukrayinskoyi-movi-yak-yedinoyi-derzhavnoyi-bude-napov-38093 (accessed 13 September 2016).
(25) See http://www.pravda.com.ua/news/2015/02/11/7058200/ (accessed 25 June 2015).
(26) Закон України про мови в Українській РСР. See http://zakon4.rada.gov.ua/laws/show/8312-11 (accessed 25 June 2015).
(27) Закон України 'Про ратифікацію європейської хартії регіональних мов або мов меншин'. See http://zakon4.rada.gov.ua/laws/show/802-15 (accessed 25 June 2015).
(28) Закон України 'Про засади державної мовної політики.' See http://zakon2.rada.gov.ua/laws/show/5029-17 (accessed 25 June 2015).
(29) Закон України 'Про порядок застосування мов в Україні'. See http://search.ligazakon.ua/l_doc2.nsf/link1/NT0332.html (accessed 25 June 2015).
(30) Закон України 'Про ратифікацію Європейської хартії регіональних мов або мов меншин, 1992 р'. See http://zakon4.rada.gov.ua/laws/show/1350-14 (accessed 25 June 2015).
(31) See http://zakon.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=v009p710-00 (accessed 25 June 2015).
(32) See http://npa.iplex.com.ua/doc.php?uid=1077.1288.0 (accessed 14 June 2015).
(33) See http://old.minjust.gov.ua/7477 (accessed 14 June 2015).
(34) See http://www.coe.int/t/dg4/education/minlang/Report/EvaluationReports/UkraineECRML1_en.pdf (accessed 23 June 2015).
(35) Закон України 'Про засади державної мовної політики'. See http://zakon4.rada.gov.ua/laws/show/5029-17 (accessed 25 June 2015).
(36) See http://zakon4.rada.gov.ua/laws/show/v010u710-13 (accessed 23 June 2015).
(37) See http://ccu.gov.ua:8080/doccatalog/document?id=290768 (accessed 13 September 2016).
(38) See http://language-policy.info/2015/06/movnyj-zakon-maje-buty-vyznanym-nekonstytutsijnym-aktyvisty/ (accessed 23 June 2015).
(39) See http://iportal.rada.gov.ua/news/Novyny/Povidomlennya/90599.html (accessed 9 June 2015).
(40) See http://iportal.rada.gov.ua/meeting/stenogr/show/5239.html (accessed 25 June 2015).
(41) See http://www.refworld.org/docid/534258f84.html and https://wcd.coe.int/ViewDoc.jsp?id=2178937&Site=CM (accessed 5 June 2015).
(42) See http://www.osce.org/hcnm/117175 (accessed 12 June 2015).
(43) In this section numbers given in square brackets refer to the article where the given law regulates a given point. For instance, [7.2] refers to Article 7, Section 2.
(44) According to the 2001 census, the representatives of more than 130 national minorities reside in Ukraine.
(45) See http://zakon.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=v010p710-99 (accessed 25 June 2015).
(46) The separatist intentions of Quebec or Catalonia as well as the Flemish–Walloon debates in Belgium strengthen Ukrainian fears of the multilingual models.

References

Andrássy, G. (2012) Freedom of language: A universal human right to be recognised. International Journal on Minority and Group Rights 19 (2), 195–232.
Besters-Dilger, J. (ed.) (2009) Language Policy and Language Situation in Ukraine: Analysis and Recommendations. Frankfurt am Main: Peter Lang.

Osnach, S. (2015) Мовна складова гібридної війни [The linguistic components of the hybrid war]. *Портал мовної політики*, 13 June. See http://language-policy.info/2015/06/serhij-osnach-movna-skladova-hibrydnoji-vijny/ (accessed 6 October 2015).

Pavlenko, A. (2008) Multilingualism in post-Soviet countries: Language revival, Language removal, and sociolinguistic theory. *International Journal of Bilingual Education and Bilingualism* 11 (3–4), 275–314.

Pavlenko, A. (2011) Language rights versus speakers' rights: On the applicability of Western language rights approaches in Eastern European contexts. *Language Policy* 10, 37–58.

Phillipson, R. (2016) Linguistic imperialism of and in the European Union. In H. Behr and Y. Stivachtis (eds) *Revisiting the European Union as an Empire* (pp. 134–163). London and New York: Routledge.

Polese, A. (2011) Language and identity in Ukraine: Was it really nation-building? *Studies of Transition States and Societies* 3 (3), 36–50.

Riabchouk, M. (2003) *Дві України: Реальні межі, віртуальні війни* [*Two Ukraines: Real Boundaries, Virtual Wars*]. Kyiv: Krytyka.

Shemshuchenko, Y. and Horbatenko, V. (2008) Законодавство про мови в Україні: хронологічний моніторинг, класифікація, понятійна основа [Legislation on languages in Ukraine: Chronological monitoring, classification, theoretical concepts]. In O. Maiboroda, M. Shulha, V. Horbatenko, B. Azhniuk, L. Nahorna, Y. Shapoval, V. Kotygorenko, M. Panchuk and V. Perevezij (eds) *Мовна ситуація в Україні: між конфліктом і консенсусом* [*Linguistic Situation in Ukraine: Between Conflict and Consensus*] (pp. 157–173). Kyiv: Istitut politychnykh i etnonacionalnykh doslidzhen imeni I. F. Kurasa NAN Ukrainy.

Shumlianskyi, S. (2006) 'Мовне питання' після парламентських виборів: від політичної кон'юктури до державної політики [The 'linguistic question' after parliamentary elections: From political conjuncture to state politics]. *Політичний менеджмент* 3, 97–104.

Stepanenko, V. (2003) Identities and language politics in Ukraine: The challenges of nation-state building. In F. Daftary and F. Grin (eds) *Nation-Building Ethnicity and Language Politics in Transition Countries* (pp. 109–135). Budapest: Local Government and Public Service Reform Initiative – Open Society Institute.

Taranenko, O. (2007) Ukrainian and Russian in contact: Attraction and estrangement. *International Journal of the Sociology of Language* 183, 119–140.

Ulasiuk, I. (2012) The Ukrainian language: What does the future hold? (A legal perspective). In A. Milian-Massana (ed.) *Language Law and Legal Challenges in Medium-Sized Language Communities. A Comparative Perspective* (pp. 25–51). Barcelona: Institut d'Estudis Autonòmics.

UN (2014) *Ukraine: UN Special Rapporteur Urges Stronger Minority Rights Guarantees to Defuse Tensions*, 16 April, Geneva. See http://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=14520 (accessed 13 September 2016).

Yushchenko, V. (2010) Чия мова – того й влада [Whose language – that power]. *День* No. 180, 6 October. See http://www.day.kiev.ua/uk/article/panorama-dnya/chiya-mova-togo-y-vlada (accessed 6 October 2015).

Zaremba, O. and Rymarenko, S. (2008) Роль зовнішніх чинників у політизації мовних проблем [The role of external factors in the politization of linguistic problems]. In O. Maiboroda, M. Shulha, V. Horbatenko, B. Azhniuk, L. Nahorna, Y. Shapoval, V. Kotygorenko, M. Panchuk and V. Perevezij (eds) *Мовна ситуація в Україні: між конфліктом і консенсусом* [*Linguistic Situation in Ukraine: Between Conflict and Consensus*] (pp. 258–280). Kyiv: Istitut politychnykh i etnonacionalnykh doslidzhen imeni I. F. Kurasa NAN Ukrainy.

Zhurzhenko, T. (2014) A divided nation? Reconsidering the role of identity politics in the Ukraine crisis. *Die Friedens-Warte: Journal of International Peace and Organization* 89 (1–2), 249–267.

8   The Reanalysis of -end as Marker for Gender-sensitive Language Use and What This Implies for the Future Expression of Gender Equality

Lars Bülow and Rüdiger Harnisch

Introduction

In this chapter, we look at future directions in undoing gender in German language use. Regarding language use and communication, undoing gender is, in sociological terms, a strategy to make gender categories irrelevant (see Hirschauer, 1994, 2001). Undoing gender is – similarly to doing gender – one possibility for meeting the criteria of gender-sensitive language use. Making language use more gender sensitive in the future is one of the major goals of gender mainstreaming language policy. First and foremost, this means in German-speaking countries that gender mainstreaming activists fight against the use of generic masculine forms in public language.

Let us make two preliminary remarks. First, gender is not necessarily a binary category. Queer linguistics in particular stress that gender is intimately connected to (many) sexualities. However, there is still a strong social ideology of gender as referred to as a binary category. As Coates (2013: 295) states: 'Even more importantly, we have come through the post-modern battles which threatened to make binaries taboo. While it is not true to say that there is now consensus, there is a sense that a more pragmatic approach needs to prevail.' Secondly, to avoid confusion, we will use the German term 'Genus' instead of 'grammatical gender'. It is important not to mix up Genus and gender. Genus is a purely grammatical category (see Leiss, 1994). However, we have to distinguish carefully between a (social) grammatical category and a category which refers to the (social/biological) gender of

persons. Therefore, we use MASC, FEM and NEUT to indicate the grammatical category 'Genus'. We use *m* and *w* to indicate the (biological/social) gender of persons: *m* is the abbreviation for *men* (or the male), *w* is the abbreviation for *women* (or the female). We decided to use *w* in order to align it more closely with the German *weiblich* and also clearly distinguish it from the grammatical category of the feminine (FEM), which is itself sometimes abbreviated as *f*.

In German, nouns have an inherent Genus: they are masculine (*der Hund* 'the dog'), feminine (*die Katze* 'the cat') or neuter (*das Schwein* 'the pig'). According to feminist theory, Genus influences the way we conceptualize the referent (see Hornscheidt, 2006; Lakoff, 1975; Pusch, 1997a). If we read a masculine role noun like *der Student* ('the student') or *der Professor* ('the professor'), we probably conceptualize the referent as a male student or a male professor, even in contexts where it is used generically. Generally, generic masculine personal nouns and pronouns are traditionally considered gender neutral. They are used in situations where the biological or social gender of the subject does not matter, is unclear, variable, or when a group to which they refer contains different genders. However, feminist and gender linguists profess that '[w]e have reached the point where the generic masculine is not what it once was. Women are refusing to feel included by words like *Wähler* ["voter"], *Ballettänzer* ["ballet dancer"], or *Zuhörer* ["listener"]' (Pusch, 1997b: 328). People who are aware of feminist thinking try to avoid such generic masculine forms. One recommended and frequently used strategy to prevent generic masculine forms is to neutralize the Genus by using nominalized participles (*die Studierenden* 'the students'). The question arises, however, as to whether this neutralization of Genus through the use of nominalized participles helps the readers and hearers to associate women and men more equally. Does this strategy really lead to more gender-fair language-based associations? There is no simple answer to this, considering that gender associations may also strongly depend on context and our knowledge about the world (see Braun *et al.*, 1998). Even if language change has an influence on our future associations, the interplay between grammar, context and cognition is complex and hard to predict.

This chapter is structured as follows. After examining the literature on how Genus may affect our conceptualizations of a referent, we briefly outline the main strategies which are used to produce gender-sensitive texts in German. Then we exemplify the emergence of a new grammatical marker in order to express gender-sensitive language use. This new marker for gender sensitivity, the participle suffix -*end*, has become dominant in specific domains such as texts at universities and in other public spaces (Bülow & Harnisch, 2015; Bülow & Herz, 2014). To investigate to what extent students actually associate such generic masculine nominalized participles with male or female gender, we then present the results of a text-based association study. Our findings suggest that Genus is a much weaker predictor in gender

association than the context in which referents are presented. Moreover, we found no significant differences between Genus-neutral nominalized participles (*die Studierenden* (PL) 'the students') and generic masculine nominalized participles (*der Studierende* (SG.MASC) 'the student'). In neutral contexts the forms associate referents with female and male gender to the same extent. Future directions would be to see to what extent the 'generic masculine' might actually be an accepted way of undoing gender.

Background

Some linguists hold that thought is shaped by language structure (von Humboldt, 1994; Pavlenko, 2014; Whorf, 1956). This idea is strongly associated with the Sapir–Whorf hypothesis. While its strong interpretation in which our thoughts are determined by our language has been abandoned in the meantime, there is nevertheless some evidence for a less deterministic reading that language somehow influences our thoughts. Boroditsky *et al.* (2003: 75) and Fausey and Boroditzky (2010), for example, found that grammatical features in language have the power to bias people's behaviour. In particular, Boroditzky *et al.* (2003) point out that the grammatical category *Genus* of inanimate objects affects people's associations. German participants were asked to write down in English (which is a 'natural gender language' with no Genus distinction) 'the first three adjectives that came to mind to describe each object' (Boroditzky *et al.*, 2003: 69) on a prepared object list. For instance, they labelled a key as *hard, heavy, metal* and *serrated* (Boroditzky *et al.*, 2003: 70). Afterwards, a group of native speakers of English had to rate the adjectives as describing feminine or masculine properties of the object. The result was that the speakers associated 'adjectives that were rated more masculine for items whose names were grammatically masculine' in German (Boroditzky *et al.*, 2003: 70).

The assumption that Genus affects our associations has also been claimed by feminist linguists (Key, 1975; Lakoff, 1975; Pusch, 1997a). Moreover, they presuppose a strong link between language use and social reality. Feminist sociolinguists and gender linguists suppose that language use has a twofold effect. On the one hand, it mirrors social reality and, on the other, it is supposed to be the central medium with which to construct this reality (cf. Bußmann, 2005: 484). Pusch (1997a: 326) reasons that '[l]anguage produces concepts, concepts influence our actions, actions influence our political and economic situation (our so-called reality), and this in turn influences language'. These scholars also assume that our gender categories are constructed (exclusively) in social interaction and that we must take language use into consideration if we want to improve conditions for women.

In the German language, the generic masculine (*Studenten* (MASC_PL) 'students') is believed to be the main culprit. It is assumed to represent the

'most far-reaching and most fundamental asymmetry' (Stahlberg *et al.*, 2007: 169) because it equates maleness with humanness. The generic masculine is believed to demarcate a masculine gender when referring to people (Braun *et al.*, 1998; Irmen & Steiger, 2005; Steiger-Loerbroks & Stockhausen, 2014). Therefore, the generic masculine 'eliminates women'; it is sexist and part of the 'patriarchal grammar' (Pusch, 1997a: 325). Furthermore, Pusch (1997a: 325) states that we are only 'picturing men', which is why 'women have all but disappeared from the human imagination'. Generic masculine forms are supposed to make it more difficult for women and girls to visualize a female role model (e.g. as a doctor or as a scientist). Consequently, feminist and gender linguists promote language change. They demand avoiding generic masculine forms in all contexts and insist on using forms that men and women can be equally associated and visualized with.

This kind of language change is politically motivated and desired (for a similar discussion, see Csernicskó, this volume). Furthermore, it is top-down, planned and intentional (cf. Hornscheidt, 2006: 288). Pusch (1997a: 323) summarizes that '[t]he process of language change set in motion by women during the last twenty years is the most significant and far-reaching linguistic innovation of the century'. At this point, it is important to evaluate the consequences of this politically motivated language change. We have to come up with empirically valid data to legitimize and challenge language changes and language change strategies envisaged and planned in the future. By now, the use of gender-sensitive language in legal and official documents 'has become a legally binding standard' (Steiger-Loerbroks & Stockhausen, 2014: 58). In 2013, for example, the German Road Traffic Act ('Straßenverkehrsordnung, StVO') was reformulated in a gender-sensitive way (see Bülow & Herz, 2014). Another prominent context in which a significant language change is still taking place is at German universities. Students and staff are keenly aware of speaking and writing in a gender-sensitive way. Today, women's and equal opportunities representatives make sure that all official documents (like study regulations or job postings) are written in line with gender-sensitive standards.

However, the great number of guidelines show that it is not intuitively clear and easy to speak and write gender sensitively. Authors also have to pay attention to several other requirements if they want to write appropriate texts. For example, '[...] legal texts must meet the requirements of being comprehensible, clear, simple, and succinct' (Steiger-Loerbroks & Stockhausen, 2014: 58).

The guidelines for using gender-sensitive language basically recommend two strategies to equally associate men and women at the same time. One is concerned with doing gender: e.g. by using splitting syntagmata like *Studentinnen und Studenten* 'students (w) and students (m)' or short splitting like *Student/in* 'student' (m/w). The other strategy is undoing gender: e.g. by using gender-unmarked role nouns like *Person* 'person' or Genus-neutral nominalized participles in the plural like *die Studierenden* 'the students' ('the studying-[one-]s').

It is assumed that using such strategies helps readers to associate a referent noun with men and women more equally. Most guidelines refer to empirical studies which, in general, worked out that the generic masculine strengthens the association of male gender (Irmen & Köhncke, 1996; Klein, 1988; Scheele & Gauler, 1993). Empirical research on the generic masculine started in the 1980s (see Stahlberg *et al.*, 2007: 176). Stahlberg *et al.* (2007: 171–180) give an overview of empirical studies concerning the question as to 'whether masculine forms are really associated with males and thus put women at a disadvantage'. However, a lot of the empirical studies have weaknesses, such as the ones summarized in Braun *et al.* (1998). Because the methodology of our own study is partly based on Braun *et al.*, we would like to give a short overview of their research design and results.

Braun *et al.* (1998) investigated the question of the extent to which people associate Genus with gender roles in German. They conducted a text-based association study, taking the context of the texts into consideration. Braun *et al.* first asked people to evaluate contexts as typically male (=+3), female (=−3) or neutral (=0). The independent variable, which was tested in two settings, was the linguistic framing by three variants in two contexts. They framed each setting version (scientific congress, meeting of a sports federation), which contains two context versions (typically female, typically male) by: (a) a generic masculine form (*Student* (MASC.SG) 'student'; *Sportler* (MASC.SG) 'athlete'); (b) a splitting form (*Studentinnen* (FEM.PL.) *und Studenten* (MASC.PL); *Sportlerinnen* (FEM.PL.) *und Sportler* (MASC.PL)); and (c) a gender neutral form (*Studierende* (PL); *sportlich Aktive* (PL)) (Braun *et al.*, 1998: 270). Each text was designed like a real newspaper article. Participants were told that they were taking part in a linguistic experiment on informational content in newspaper articles. They used a paper questionnaire including open-ended and closed-ended questions. Some questions were related to the use of Genus in the newspaper articles and some served for the credibility of the cover story. Integrated in these questions was the following target question: 'Wieviel Prozent waren Frauen (bzw. Männer)?' ('How many percent were women (or men)?'). The authors predicted 'a continuous increase of associated women over the linguistic form from the generic masculine over the neutral forms to the splitting syntagmata' (Braun *et al.*, 1998: 274, trans. LB, for German original quote see Appendix Quote 1).

Braun *et al.* observed that the context has a huge impact on the extent to which women and men will be associated with a referent, which is not surprising, considering that the context such as the described social situation or evaluating adjectives shape our associations. Context is here seen as both a linguistic configuration and a social constellation. However, the context factor has so far often been neglected (see, for example, Irmen & Köhncke, 1996; Scheele & Gauler, 1993). Furthermore, Braun *et al.* (1998) found that

women were most frequently associated in the splitting syntagmata frame versions. Regarding the two other conditions, they found no clear effect. In the scientific congress setting, more women were associated in the generic masculine frame version than in the neutral frame version. In the sports federation setting, more women were associated in the neutral frame version than in the generic masculine frame version. That is why they conclude their argumentation as follows:

> A gender-neutral linguistic form is no alternative to the generic masculine, because it has hardly any influences on the increase of association of 'female'. If language should be used as a medium to reach gender equality, then splitting syntagmata is the best strategy to realize it. (Braun *et al.*, 1998: 281, trans. LB, for German original quote see Appendix Quote 2)

Braun *et al.* (1998: 281) added to their results that splitting syntagmata like *die Studentinnen und Studenten* can be 'schwerfällig' ('cumbersome') and 'schwer verständlich' ('difficult to understand'). Moreover, (short) splitting and using gender-unmarked role nouns are not the most aesthetically pleasing solutions. They may be detrimental to the text in that they take too much time to read or produce because they are not standard solutions for resolving communicative problems (e.g. short-splitting is a particular problem for machine readability). Still, feminists and gender linguists demand creative solutions to these issues (see Hellinger, 2004; Steiger, 2008). However, creative solutions require more processing effort as it is cognitively expensive to find a creative strategy to avoid the generic masculine.

A less creative but highly economic solution is to use the Genus-neutral nominalized participles like *Studierende* ('students') to write and speak gender sensitively. Nominalized participles are conspicuously marked by the *-end* suffix (*Studier-end-e*), whose usual function is to express that someone or something is 'in the process of doing something' (cf. Elsen, 2011: 106, trans. LB). However, there is evidence that the *-end* suffix has now adopted a new function in gender-sensitive writing. Bülow and Herz (2014), for instance, found that the authors of the reformulated German Road Traffic Act (2013) transformed more than 30 generic masculine forms like *Fußgänger* ('pedestrians') or *Radfahrer* ('cyclists') into nominalized participles such as *zu Fuß Gehende* ('those who go by foot', 'pedestrians') and *Rad Fahrende* ('those who cycle by bike', 'cyclists') to meet the requirement of gender sensitivity. What is more, Harnisch (2016) found that in many study regulations, generic masculine forms like *Studenten* ('students') or *Dozenten* ('lecturers') had been transformed into Genus-neutral nominalized participles like *Studierende* ('students') and *Dozierende* ('lecturers') to make the texts gender sensitive. What is even more interesting is that Harnisch (2016), Bülow (2016) and (Bülow & Harnisch, 2015) have found that many authors use nominalized participles in the singular (*der Studierende* (MASK.SG) 'the student').

Apparently, authors may not be aware that nominalized participles in the singular are actually Genus-specific rather than Genus-neutral. To give an example from a reformulated study regulation (Harnisch, 2016):

> **Original text**: *Auf Anfrage erhält der Student* ['the student' (**generic masculine** form for (m/w/x)] *Auskunft über den Stand seiner* ['his' (**generic masculine**)] *Leistungspunkte*.
> **Updated text**: *Auf Anfrage erhält der Studierende* ['the student' (**generic masculine** form for (m/w/x)] *Auskunft über den Stand seiner* ['his' (**generic masculine**)] *Leistungspunkte*.
> **Meaning**: 'The student can ask questions regarding the status of his credit points.'

Also, the authors of the new version of the German Road Traffic Act used nominalized participles in the singular (Bülow & Herz, 2014). For example:

(a) *als Fußgänger* ['as pedestrian' (MASC.SG)] is replaced by *als zu Fuß Gehender* ['as somebody going by foot' (MASC.SG) (StVO § 49 (1))]
(b) *als Veranstalter* ['as organizer' (MASC.SG)] is replaced by *als Veranstaltender* ['as somebody organizing' (MASC.SG) (StVO § 49 (2))]
(c) *der zu Überholende* ['the one who is to be overtaken' (MASC.SG)] remains *der zu Überholende* (StVO § 5 (2))

Harnisch (2016) who began a body of work in the early 1990s and Bülow (2016) who has continued this line of work since 2013 collected more than a hundred cases from different types of texts in which a traditional generic masculine was replaced with another more complex generic masculine nominalized participle in order to comply with the requirements of gender-mainstreaming ideology. This is a proto-example of a process of reanalysis which can be interpreted as exaptation in the evolutionary sense of the word's meaning (Bülow, 2016; Bülow & Harnisch, 2015). Exaptation means the evolution of a new function on the basis of an old form (cf. Gould & Vrba, 1982; Lass, 1997; Simon, 2010: 52). Bülow and Harnisch (2015) claim that the participle suffix *-end* is reanalyzed by language users as a marker for gender-sensitive language use.

However, in the process of exaptation, non-intended language change also has to be considered. Non-intended language change is the result of human (inter)action rather than deliberate planning and leads to spontaneous orders. These orders on the macro-level are the result of maxim-guided performances by individuals on the micro-level, as Keller (1997) points out:

> The spontaneous order [...] comes into existence due to certain forces which, under certain framing conditions, influence individual elements.

If we call the level of the individual elements the system's microstruc-ture and the emerging order itself the macrostructure, we arrive at the following definition: spontaneous order is a macrostructural system that emerges under certain framing conditions, due to microstructural influences that are not directed toward the system's formation. (Keller, 1997: 4)

Keller (1994: 102) summarizes the 'hypermaxim' as follows: 'Talk in such a way that you are socially successful, at the lowest possible cost'. For the language user, following this maxim is usually a subconscious process. The selection of the linguistic elements happens in the triangle of tension between: (a) inner systematic relations (e.g. morphological versus phonologi-cal constraints); (b) factors of cognitive language processing; and (c) striving towards success in one's social life (Bülow, 2016).

Bülow and Harnisch (2015) and Bülow (2016) argue that with respect to Keller (1994) we can observe that non-intended language change is stronger and more sustainable than language planning. Language planning is target oriented. Language is not teleological. It is impossible to imagine that lan-guage can reach a perfect, static and finite state. Language planning under-estimates the complexity and role of interacting factors that exist in language change. Both idiolectal linguistic knowledge and communal language are dynamic and complex adaptive systems (e.g. Ellis 2011), which are in perma-nent synchronization on all linguistic levels.

The current use of the masculine nominalized participles in the singular (e.g. *der Studierende, als zu Fuß Gehender*) as a gender-sensitive form is worth further investigation. If it is true that people reanalyze and accept the generic masculine nominalized participles as gender-sensitive forms, we have to ask whether it is really the masculine Genus in the first place that is responsible for the fact that more men than women are associated when such forms are used. This is the main question in our study.

Method

Hypotheses

To find out whether Genus can influence our associations, we conducted a text-based association study with students under the age of 30. The current study examines to what extent they associate generic masculine nominal-ized participles including the *-end* suffix with male or female gender. We compared the effect of generic masculine nominalized participles in the sin-gular with the effect of genus-neutral nominalized participles in the plural.

Feminist and gender linguists would assume that participants associate significantly more women in Genus-neutral framings (*die Studierenden*)

than in generic masculine framings (*der Studierende, der Student*). Regarding the anticipated function of the *-end* suffix to signal gender-sensitive lan-guage use it might be that the participants associate more women when they are faced with generic masculine nominalized participles (*Studierender*) than when they are faced with the traditional generic masculine forms (*Student, Studenten*). However, there are no reliable data to date that shed light on how the *-end* suffix affects our associations in generic masculine nominalized participles. Besides, we need to consider the context as an independent variable, in line with Braun *et al.* (1998). For this reason we assume that:

> **H1**: Participants associate significantly more women in neutral contexts than in male contexts.

Furthermore, regarding the postulates and theories of feminist and gender linguists, which predict that the generic masculine makes women invisible, we have to check the following hypotheses:

> **H2**: Participants associate significantly more women in neutral contexts that are framed by Genus-neutral nominalized participles than in neu-tral contexts that are framed by generic masculine nominalized participles.

> **H3**: Participants associate significantly more women in male contexts that are framed by Genus-neutral nominalized participles than in 'male' con-texts that are framed by generic masculine nominalized participles.

According to our own observations – that *-end* is reanalyzed and exapted as a marker for gender-sensitive language use in general – we hypothesize the following:

> **H4**: Participants equally associate women and men in neutral contexts that are framed by Genus-neutral nominalized participles and by generic masculine nominalized participles.

> **H5**: Participants equally associate women in male contexts that are framed by Genus-neutral nominalized participles and by generic mas-culine nominalized participles.

H4 and H5 mean that we expect no Genus framing either in the male or in the neutral contexts. Furthermore, we expect that participants equally asso-ciate women and men in the neutral contexts. We decided to use an experi-mental, text-based association study which is partly oriented on the Braun

et al. (1998) study in order to provide evidence for or against the above-mentioned hypotheses.

Participants

In total, a sample of 236 students from the University of Passau and the University of Munich participated in our text-based association study. Agreements from the students were obtained prior to the lessons. It was emphasized that participation was anonymous and voluntary. We made the sample of university students more representative by testing participants from four study courses: *Teacher Training* for all levels of education (German); BA *International Cultural and Business Studies*; *Law*; and BA *Government and Public Policy*. For various reasons, we had to exclude 36 participants (aged over 30, did not answer the questions, did not have German as mother tongue, etc.). One hundred and ninety-six datasets entered the first analysis. All students in this dataset were native German speakers and between the ages of 19 and 30 (mean age 21). We tested many more women ($n = 155$) than men ($n = 41$) (see Table 8.1).

Variables

Various studies have found evidence that the gender of the participants can have an effect on the results (see Heise, 2003; Klein, 1988). Because 41 male participants distributed over seven texts is not enough to have robust results with gender as independent variable, we excluded all male participants in the second data analysis. Age might also be a variable (see Schröter *et al.*, 2012; Switzer, 1990), but we controlled for that by excluding participants over the age of 30.

The current study investigates the influence and interaction of two independent variables. The first was the linguistic form of the referent (referred to as 'role nouns'). We framed three different newspaper articles with either Genus-neutral nominalized participles in the plural (*die Studierenden, die Fahrradfahrenden*), generic masculine nominalized participles in the singular (*der Studierende, der Fahrradfahrende*), or traditional generic masculine forms in plural and singular (*Student, Studenten*). The second independent variable was the text setting (male or neutral) of the three newspaper articles (A, B, C).

Procedure and materials

The study was conducted in German, and none of the participants was aware of the purpose of the study. As in the Braun *et al.* (1998) study, participants were told that they would take part in a study on information structure in online newspaper texts. The task was administered in classrooms of 10–80 students with various academic backgrounds. Instructions were given orally by the investigator. The experimenter was present in the classroom during the data collection, which lasted about 10 minutes.

The students first had to read one of the seven different text versions. The framing variable was tested in three settings that were all pre-tested. This means that we framed two texts (A, B) in two frame versions (A1, A2, B1, B2), and one text (C) in three frame versions (C1, C2, C3). Text A was about drinking alcohol and cycling, which proved to be a 'male context', while Texts B and C both were 'neutral contexts': Text B was about bicycle paths in cities, and text C was about a study regulation at the Ludwig Maximilian University Munich. A1, B1 and C1 were framed with the Genus-neutral nominalized participles in the plural. A2, B2 and C2 were framed with generic masculine nominalized participles in the singular. C3 was framed with the traditional generic masculine forms in singular and plural. Participants either read the text including nominalized participles in the plural, or the text including nominalized participles in the singular, or the traditional generic masculine forms (see Appendix Text Box 8.1).

Participants were randomly assigned to one of the seven text versions. They had two minutes to read 180 words. They were asked to remember as much information as possible. Notes and annotations were not allowed. After the reading, the texts were collected, and two questionnaires designed with Unizensus Software were handed out. The first questionnaire contained eight items. The first two were yes/no questions about the content of the text. They were also framed with the stimuli words used in the texts (Table 8.2). Question (1), for instance, read as follows: 'Does the text answer the question whether the student ("der Studierende" (MASC. SG.NOM)) has a disadvantage that there is no study regulation until now?' (see Appendix Text Box 8.2). Question (3) was the target question: 'Please estimate how many female and male persons might be concerned about

Table 8.1 Distribution of the participants over the texts and their mean age

Text/context	No. participants	No. females	No. males	Mean age
A1	32	27	5	21,9
A2	33	27	6	21,3
B1	27	19	8	21,2
B2	30	23	7	21,3
C1	27	22	5	21,8
C2	22	17	5	21,2
C3	25	20	5	20,7
Total	**196**	**155**	**42**	$\bar{x} = 21,35$

Table 8.2 Stimuli word forms used in the study

Text version	Stimuli word forms used in the text	Stimuli word forms used in the questionnaire
A1	Radfahrende (ACC), Radelnde (ACC), Fahrradfahrende (NOM) (all PL)	Fahrradfahrende (ACC.PL)
A2	Der Radfahrende (NOM), den Radfahrenden (ACC), den Radelnden (ACC), der Fahrradfahrende (NOM) (all MASC.SG)	einen Fahrradfahrenden (ACC.MASC.SG)
B1	aller Zweiradfahrenden (GEN), Fahrradfahrende (ACC), Radfahrende (NOM) (all PL)	Fahrradfahrende (ACC.PL)
B2	Zweiradfahrenden (GEN), den Fahrradfahrenden (ACC), der Radfahrende (NOM) (all MASC.SG)	den Fahrradfahrenden (ACC.MASC.SG)
C1	die Studierenden (ACC), Studierende (ACC), Studierende (NOM), Studierende (DAT) (all PL)	die Studierenden (NOM.PL)
C2	den Studierenden (ACC), der Studierende (NOM), dem Studierenden (DAT) (all MASC.SG)	der Studierende (NOM.MASC.SG)
C3	den Studenten (ACC.SG), dem Studenten (DAT.SG), Studenten (NOM.PL) (all MASC)	der Student (NOM.MASC.SG) die Studenten (NOM.MASC.PL)

the problem that no study regulation exists so far? Express your rating in percent (%). Please make sure that the sum is 100%' (see Appendix Text Box 8.3).

The next five questions were again single-choice questions about the text. They were meant to ensure the plausibility of the instructions and to check whether the participants had read and understood the texts. Participants who had made more than two mistakes were excluded. The students then completed a two-page questionnaire about their sociodemographic backgrounds. On this basis we excluded participants who were older than 30 or who were not native speakers of German.

Design and analysis

SPSS was used to run the statistics. To find significant differences ($p < 0.05$) regarding the influence of the variables in the dataset of the female participants we conducted a two-way ANOVA. Then, to find significant differences ($p < 0.05$ again) between the levels we conducted a post hoc test (Tukey; Hochberg's GT2).

Results

The data of the male and female participants (see Table 8.3) combined show a very similar mean to the data in which the male participants have been excluded (see Table 8.4).

Regarding the adjusted data (155 female participants) the statistical analysis shows that there was no significant effect on the levels of framing from how many women were associated (F (2, 148) = 0.553, $p = 0.576$). However, Figure 8.1 shows that there was a significant effect of context. A was the male context and B and C were neutral contexts (F (2, 148) = 16,81, $p < 0,001$). A post hoc analysis revealed that only the difference between texts A and B ($M = 11.89$, $SD = 2.31$) and A and C ($M = 10.22$, $SD = 2.106$) were significant at $p < 0.001$. The difference between the neutral context texts B and C ($M = 1.67$, $SD = 2.26$) was not significant ($p = 0.841$). Furthermore, there was no significant interaction between levels of framing and levels of texts. We can conclude that the Genus framing had no influence on how many women were associated. The variation in the data can be explained by the different contexts.

Table 8.3 Distribution over both female and male participants

Text (context)	Framing	No. participants	Associated women (%)	Associated men (%)
A1 (male)	1	32	37.0	63.0
A2 (male)	2	33	35.7	64.3
B1 (neutral)	1	27	52.9	47.1
B2 (neutral)	2	30	49.5	50.5
C1 (neutral)	1	27	49.6	50.4
C2 (neutral)	2	22	46.9	53.1
C3 (neutral)	3	25	46.8	53.2
Total		196	$\bar{x} = 45.0$	$\bar{x} = 55.0$

Table 8.4 Distribution over female participants

Text (context)	Framing	No. participants	Associated women (%)	Associated men (%)
A1 (male)	1	27	37.8	62.2
A2 (male)	2	27	36.3	63.7
B1 (neutral)	1	19	48.4	51.6
B2 (neutral)	2	23	49.4	50.6
C1 (neutral)	1	22	49.6	50.4
C2 (neutral)	2	17	46.4	53.6
C3 (neutral)	3	20	45.5	54.5
Total		155	44.2	55.8

Figure 8.1 The association of women over the different framings
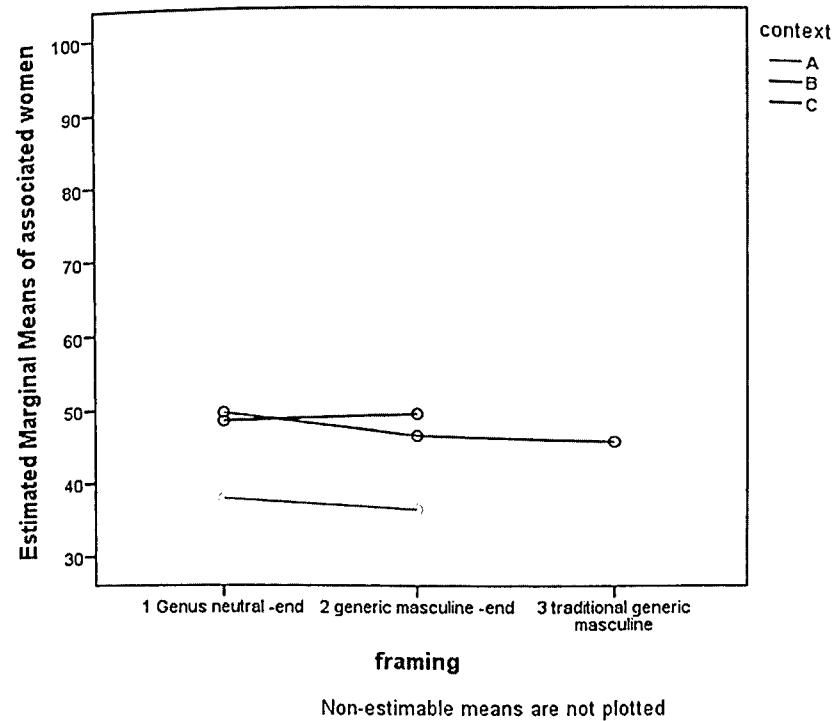

Figure 8.2 The association of women over the different framings

As far as framing is concerned, Figure 8.1 shows that only a few more women tend to be associated in the plural versions of Texts A and C, which were framed with Genus-neutral nominalized participles, than in the singular versions which were framed with generic masculine nominalized participles. For Text B it was the other way around. However, this was not a significant tendency. Comparing C3 with C2 and C1 we found no significant differences.

The boxplot (Figure 8.2) shows variance in the data. In both nominalized participle framings (Genus-neutral and generic masculine) the range of variation is similar, from 20% to 70% associated women. The traditional generic masculine framing shows less variation, from 35% to 60% of associated women, with one outlier with 70%.

Figure 8.3 shows the variation of associated women over each text version. The most variation of associated women is found in text C2 (20%–70%). All other texts have similar variation ranges. The figure illustrates the context effect and explains why we have such a range of variation over the different framing versions.

Figure 8.3 Associated women in each text version

Discussion

Feminists and gender experts have argued that we should use gender-sensitive forms for referents that include both men and women equally. Several studies showed that forms may indeed affect conceptualizations (see literature review above), but others have argued that the results may be misleading because the context was not controlled for. Moreover, one strategy to use gender-sensitive forms, viz. plural nominalizations, is now often replaced with a singular form which, in effect, is masculine. The current study examined whether the grammatical form or the context is more influential in conceptualizing the number of women.

We found that our participants associate significantly more women in neutral contexts than in male contexts regardless of Genus framing. They equally associate women in male contexts regardless of Genus framing, and they associate women and men nearly equally in 'neutral' contexts regardless of Genus framing. The observation that the data of the male and female participants combined (see Table 8.3) show a very similar mean to the data in which the male participants have been excluded (see Table 8.4) is regarded as evidence for the validity of our measurement. Furthermore, it seems that we did not have an effect on the gender level in our data. However, it needs to be borne in mind that, in order to make reliable statements on how gender affects the associations, we did not have enough male participants in our dataset. Future studies need to fill this gap.

In part, the results run counter to commonly held views in feminist and gender theory, which raises several questions. Regarding the results from the neutral contexts B1, B2, C1 and C2, one could get the impression that the nominalized participle suffix -end has the power to infer a mostly equal association of women and men. At the same time, there was no significant contrast between the text versions which contained a Genus-neutral nominalized participle (B1, C1) and the form which contained the traditional generic masculine forms (C3). So why do Genus-neutral forms in our study conceptualize women in the same way as generic masculine forms? We suggest that this result is in line with theories other than gender mainstreaming ideology. Braun et al. (1998) concluded, for example, that a 'gender-neutral linguistic form is no alternative to the generic masculine, because it has hardly any influence on the increase of association of "female"' (Braun et al., 281, trans. LB, for German original quote see Appendix Quote 2). Furthermore, they also found a strong context effect. Stahlberg et al. (2007: 176) try to give a preliminary explanation for this phenomenon: it 'might be that the effect of neutral forms is especially context-sensitive and as a consequence extra-linguistic factors such as the expected base rates of women and men in a given context may moderate their effects.' Furthermore, we have to pay attention to the fact that we used generic masculine forms in the singular and in the plural in C3. There is some evidence that number can affect the results (Rothermund,

1998). It is assumed that plural forms evoke less gender-specific associations than singular forms. This is an empirical question that we urge applied linguists to take into consideration in the future.

Our results also indicate that it is not necessary to avoid generic masculine forms in future language use, neither the nominalized participles (der Studierende 'the student') nor the traditional ones (der Student 'the student'). It seems that generic masculine forms conceptualize women and men equally in neutral contexts. The generic masculine is in itself not a strong factor for the number of women conceptualized. This means that there is no need to change language in neutral contexts. It is, for example, obvious that women and men equally participate in traffic. It would thus make no sense to change words like Fußgänger ('pedestrian') into zu Fuß Gehende ('pedestrian', 'those who go on foot'), except for ideological reasons, of course. If it is important to emphasize the presence of women in a 'male context' in future language policy then it might be helpful to make women more visible through the use of gender-sensitive language. However, our findings suggest that the Genus-neutral or the generic masculine nominalized participles are not necessarily helpful in reaching that goal. This is why we need a non-ideological discourse in the future that takes account of the context and is based on valid empirical research and not on personal attitudes.

Coming back to Pusch's statement that 'women are refusing to feel included by words like Wähler ["voter"], Ballettänzer ["ballet dancer"], or Zuhörer ["listener"]' (Pusch, 1997b: 328): recent empirical research shows that this view is not shared by all women. Wetschanow and Doleschal (2013: 323, trans. LB, for German original quote see Appendix Quote 3) state that '[m]ost recent studies indicate that people under 25 use gender-sensitive formulations less frequently'. In particular, most young women under the age of 30 emphasize that they do not like to be patronized by feminist language policy. This trend was also recognized in other studies (e.g. Irmen & Steiger, 2005: 231; Schröter et al., 2012: 374–375; Spieß et al., 2012: 1; Wetschanow & Doleschal, 2013: 322). Most young women accept and use generic masculine forms as gender-neutral forms (Bülow & Herz, 2015). They do not necessarily mix up Genus and gender. Using generic masculine forms is in accordance with undoing gender for these women. Wetschanow and Doleschal (2013: 322, trans. LB, for German original quote see Appendix Quote 4) speculate, for example, that '"undoing gender" [...] could be the main motivation for turning away from splitting [Studentinnen und Studenten "students (w) and students (m)"; LB] and for returning to the generic masculine [Studenten "students" (w/m/x); LB]'. As already pointed out, undoing gender is a conscious decision to make gender categories irrelevant. The problem is, however, that the generic masculine is identical in its form to role nouns that can refer to a single male person (der Student = Peter). Therefore, it might be not unlikely that its form can become reanalyzed in the future again. Some linguists are convinced that gender 'is so fundamental to social organization and social

structure that linguistic means to refer to this category are indispensable for speech communities' (Stahlberg *et al.*, 2007: 163). However, there is a scarcity of research on how young women and men who grew up in an educational system which provides equal opportunities for them conceptualize generic masculine forms. This gap needs to be filled.

Obviously, there are many factors involved in the process of language change. The question of how gender is associated with linguistic structure is complex and interacts with psychological, social and linguistic issues. On the one hand, we can attest to a strong influence of gender-mainstreaming ideology and policy, which leads to a kind of language change. On the other hand, we are confronted with people's language use in daily life which undermines the language policy efforts. Nevertheless, there seems to be a comeback of generic masculine forms on the horizon.

Conclusion

It was our goal in this chapter to provide empirical evidence as to how young people conceptualize Genus-neutral and generic masculine *-end* participle forms in comparison to traditional generic masculine forms. The assumption that the generic masculine disadvantages women is a powerful one in political and academic discourse because people easily mix up Genus (a grammatical category) and gender (a social and biological category). However, this study demonstrated that Genus framing was no factor in our association study, as it made no difference whether the text was framed by a Genus-neutral form or a generic masculine form. Also, it seems to be irrelevant whether the text is framed by a nominalized participle (SG/PL) or a traditional generic masculine form (SG/PL).

A further goal was to outline some fundamental implications for future applied linguistics studies as well as for future language policy. Our study gives rise to an array of concrete questions regarding upcoming language change projects in German-speaking countries, such as the following. Is it really necessary at this point to revise every legal text even if men are more often accused as murderers and even if women and men equally participate in road traffic? Do we have to change compounds such as *Studentenwerk* ('student union', 'student service') into *Studierendenwerk* ('student union', 'student service') even if we have the same proportion of female and male students in German universities? Do we have to gender every research proposal even if our research has nothing to do with human interaction?

The current study suggests that we might have to be more cautious when we accuse the generic masculine of not doing justice to women adequately. Quite to the contrary, the generic masculine has the potential to become the most economical way not to express gender if we do not confuse 'Genus' with 'gender' in the future. It is a conscious act of undoing gender. Whether equal

visibility of women can be achieved merely by changing grammar is doubtful. It seems that the generic masculine is a basic and very economical grammatical category, not only for nouns, but also for pronouns (see Harnisch's 2009 programmatic heading). The dominance of generic masculine forms also presents evidence for the great importance of generic language use in general. In the future, applied linguistic research needs to focus more on the interplay between contextual factors and inherent grammatical structures and features. Furthermore, we need to study whether innovations in gender-neutral language have long-term effects in real time rather than only immediately measurable apparent time effects across different age groups. We have to carefully detect where and when it is necessary to make women more visible languagewise – ideally by doing gender rather than by undoing gender.

Acknowledgements

References

Boroditsky, L., Schmidt, L.A. and Phillips, W. (2003) Sex, syntax, and semantics. In D. Gentner and S. Goldin-Meadow (eds) *Language in Mind. Advances in Study of Language and Thought* (pp. 61–79). Cambridge. MA: MIT Press.

Braun, F., Gottburgsen, A., Sczesny, S. and Stahlberg, D. (1998) Können Geophysiker Frauen sein? Generische Personenbezeichnungen im Deutschen. *Zeitschrift für germanistische Linguistik* 26, 265–283.

Bülow, L. (2016) Sprachdynamik im Lichte der Evolutionstheorie – Für ein integratives Sprachwandelmodell. As yet unpublished doctoral thesis, Passau.

Bülow, L. and Harnisch, R. (2015) The reanalysis of German *-end* as a marker of gender-sensitive language use. A process of exaptation. *JournaLIPP* 4, 85–96.

Bülow, L. and Herz, M. (2014) Semantische Kämpfe um Wissenschaftlichkeit und Ideologie. Gender studies, ihre Gegner und die Konsequenzen für den Sprachgebrauch und das Sprachsystem am Beispiel des generischen Maskulinums. *Linguistische Berichte* 240, 471–516.

Bülow, L. and Herz, M. (2015) Undoing gender? Ein Abgleich sprachpolitischer Maßnahmen in Rechtstexten mit dem tatsächlichen Sprachgebrauch junger Frauen. *Muttersprache* 2/2015, 133–155.

Bußmann, H. (2005) Haben Sprachen ein Geschlecht? – Genus/gender in der Sprachwissenschaft. In H. Bußmann and R. Hof (eds) *Genus. Geschlechterforschung/ Gender Studies in den Kultur- und Sozialwissenschaften. Ein Handbuch* (pp. 483–518). Stuttgart: Kröner.

Coates, J. (2013) *Women, Men and Everyday Talk*. Houndmills: Palgrave Macmillan.

Ellis, N.C. (2011) The emergence of language as a complex adaptive system. In J. Simpson (ed.) *The Routledge Handbook of Applied Linguistics* (pp. 654–667). London: Routledge.

Elsen, H. (2011) *Grundzüge der Morphologie des Deutschen*. Berlin: de Gruyter.

Fausey, C.M. and Boroditzky, L. (2010) Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review* 17, 644–650.

Gould, S.J. and Vrba, E. (1982) Exaptation – a missing term in the science of form. *Paleobiology* 8, 4–15.

9   Analytic Framework of the Critical Classroom: Language and Beyond

Gyöngyi Fábián

Introduction

Global education requires critical thinking. In an age in which information is selected and presented in increasingly sophisticated ways, when the only constant feature is change and when problems must be solved quickly, thinking is a survival skill for the individual and the society alike. The world summit on teaching (Asia Society, 2012) claims that only providing higher order skills for the majority of students is an adequate goal and, in addition, future world citizens not only need subject knowledge, but also a wide range of skills and attitudes. However, how these higher order skills are defined, and the balance among various abilities, knowledge and values, varies from country to country. The Assessment and Teaching of 21st Century Skills consortium, which includes countries from across the globe, provides one widespread definition.[1] This definition divides 21st century skills, knowledge and attitudes into four categories. Among these categories, ways of thinking comprising elements of creativity/innovation, critical thinking, problem-solving, decision-making and learning to learn completed with a positive attitude to learning are considered to be key factors.

Regarding its social aspects, critical thinking is also a key competence in performing citizenship activities, which is also a reason why researchers and educators worldwide have recently focused on identifying and developing the strategies, competencies and attributes of critical thinkers.

Despite the widespread effort, we still face a lack of consensus regarding the definition of critical thinking. The purposes of this paper are to: (a) explore the ways in which critical thinking has been defined by researchers; (b) investigate how critical thinking definitions can be extended into a more comprehensive framework for the analysis of theory and practice in the classroom; (c) learn what role language may have in the research of classroom

critical thinking; and (d) draw conclusions and further implications for future development and research in the subject.

Current Approaches to the Concept of Critical Thinking

Although the growing body of literature in critical thinking demonstrates the widespread recognition of its importance, there is a notable lack of consensus regarding the definition of the concept. The main purpose of this section is to provide a brief outline of the development of the concept by highlighting the turning points of its evolution in the past decades until the present time.

The literature on critical thinking is rooted in the traditional academic discipline of philosophy, where the 'rich concept'[2] has been developed into a coherent and sound framework of thinking about critical thinking by emphasizing the perfection of thought, and by focusing on the qualities of the ideal thinker.

Towards the end of the 20th century the rather vague interpretation of critical thinking characterized by 'reflective and reasonable thinking' (Ennis, 1985: 45) was soon completed, with the criteria of being goal directed and purposive (Ennis, 1985; Facione, 1990) and with the standards quality thinking is required to meet (Bailin et al., 1999; Lipman, 1988; Paul & Elder, 2008). More recently the concept has been developed into a more detailed description of the good thought, the purpose of thinking, and the interaction between the thought and the thinking person.

At present a widely held view of the philosophical school suggests that critical thinking maintains the fundamental intellectual standards of the good thought, which are clarity, accuracy, precision, relevance, depth, breadth, logic and significance (Paul & Elder, 2008). Scriven and Paul's statement presented at the 8th Annual International Conference on Critical Thinking and Education Reform as early as 1987 extends the critical thinking concept with two components related to the thinker himself. On the one hand, they comprise a set of information and belief generating and processing skills and, on the other hand, they involve the habit, based on intellectual commitment, of using those skills to guide behavior.[3] While this view has become widely accepted in critical thinking literature, the latter component has generated further ideas as to what this affective dimension of habit might mean.

The affective dispositions dimension of critical thinking is made clearer in the Delphi Report (Facione, 1990), which aims to promote success in educational instruction and assessment through providing a detailed description of the skills, subskills and dispositions characterizing critical thinking. The consensus statement of the document, created by a panel of experts, articulates an ideal of critical thinking and portrays the ideal critical thinker as follows:

We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriologi-cal, or contextual considerations upon which that judgment is based. CT is essential as a tool of inquiry. As such, CT is a liberating force in educa-tion and a powerful resource in one's personal and civic life. While not synonymous with good thinking, CT is a pervasive and self-rectifying human phenomenon. The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, dili-gent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit. (Facione, 1990: 2)

The psychological school of thought often intends to move away from this ideal, and focuses on how people actually apply critical thinking. It means that researchers intend to describe the types of actions and behaviors critical thinkers can perform, which involves defining a list of essential skills or procedures completed by the critical thinking person.

According to the Delphi Report (Facione, 1990), the six essential skills include the following:

(1) *Interpretation*: the ability to understand information.
(2) *Analysis*: the ability to identify the main arguments.
(3) *Evaluation*: the ability to judge whether this argument is credible and valid based on the logic and evidence given.
(4) *Inference*: the ability to decide what to believe based on solid logic, and to understand the consequences of this decision.
(5) *Explanation*: the ability to communicate the process of reasoning to others.
(6) *Self-regulation*: the ability to monitor one's own thinking and correct flaws in logic.

Seven dispositional elements, or habits of mind, are also identified by this panel of experts and are included in the document:

(1) *Inquisitiveness*: concern to become and remain well-informed.
(2) *Truth-seeking*: willingness to face one's own biases and reconsider views.
(3) *Critical thinking self-confidence*: trust in one's ability to reason.
(4) *Open-mindedness*: flexibility in considering alternative viewpoints.
(5) *Systematicity*: systematic thinking that follows a linear process.
(6) *Analyticity*: the willingness to pick apart your own and others' logic.
(7) *Cognitive maturity*: being persistent in seeking the truth.

While The Delphi Report (Facione, 1990) provides the most common defini-tion of critical thinking procedural elements to date, philosophers have often criticized what they call the reductionist approach, that is, reducing a com-plex system of knowledge and skills into a series of disconnected steps or skills which one can move through without actually engaging in critical thought (Bailin, 2002; Sternberg, 1986).

Furthermore, the two schools of philosophy and psychology disagree con-cerning the importance of criteria in thinking, as well. In other words, the debate between them seems to be circled around the question of whether criti-cal thinking should meet certain standards or not. While philosophers main-tain a normative approach to achieving the goal, that is, the perfection of thinking and thought according to fundamental criteria, psychologists are more concerned with observing and developing the procedures and the ways of per-forming the thinking process. Bailin (2002) criticizes psychologists' attitudes arguing that a justifiable conception of critical thinking must be explicitly nor-mative, focusing on the adherence to criteria and standards. She reasons, claim-ing that criteria cannot be completely proceduralized, as follows:

Either the procedure is described non-normatively, in which case there is no assurance that it will be carried out in a critical manner; or, in order to ensure criticality, the normative criteria must be built into the description of the procedure. A non-normative description of a procedure fails to cap-ture what is most essential about critical thinking. (Bailin, 2002: 364)

From a meta-theoretic point of view, another instance of disagreement in critical thinking studies can be mentioned. Although much valuable work has been completed in order to formulate a consensus of critical thinking concepts, the field still seems to view the problem in slightly different ways. One example of the critical thinking conceptual differences is shown in the manner in which permanence in individual critical thinking is handled in the various theories (see Figure 9.1).

The most traditional approach to critical thinking intends to maintain the quality of the product of thinking. It investigates the thought, its internal cohesion, the relationship between the thought and reality, and the relation-ship between the thought and the language expressing the thought according to universal intellectual standards. The main aim of the approach is to identify flaws of thinking or expression and to introduce corrections to develop the
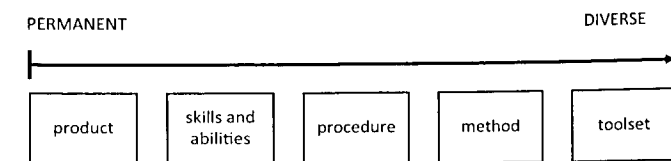
PERMANENT   DIVERSE

product	skills and abilities	procedure	method	toolset

Figure 9.1 The concept of critical thinking according to permanence

thought and thought utterances, or language, for perfection. At the same time, the psychological approach intends to identify a set of skills and abilities required to complete the process of critical thinking. A set of skills or abilities will enable the individual to proceed, but might not be demonstrated in the process observed, which may be the reason why recommendations for efficient procedures and methods have also been developed. Still others argue that instead of a single way to think critically, there is a variety of tools one can choose from when completing the thinking process, which implies that a variety of routes are available in successful critical thinking, too.

Despite the criticism and the differences between the approaches to critical thinking it is clear that there are a number of areas of agreement between the philosophical and the psychological schools of thought. Some overlapping areas can be pinned down in the essential skills and abilities, such as inductive and deductive reasoning, making decisions or solving problems, or asking and answering questions for clarification (Ennis, 1985; Facione, 1990, 2013; Paul, 1990). Most researchers also agree that, in addition to skills and abilities, critical thinking requires certain dispositions of the thinker (Bailin *et al.*, 1999; Ennis, 1985, 1991, 1996; Facione, 1990; Paul, 1990), of which the most often cited are open-mindedness, fair-mindedness, the propensity to seek reason, inquisitiveness, flexibility and respect for others' opinions.

Thus, at present it seems that the philosophical and psychological schools have been highly successful in drawing a thorough picture of the domains within which the critical thinking of the individual can be identified either for research or for developmental purposes. They have a clear understanding of the quality of the critical thought on the one hand, and the quality of the critical thinker in terms of the dispositions, skills and abilities critical thinking requires on the other.

The two schools provide us with valuable findings. The investigation of the three domains – (1) the product (the good thought) best developed by philosophy; (2) the person (the ideal dispositions of the thinker) best developed by personal psychology; and (3) the process (the skills and abilities of the thinker) best developed by cognitive psychology – will help us better understand the critical thinking concept at the level of the individual. Since the three domains describe individual performance, that is, the individual demonstrating certain dispositions and applying particular skills and abilities during the process of thinking, and because at the nexus of the three domains we will find the individual himself, we will integrate them in what we call the individual factor of our critical thinking framework (see Figure 9.2).

The Rise of the Educational School of Thought

Although at present it is a widely held view that developing the critical thinking of students is one of the main goals of education, the educational
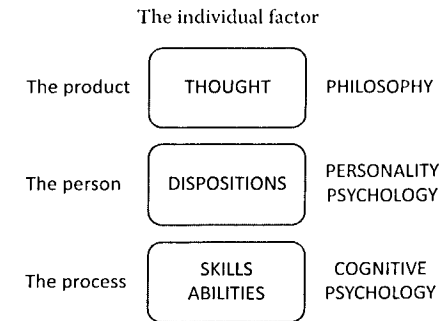

Figure 9.2 The conceptual area of critical thinking in philosophy and psychology

academic strand was not recognized in the literature of critical thinking until the late 20th century (Sternberg, 1986). The educational school appears to draw largely on what has already been developed by the philosophical and psychological schools, while extending their theories with theories of developmental psychology. However, the educational school of thought has failed to come up with a consensus of definitions of the concept of critical thinking and, as a consequence, competing definitions of critical thinking are still vague and teachers are in need of a clear and tangible definition of the concept (Bailin, 2002; Willingham, 2007).

This can be illustrated by the fact that the concepts of critical thinking and higher order thinking are often used synonymously in the literature on critical thinking in education. Higher order thinking was already addressed in an early study on cognitive development (Bloom, 1956). Because the concept of critical thinking is also implied in the three highest levels, namely analysis, synthesis and evaluation, of the taxonomy, they are frequently quoted as critical thinking procedure. Still, many agree that the concepts within the taxonomy lack the clarity necessary to guide critical thinking instruction and assessment in a useful way and, furthermore, the frameworks developed in education have not been tested as vigorously as those developed within either philosophy or psychology (Lai, 2011).

In contrast to the lingering ambiguity of the concept, the general interest in the education scene has stimulated the rise of a new aspect of critical thinking theory among philosophers and psychologists. Some contextual features of the educational situations have emerged as crucial factors in influencing the individual's critical thinking processes (see Figure 9.3). Fulfilling the goal of educating critical thinkers may be achieved with the direct assistance of the teacher and, more indirectly, through the subject matter of teaching. Interestingly, it is the indirect impact of some subject-related aspects that has attracted researchers' attention. Among these aspects the role of background knowledge and the qualities of the learning task have raised particular interest.
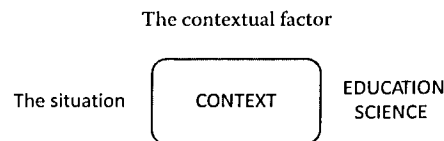
The contextual factor

Figure 9.3 The contextual area of critical thinking in education

Most researchers of critical thinking agree on the important role of background knowledge (McPeck, 1990; Willingham, 2007), meaning that the student needs something to think about. Some go even further (Bailin *et al.*, 1999; Facione, 1990; Paul, 1992), claiming that critical thinking is domain specific. Willingham (2007) argues that critical thinking is fundamentally intertwined with domain knowledge and, as such, is subject specific. Similarly, Bailin (2002) states that domain-specific knowledge is necessary for critical thinking, as what constitutes valid evidence, arguments and standards tends to vary across domains.

Norris (1985), on the other hand, analyzes the relationship between the task provided and the thinking response. He explains that the inferences and the appraisals of inferences that a person can justify making depend on the background assumptions, the level of sophistication and the concept of the task (Norris, 1985). His idea is further enhanced by Bailin (2002), who claims that critical thinking always takes place in response to a particular task, question, problematic situation or challenge, which always arise in particular contexts.

She also expands on what constellation of intellectual resources is required in particular contexts in response to particular challenges, and adds that their application depends on the context features. Some resources seem to be particular to particular contexts. She illustrates this argument through the example of the principles which govern the conduct of enquiry and the criteria for evaluation in specific disciplines. The principle requiring experimental control, for example, applies in a variety of science areas including physics, chemistry and biology, but is irrelevant in such areas as literary criticism. Similarly, the criteria for evaluation of sources in historical inquiry have no relevance to the evaluation of philosophical arguments but have some overlap with the criteria for the assessment of sources in science (Bailin, 2002).

While recent philosophical and psychological critical thinking theories have drawn educators' attention to the power that educational contexts exercise while practising and improving critical thinking in the classroom situation on the one hand, the development in education sciences has challenged the traditional ways of teaching on the other. This challenge has inspired many to develop new critical thinking development and assessment methods for classroom application all around the world. An abundance of materials is aimed at fostering critical thinking, and thus suggests
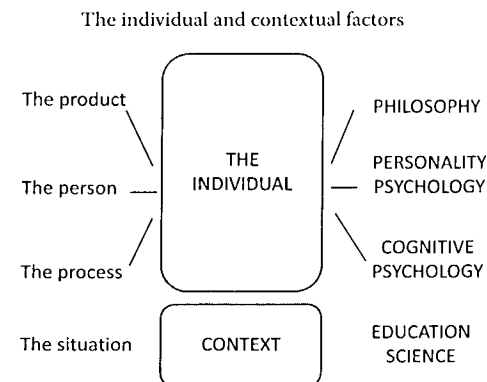
The individual and contextual factors

Figure 9.4 The conceptual areas of current critical thinking research

ways, procedures and techniques for improving students' critical thinking, as well as providing materials for application in teaching and assessment classroom procedures.

Although the elements of the context relevant to critical thinking in the classroom still need to be elaborated, the aspect of context itself has earned a place in the framework of the critical thinking concept (Figure 9.4). While a lot of valuable work has been done in the area of teaching and assessing critical thinking in the classroom, there seems to be a lack of systematic and clear description regarding what variables the teacher or the researcher of the critical thinking class needs to consider. In what follows we shall attempt to demonstrate what is problematic about many of the prevalent assumptions about critical thinking and shall propose, as an alternative, a conception of critical thinking that rests on more educationally defensible assumptions.

An Alternative Approach to Critical Thinking Concept in Education

In order to develop the concept of critical thinking in education, it is necessary to demonstrate what is problematic about many of the prevalent assumptions about critical thinking, which are linked to the ambiguity of the concept on one hand, and the lack of an interactional approach to the issue on the other.

While over the past decades ample intellectual effort has been invested in elaborating the qualities of the critical thought, the most efficient thinking process and the necessary skills, abilities and dispositions of the thinking individual, we are still far from a consensus on what critical thinking means in general and, in particular, what it means in education.
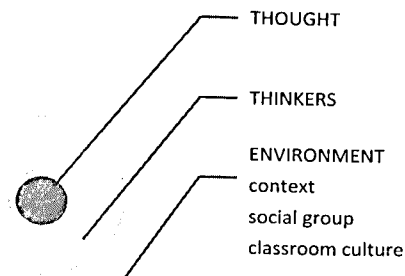
Figure 9.5 The educational environment of critical thinking

As opposed to the narrow framework of critical thinking focusing on the individual, extending it to the environment where the teaching and learning process takes place creates a wider framework for critical thinking in the formal educational environment (see Figure 9.5). While some research addresses the sensitivity of critical thinking to some features of the learning context, the issue which has hardly been discussed in the literature seems to be the immediate environment of the thinker and the impact that the affective and the sociocultural factors in this environment might have on the thinker's behaviour. In other words, the social environment of the thinker and the thinking process on the one side, and the education culture within which the interaction takes place on the other side are often left out of the debate over critical thinking.

To respond to this need, below we will promote an interactional perspective by which to develop the concept of critical thinking in educational theory and practice by discussing these two distinct environmental factors. Besides the contextual factor, we will attempt to integrate the social and educational-cultural factors into our proposed conceptual framework of critical thinking.

The Social Factor in Classroom Critical Thinking

First of all, it will be necessary to differentiate the concept of context in current critical thinking literature from what we mean by the social factor in critical thinking. Ennis (1998) suggests that critical thinking practice is more readily accepted through shared decision-making processes. He promotes practices of critical thinking at group level, if this is possible at all, as a replacement for the individual level. Classrooms that encourage critical thinking possess distinguishing features (Browne & Freeman, 2000). A critical thinking classroom commonly reflects the attributes of frequent

questions, developmental tension and fascination with the contingency of conclusions and active learning, which reinforce one another to provide developmental stimuli for enhanced critical thinking. The conversational imperative of critical thinking in the class is observable mainly in the form of asking a series of critical questions both on behalf of the teacher and the students, where the questions may have the aim of clear comprehension, as well as of furthering the discussion towards reasoning, argumentation and conclusion.

From an educational perspective, we do agree that critical thinking is very much an interactional participant activity. Having said that, we would like to emphasize the conversational imperative of critical thinking in education through looking at the partners of the individual thinker during the classroom interaction, that is, the class. However, we believe that the class is a social group where a paraphernalia of motivating and demotivating cognitive and affective factors will enhance, prevent or disrupt the development of critical thinking. To illustrate this, we will expand below on some of the social factors operating in this social environment.

Norris (1985) claims that a threat-free environment is indispensable for stimulating the expected thinking process. As within the educational context the main source of stress and threat is the teacher herself, recent literature has paid more attention to preferable teacher behaviour stimulating critical thinking in the classroom. However, we believe that threat or, more often, peer pressure on behalf of the group might exert an even stronger influence on individual behaviour in the classroom situation.

From an attitudinal approach, experience shows that critical thinking threatens the amiable calm of much of our interactions with one another, which may be threatening people away from revealing their ideas. The reason is that, while controversy creates intellectual challenge, it may create cognitive dissonance in the minds of the participants, which may make them feel uncomfortable. Students in class are no exceptions.

However, it needs to be emphasized that the teacher can make an important contribution to the development of critical thinking of the individual by promoting a mind-set of free and critical thinking within the group where failure is not criticized.

Although there is much more to be discussed concerning the classroom social environment impact on critical thinking development, the reasoning in the current literature seems to be firm enough for us to add a social component to our framework, and furthermore to provide researchers with a wider perspective of classroom behaviour study (see Figure 9.6). Our decision is further supported by social psychology theories, which warn that some phenomena featuring the climate in the social group may influence individual behaviour. Based on empirical experience, below we would like to draw the attention to just a few more examples of the influencing factors that might be of interest to researchers of critical thinking in classroom environments.
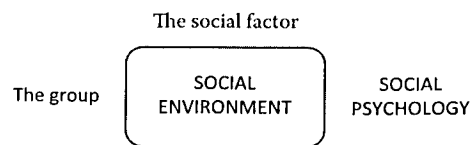
The social factor

```
                        ┌──────────────┐
                        │   SOCIAL     │    SOCIAL
        The group       │  ENVIRONMENT │  PSYCHOLOGY
                        └──────────────┘
```

Figure 9.6 The social area of critical thinking in education

Conformity helps the group to introduce shared rules to the members, which simplifies the life of the individual and group alike. At the same time, it exercises pressure on the members' behaviour and thinking. While conformity may reduce the sense of responsibility of the individual, it may limit critical thinking within the group.

A strong sense of 'my way is the only way' in some of the group members' minds, often called ethnocentricity, will be a serious barrier to fair debates and, furthermore, may be coupled with stereotyping. The reason is that people who stereotype tend to have unbending opinions. It means that in order to avoid cognitive dissonance in extreme situations they even respond to challenges by distorting reality rather than reconsidering the assumptions and conclusions, which they maintain as strong beliefs. The facts are compared to the stereotype and often rejected because they fail to fit the stereotype. Stereotypes in the group will not only hinder critical discussions but will force the members to take sides according to group preferences, will limit individual achievements, and destroy cooperation through embarrassment, fear and blame triggered within the group.

Face saving means the effort invested into preserving our image in situations which seem to threaten it. A classroom provides myriad situations where this kind of threat becomes reality, and where face saving appears to be a necessity as it acts as self-protection. On the other hand, face saving may block the growth of self-awareness and, from our point of view, may act as a barrier to interaction where the face is in danger of being lost. Instances of debates employing critical thinking are threatening scenes for the face.

Cultural Variables of the Classroom Environment

In the majority of cases, the culture where an educational institution is established will have a direct influence on the culture of the school. It means that the educational institution develops the curriculum, methods and teaching strategies in line with the specific cultural traditions of the wider environment.

Cultures differ in a number of ways. One way of completing explorations into critical thinking related cultural differences in education is through exploring the existing literature on the bias of critical thinking, which has,

however, raised little in the way of revolutionary thoughts during the course of the past decades.

As a response to earlier work on the topic Ennis (1998) raises the issue once more with the aim of generating a discussion and suggesting strategies for approaching cultural bias. Among the bias challenges discussed, he elaborates on a number of behaviours as potentially problematic, some of which might be relevant to the further study of educational cultural contexts. One of his claims is that asking for reasons and being open to alternatives may disrupt some culturally accepted behavioural patterns in the family and in the school, especially in more traditional contexts. He believes that in this situation promoting critical thinking practice is preferable rather than holding it as an ideal, which he considers unfair.

Ennis's claims seem to be in line with our own recent findings (Fábián, 2015). Our investigation was inspired by our experience collected on a teacher training course. We hypothesized that the trainees' reluctance to actively participate in discussions and to perform critical thinking demonstrated a learned behaviour which is rooted in the behavioural traditions in their previous educational contexts. Thus the students ($N = 40$) aged between 19 and 22 were asked to reflect on their previous education experiences, and to discuss and describe the typical features of the Hungarian educational cultural context based on the individualistic and collective dimensions of cultural differences (Hofstede, 1986). Even though there are signs of a minor shift towards the individualistic cultural features, the results clearly show that the collectivist characteristics of the Hungarian classroom outweigh the individualistic characteristics in many aspects. From a critical thinking point of view, this means that the class discussions are often conducted in a milieu characterized by formal harmony in learning situations, which should be maintained at all times. The strategies of doing so involve individual students being called upon by the teacher to speak personally or individuals only wishing to speak up in small groups. What is more, large classes split socially into smaller cohesive subgroups based on criteria such as ethnic affiliation, and these groups are only ready to conduct discussions out of class.

Based on our findings, we can hypothesize that the dominance of the collectivist features of the classroom learning environment will prevent the individuals from practising critical thinking in class and, what is more, will provide little opportunity for students to observe critical thinking behaviours in the classroom. These conclusions may lead us to the assumption that the cultural bias of critical thinking might also persist in a collectivist environment and, furthermore, it can be feared that in these classrooms resistance to change is prevalent.

Resistance to change naturally comes from fear of insecurity. In order to maintain cultures, tradition and social patterns are often created to give security and to resist change at the same time. In this way, tradition can be used to the advantage of the members of the cultural group. However,
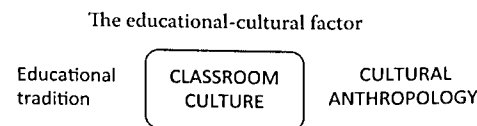
The educational-cultural factor

Educational   CLASSROOM   CULTURAL
tradition   CULTURE   ANTHROPOLOGY

Figure 9.7 The cultural area of critical thinking in education

resistance to change may make impartial judgement, which is considered to be an essential precondition of critical thinking, impossible.

Based on the reasons discussed above, we would like to propose an extension of the analytic framework of critical thinking with a fourth component, which is the educational-cultural factor of the classroom (Figure 9.7). Obviously, change in the cultural traditions of the school and education, a shift towards a more active way of learning boosted by critical thinking in the class, can only be initiated by teachers themselves.

Teachers' Critical Thinking Practices

Teachers are key figures in education and in developing critical thinking because of a number of reasons. First of all, they need to act as critical intellectuals and knowledge creators who respond to context and student needs. Furthermore, they might enhance critical thinking by providing tasks that require critical thinking. They can act as stimulants by providing learners with frequent opportunities for direct practice of evaluation skills and attitudes, which allow them to experiment with critical thought in an active manner. However, as Browne and Freeman (2000) suggest, active learning is not merely a matter of the instructor provoking students to engage with the material. A student experiencing active learning impels herself to consider and carefully evaluate the arguments of both her teacher and her peers, and eventually the arguments of her own construction. This latter statement also implies that, ideally, teachers themselves are challenged in the critical thinking class. This implies that critical thinking is a competence which teachers and teacher educators themselves need to master as models for their students.

Although little evidence of critical thinking classroom practices of teachers and teacher educators has been gathered, a wide gap seems to exist between education policy expectations and teachers' classroom behaviour, and fears are growing that the expectations of critical thinking theorists or of education policy makers are rarely met in classroom practice.

Paul (1992) argues that typical school instruction does not encourage the development of higher order thinking skills like critical thinking; instead, with its emphasis on the coverage of content, it is designed as though recall were equivalent to knowledge. He describes this type of learning as simply learning by rote or association, with the end result that students memorize material without understanding the logic of it. It seems that for most students believing, not thinking, is knowing, while they tend not to recognize that their assertions, beliefs and statements have implications, and thus require evidence to support them.

Some argue that the reasons for teachers' inadequate behaviour are related to a lack of information on the concept, which may constitute a significant obstruction to the successful implementation of critical thinking in curriculum, instruction and school culture (Bailin et al., 1999; Browne & Freeman, 2000). Others warn that thinking begins only when a state of doubt about what to do or believe exists (Browne & Freeman, 2000), which allows students to have doubts about what is being taught to them and, what is more, allows the teacher to have doubts about what she is teaching. It seems that doubt has become an important element of appropriate teacher behaviour.

What is more, doubt is considered to be an inherent element of building teacher identity, which is a complex matter of the social and the individual. According to some (Pearce & Morrison, 2011; Shuck & Buchanan, 2012) the shaping of a professional identity takes place during teachers' social exchanges and as a result of interactions with other members of the school community, such as students and parents. Such interactions take various forms. They include the conversations and emails teachers have about teaching, the seeking of support from critical friends about their practice, and endless other interchanges between teachers. Through these exchanges, many of which are triggered by doubt about teaching and self, teachers are modelling and shaping the self.

Initial teacher education programmes have a great responsibility for developing teacher identity. Based on recent research, the European Commission (EC, 2013) recommends a number of elements of teachers' core competence requirements for effective initial teacher education. The recommendations include a number of areas where critical thinking counts, such as engagement in innovation, critical and responsive attitudes to innovation, and professional improvement.

On the other hand, some research findings warn that the social environment of the workplace may set up barriers to the development of critical reflection in initial teacher training programmes. Down and Hogan (2000) consider some of the impediments to critical reflective practice in the context of a workplace-based professional extension programme (PEP). The authors discuss conservatism in teacher education and the pressures of the new corporate workplace culture as influential frameworks in the development of interns. Concerns about the direction of workplace learning and its emphasis on credentialism, technical training and worker compliance are revealed in order to pinpoint some of the tensions and dilemmas involved in attempting to develop more thoughtful socially critical practice in a context that reinforces a view of the teacher as practical professional.

In conclusion, we can state that among teachers' dispositions, critical attitudes to one's own teaching comprising doubt, examining, discussing, questioning practices are considered to be essential components of 21st-century teacher competencies. However, some of the current practice of initial teacher education and some evidence collected from classroom research do not seem to report on efficient practices of developing critical thinkers either among teacher trainees or students.

Conclusions and Implications for the Future

In our paper, one of our intentions was to integrate and extend the current theories on critical thinking concepts into a wider conceptual framework used for analyzing and developing critical thinking theory and practice in education. We found that critical thinking in the classroom can be analyzed from the perspective of the individual thinker and from that of the environment surrounding the thinker. The analysis can be completed through examining the component of the individual and the three factors of the context, the social environment and the classroom culture (see Figure 9.8).

The theory and practice of investigating the individual components of critical thinking have successfully been developed by and applied in the philosophical and psychological schools of thought. The individual component comprises elements within the area of the thought, the person and the process. It involves the elements of good thought formulated according to universal intellectual standards, the thinker's behaviour demonstrating essential intellectual traits and dispositions, and furthermore the skills and abilities the ideal thinker employs during thinking.

A more comprehensive and valid analysis of critical thinking in the classroom, however, requires extension of the current analytic framework by adding the components of the context, the social environment and the educational culture, which together are referred to as the environmental component in our paper.
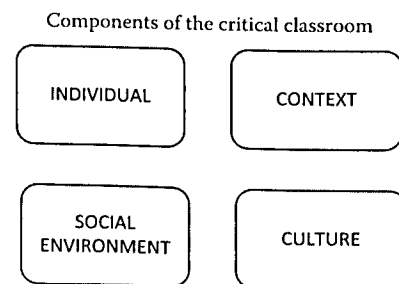
Components of the critical classroom

Figure 9.8 Individual and environmental aspects of critical thinking in education

The environmental component comprises elements within the three distinct areas of the situation, the learning group and the traditions of the educational environment. The area of the situation, known as the contextual factor, involves elements in the immediate context of the specific learning and teaching process in formal education. Some examples of the elements within this factor are the task-related features of learning, such as the language of instruction within the task, or the conditions of completing the task, such as deadlines for student assignments. The area of the learning group, known as the social factor, involves elements in the immediate human environment of the thinker. Some examples of the elements of the social factor include the social psychological features of the group, such as the attitudes and values within the group, or dynamics of group processes such as group cohesion or climate. The area of the educational traditions, known as the educational-cultural factor, involves elements in the wider educational environment of the thinker. Some examples of the elements of the educational-cultural factor are the methods of teaching and assessment applied in the educational institution or in the subject, the beliefs related to learning and teaching, or the subject-related goals and aims of the education.

Based on the literature consulted, we can conclude that the individual component of the critical thinking concept is well defined in terms of its elements and the forms and the methods of analysis, as well as the methods of development. On the other hand, the environmental components of the contextual, social and educational-cultural factors of the critical thinking concept require further research and improvement in terms of their elements, the forms and methods of analysis, or the development of critical thinking in the classroom.

It seems vital to collect more evidence concerning the critical thinking related social factors of the classroom. We believe that in this work some existing findings concerning the barriers to critical thinking rooted in the social environment of the workplace (Down & Hogan, 2000) might also carry considerable implications for classroom research.

We found that one shocking difference between cultures is that critical thinking may not be seen as positive behaviour of the individual in some places. This implies the necessity for our better understanding of how classroom cultures are related to cultures, how educational cultures might be distinguished from one another, and what features characterize the classroom culture that is supportive in developing students' critical thinking.

Contemporary research suggests that the environmental aspects of classroom critical thinking may be revealed through the close investigation of the discourse in the classroom as well. This exploration can follow diverse routes, out of which some are readily available in literature, which fact is supported by the growing body of research in a variety of features of classroom interaction. Here we will mention two of these.

Due to the fact that classroom learning and teaching is a social action, conversation analysis methods focusing on the organization of classroom activities in interactions may assist the analysis of critical thinking instances of the participants of these actions in the same way. On the other hand, critical discourse analysis methods have been widely used in both developing and researching methodology for thinking improvement in social studies through the study of ideology and power relations, with the intention of revealing the special features of the text, the interaction and the social context of discourse. We believe that the application of critical discourse analysis methods may inform the researcher of the critical thinking classroom language about the relevant aspects in the same way.

However, we intend to go even further than that. We also hypothesize that the underlying educational-cultural aspects of the classroom deep behind the surface factors of the individual, the social and the contextual aspects will also lend themselves to analysis through a thorough investigation of the language of the classroom. The methods of the exploration of the cultural-educational aspects through the study of language, however, still require further efforts.

Because all the four factors are functionally interrelated, it is crucial to understand their relationship to one another within the whole componential system of critical thinking. Figure 9.9 demonstrates a possible operational structure of the componential system discussed above. The educational-cultural factor acts as a fundamental basis for the practice and the development of the critical thinking of the individual, and for the practice and the development of critical thinking within the group during the teaching and learning process. It influences, extends or limits thinking processes based on beliefs, attitudes and traditions held in the educational context. It has immediate impact on the behaviours of individuals and learning groups, as well as on the teaching curriculum, methods, materials and tasks provided during the

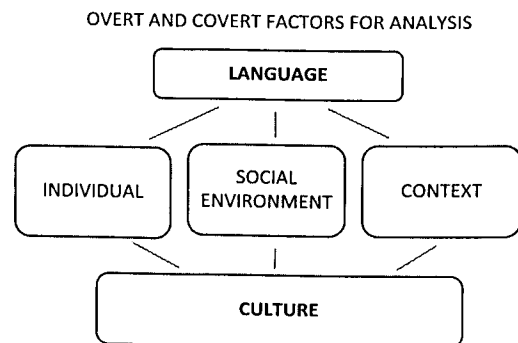OVERT AND COVERT FACTORS FOR ANALYSIS

Figure 9.9 The analytic framework of critical thinking in education

teaching process. However, the study of the educational-cultural factor may require a variety of novel research methods adopted from cultural anthropology, management, social psychology or various other disciplines.

The three factors of the individual, social environment and context require different research methodologies, designs and methods of different complexity based on a variety of disciplinary conventions and traditions. Research into the critical thinking of the individual is well established in philosophy and psychology. A wide selection of research instruments developed for a variety of purposes are available within the field. The study of the social environment, however, may need the development of new research instruments borrowed primarily from social psychology and adopted for critical thinking research purposes in the classroom. Some of the results of research in education may provide implications for the investigation of the teaching and learning context; the aims of the research, however, require some adaptation of the existing methods and techniques of investigation in order to reveal the critical thinking related features of the context.

All the three factors of the individual, the social environment and the context may involve research into the language in which critical thinking is overtly realized, that is, into the language of the thought, the language of the interaction within the thinking group, and the language of classroom instruction and of the learning task, as well as the language of student responses. Within this framework, while investigating the critical thinking related language in the classroom will provide access to overt information on the surface level, it may also provide some information related to the covert elements in the deep structure of the conceptual framework of critical thinking – in other words, information about the individual, the context, the social environment of thinking or, even deeper, the culture of education in the classroom.

Our findings concerning the necessity of the extension of the conceptual framework of critical thinking and the framework of researching critical thinking in the classroom suggest that much theoretic and practical scientific work will be needed in order to draw a more comprehensive picture of classroom critical thinking in education. This will also include adapting existing designs of research methodology, such as those related to cognition in conversation analysis or critical discourse analysis, as well as novel methods and instruments applied for classroom language research developed for the particular purpose of revealing the individual, social, contextual and cultural features of critical thinking classroom practices.

Our recommendation related to developing critical thinking in the classroom concerns the educators of teacher training institutions and venues. Teachers' professional identity development is influenced during the training programmes of higher education institutions, and later in internship programmes by education professionals. Educators, supervisors or mentors are able to create an appropriate environment for interns' identity development

by modelling critical thinking, and maintaining a climate where critical thinking is supported. Thus, it is vital that educators, supervisors and mentors working with trainee teachers have a clear understanding of what the critical thinking concept means, and furthermore are capable of maintaining critical thinking attitudes and of developing critical thinking skills and abilities in the work environment.

Notes

(1) See http://www.atc21s.org/ (accessed 29 October 2015).
(2) See https://www.criticalthinking.org/pages/defining-critical-thinking/766 (accessed 16 July 2015).
(3) See https://www.criticalthinking.org/pages/defining-critical-thinking/766 (accessed 16 July 2015).

References

Asia Society (2012) Teaching and leadership for the twenty-first century. Report. See http://asiasociety.org/files/2012teachingsummit.pdf (accessed 13 July 2015).

Bailin, S. (2002) Critical thinking and science education. *Science & Education* 11 (4), 361–375.

Bailin, S., Case, R., Coombs J.R. and Daniels, L.B. (1999) Conceptualizing critical thinking. *Journal of Curriculum Studies* 31 (3), 285–302.

Bloom, S.B. (1956, 1984) *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain.* New York: Longman.

Browne, M.N. and Freeman, K. (2000) Distinguishing features of critical thinking classrooms. *Teaching in Higher Education* 5 (3), 301–309.

Down, B. and Hogan, C. (2000) Critical reflective practice and workplace learning: Impediments and possibilities in teacher education. *Australian Journal of Teacher Education* 25 (2). See http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1292&context=ajte (accessed 13 July 2015).

EC (2013) *Supporting Teacher Competence Development for Better Learning Outcomes.* Brussels: European Commission. See http://ec.europa.eu/education/policy/school/doc/teachercomp_en.pdf (accessed 13 July 2015).

EC (2014) *Initial Teacher Education in Europe: An Overview of Policy Issues.* Brussels: European Commission. See http://ec.europa.eu/education/policy/strategic-framework/expert-groups/documents/initial-teacher-education_en.pdf (accessed 13 July 2015).

Ennis, R.H. (1985) A logical basis for measuring critical thinking skills. *Educational Leadership* 43 (3), 44–48.

Ennis, R.H. (1991) Critical thinking: A streamlined conception. *Teaching Philosophy* 14 (3), 5–22.

Ennis, R.H. (1996) Critical thinking dispositions: Their nature and assessability. *Informal Logic* 18 (2–3), 165–182.

Ennis, R.H. (1998) Is critical thinking culturally biased? *Teaching Philosophy* 21 (1), 15–33.

Fábián, G. (2015) Non-critical thinking: What if not thinking? *Procedia – Social and Behavioral Sciences* 186, 699–703.

Facione, P.A. (1990) *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction.* Millbrae, CA: California Academic Press.

Facione, P.A. (2013) Critical thinking: What it is and why it counts. *Insight Assessment.* See www.insightassessment.com/pdf_files/what&why2006.pdf? (accessed 31 July 2015).

Hofstede, G. (1986) Cultural differences in teaching and learning. *International Journal of Intercultural Relations* 10 (3), 301–320.

Lai, E. (2011) Critical thinking: A literature review. Pearson Report Series. See http://images.pearsonassessments.com/images/tmrs/CriticalThinkingReviewFINAL.pdf (accessed 13 July 2015).

Lipman, M. (1988) Critical thinking – what can it be? *Educational Leadership* 46 (1), 38–43.

McPeck, J.E. (1990) Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher* 19 (4), 10–12.

Norris, S.P. (1985) Synthesis of research on critical thinking. *Education Leadership*, 40–45. See http://www.ascd.org/ASCD/pdf/journals/ed_lead/el_198505_norris.pdf (accessed 23 January 2016).

Paul, R. (1990) *Critical Thinking: What Every Person Needs To Survive in a Rapidly Changing World.* Rohnert Park, CA: Sonoma State University: Center for Critical Thinking and Moral Critique.

Paul, R.W. (1992) Critical thinking: What, why, and how? *New Directions for Community Colleges 77,* 3–24.

Paul, R. and Elder, L. (2008) *The Miniature Guide to Critical Thinking Concepts and Tools.* Tomales, CA: Foundation for Critical Thinking. See http://www.criticalthinking.org (accessed 20 April 2015).

Pearce, J. and Morrison, C. (2011) Teacher identity and early career resilience: Exploring the links. *Australian Journal of Teacher Education* 36 (1). See http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1513&context=ajte (accessed 13 July 2015).

Schuck, S. and Buchanan, J. (2012) Dead certainty? The case for doubt in teacher Education. *Australian Journal of Teacher Education* 37 (8). See http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1889&context=ajte (accessed 13 July 2015).

Sternberg, R.J. (1986) *Critical Thinking: Its Nature, Measurement, and Improvement.* Washington, DC: National Institute of Education. See http://eric.ed.gov/PDFS/ED272882.pdf (accessed 13 July 2015).

Willingham, D.T. (2007) Critical thinking: Why is it so hard to teach? *American Educator,* Summer, 8–19.

10 Strategies in Multilingual Learning: Opening New Research Avenues

Ulrike Jessner and Valentina Török

Introduction

The last few decades have revealed an increased interest in research on third language acquisition (TLA) and multilingualism. Ever since Bialystok in the late 1980s showed that speakers of more than one language show a heightened cognitive advantage, it has become obvious that these individuals also show a higher capacity for problem solving. Over the last 15 years, research on TLA and multilingualism has progressively intensified. One of the main goals has been to describe multilingual phenomena in order to investigate differences and similarities between second and third language acquisition. Research on multilingual learners has yielded consistent evidence that both metalinguistic and metacognitive awareness are crucial for the enhancement of language learning strategies in multilingual learners, i.e. users (Jessner, 2006; Moore, 2006). For instance, Kemp (2001) and Klein (1995) were able to show that multilingual learners are more efficient in decoding the grammatical structure of another language by using more grammar learning strategies.

In this chapter, which will be informed by a dynamic systems and complexity theory (DCT) perspective, as introduced by Herdina and Jessner (2002) in their *Dynamic Model of Multilingualism* (DMM), we will discuss strategies in their interdependence on crosslinguistic interaction. We will start our discussion by focusing on metalinguistic awareness in multilingual learning or rather multilingual awareness as an emergent property of a multilingual system. In support of the crucial role that multilingual awareness plays in the complex processing mechanisms found in third language learning and use, we will present a number of examples of multilingual awareness from recent studies carried out at the University of Innsbruck. Future research challenges will be presented at the end of the chapter.

Multilingual Awareness in Learning and Use

Throughout recent decades, metalinguistic awareness has developed into a popular subject in multilingualism research. Metalinguistic awareness in TLA has been of particular interest to researchers. Metalinguistic knowledge or awareness of this knowledge influences further language learning or learning a second foreign language, which was later termed multilingual awareness (see Jessner, 2006, 2008). Evidence stems from a number of studies of TLA in bilingual children and the learning of artificial language in expert language learners.

There has also been an increased focus on the bilingual advantage in recent years (see de Bot, this volume), in particular in educational studies. The work by Bialystok and her collaborators on bilingualism across the lifespan has to be quoted here as most influential. Bialystok (e.g. 2001, 2005, 2009) concludes from her studies that there are no universal advantages for bilinguals but that high levels of proficiency in both languages lead to advantages within tasks requiring more analytic linguistic knowledge, as also indicated by Mohanty (1994) in his study of the Kond tribal children in India. In a very recent study focusing on immersion programmes, Bialystok and Barac (2012) found that the level of proficiency in the language of testing was related to performance on metalinguistic tasks, and that the performance on executive control tasks was related to the length of time in the immersion programme; that is, a distinction was found between representational structure and executive control. From a lifespan perspective, 'the bilingual profile for executive control in both children and adults emerges with experience in a bilingual environment' (Bialystok & Barac, 2012: 71).

Over the last few years the exploration of the cognitive aspects of bi- and multilingualism has become a major research theme. As a result of the development of the concept of multicompetence by Cook (e.g. 2012), an increasing number of researchers have been engaged in investigations into the nature of the L2 user and her/his cognitive qualities, which differ from those of a monolingual person (see, for example, Pavlenko, 2005). In recent publications, for example, Athanasopoulos (2006) focuses on the effects of grammatical representation of number on cognition, while Kharkhurin (2007) analyzes divergent thinking in bilinguals.

On the M-Factor

The M(ultilingualism)-factor is an emergent property, which can contribute to the catalytic or accelerating effects in TLA. The multilingual system is not only in constant change but the multilingual learner also develops certain skills and abilities that the monolingual speaker lacks. These are language-specific and non-language specific or sociocognitive skills used in

language learning, language management and maintenance. In particular, in the case of typologically related languages, a catalytic effect, that is a qualitative change in further language learning, has been detected in experienced language learners. What these skills and abilities have in common is their relatedness to a heightened level of metalinguistic awareness in multilingual learners and users or multilingual awareness, which is the most pivotal underlying emergent property of the multilingual's cognitive system enabling the function of the interaction between the complex, dynamic systems.

Whereas in second language acquisition (SLA) the learner refers to a monolingual norm, that is, she or he bases SLA on the acquisition of the first language (Herdina & Jessner, 2002: 131), in TLA a metasystem in multilinguals, which is the result of a bilingual norm, is developed. Following Lorenz (1972), who used the butterfly effect or sensitive dependence on initial conditions to refer to the unpredictability of dynamic systems, in DMM an M-effect, which refers to this qualitative change in third language learning, is assumed to exist in multilingual systems. In DMM it is argued that the multilingual system is in constant flux and that the complexity and dynamics can only be explicated by a holistic standpoint, which represents a fundamental condition of a DCT approach. In this perspective, emergent properties of the open multilingual system (skills and abilities developed by multilingual users which are not to be found in monolinguals) have to be focused on, as does the interdependence of all parts of the system.

Such an approach not only emphasizes the importance of a definition of multilingual proficiency based on a holistic understanding of the diverse components of the construct, but also stresses the interrelation between neuro-, socio- and psycholinguistic aspects of multilingualism. From a DCT perspective, the idea of nested systems contributes to an understanding of the relationship between two seemingly different approaches to the same phenomenon. In the case of language attrition in multilinguals, the interdependence between these two aspects is obvious, since it is the changes in environment that eventuate in changes in linguistic knowledge (typically affecting the L1 of migrant children).

Defining multilingual proficiency

In the DMM, multilingual proficiency is defined as the dynamic interaction between the various psycholinguistic systems (LS1, LS2, LS3, LSn) in which the individual languages (L1, L2, L3, Ln) are embedded, crosslinguistic interaction, and what is called the M-factor. The latter refers to all the effects in multilingual systems that distinguish a multilingual from a monolingual system, that is, all those qualities that develop within a multilingual speaker/learner due to the increase in language contact(s). As mentioned above, language contacts through language choice depend on the perceived communicative needs of the individual. In other words, the neuro-, socio- and psycholinguistic

systems of the multilingual individual, which are in constant change, interact with each other in a non-additive but cumulative way.

Crosslinguistic influence (CLI) in a multilingual system does not only take place from the first language to the second and vice versa. Further influence has been detected from the L1 to the L3 and from the L2 to the L3 and vice versa. This expansion of transfer possibilities demonstrates that multilingual acquisition is a far more complex process than SLA, where the role of the L1 in the development of the L2 has been researched extensively. De Angelis (2007: 20ff.) proposed that combined CLI is 'the simultaneous influence of more than one language upon a target language, i.e. a many-to-one-type' – expanding on the idea of transfer to multidirectional transfer. Still, in most cases, it will be rather difficult to attribute CLI to only one specific SL amongst several that are interacting. The more languages are involved, the more relationships between the language systems have to be considered, since the number of possible interactions in multilinguals increases with every additional language and sociopragmatic circumstantial conditions.

This discussion also makes evident that learning another language (e.g. an L3) can counteract the maintenance of an L2 or L1. In other words, language attrition or loss appear more often in multilingual than in bilingual contexts. In this case the L3 will become more dominant than the L2 due to the limitations of resources for languages, as defined in Zipf's law of least effort (Zipf, 1968). Consequently, using an L1 an as indicator for 'permanent' language dominance in life-course will turn out to be problematic in a multilingual context (see also Jessner, 2003a). Language attrition processes also point to the fact that language learning consists of non-linear and reversible processes; that is, development refers to both acquisition and attrition (Cook, 2003; de Bot & Clyne, 1989; de Bot et al., 2007; Jessner, 2003a). Even if parts of the multilingual system can become fossilized, and will in very general terms stop growing, they will still be able to exert influence on other parts of the system.

Of major interest in CLI research is the identification and study of factors likely to activate/stimulate or to decelerate and/or inhibit crosslinguistic interactions (see De Angelis et al., 2015a). These factors can be related to the language learner (e.g. her or his proficiency in the acquired languages; see Lindqvist, 2010; Williams & Hammarberg, 1998), to the languages (e.g. language distance; see Cenoz, 2001; Ringbom, 1987) or to the context (e.g. formality of the situation; see Dewaele, 2001). Researchers focus on different aspects and suggest different categorizations of such factors. De Angelis (2007: 19-40), for instance, places special emphasis on factors that trigger CLI from a non-native language, comprising language distance/proximity, target and source language proficiency, recency of use, length of residence and exposure to a non-native language environment, order of acquisition and formality of context (for a recent account of work on crosslinguistic phenomena in SLA see Alonso Alonso, 2016). But also new areas of transfer occurrence have been examined by Gabrys-Barker (2011: 97), who highlights the importance of

'transfer of learning and transfer of training' in TLA, which involves '[e]xplicit learner training in learning based [...] on the learners' former L2 experiences' (Gabrys-Barker, 2011: 97; see also Allgäuer-Hackl & Jessner, 2014).

The paradox of transfer

Crosslinguistic interaction (CLIN) in multilinguals, seen as a wider concept than Kellerman and Sharwood Smith's (1986) crosslinguistic influence, is described as an umbrella term including not only transfer and interference but also code switching and borrowing. Furthermore, it is also meant to cover another set of phenomena that includes the cognitive effects of multilingual development. These are non-predictable dynamic effects that determine the development of the systems themselves (Jessner, 2003b; Kellerman, 1995). Such a view is also related to, but not identical with, Cummins' Common Underlying Proficiency (e.g. 1991) and Kecskes and Papp's Common Underlying Conceptual Base (2000) (see also Cook, above). According to the DMM, seemingly identical phenomena of transfer can lead to divergent results in different multilingual systems. This phenomenon was referred to as the paradox of transfer in Herdina and Jessner (2002).

Prior language experience and knowledge play a crucial role in understanding crosslinguistic interaction and what makes multilingual processing differ from monolingual processing. Multilingual awareness as the awareness of the experience, and the resulting knowledge shapes cognitive mechanisms in the multilingual when being confronted with new tasks.

Components of multilingual awareness

When linguists talk about awareness, they do not only refer to awareness in/of a single language, but also to the relationship between the languages in a learner's mind. James (1996: 138ff.), for instance, describes different types of knowledge with regard to language. One can know different languages, for instance German, Hungarian and English, but one can also know about the *relation* between these languages. According to him, this knowledge can be found at the procedural level of performance when elements from the L1 are transferred to the target language, or at the cognitive level of intuition, in which case he talks of 'crosslinguistic intuition'. Knowledge can also be held at the explicit (declarative) level of metacognition, which James (1996: 139) refers to as 'crosslinguistic awareness' (XLA).

The cognitive turn in the study of language contact was introduced by James (1996), who described it as 'the language transfer issue of classical Contrastive Analysis', which 'becomes a new issue of metalinguistic transfer – and its relationship to crosslinguistic awareness' (James, 1996: 143). The search for crosslinguistic equivalents is marked by the search for similarities, which forms a considerable part of metalinguistic thinking going on during L3 production and thus refers to the relationship between CLIN and

metalinguistic awareness, pointing to the dynamic interplay which sheds light on key variables that form part of the M-factor, as described above (see also Jessner, 2008: 279).

According to Jessner (2008: 279), in reference to her 2006 study, XLA 'in multilingual production is described as (a) tacit awareness shown by the use of cognates in the supporter languages (mainly in the use of combined strategies) and (b) explicit awareness in the case of switches that are introduced by meta-language'. In another study, Jessner (2005) found that the application of metalinguistic knowledge and also the application of metalanguage can influence multilingual processing, as already mentioned. This was supported by her student Graus in her study on crosslinguistic lexical influence from English (L2) on Italian (L3) in spontaneous written production.

It has to be noted that MLA and XLA appear to be difficult to disentangle. It has become clear that not only do the two components interact but the levels of awareness also exert influence on the organization of the multilingual mental lexicon as they show influence on the activation of the individual languages in multilingual production (Jessner, 2006: 116). Whereas crosslinguistic awareness can be defined as the awareness (tacit and explicit) of the interaction between the languages in a multilingual's mind, metalinguistic awareness adds to crosslinguistic awareness insofar as it makes objectification possible (see Jessner *et al.*, 2016).

Current Research on Strategies

Current knowledge on learning strategies is widely based on work by O'Malley & Chamot (1990), Oxford (1990) and Wenden (1991). Gregersen and MacIntyre, for instance, define language learning strategies as having been 'characterized as attempts, actions, steps, thoughts and behaviours, methods and techniques and as learner's contributions among others' (Gregersen & MacIntyre, 2014: 147). Language learning strategies include both mental and physical activities, creating a kind of metaphorical chart or image in the learner's brain. Gregersen and MacIntyre highlight the fact that the range of strategies (both mental and physical) depends on the will and motivation of the individual learner, adding that '[t]his expansive list of nouns used to define language learning strategies is accompanied by an equally wide-ranging inventory of descriptors that modify them: deliberate, intentional, goal-directed, chosen, teachable, conscious or semi-conscious, purposeful, at times automatic' (Gregersen & MacIntyre, 2014: 149). The authors then differentiate between four different types of language learning strategies:

(1) *Cognitive strategies*: activating knowledge by using the senses, reasoning, conceptualizing with details and conceptualizing broadly and going beyond immediate data.

(2) *Social strategies*: dealing with sociocultural contexts and identities, inter-
acting in order to learn (overcoming fears and knowledge gaps).

(3) *Affective strategies*: activating supportive emotions, beliefs and attitudes,
generating and maintaining motivation.

(4) *Metacognitive strategies*: allowing learners to invoke higher order control
functions to analyze, monitor, evaluate, plan and/or organize TL perfor-
mance); this would mean to plan, obtain, use resources, organize, imple-
ment plans, orchestrate strategy use, monitor and evaluate.

As early as 2004, Chamot (2004: 15) pointed out critical issues that arise
from this body of research. These concern the identification procedures of
language learning strategies, strategy terminology and classification, and the
effects of learner characteristics, culture and language learning context on
strategy use. In particular, the term 'strategy' serves a wide range of defini-
tions. Very often the term is confounded with terms such as 'learning behav-
iours', 'tactics' or 'techniques'. Historically speaking, the notion of strategy
was originally used as a military term in order to win a war. Tactics would
then lead as supportive methods in order to fulfil the overall strategy of win-
ning. 'Tactics' is employed by some writers to denote a specific activity
within an overall 'strategy' (see Griffiths, 2013).

One of the issues scholars have been concerned with is the distinction
between cognitive and metacognitive strategies (Chamot, 1994; O'Malley &
Chamot, 1990; Oxford, 1990; Wenden, 1991). Wenden (1998), by basing her
work on Flavell (1979), clearly made a similar distinction between 'knowl-
edge' and 'self-management', while Rubin (2001), following the cognitive
psychologist Butler (1997), referred to them as 'knowledge' versus 'proce-
dures'. According to the latter, knowledge (of strategies, self or background)
will vary by learner. However, procedures are rather the overarching manage-
ment process which all expert learners use to regulate/manage their learning.
These do not vary by learner but rather by task, learner goal and learner
purpose (Griffiths, 2008: 11). Oxford (1989) defines metacognitive awareness
as a complex cluster of factors. According to Wenden (cited in Oxford, 1989:
237) those factors are 'what learners know about themselves and about their
own learning process – for instance, kinds of language used, proficiency level,
the outcomes of learning, and learners' own proficiency, feelings, aptitude,
physical state, age, learning style, social role, character, and personal theory
of language'. It thus becomes clear that the field of language learning strategy
is characterized by conflicting and competing terms, definitions and classi-
fication systems, as rightly pointed out by Mitits (2015: 56–58).

Whether strategies require a consciousness on behalf of the learner has
also been discussed in the academic community. It has been argued that, if
learning strategies aim to facilitate learning, they require a certain level of
consciousness on behalf of the learner. Rabinowitz and Chi (1987) suggest
that strategies cannot be strategic without a certain degree of consciousness.

Therefore, they can no longer be considered as strategies as soon as they are
implemented automatically. It seems that many scholars nowadays comply
with a continuum from conscious to unconscious (see Jessner, 2006). Hence
the role that multilingual awareness plays in strategy use needs to be studied
from a different theoretical background, as offered by DCT. In the following
the fairly new area of research on multilingual strategies and its importance
for research on strategies will be discussed.

Multilingual Strategies: Beyond SLA Research

Although research on strategies has been carried out in SLA for over half
a century now, strategy use in multilingual learners presents itself as a rather
under-researched field. Although there is a long research tradition in investi-
gating language learning strategies and an increasing body of research in
multilingualism, studies comparing monolinguals' and multilinguals' lan-
guage learning strategy use are rare. Hence there is very little literature that
brings together multilinguals, language learning strategies and additional
language learning.

The most well-known investigation focusing on the good language
learner is the large-scale interview study carried out by Naiman *et al.* (1996
[1978]), which showed that the learning success of good learners can be
attributed to the use of the following strategies: an active learning approach,
realization of language as a system, realization of language as a means of
communication, handling of affective demands, and monitoring of progress
(see also Oxford, 2011). Similarly, in her study of language learning styles in
adults, Ramsey (1980) found that multilinguals dominated in the group of
successful learners. In a limited time period, they outperformed their mono-
lingual counterparts in learning a foreign language that had been previously
unknown to them. Along similar lines, in his investigation into the strategy
use of Irish learners of German and French O'Laoire (2001) described how
those learners who were bilingual in English and Irish made more use of
strategies than those who were dominant in English. In a later study of the
same population (O'Laoire, 2004), he found that the metalinguistic knowl-
edge that was conferred on learners of L3/L4 by the study of Irish was sig-
nificant even in the context of underachievement. Similarly, Yelland *et al.*
(1993) reported on the metalinguistic benefits of limited contact with a
second language with regard to reading acquisition.

The use of language learning strategies is dependent on language learning
awareness, which guides the learner's learning process, language perception
and production (Rampillon, 1997; Wolff, 1993). Language learners develop
their own beliefs (Kalaja, 1995) or subjective theories (de Florio-Hansen, 1998;
Kallenbach, 1996; Knapp-Potthoff, 1997) about language learning. Hufeisen
(1998) intended to analyze how learners evaluate their own multilingualism,

(2) *Social strategies*: dealing with sociocultural contexts and identities, inter-
     acting in order to learn (overcoming fears and knowledge gaps).
(3) *Affective strategies*: activating supportive emotions, beliefs and attitudes,
     generating and maintaining motivation.
(4) *Metacognitive strategies*: allowing learners to invoke higher order control
     functions to analyze, monitor, evaluate, plan and/or organize TL perfor-
     mance); this would mean to plan, obtain, use resources, organize, imple-
     ment plans, orchestrate strategy use, monitor and evaluate.

As early as 2004, Chamot (2004: 15) pointed out critical issues that arise
from this body of research. These concern the identification procedures of
language learning strategies, strategy terminology and classification, and the
effects of learner characteristics, culture and language learning context on
strategy use. In particular, the term 'strategy' serves a wide range of defini-
tions. Very often the term is confounded with terms such as 'learning behav-
iours', 'tactics' or 'techniques'. Historically speaking, the notion of strategy
was originally used as a military term in order to win a war. Tactics would
then lead as supportive methods in order to fulfil the overall strategy of win-
ning. 'Tactics' is employed by some writers to denote a specific activity
within an overall 'strategy' (see Griffiths, 2013).

One of the issues scholars have been concerned with is the distinction
between cognitive and metacognitive strategies (Chamot, 1994; O'Malley &
Chamot, 1990; Oxford, 1990; Wenden, 1991). Wenden (1998), by basing her
work on Flavell (1979), clearly made a similar distinction between 'knowl-
edge' and 'self-management', while Rubin (2001), following the cognitive
psychologist Butler (1997), referred to them as 'knowledge' versus 'proce-
dures'. According to the latter, knowledge (of strategies, self or background)
will vary by learner. However, procedures are rather the overarching manage-
ment process which all expert learners use to regulate/manage their learning.
These do not vary by learner but rather by task, learner goal and learner
purpose (Griffiths, 2008: 11). Oxford (1989) defines metacognitive awareness
as a complex cluster of factors. According to Wenden (cited in Oxford, 1989:
237) those factors are 'what learners know about themselves and about their
own learning process – for instance, kinds of language used, proficiency level,
the outcomes of learning, and learners' own proficiency, feelings, aptitude,
physical state, age, learning style, social role, character, and personal theory
of language'. It thus becomes clear that the field of language learning strategy
is characterized by conflicting and competing terms, definitions and classi-
fication systems, as rightly pointed out by Mitits (2015: 56–58).

Whether strategies require a consciousness on behalf of the learner has
also been discussed in the academic community. It has been argued that, if
learning strategies aim to facilitate learning, they require a certain level of
consciousness on behalf of the learner. Rabinowitz and Chi (1987) suggest
that strategies cannot be strategic without a certain degree of consciousness.

Therefore, they can no longer be considered as strategies as soon as they are
implemented automatically. It seems that many scholars nowadays comply
with a continuum from conscious to unconscious (see Jessner, 2006). Hence
the role that multilingual awareness plays in strategy use needs to be studied
from a different theoretical background, as offered by DCT. In the following
the fairly new area of research on multilingual strategies and its importance
for research on strategies will be discussed.

Multilingual Strategies: Beyond SLA Research

Although research on strategies has been carried out in SLA for over half
a century now, strategy use in multilingual learners presents itself as a rather
under-researched field. Although there is a long research tradition in investi-
gating language learning strategies and an increasing body of research in
multilingualism, studies comparing monolinguals' and multilinguals' lan-
guage learning strategy use are rare. Hence there is very little literature that
brings together multilinguals, language learning strategies and additional
language learning.

The most well-known investigation focusing on the good language
learner is the large-scale interview study carried out by Naiman et al. (1996
[1978]), which showed that the learning success of good learners can be
attributed to the use of the following strategies: an active learning approach,
realization of language as a system, realization of language as a means of
communication, handling of affective demands, and monitoring of progress
(see also Oxford, 2011). Similarly, in her study of language learning styles in
adults, Ramsey (1980) found that multilinguals dominated in the group of
successful learners. In a limited time period, they outperformed their mono-
lingual counterparts in learning a foreign language that had been previously
unknown to them. Along similar lines, in his investigation into the strategy
use of Irish learners of German and French O'Laoire (2001) described how
those learners who were bilingual in English and Irish made more use of
strategies than those who were dominant in English. In a later study of the
same population (O'Laoire, 2004), he found that the metalinguistic knowl-
edge that was conferred on learners of L3/L4 by the study of Irish was sig-
nificant even in the context of underachievement. Similarly, Yelland et al.
(1993) reported on the metalinguistic benefits of limited contact with a
second language with regard to reading acquisition.

The use of language learning strategies is dependent on language learning
awareness, which guides the learner's learning process, language perception
and production (Rampillon, 1997; Wolff, 1993). Language learners develop
their own beliefs (Kalaja, 1995) or subjective theories (de Florio-Hansen, 1998;
Kallenbach, 1996; Knapp-Potthoff, 1997) about language learning. Hufeisen
(1998) intended to analyze how learners evaluate their own multilingualism,

how they view the interaction of their different languages, and whether they think their different languages help or hinder them when speaking, listening, understanding or writing their different foreign languages. It turned out that strategies were considered to be of substantial help in learning a new language and that multilingual learners employed them for many different kinds of tasks in their foreign language production and comprehension. For a solid and detailed overview of learning strategy research on multilingualism the reader is referred to Mißler (1999: 200), who carried out a large-scale investigation based on a German version of the Strategy Inventory for Language Learning (SILL) developed by Oxford (1990). Mißler researched the use of strategies of multilingual students who had acquired linguistic knowledge in an average of four languages before they started learning the target language. She found that, while the number of strategies increased with language learning experience, strategy use also depended on individual factors. Along those lines, Müller-Lancé (2003a, 2003b) developed a strategy model of multilingual learning, in which he pointed out that monitoring in inferencing processes is concerned with the success of strategies.

In her 2006 study, Jessner evidences the simultaneous activation of the languages in the subject's repertoires while searching for words. This finding is in line with Kellerman and Bialystok (1997: 37), who maintain that multilinguals use communication strategies that are related to the metalinguistic dimensions of the processes of analysis and control. These processes include monitoring functions such as error detection and correction, and when there is a linguistic deficit the balance between the two processes is disturbed. As a result, multilingual users resort to strategic behaviour to restore communication. Such strategies can be conscious or unconscious (Faerch & Kasper, 1983: 36), automatic or non-automatic switches (Vogel, 1992), intentional or non-intentional (Poulisse & Bongaerts, 1994), etc.

The findings of the Tyrol study revealed that there is a relationship between crosslinguistic interaction and linguistic awareness in the use of multilingual compensatory strategies (see Jessner et al., 2016). With respect to strategy form, distinctions were made between German-based strategies, Italian-based strategies, and combined strategies in which learners made use of both languages to retrieve an expression in English. As for the functions of strategies, two types of functions of the various strategies were identified: strategies which served to compensate for lexical insecurity, i.e. for a total lack of target language knowledge, or strategies that were employed in the search for alternatives. Finally, the data analysis also showed that the multilingual students made use of facilitation, simplification and/or avoidance as part of their strategic behaviour.

Kemp (2007) studied the use of grammar learning strategies by 144 multilinguals who had learnt or were learning between two and 12 languages (indigenous, foreign, heritage or dead languages). Results showed that the more languages the participants knew, the greater the number and frequency of grammar strategies as well as the number of grammar learning strategies that they themselves reported using. Moreover, this tendency increased when knowledge of languages exceeded a third language.

Psaltou-Joycey and Kantaridou (2009) investigated the possible relations between degrees of plurilingualism and strategy use. The subjects were 1555 Greek university students learning foreign languages in an academic context. The results of the study indicated that the trilingual students used more strategies more frequently than the bilingual participants, especially those strategies that promote metalinguistic awareness, and that more advanced trilinguals made more frequent use of cognitive and metacognitive strategies. Another recent study involving foreign language learners and their strategy use was conducted by Sung (2011), who investigated the influence of the number of foreign languages studied on the frequency of the strategy categories used and found that there is a positive correlation between the two factors. The participants who had previously studied one foreign language used cognitive, metacognitive, affective and social strategies less frequently than those who had studied two or more languages.

Mitits (2015) carried out a large-scale study with over 1200 participants attending junior high schools in Komotini, Thrace. She concentrated on the question of whether multilingual early adolescent language learners transfer language learning strategies from their L2 Greek to FL English. The participants were students aged 12–15 who were selected for the particular research because a large yet unidentified number of them spoke more than one language on a daily basis besides learning English at school. Other languages taught as school subjects were not considered. Apart from the monolingual Greek-speaking students, the multilingual learners mainly belonged to the Muslim minority of Thrace, which is either Turkish-speaking or Pomak-speaking and, in fewer cases, Romani-speaking, or they were in immigrant families from countries belonging to the former Soviet Republics, Albania, Bulgaria, among others. In this study, 17 languages or combinations of languages as L1, and 26 languages or combinations of languages as languages the multilinguals speak at home, were identified. The results showed that the multilinguals exceeded the monolinguals in the use of strategies for learning EFL, which can be attributed to prior language learning being a benefit for the multilingual learners in that they tend to transfer the strategies they already employ in the languages they have been using and developing. Moreover, there was a positive correlation between L2 Greek and FL English in terms of frequency of strategy use, which implies that those learners who use more strategies more often when learning Greek also do so when learning English, and vice versa. In other words, those learners who use fewer strategies when learning the second language also use fewer strategies when learning the foreign language. On the whole, a positive CLI in L3 acquisition is apparent here with respect to transfer of strategies. It is generally assumed that such transfer of strategies from one additional language to another is a

typical feature of a multilingual learner. The most important finding of the said study is that, despite the fact that the multilingual teenagers in Komotini junior high schools generally come from underprivileged socio-economic backgrounds and their multilingualism can be described as subtractive rather than additive, they nonetheless outperform the monolinguals in strategy use. This result reminds us of early research on metalinguistic awareness in India by Mohanty, who, in several investigations carried out between 1978 and 1987 (as summarized in Mohanty, 1994), made evident that bilingual Kond tribal children proficient in Kui and Oriya were significantly better than unilinguals (Kui) on a variety of metalinguistic tasks. In another line of research on multilingualism, subjects have been confronted with artificial language learning and unknown languages.

Decoding an unknown language

In a comparative study in which participants learnt a miniature linguistic system, Nation and McLaughlin (1986) came to the conclusion that multilinguals employed strategies to help them find resources to process linguistic information more efficiently in a situation when they were not given explicit instructions to learn. The most often cited early study comparing monolingual and bilingual strategy use is by Nayak et al. (1990), who investigated the language learning skills of 48 monolinguals and multilinguals (aged 16–42) during the learning of an artificial language. They concluded that multilinguals could adjust their learning strategies to the requirements of an implicit learning task more effectively than monolinguals, although no differences were found between multilingual and monolingual on an explicit learning task in which participants were told to find the rules. The authors suggested that the multilinguals were more capable of structuring their strategies to the requirements of the task, which leads to the conclusion that one reason for the superior performance of the multilingual participants is their greater flexibility in switching strategies. This early line of research has turned out to be of crucial importance for our current knowledge of the good language learner (see also McLaughlin & Nayak, 1989).

In more a recent work, Dahm (2015) reported on a strategy study that is part of a large-scale classroom investigation on Pluralistic Approach to Unknown Languages (PAUL) sessions, in which students were confronted with three unknown languages: Dutch, Italian and Finnish. The three successive sessions focused on metasemantic, metasyntactic and metaphonological activities. The findings of this highly innovative multilingual strategy training showed that the choice of strategy mainly depends on the perceived linguistic distance between the source language and the target language. The author found that the most frequently implemented strategies were comparison and translation, while the least readily implemented strategy was inferencing.

The results of the study contribute to our understanding of crosslinguistic interaction in TLA with respect to the use of metalinguistic and metacognitive strategies in language learning. It highlights the necessity to introduce strategy training in L2 English in order to benefit from the transferability of strategies and increase creative transfer (see also Allgäuer et al., 2016).

Furthermore, following Haskell (2001), Dahm presents a taxonomy for levels of transfer to determine the impact of connecting past learning to new situations. The taxonomy distinguishes between different levels of transfer, ranging from near transfer to creative transfer. Strategic knowledge involves knowledge of our own cognitive processes, occurring during self-monitoring of our progress when attempting to learn. According to Mehrnoosh et al. (2013: 2), declarative knowledge is added into the memory through chunks and is activated and subsequently increases in strength as a result of frequent use, i.e. practice. Proceduralization takes place when declarative knowledge is transformed into procedures, i.e. production rules, for performing a skill. Production rules also require practice to increase activation and strength. Practice at this stage results in the automatization of skill. However, declarative knowledge seems to be essential for successful transfer as it establishes the preconditions for the four other types. This is confirmed by Mehrnoosh et al. (2013: 14) when they consider that 'availability of speeded explicit declarative knowledge, or at least partially acquired proceduralized knowledge can enable learners to generalize it to unfamiliar contexts'.

In the following examples, which stem from a large-scale study on linguistic awareness in language attrition carried out at Innsbruck University by the DyME-research group (http://www.uibk.ac.at/anglistik/dyme), the subjects were given a text in a hitherto unknown language which turned out to be Romanian. Think-aloud protocols (TAP) were used to get some insights into the processing and production mechanisms applied by the young adult students. At the time of the study the Tyrolean students had been in touch with at least three languages (German/English/Latin and/or Italian or French/Spanish) during their school career. The following two TAP extracts give evidence of two participants' metalinguistic knowledge, that is, mainly grammatical knowledge, and the use of supporter languages such as German, English, French, Italian and Spanish, as well as Latin. At the same time the informants apply their world knowledge while trying to figure out the meaning of the text. We can also see from the following TAP examples that the students use compensatory strategies and show a high degree of creativity in the application of problem-solving activities. The main part of the translations added after the examples are translations from German to English. Original words in English are marked through capitals. The Romanian words (or morphemes) mentioned are underlined.

PAR-203-T1-12 *halt aus **Englisch** wie zum Beispiel das disabilitati das irgendwo und aus **Französisch** genauso, zum Beispiel das este este äh ja und*

*auch aus dem **Deutschen** in centrul und dann halt auch aus dem **Lateinischen** zum Beispiel das fumatorilor oder -or das keine Ahnung -ior ist ja eine Steigerung, ja das si ist auch **Französisch**, könnt auch **Italienisch** sein*

[Translation: PAR-203-T1-12 *well from **English** such as disabilitati somewhere and from **French** as well, for example the este este uh yes and also from **German** in centrul and then from **Latin** for example fumilator or the –or, no idea, -ior is an augmentation, yes and si is also **French**, could also be **Italian**]*

PAR-202-T2-12 ... *steht Hotel und da steht Hotelul dann steht da zum Beispiel Hotelului, das gleiche wie Restaurant oder Restaurant, da steht Restaurantul mit dem -ul dran nachher und die werden direkt angehängt und im Vergleich zu dem das ist nachher durch einen Bindestrich verbunden was in dem Fall nicht da ist und auch wie hier centrul, da hab ich mir gedacht vielleicht sind das bestimmte Präpositionen oder so etwas die dran gehängt werden so wie im **Lateinischen**, oder sonst was oder cum das teilweise angehängt wird*

[Translation: PAR-202-T2-12 ... *it says hotel and here it says hotelul and then it says here for example hotelului, the same as restaurant or restaurant, it says restaurantul with the –ul added to it and these are directly added and compared to that one it is linked with a hyphen afterwards which is not there in that case and also as here with centrul, I thought maybe these are particular prepositions or something like that, which are added as in **Latin**, or else or cum which is sometimes added]*

At the same time, the examples evidence strategies which show either tacit or explicit awareness, as discussed above. The next three examples illus-trate unconscious/tacit crosslinguistic awareness, while the two subsequent ones show conscious/explicit crosslinguistic awareness:

Unconscious XLA/CLIN

PAR-581-T1-13 ... *es ist am (...) in einer geschichtli [Wort abgebrochen] also in ein der historischen Umgebung weil istorica zona [spanisch ausgesprochen] (...) in (...) der Stadt (...)*

[Translation: PAR-581-T1-13 *..it is on (..) in a historic [word cut off] so in a historical setting because of istorica zona [Spanish pronunciation] (...) in (...) the city (...)]*

PAR-581-T2-13 ... *mit einem schönen Ambiente ... ambiente perfecta [spanisch ausgesprochen]*

[Translation: PAR-581-T2-13 ... *with a nice ambience ... ambiente perfecta [Spanish pronunciation]]*

In the examples above, the study participant activates her L4 (in this case Spanish) during the problem solving process by applying her knowledge of

Romance languages, in this case by intuitively sorting to Spanish pronunciation.

PAR-487-T1-13 ... *disabilitati (...) und ich glaube das heißt irgendwas mit DISABILITY [englisch ausgesprochen] und dass man da irgendwie nicht fähig ist, irgendetwas zu machen, also denke ich schon, dass es für Behinderte geeignet ist*

[Translation: PAR-487-T1-13 ... *disabilitati (...) and I think it means something like DISABILITY [English pronunciation] and that a person is somehow not able to do something, therefore I really believe that it is suitable for handicapped people]*

In this example the participant uses his L2 English to arrive at lexical understanding of the unknown word.

Conscious XLA/CLIN

In the following two examples a conscious link to the supporter languages is established by explicitly referring to language choice. In the first example we find explicit reference to L4 Spanish, in the second to L3 Italian.

PAR-487-T1-13 ... *also da steht irgendwas von rauchen, weil da fumilator steht und irgend [Wort abgebrochen] fumar [spanisch ausgesprochen] heißt ja auf Spanisch rauchen*

[Translation: PAR-487-T1-13 ... *ok it says something about smoking, because it says fumilator and som [word cut off] fumar [Spanish pronunciation] means smoking in Spanish]*

PAR-487-T1-13 ... *ah okay also im Italienischen heißt ja camera [italienisch ausgesprochen] Zimmer und da steht dass es also äh [stottert] Einzelzimmer gibt, Doppelzimmer (...) oh matrimonial [italienisch ausgesprochen]*

[Translation: PAR-487-T1-13 ... *ah ok in Italian it's camera [Italian pronunciation] room and it says that there are also ah [stutters] single rooms, double rooms (...) oh matrimonial [Italian pronunciation]]*

These examples show that, despite the well-known methodological flaws of TAP, a closer look at consciousness levels in multilingual processing presents a necessary step towards better insights into multilingual strategies. In many cases the multilingual subjects make informed and explicit guesses based on their multilingual learning experience and multilingual knowledge. However, there are also many other instances which show that our knowledge on multilingual strategies is still very limited.

Promising Research Avenues

Current work on strategies makes clear that in terms of multilingual learning we have very limited knowledge of the actual processes and their nature. There is no doubt that work on multilinguals can help to disentangle the interweaving and interdependence of metalinguistic and crosslinguistic awareness in processes of crosslinguistic interaction and offer food for further thought. However, what is clear is that the applied strategies clearly differ from mono- or even bilingual users in quality, thereby providing evidence in favour of the M-factor. More work on the link between CLIN and awareness has been very recently expressed by De Angelis *et al.* (2015b: 257) when they state that '[a]n interesting and certainly insightful direction of future research would be to apply similar research methods to other language combinations, age groups and (meta-)linguistic levels, as well as to extend some of the research designs so that they include the phenomenon of metalinguistic awareness' (see also Jessner, in press).

A crucial part of the emergent qualities in a multilingual user is cognitive flexibility, which is a feature of bi-/multilingual learners which has been recorded in a number of investigations comparing monolinguals with bilinguals. Specifically, it has turned out that bilinguals are more divergent, creative, original and flexible learners who are more fluent and elaborate. For instance, in his study on the cognitive development of Italian-English bilinguals and Italian monolinguals, Ricciardelli (1992) found that bilinguals who were more proficient in both Italian and English performed significantly better on creativity, metalinguistic awareness and reading than their monolingual controls. It should be noted, though, that in order for bi-/multilinguals to be able to profit from their knowledge of several languages and, by extension, from increased cognitive flexibility, they need high proficiency in both languages.

Further work regarding CLIN, XLA and MLA must also be recommended here, especially regarding cognitive flexibility, which, as the most important prerequisite mental ability, underlies the heightened creativity in multilingual language users and therefore enhances not only multilingual awareness but also provides more evidence of the M-factor (see Jessner & Török, in prep.).

There are a number of problems related to research on strategies, and it is argued that work on multilingual strategies can contribute to a better understanding of their nature and origin. DCT-based thinking and its application to multilingualism certainly alters research approaches to the field of strategy use in multilingual learning and thus learning in general. This implies that the current knowledge basis provided by work on SLA is simply not sufficient to cover all the eventualities of multilingual learning and use, particularly if we consider the complexity and dynamics of multilingual development.

The following quote from an experienced language learner illustrates our suggestions:

> I rather enjoyed getting acquainted with language archaeology. Was this only learning for learning's sake? I don't think so. Generally speaking, old languages helped me understand language change, seeming inconsistencies and illogical paradigms. Also, at some point all these languages came to support each other. English was a main source for German, my French benefited from Latin and English, Old English was not difficult at all because I could make use of Latin, English and German, whenever in trouble. (Popovic, 2009: 38)

Acknowledgements

References

Allgäuer-Hackl, E. and Jessner, U. (2014) Und was sagt die Mehrsprachigkeitsforschung dazu? Neue Perspektiven in der Mehrsprachigkeitsforschung und deren Relevanz für Unterricht und LehrerInnenbildung. In A. Wegner and E. Vetter (eds) *Mehrsprachigkeit und Professionalisierung in pädagogischen Berufen: Interdisziplinäre Zugänge zu aktuellen Herausforderungen im Bildungsbereich* (pp. 125–145). Opladen, Berlin and Toronto: Budrich.

Allgäuer-Hackl, E., Hofer, B. and Jessner, U. (2016) Emerging multilingual awareness in educational contexts: From theory to practice. *Canadian Modern Language Review* 72 (2), 157–182.

Alonso Alonso, R. (ed.) (2016) *Crosslinguistic Influence in Second Language Acquisition.* Bristol: Multilingual Matters.

Athanasopoulos, P. (2006) Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition* 9 (1), 89–96.

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

Bialystok, E. (2005) Consequences of bilingualism for cognitive development. In J.F. Kroll and A.M.B. de Groot (eds) *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 417–432). Oxford: Oxford University Press.

Bialystok, E. (2009) Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition* 12, 3–11.

Bialystok, E. and Barac, R. (2012) Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition* 122, 67–73.

Butler, D.L. (1997) The role of goal setting and self-monitoring in students' selfregulated engagement in tasks. Paper presented at the American Education Research Association Meeting, Chicago, IL.

Cenoz, J. (2001) The effect of linguistic distance, L2 status and age on cross-linguistic influence in third language acquisition. In J. Cenoz, B. Hufeisen and U. Jessner (eds) *Cross-Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives* (pp. 8–20). Clevedon: Multilingual Matters.

Part 3

Future Implications for Instructed Second Language Acquisition: Empirical Evidence

11 Academic Writing Development: A Complex, Dynamic Process

Wouter Penris and Marjolijn Verspoor

Introduction

Traditionally we look at learning outcomes by examining single outcomes. A new and future direction is to look at the actual process of development. Imagine an advanced 17-year-old student of English (L2) who has just finished secondary school in the Netherlands and wants to become an English teacher. He first completes a teacher training programme, and later at age 30 he obtains a university master's degree in the Netherlands. After high school he is quite advanced already (estimated low B2 level), and when he finishes his MA thesis he is able to write an academic research paper with the proper academic register (estimated C2 level). The purpose of the present chapter is to gain insight into the linguistic developmental process of his academic writing from a dynamic perspective. Over the course of 13 years (with a gap of five years), he writes many texts, 49 of which are selected to be examined in detail. The analyses show that his writing development is a long, complex, dynamic process, in which different subcomponents of the language change in interaction with each other. During his teacher training programme the language develops substantially differently from his development during his university programme, where more of an academic register is expected. As the language develops, longer noun phrases occur, and more academic words appear, as reflected in a longer average word length. The linguistic system becomes more accurate as the process of acquisition continues, at one point quite abruptly, but even at the end of the participant's studies, the writing still contains some errors.

This study not only gives insight into the differences between the characteristics of advanced formal writing and academic writing, but also has implications for the assessment and measurement of linguistic development. It turns out that not a single dependent variable develops linearly, and they all may level off during development. However, at the end, we will suggest

that the finite verb–token ratio is the best overall complexity and sophistication developmental measure, as it correlates highly with all other variables.

Background

A perspective from complex dynamic systems theory (CDST) – in other chapters simply referred to as dynamic systems theory (DST) – holds that when target behaviour is practised, the result is not always positive (de Bot & Larsen-Freeman, 2011; see also de Bot, Jessner & Török, Cergol Kovačević, and Bátyi, all this volume). Sometimes there is success and sometimes failure before actual mastery, resulting in variability, non-linear development, and progression in iterative steps. There is continuous movement in the system due to self-reorganization of variables or subsystems, interacting with one another and with outside factors, and all variables are dependent on their initial conditions. As a result of the availability of internal and external resources, these systems develop over time, which may occur at the expense of each other (van Dijk et al., 2011). Van Dijk and van Geert (2007) and Spoelman and Verspoor (2010) show that a period of increased variability may indicate a phase shift: a period of overuse or overgeneralization is often followed by a more stable phase with less variability, known as an attractor state. When multiple logically-connected linguistic subsystems develop, they may have different relations at any different phases over time (Spoelman & Verspoor, 2010; van Geert, 2008): they may develop synchronously and support each other, or asynchronously and compete with each other. Alternatively, one system may have to be in place before another can start developing.

To trace the linguistic development of different subsystems in an advanced writer over time, we have to identify and operationalize appropriate constructs, representing subsystems in the language. Ever more complex and sophisticated utterances and fewer simple utterances are expected to occur (Tomasello, 2003; van Geert, 2009), moving from 'the simplest and most frequent constructions to more complex and less frequent ones' (Verspoor et al., 2012: 256). The number of errors should also decrease (Verspoor & Behrens, 2011). However, we do not expect the same subsystems to change over time. As Verspoor et al. (2012) show with learners across five levels of proficiency (from absolute beginners to high intermediates), different subsystems seem to change from one level to the next: between Levels 1 and 2 they report significant differences mainly with respect to lexical measures, pointing to a lexical precursor relation for syntactic growth, in line with Caspi (2010). They also describe that between Levels 2 and 3 growth occurs mainly in syntactic measures, pointing to a competitive relation between lexical and syntactic subsystems. Between Levels 3 and 4 growth occurs in both lexical and syntactic measures, pointing to a supportive relation between them, and between Levels 4 and 5 growth is again

mainly observed in lexical measures, specifically in formulaic sequences, suggesting this subsystem of the lexical subsystem was focused on.

Three important studies report on dynamic relations between linguistic variables. From a 36-week CDST study using L2-English subjects with differing L1 backgrounds, Caspi (2010) concludes that lexical complexity is a precursor to lexical accuracy, which in turn is a precursor to syntactic complexity, which in turn is a precursor to syntactic accuracy. Verspoor et al. (2008) report on the developing system of an advanced L2 English learner who was studied longitudinally over a period of three years, discovering evidence for a competitive relationship between type–token ratio (TTR) and average sentence length (ASL), suggesting that a focus on word diversity took a toll on syntactic complexity and vice versa. Moreover, they suspect that a supportive relationship between average noun phrase length (ANPL) and finite verb–token ratio (FVTR) existed (Verspoor et al., 2008), possibly suggesting that increased noun phrase complexity also led to increased sentence complexity. Further investigating the same corpus as Verspoor et al. (2008), Schmid et al. (2011) focus on syntactic complexity, looking at the dispersion of sentence types and reporting that syntactic complexity first evolves through increased use of finite dependent clauses (adverbial, relative and nominal dependent clauses), which is later replaced by longer non-finite constructions and longer noun phrases. Additionally, after having constructed a customized lexical corpus for the subject, they observe that unique word use (occurring once in the corpus) increases over time, and frequent word use (20% most frequently occurring words) decreases, ultimately balancing out.

When the present study started, our learner was at about Level 5 of Verspoor et al. (2012), and he gradually turned into a highly proficient academic writer over a period of 13 years. This process culminated in a highly appraised MA thesis in Applied Linguistics, which was written in English, thus also reflecting the learner's L2 writing skills. In other words, the final writings were in an academic register. Because different subsystems seemed to develop at altering gradations at different proficiency levels, it would be interesting to see what general and specific subsystems developed at these higher levels of proficiency, especially at an academic level.

According to the literature, academic writing is characterized by structurally compressed constructions. Comparing spoken language to academic written language, Biber and Gray (2010) report that academic writing is structurally more 'elaborated', mainly with longer noun phrases and fewer finite dependent clauses. However, finite dependent clauses are commonly mentioned in the L2 literature on complexity and development (e.g. Wolfe-Quintero et al., 1998) and generally have been found to be good indicators of more advanced proficiency levels. This may be explained by the fact that most L2 studies do not include very advanced students, and it is very possible that at the most advanced levels the number of finite clauses decreases and noun structures are elaborated more. This is in line with Halliday et al.

(1999), who state that linguistic development proceeds from mostly parataxis (i.e. coordination) to hypotaxis (i.e. subordination), to language with much higher levels of lexical density (D) and more complex phrases (as opposed to more clauses), which we will call structurally elaborated.

We can illustrate our hypothesized stages in the development of academic writing by means of an example sentence originally written by our participant and rewritten for the sake of the argument. Note, however, that our third example is not necessarily a more felicitous sentence, but rather represents a fictitious example of a structurally elaborated sentence. The numbers below the examples are given to illustrate how the different constructs in our own study are operationalised and counted. They will be discussed in more detail in the methods section.

Example 1 Parataxis

[*The results of experiment 1*] **seem** to strongly support [*the existence of homoiophobia.*] [*All 22 expressions*] **were** acceptable, but [*they*] **were** all tested for [*acceptability*]. [*Only the first two of the 22*] **were** accepted by [*all judges*].

The text consists of a series of simple and compound sentences. There are no finite dependent clauses. (37 words, 3 sentences, 4 finite verbs, 7 noun phrases; ASL = 12.33, FVTR = 9.25, ANPL = 3.28 words.)

Example 2 Hypotaxis

[*The results of experiment 1*] **seem** to strongly support [*the existence of homoiophobia*]; even though [*all expressions that* **were** *tested for acceptability*] **were** in reality acceptable, only [*the first two ranked out of the 22*] **seem** to be completely accepted.

The text consists of a compound-complex sentence with two main clauses and two finite dependent clauses. (40 words, 1 sentence, 4 finite verbs; ASL = 40, FVTR = 10, ANPL = 6 words.)

Example 3 Structurally elaborated

[*The results of experiment 1*] **strongly** support [*the existence of homoiophobia with only two of the 22 constructions, all acceptable in the target language, accepted by the L2 judges*].

The text consists of a simple sentence with only one main clause and no finite dependent clauses, but it includes a structurally compressed construction, in this case a very long noun phrase. (29 words, 1 sentence, 1 finite verb; ASL = 29, FVTR = 29, ANPL = 13.5 words.)

To be able to trace syntactic development, different kinds of constructs can be used. However, we need to be aware of what these represent, to what extent they show advancement in linguistic development, and how they relate to one another in terms of overlap. One of the most commonly used measures is sentence- or T-unit length. Length measures are accepted as reliable indicators of general syntactic complexity, but they do not indicate which elements have made the sentences more complex (Wolfe-Quintero et al., 1998). For instance, our Example 2 is the longest sentence, but the length measure does not tell us that dependent finite clauses occur in it. Moreover it may be argued that Example 3, with its longer noun phrase, is more complex or sophisticated. That is why FVTR may be a better measure, as it reflects the internal complexity of the sentence better than ASL. In the current study, we will look at both ASL and ANPL. If there is a high correlation between the two measures, we may conclude that longer sentences at this level are partly due to the longer noun phrases or other non-finite constructions.

On the whole, ratios are considered better markers for proficiency growth (Wolfe-Quintero et al., 1998). In the current study, we will trace finite verb–token ratio (FVTR: all tokens in the text divided by the number of finite verbs), which will indirectly tell us whether the finite clause includes longer noun phrases and/or other non-finite constructions. Once an academic level of writing has been acquired, the FVTR should correlate highly with ASL and ANPL. All three are expected to increase, but as different aspects of the system may develop at different times, these three measures may not develop synchronously, and by tracing them, we may see which constructions develop at different times.

The length and ratio measures still do not tell us exactly what changes when. Two more variables we will trace are the type of sentences and types of dependent clauses. Traditionally, sentences can be classified as simple, compound, complex, compound-complex and finally incomplete sentences, known as fragments (see Appendix A for our exact definitions). We will trace the development of these sentence and clause types over time and expect especially the simple sentences to decrease over time early on, as it did in, for example, Verspoor et al. (2012) and Schmid et al. (2011). However, as Biber and Gray (2010) point out, sentences in academic writing are generally simple in nature (with one finite verb), although they may contain elaborated structures such as longer noun phrases, which means that in this study the number of simple sentences could actually increase at the very end.

The lexicon is also expected to become more sophisticated over time (Leki et al., 2008). In general, more proficient writers will use words that are less frequent, they will use a greater number of different words, and in academic writing the learner is expected to use more academic words. In English, academic words and less frequent words are generally longer words, so average word length (AWL) is also an indirect measure of lexical sophistication (Grant & Ginther, 2000). Indeed, more advanced English words are longer and lower in frequency (Wolfe-Quintero et al., 1998). AWL is an accurate indicator of essay complexity (Jarvis et al., 2003). However, the high

occurrence of function words in English is known to negatively influence AWL, as these are short and frequent, leading to a possible deflation (Schmid et al., 2011). Therefore, the average length of lexical words may be a better measure (Schmid et al., 2011).

Another lexical sophistication variable is the proportion of tokens belonging to the Academic Word List (Coxhead, 2000); when this proportion is high, the text is more lexically sophisticated. Recently, this list was updated through corpus research, resulting in a more precise and comprehensive list (COCA; Davies, 2008–). This variable is especially relevant within the context of the present study, in which academic register is expected to develop. In our study we will trace the AWL of lexical words and the number of words on the latter academic word list.

The two measures of lexical diversity employed in this study are TTR and D. TTR (number of tokens divided by the number of different words or types) indicates how diverse word use is. TTR is known to become unreliable for larger samples due to the high number of function words present in the English language (Schmid et al., 2011). However, in our samples we kept the text length the same, so this may not be an issue here. MacWhinney (2000) notes that D, an alternative version of TTR, proposed by McKee et al. (2000), is a better measure because of mathematical curve fitting; the function word declining curve in TTR is mathematically compensated for (Johansson, 2008). Miralpeix (2006) mentions that D is accurate across various languages and contexts, and that D indeed circumvents the known problems with TTR.

Finally, all L2 learners are known to produce errors, which decrease in frequency as proficiency increases (e.g. Leki et al., 2008; Wolfe-Quintero et al., 1998). Verspoor et al. (2012) suggest that beginners up to intermediate learners show fewer grammar than lexical and spelling errors, reporting high levels of variation across error types and proficiency levels, which, as the authors speculate, might decrease at higher levels. Finally, in a similar case to this one, Verspoor et al. (2008) report on the case of an advanced L2 subject who produced too few errors for error variables to become informative. In this study, we intend to look into errors, but because there are not enough specific types to trace, we will combine them into two general categories: lexical and syntactic errors.

Research questions

Thus, the aims of the present chapter are to find out how an advanced writer develops his language over time. The following research questions will be addressed:

(1) Does the advanced writer start to write longer sentences over time, using more complex and compound-complex structures, more finite

dependent clauses, longer noun phrases and other non-finite constructions, does he use fewer frequent words and more unique words, does he diversify his vocabulary more, does he start using longer and more academic words and, finally, does he make fewer errors over time?
(2) Will the number of simple sentences increase again at the end?
(3) Are there any developmental jumps in any of the variables?
(4) How do the main variables interact over time, and are there clear moments of interaction among the variables?

Method

By tracing various syntactic, lexical and accuracy variables of one advanced L2 English writer in 49 samples written over a period of 13 years, we will try to gain insight into the developmental process in his writing. From the analyses, we will also try to determine the best general measures of syntactic and lexical complexity and see how they interact with each other and with general accuracy over time.

Participant

The participant is a 17-year-old Dutch L1 learner of English as a Foreign Language, who finished his high school education at HAVO level (Higher General Secondary Education) with English at a low B2 level (CEFR; Council of Europe, 2001). For four years, he followed a teaching training professional programme for English, where he reached an estimated proficiency level of a high B2 or low C1 at the end. He subsequently taught English at high school for five years. Then he started studying at a research university and received an MA degree in Applied Linguistics after three years, where he reached an estimated C2 at the end. At the end of the study the subject was 30 years of age.

Texts

During his two studies (teacher training and university) he produced roughly 300 texts on various topics in different genres. No texts were produced during the intermission of five years. The teacher training prepared him for teaching and was generally less academic in nature than the university programme.

The first step was to select texts of a similar genre as assessed and agreed upon fully by the two authors. Only formally written texts were included (i.e. official assignments rather than journals or personal reflection). The production date of these texts was verified through study guides, emails and diaries. If unverifiable, the text was excluded, and when two texts had been produced in the same week, only one was randomly included. Thus, 49 texts

were included on grounds of formality and academic style. These were consecutively numbered by production date, and from each text a sample of 200 words was randomly chosen. All texts had been produced at home, based on homework assignments, without immediate time pressure and with free access to reference materials. Texts 1–31 were produced during teacher training (Phase 1), and Texts 32–49 during university (Phase 2).

Coding the texts

Each text was coded and analyzed using Codes for the Human Analysis of Transcripts (CHAT) and Computerized Language Analysis (CLAN), both created by MacWhinney and Snow (1990) for the Child Language Data Exchange Systems (CHILDES) project on a number of hand-coded features and a number of automated ones. The coding was done by the first author and checked by the second author. Any problematic analyses were discussed until agreement was reached.

A number of modifications to the texts were made, ensuring consistent and accurate quantification. Normally varying greatly in length, all proper names were replaced with *name*, all numbers with *numb*, and all geographic names were replaced with *place* – all four- or five-letter words. During university the original writings were interlaced with many quotes, so these were deleted or replaced with *quote* if the sentence structure allowed it. All enumerations exceeding three words were cut, avoiding overestimating ASL. Further alterations were made as required by CHAT, which are discussed in detail on the relevant website (MacWhinney, 2000).

Variables

In total, 13 variables measuring lexical and syntactic complexity were extracted, presented in Table 11.1, along with underlying calculations. Furthermore, lexical accuracy and syntactic accuracy were operationalized into three and six variables, respectively (see Table 11.2). However, the numbers in each separate category were too low to be analyzed separately, so the three variables for lexical accuracy and the six variables for syntactic accuracy were grouped together into two variables, referred to as lexical errors and syntactic errors, respectively.

Design and analysis

This is a longitudinal case study in which a number of variables were traced over time. To find general differences between the two phases, independent samples T-tests, with means and standard deviations were conducted (α was at 0.05). The raw data were plotted and polynomials of the second degree were added to visualize general trends. Pearson correlations (two-tailed) were calculated to check for general connections between

Table 11.1 Variables measuring syntactic and lexical complexity

Name	Operationalisation
Syntactic complexity variables	
Average sentence length	Total number of tokens divided by number of sentences
Finite verb–token ratio	Total number of tokens divided by the number of finite verbs
Average noun phrase length	Number of words in the longest possible noun phrases (with embedded noun phrases and clauses) divided by the number of noun phrases
Sentence type (S + C + F, CCX + CX)	Number of sentences per text that are: fragment (F; no finite verb)/simple (S; one finite verb)/ compound (C; two or more main clauses)/ complex (CX; one main clause and one or more finite dependent clauses)/compound complex (CCX; a complex sentence with compound elements)
Finite dependent clauses	Number of finite dependent clauses (relative, nominal and adverbial)
Non-finite clauses	Number of non-finite clauses
Lexical complexity variables	
Average word length	Total number of lexical items (nouns, verbs, adjectives and adverbs; function words not included) divided by the number of letters therein
Academic word list	Proportion of tokens in a text belonging to the academic word list, as supplied by the COCA website (Davies, 2008–)
Frequent lexical items	Proportion of tokens in a text belonging to the 50 most frequent words in the subject's own corpus, formed on the basis of the subject's own 49 texts
Unique lexical items	Proportion of tokens in a text that occur only once in the subject's own corpus
Type–token ratio	Types divided by tokens
Lexical density	TTR = D/N [(1 + 2N/D)$^{1/2}$ – 1] (built-in *vocd* command in CLAN (MacWhinney, 2000))

variables (α at 0.05 again). Because the texts were written in two phases, the analyses were done on the data of both phases, and then separately for each phase. Strong peaks or dips in variability were tested for significance through resampling techniques and Monte Carlo analyses (van Dijk *et al.*, 2011), using Poptools (Hood, 2004). By randomly reshuffling the data 5000 times, a

Table 11.2 Overview of the errors belonging to the lexical, syntactic and sentence level categories

Error type	Problem
Lexical errors	
Lexical error	Incorrect word use due to literal L1 translations of words or expressions, errors in function words, errors caused by the incorrect use of a word semantically related to target form, blends of English and Dutch
Spelling error	Incorrect spelling due to L1 interference, phonetic spelling, homophone spelling of target form, typos
Authentic error	Incorrect use of chunks/expressions beyond the word level
Syntactic errors	
Verb error	Incorrect predicate form or predicate use
Grammatical error	Incorrect use of apostrophe, congruence, word class, number, articles
Mechanical error	Incorrect use of capitals or spaces
Word order error	Word order (L1 interference or not)
Syntactic subcategory errors: punctuation at sentence and clause level	
Punctuation error at sentence level	Incorrect use of comma, full-stop, colon, semi-colon, leading to problems such as comma splice, fused sentence, fragments, or a restrictive or non-restrictive modifier punctuated incorrectly
Punctuation error at clause level	No comma before a conjunction that separates two clauses

Source: de Vries (2009).

Monte Carlo analysis calculates how often a similar peak occurs in the dataset when shuffled. If a peak occurred less than 250 times, it was deemed significant (α at 0.05).

To visualize interactions between variables, they were first normalised so they would be on the same scale. To neutralise the variability to some extent, they were supplanted with LOESS curves, which were calculated using Table Curve 2D (Systat Software, 2015).

Results

In this section, we will discuss syntactic development, lexical development and accuracy development separately. In each figure, related variables are presented side by side. The open space designates the five-year intermission between teacher training and university. At the end of the section, we examine the interactions between syntactic, lexical and accuracy variables.
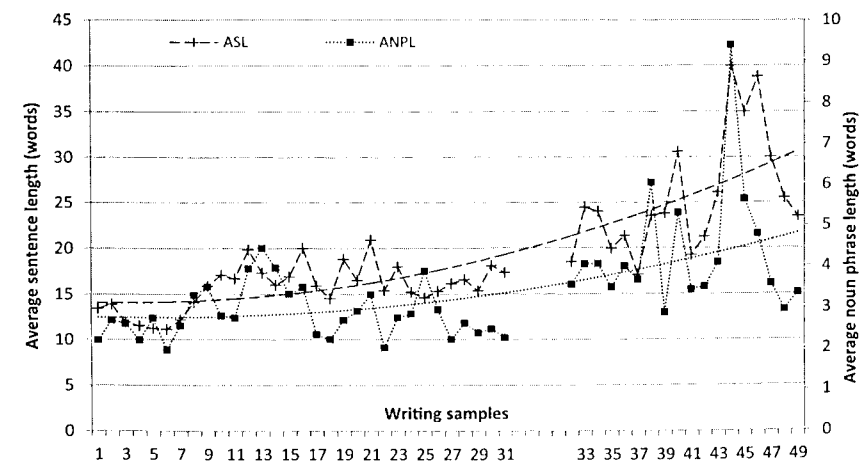

Figure 11.1 The development of average sentence length (ASL) and average noun phrase length (ANPL) over time, including polynomial trendlines (2nd)

Syntactic development

In Figure 11.1, ASL and ANPL have been plotted. On average, early sentences were between 10 to 15 and noun phrases between two and three words long. Towards the end there were texts with sentences of 40 words and noun phrases of nine words. However, no linear growth is visible in either variable. Polynomials indicate that both variables show relative stability in Phase 1, but increased growth in Phase 2. On average, ASL was significantly longer during Phase 2 ($M = 25.69$, $SD = 6.706$) than during Phase 1 ($M = 15.73$, $SD = 2.550$; $p < 0.001$). Likewise, ANPL was significantly longer during Phase 2 ($M = 4.32$, $SD = 1.543$) than during Phase 1 ($M = 2.85$, $SD = 0.631$; $p < 0.001$). A resampling and Monte Carlo analysis was not significant for ANPL ($p = 0.260$) or ASL ($p = 0.362$).

There was a very strong positive relationship between ASL and ANPL ($r = 0.769$, $p < 0.001$), although correlation strengths per phase differ: it is weaker in Phase 1 ($r = 0.421$, $p = 0.018$) as compared to Phase 2 ($r = 0.690$, $p = 0.002$), suggesting that increased sentence length in Phase 1 was not necessarily instigated by longer noun phrases. During Phase 2, the two developed more synchronously.

In the next step, FVTR was compared with ANPL in Figure 11.2. Early on, the FVTR was between six and nine and noun phrases between two and three words. Towards the end there were texts with an FVTR of around 13 and noun phrases of nine words. Again, no linear growth is visible in FVTR. On average, FVTR is significantly higher during Phase 2 ($M = 10.61$, $SD = 1.327$) than during Phase 1 ($M = 8.38$, $SD = 1.434$; $p < 0.001$). FVTR
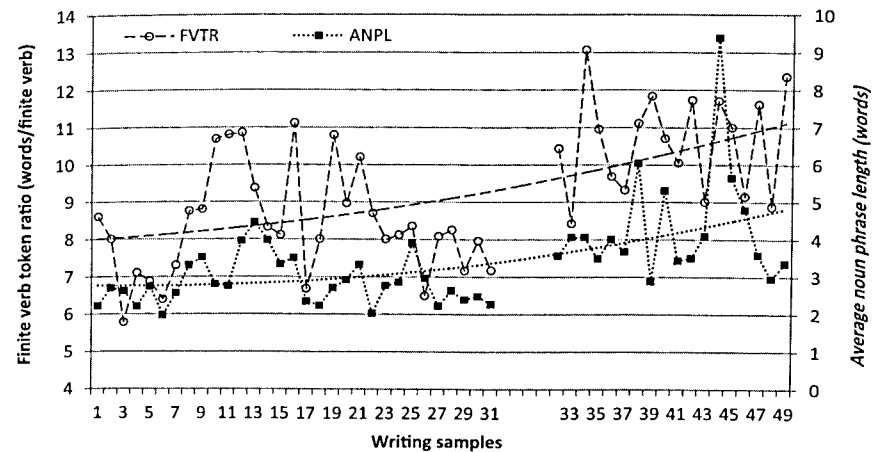
Figure 11.2 The development of finite verb–token ratio (FVTR) and average noun phrase length (ANPL) over time, including polynomial trendlines (2nd)

Figure 11.3 The smoothed (LOESS, 30%), normalised (0–1) development of finite verb–token ratio, average noun phrase length and average sentence length

Figure 11.4 The development of finite and non-finite dependent clauses with LOESS curves

seems to show stronger peaks and dips, but a resampling and Monte Carlo analysis showed that the observed peaks were not significant ($p = 0.48$).

There was a strong positive relationship between FVTR and ANPL ($r = 0.515, p < 0.001$), although correlation strengths differed per phase: it was moderate and significant in Phase 1 ($r = 0.455, p = 0.010$), suggesting that the higher FVTR was related to longer noun phrases, but during Phase 2 it was weak and insignificant ($r = 0.155, p = 0.538$), suggesting that at this stage other non-finite constructions were used to cause the higher FVTR.

When correlating FVTR with ASL, there also was a strong positive relationship ($r = 0.611, p < 0.001$), although a marked difference was found in correlation strength per phase: it was very strong and significant during Phase 1 ($r = 0.712, p < 0.001$), suggesting that at this time higher FVTR was related to longer sentences. However, the correlation was very weak and insignificant in Phase 2 ($r = 0.059, p = 0.816$), suggesting that there was no longer a relation between FVTR and sentence length in Phase 2, which indicates that longer sentences were not internally more complex.

Next, for easy comparison, the three variables FVTR, ANPL and ASL were normalized, smoothed and plotted in Figure 11.4 to examine changing relations among them.

In Figure 11.3, the FVTR takes off the fastest, showing a hump around Point 12, while the ANLP shows a hump at the same time, suggesting sentences become internally more complex by means of longer noun phrases. The FVTR goes down and then up again towards the end and noun phrases level off, suggesting there is a rather strong relation between FVTR and ANPL. The ASL, which is the more general measure, shows fewer curves but
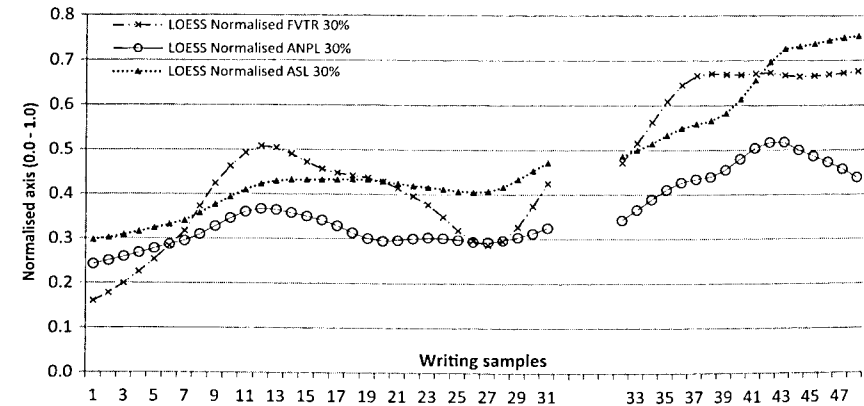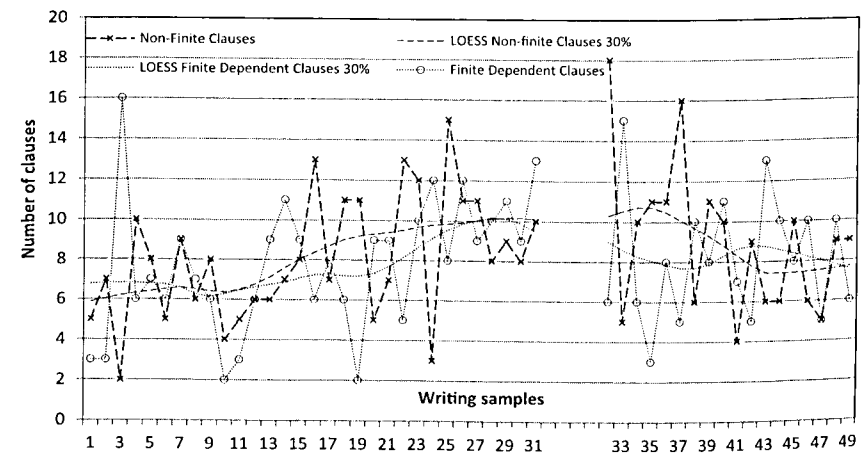
rises gently. In Phase 2 all three measures show a steep incline, with FVTR levelling off around Sample 37. ANPL rises too, but goes down around data point 41 when ASL and FVTR intersect, suggesting that sentences become longer, but no longer by means of longer noun phrases but by means of relatively more finite dependent clauses.

To see what may have caused the FVTR to level off, the number of finite dependent clauses was compared with the number of non-finite clauses per text in Figure 11.4. Both the number of finite and non-finite dependent clauses show

a high degree of variability over time, albeit with a general upward trend in Phase 1 and a downward trend in Phase 2. In Phase 1 there was a peak in finite dependent clauses at point 3, and in Phase 2 there were two high peaks in non-finite dependent clauses, none of which was significant ($p = 0.521$, $p = 0.919$, respectively). At around point 40 the LOESS curve shows an interesting shift: non-finite clauses go down and finite clauses go up slightly. Apparently, early on in Phase 2, non-finite clauses were overused to some extent, which explains the higher FVTR early on in Phase 2, which levelled off around point 40.

On average, the number of non-finite dependent clauses is higher in Phase 2 ($M = 9.00$, $SD = 3.742$) than in Phase 1 ($M = 8.06$, $SD = 3.108$), but this difference is not significant ($p = 0.351$). Similarly, the number of dependent clauses is higher in Phase 2 ($M = 8.11$, $SD = 3.104$), as compared to Phase 1 ($M = 7.81$, $SD = 3.341$), but this difference is not significant ($p = 0.754$). A Pearson correlation between the finites and non-finites was negative, but it was rather weak and did not reach significance ($r = -0.213$, $p = 0.142$). Similarly, correlation strengths per phase did not reach significance: the strength of the correlation was very weak in Phase 1 ($r = -0.069$, $p = 0.711$), suggesting the variables were unrelated to each other, and during Phase 2, although much stronger, it still did not reach significance ($r = -0.466$, $p = 0.051$).

In Figure 11.5, the sentence types have both been plotted; simple and compound sentences and fragments are added together as they contain main clauses only. Complex and compound complex sentences are added together, as they both contain finite dependent clauses. Note that the figures show the raw numbers of occurrences.

The polynomials indicate that the learner first used relatively many S + C + F, which gradually declined over time. Meanwhile, the number of CCX + CX seemed to stay relative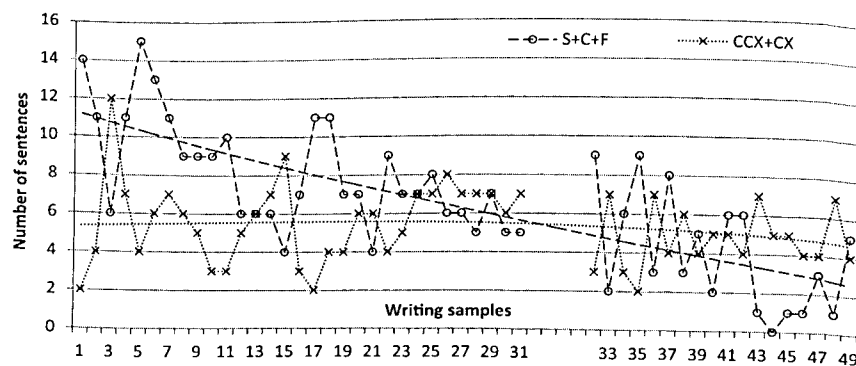ly stable, showing only a slight decline at the end of Phase 2. An interesting peak occurred in CCX + CX at data point 3, which was significant, suggesting it is a developmental peak: the learner is clearly overusing complex sentences from an early stage.

On average, S + C + F was significantly higher in Phase 1 ($M = 8.13$, $SD = 2.895$) as compared to Phase 2 ($M = 3.94$, $SD = 2.900$; $p < 0.001$). The CCX + CX was also slightly higher in Phase 1 ($M = 5.68$, $SD = 2.135$) than in Phase 2 ($M = 4.78$, $SD = 1.517$), but this difference was not significant ($p = 0.123$). This also implies a relative increase of complex constructions over time. Neither variable displays linear behaviour.

There was a weak negative relationship between S + C + F and CCX + CX ($r = -0.326$, $p = 0.022$), although correlation strengths differed per phase: both were strongly negative and significant (Phase 1, $r = -0.532$, $p = 0.002$; Phase 2, $r = -0.698$, $p < 0.001$). Logically, an increase in complex constructions occurred at the expensive of simple ones.

To see if the number of simple sentences increased at the end, simple sentences were plotted separately in Figure 11.6, with two LOESS curves, with the 24% showing more detail than the 30 curve. On average, significantly more simple sentence constructions were used during Phase 1 ($M = 6.45$, $SD = 2.249$) than during Phase 2 ($M = 2.89$, $SD = 2.139$; $p < 0.001$). However, if we look closely at the raw data and the 24% LOESS, there is a slight increase in the number of simple sentences around data point 47, suggesting that the writer started using more simple sentences towards the very end. Because we know that ASL kept increasing at that time in contrast to ANPL, the writer must have used other structurally elaborated constructions.
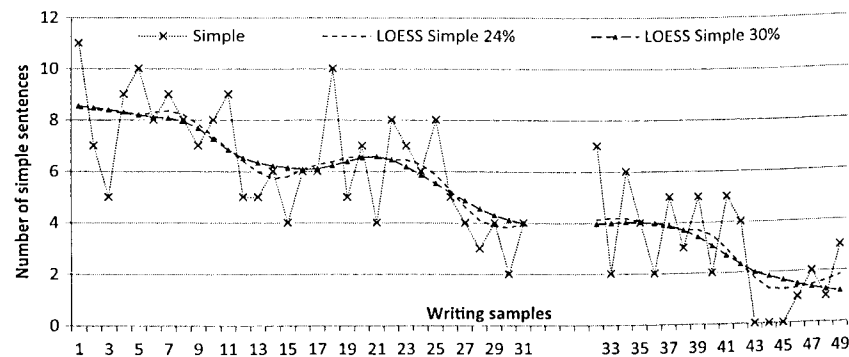

Figure 11.6 The development of the use of simple sentences, including smoothed LOESS at 24% and 30%

Lexical development

Lexical development was measured in terms of frequency (by comparing the 20% most frequently used words with words that occurred only once in
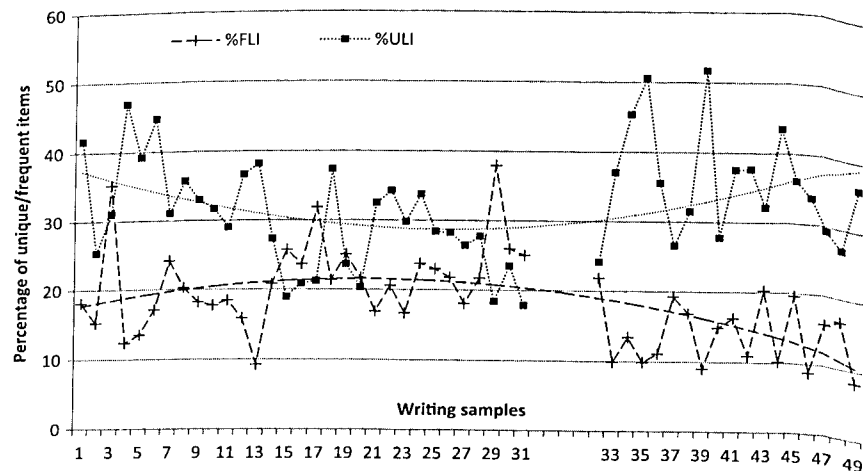

Figure 11.5 The development of the combined simple, compound sentences and fragments (S + C + F) and the combined compound complex and complex sentences (CCX + CX), including polynomial trendlines (2nd)

Figure 11.7 The development of the proportions of frequent lexical items (%FLI) and unique lexical items (%ULI), including polynomial trendlines (2nd)

the current corpus), diversity (operationalized as TTR and D) and sophistication (by comparing AWL and the proportion of academic words).

Figure 11.7 shows the proportions of frequent lexical items (%FLI) and unique lexical items (%ULI) in the student's corpus. There is variability in both measures, but the polynomials show that, initially, the %ULI starts out high and then steadily declines in Phase 1, while steady growth is visible in Phase 2. Actually, the highest %ULI peaks in the data were significant ($p = 0.041$), suggesting they are developmental peaks; the learner is overusing unique items, especially in Phase 2. The inverse occurs with %FLI, steadily increasing in Phase 1 with even a peak at the end, but it is not significant. In Phase 2 there is a steady decline.

On average, the %ULI was significantly lower during Phase 1 ($M = 30.13\%$, $SD = 7.682$) than during Phase 2 ($M = 35.72\%$, $SD = 7.947$; $p = 0.019$). Inversely, the %FLI was significantly higher during Phase 1 ($M = 21.17\%$, $SD = 6.157$) than during Phase 2 ($M = 14.04\%$, $SD = 4.451$; $p < 0.001$).

Frequently used words and unique words are natural competitors, because if you use the one, you cannot use the other. A Pearson correlation revealed that, indeed, there was a strong negative relationship between %FLI and %ULI, $r = -0.673$, $p < 0.001$; the higher the %ULI, the lower the %FLI, and vice versa. Also in the separate periods, the negative correlations remained strong (Phase 1: $r = -0.621$, $p < 0.001$; Phase 2: $r = -0.647$, $p = 0.004$).

Lexical diversity was traced by means of TTR and D, as illustrated in Figure 11.8. Although there is variability in both measures, neither the TTR nor D show much growth. However, on average, the TTR was significantly lower during Phase 1 ($M = 0.58$, $SD = 0.039$) than in Phase 2 ($M = 0.61$,
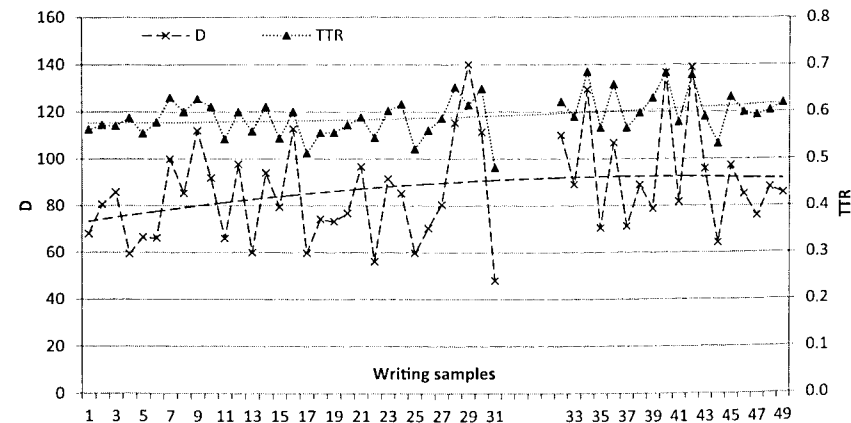

Figure 11.8 The development of lexical density (D) and type–token ratio (TTR), including polynomial trendlines (2nd)

$SD = 0.043$; $p = 0.009$). There was no difference in D between the two phases (Phase 1: $M = 82.63$, $SD = 21.148$; Phase 2: $M = 93.99$, $SD = 22.269$). As both variables measure lexical diversity, they should correlate highly. The Pearson correlation indeed revealed a very strong positive relationship between D and TTR, $r = 0.857$, $p < 0.001$ (two-tailed). When looking at the separate periods, correlation strengths were different, i.e. very strong in Phase 1 ($r = 0.817$, $p < 0.001$) and even stronger in Phase 2 ($r = 0.901$, $p < 0.001$).

Finally we looked at word sophistication. Figure 11.9 shows average length of lexical words and the proportion of academic words. The two lexical sophistication variables show very similar developmental patterns. There are high levels of variability, but the polynomials show little growth in Phase 1. In Phase 2, however, both variables show a sharp increase and both variables were significantly higher in Phase 2 ($p < 0.001$) (AWL in Phase 1: $M = 6.54$, $SD = 0.246$ and in Phase 2: $M = 7.22$, $SD = 0.425$; %ACWL in Phase 1: $M = 6.55$, $SD = 0.246$ and in Phase 2: $M = 17.74$, $SD = 7.540$).

A Pearson correlation analysis revealed that there was a very strong positive relationship between the two variables, $r = 0.849$, $p < 0.001$ (two-tailed), which, looking at the separate phases, was even stronger in Phase 2 than in Phase 1 (Phase 1: $r = 0.622$, $p < 0.001$; Phase 2: $r = 0.722$, $p < 0.001$). From these strong correlations, we may conclude that they partially tap into the same construct, and the use of academic words is also reflected by AWL.

Accuracy

The two language accuracy variables that were studied were the number of lexical errors and the number of syntactic errors. An average of 44% of the syntactic errors were caused by sentence-level errors in punctuation and
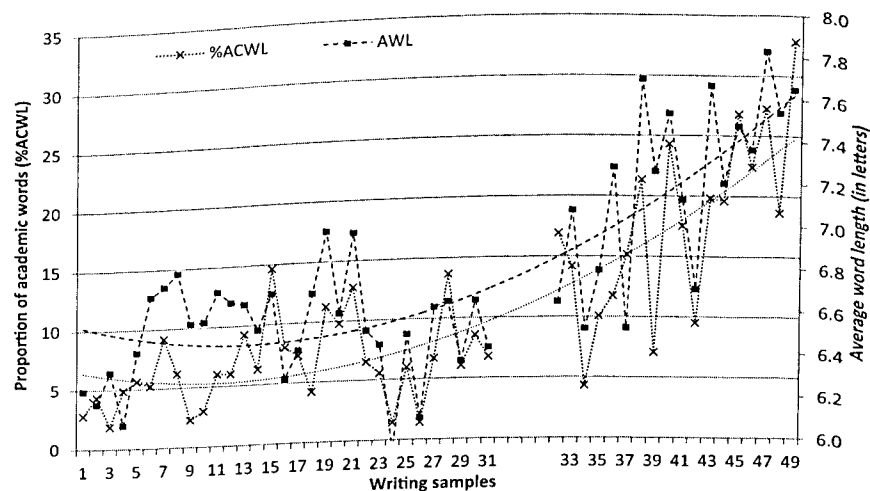
Figure 11.9 The development of average length of lexical words (AWL) and the proportion of tokens belonging to the COCA academic word list (%ACWL), including polynomial trendlines (2nd)

mechanics, which could be argued as less serious. Therefore, Figure 11.10 shows lexical errors and syntactic errors without punctuation and mechanical errors.

Over the course of the study, both variables show a steady decline, eventually leading to practically no errors at the end. The polynomials show that there are slightly more lexical than syntactic errors in the beginning, but the
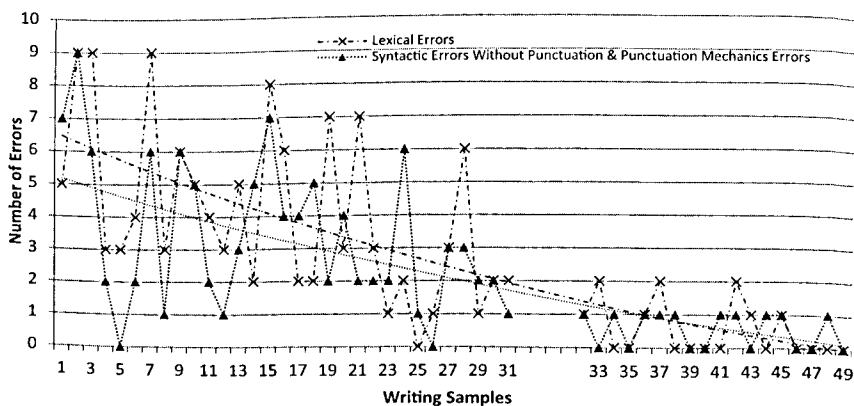
Figure 11.10 The development of lexical and syntactic errors (excluding punctuation and punctuation mechanics errors), including polynomial trendlines (2nd)

decline over time is also steeper. On average, there were significantly ($p < 0.001$) more lexical and syntactic errors in Phase 1 than in Phase 2 (Lexical Phase 1: $M = 4.06$, $SD = 2.568$ and Lexical Phase 2: $M = 0.56$, $SD = 0.784$; Syntactic Phase 1: $M = 6.06$, $SD = 2.682$ and Syntactic Phase 2: $M = 1.22$, $SD = 1.555$).

A Pearson correlation revealed that there was a strong positive relationship between lexical errors and syntactic errors ($r = 0.550$, $p < 0.001$). When looking at the separate periods, correlation strengths did not reach significance (Phase 1: weak, $r = 0.174$, $p = 0.350$ and Phase 2: moderate, $r = 0.279$, $p = 0.263$). From this we may conclude that lexical and syntactic errors both decrease, but not always synchronously, suggesting some degree of competition.

Interactions between the three main subsystems

Since the different subsystems under investigation may interact with each other during their development, it would be interesting to have a closer look at these interactions. However, we will limit ourselves here to examining the most general relation between syntactic, lexical and accuracy development. Before doing so, we need one measure to represent each subsystem. To this end we will assume that the measure that correlates best with the other measures in one subsystem is the one that subsumes those other measures and is therefore the most suitable to represent the subsystem as a whole.

For the syntactic measures, Table 11.4 in Appendix B shows that general sentence length did not correlate significantly with the number of dependent clauses, nor with the number of complex sentences. The FVTR shows significant ($p < 0.05$) correlations with all the other syntactic measures except simple sentences. Moreover, the FVTR correlates significantly ($p < 0.05$) with all lexical variables. Therefore, we conclude that the FVTR, a syntactic complexity measure, is the best general variable to represent syntactic growth.

For lexical measures, Table 11.5 in Appendix C shows that no single lexical variable correlated significantly with all lexical variables. AWL is considered the best measure for lexical complexity, as it showed the most highly significant ($p < 0.01$) correlations. Finally, for accuracy development, we propose to use the total number of errors, including punctuation and mechanical errors.

Figure 11.11 shows how the three variables interact over time. Note that the variables are normalized for easy comparison, and for accuracy the number of total errors (a negative score) have been subtracted from 1 to create a positive accuracy score. Figure 11.11 shows that, in Phase 1, all three variables first grew synchronously, then word complexity levelled off, while accuracy and sentence complexity kept growing. At around Sample 11, syntactic complexity improved at the expense of accuracy. After Sample 19, accuracy increased at the expense of both lexical and syntactic complexity. In Phase 2, both lexical and syntactic complexity grew synchronously, with
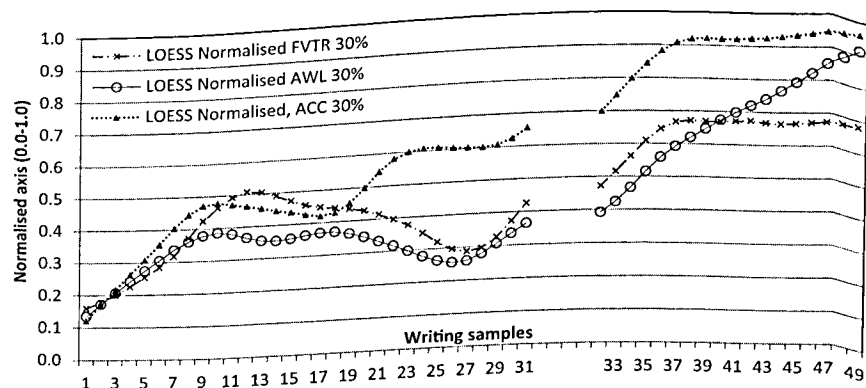
Figure 11.11 The smoothed (LOESS 30%) development of the syntax (FVTR), the lexicon (AWL) and accuracy (ACC)

accuracy lagging behind. At about Sample 37, syntactic complexity and accuracy levelled off, and lexical complexity kept increasing.

Discussion

In this case study we traced the writing development of an advanced learner over a period of 13 years with a gap of five years. In the first phase, the writer attended a teacher training college and was asked to write formal essays, progressing from an estimated low B2 level to a low C1 level. In the second phase, he attended university and was expected to write more academically, progressing from an estimated low C1 to C2 level. The case study is interesting because we were able to compare development at two proficiency levels within the same learner over time to see what subsystems of the language develop at each stage and how the variables compare.

We observed that all variables showed high degrees of variability, so we may conclude that linguistic development is non-linear. At present, the underlying causes of the variability are unknown, but it is highly likely that some are due to task effects or the selected samples of the texts. Moreover, some variability is due to language internal dynamics; for example, when longer noun phrases are used, fewer finite dependent clauses are needed to express the same notion. But as a CDST approach would argue, some variability is developmental in nature. Indeed, as our Monte Carlo analyses showed, two peaks were so odd that they could not have been due to chance: one in the use of complex sentences and one in the use of unique words, suggesting that both of these variables – one in syntax and one in the lexicon – were overused before settling down in a less variable state. Furthermore, the finite and non-finite dependent clauses showed moments of overuse. The

peaks were not significant in the Monte Carlo analyses, but the raw data and the LOESS curves strongly suggest that finite dependent clauses were overused early on in Phase 1, and non-finite dependent clauses were overused early on in Phase 2.

We also wanted to know what variables seem to grow at which stage. Table 11.3 shows that almost all variables have a significantly higher average in Phase 2. Overall, sentences became longer (in line with Wolfe-Quintero *et al.*, 1998), with a higher FVTR, longer noun phrases (in line with Biber & Gray, 2010), longer words (in line with Jarvis *et al.*, 2003), more academic words (in line with Leki *et al.*, 2008), more unique words (in line with Schmid *et al.*, 2011) and more lexical diversity through TTR (in line with Halliday *et al.*, 1999).

Variables that did not grow significantly were finite dependent clauses and non-finite clauses, which is in line with Schmid *et al.* (2011), who report a growth in dependent clauses, which later decline as non-finite

Table 11.3 Overview of peaks of overuse and differences between phases per variable

Variable	Peaks of overuse	Significant difference between Phase 1 and 2
Syntactic complexity variables		
Average sentence length		Higher in Phase 2
Finite verb–token ratio		Higher in Phase 2
Average noun phrase length		Higher in Phase 2
Combined fragments, simple, and compound sentences		Lower in Phase 2
Combined compound-complex and complex sentences	In Phase 1	No difference
Simple sentences		Lower in Phase 2
Finite dependent clauses		No difference
Non-finite clauses		No difference
Lexical complexity variables		
Average word length		Higher in Phase 2
Academic word list		Higher in Phase 2
Frequent lexical items		Lower in Phase 2
Unique lexical items	In Phase 2	Higher in Phase 2
Type–token ratio		Higher in Phase 2
Lexical density		No difference
Error variables		
Syntactic errors		Lower in Phase 2
Lexical errors		Lower in Phase 2

constructions take their place. This means that the number of finite dependent clauses, usually found to be a good indicator of development (Leki *et al.*, 2008), are not a very good measure at this advanced level. The dependent clauses show best what we mean by some subsystems developing at different times. Early on in Phase 1 there is a clear overuse of finite dependent clauses, while there is a rather clear overuse of non-finite dependent clauses early on in Phase 2 – even though the peaks were not significant. Towards the end, the two types of dependent clauses are much more balanced, resulting in a lower FVTR, which we saw in Figure 11.11. These findings are not quite in line with our expectations. Based on Biber and Gray (2010), we expected the number of finite dependent clauses to decline over time and the number of simple sentences (those containing one finite verb) to increase towards the end. We did see a slight increase of simple sentences towards the very end in the raw data and the 24% LOESS curve.

As far as lexical measures are concerned, Schmid *et al.* (2011) found a strong overlap between TTR and D, which is in line with our findings. However, when comparing Phase 1 to Phase 2, D remained the same, whereas TTR showed a significant difference. We conclude that D is not a useful measure at this advanced level, but TTR is.

As far as interactions are concerned, Figure 11.3 is interesting, as it shows that the three interrelated syntactic measures with partially overlapping constructs – ASL, FVTR and ANPL – do not develop synchronously the whole time. The ASL rises gently with few curves, which means it is a very general measure that does not really show when and how sentences become more complex. The FVTR and ANLP show more ups and downs in the LOESS curves, and they develop rather synchronously, which suggests that both the increase and decrease in FVTR is mainly due to noun phrase length. In Phase 2 all three measures show a steep incline, but what is interesting is that ASL keeps growing until the end, whereas the FVTR levels off around data point 37, and the ANPL even goes down around data point 41. From Figure 11.4 we can see that the decline of the FVTR is mainly due to the fact that the number of non-finite dependent clauses goes down. The fact that sentence length increases therefore cannot be due to the length of noun phrases, which actually decreases, or the number of non-finite constructions, as the FVTR goes down. A logical explanation might be the number of complex and compound-complex constructions with finite dependent clauses. As a matter of fact, examining these constructions separately supports this hypothesis; consider Figure 11.12, in which the proportions of the different sentence types are plotted. The relative number of simple sentences goes down, and both complex and compound complex constructions increase to the very end, explaining the longer sentences.

Finally, we wanted to see the interactions between the development of the three main subsystems in question: syntactic complexity, lexical complexity and accuracy. To be able to trace general syntactic and lexical
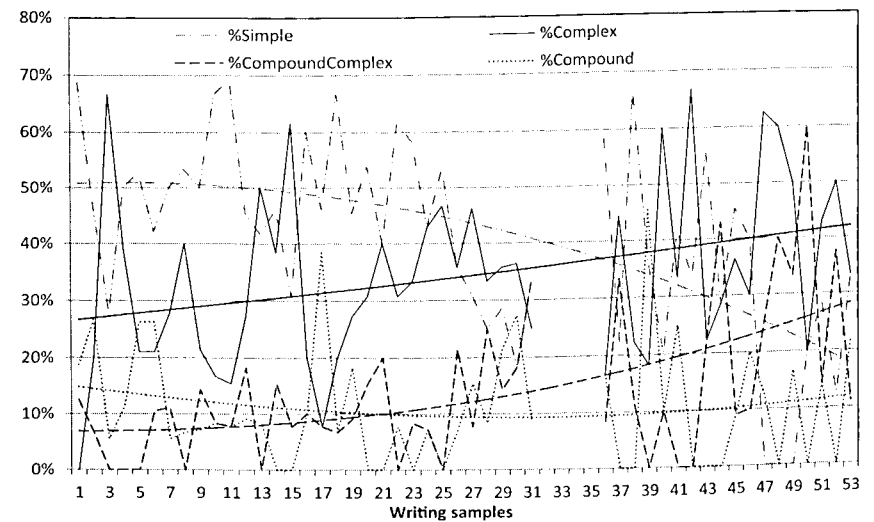

Figure 11.12 Development of proportionate sentence-type use with polynomials to the 2nd degree

complexity, we determined that FVTR and AWL were most representative of their respective subsystems, as they correlated best with the other measures in their own category. For accuracy, we simply took the inverted total number of errors.

One interesting interaction of the three variables could be seen in the LOESS curves in Figure 11.11. For this advanced writer, the first hump in development was in accuracy, then in syntactic complexity and then a small hump in lexical complexity. Accuracy kept improving over time, but both syntactic and lexical complexity decreased towards the end of the teacher training programme. During the university programme, both syntactic and lexical complexity increased steadily, but at one point syntactic complexity levelled off and lexical complexity kept increasing. The LOESS curve clearly showed that the different constructs at some points in time did not develop synchronously. These findings contrast somewhat with Caspi (2010), who reported on learners in a similar academic phase as the one our subject was in. She found the following order in development by means of modelling for three of the four subjects: lexical complexity > lexical accuracy > syntactic complexity > syntactic accuracy. The reason for the differences could be the actual proficiency level of the participants at the beginning of the study, as they had different first languages from our subject. Moreover, as several studies have shown, there is variation among learners. For instance, we can read in Chan *et al.* (2015) that even identical twins in very similar circumstances do not develop in exactly the same manner.

The implications for future directions are the following. The detailed analysis of one writer over the course of 13 years shows that linguistic development in writing is a rather erratic process when examined up close, with lots of variability in almost every measure we traced. Therefore, if we want to know about the actual process of development, we must look at these many different related variables to ascertain what changes occur in what manner and what order. At the same time, we must realise that most of these individual measures, such as finite or non-finite dependent clauses, are useful for general developmental purposes, as they will change over time. However, if we are interested in development in very general terms, we should look at ASL and AWL, since both showed rather continuous growth until the end and might therefore be considered overall as broad developmental measures. This can be explained by the fact that they average over many instances (sentences and words). However, we would argue that another excellent general measure at these advanced levels would be the FVTR because it significantly correlated not only with all syntactic variables but also with all lexical variables. We think it is such an excellent measure because it takes the internal complexity of a clause in terms of non-finite constructions including longer noun phrases into account. However, it must be noted that this particular measure does not show development at lower levels (Verspoor et al., 2011).

References

Biber, D. and Gray, B. (2010) Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes* 9, 2–20.

Caspi, T. (2010) A dynamic perspective on second language development. Unpublished doctoral dissertation, University of Groningen.

Chan, H., Verspoor, M. and Vahtrick, L. (2015) Dynamic development in speaking versus writing in identical twins. *Language Learning* 65 (2), 298–325.

Council of Europe (2001) *A Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Coxhead, A. (2000) A new academic word list. *TESOL Quarterly* 34, 213–238.

Davies, M. (2008–) The corpus of contemporary American English: 450 million words, 1990–present. See http://corpus.byu.edu/coca/.

de Bot, K. and Larsen-Freeman, D. (2011) Researching SLD from a DST perspective. In M. Verspoor, K. de Bot and W. Lowie (eds) *A Dynamic Approach to Second Language Development: Methods and Techniques* (pp. 5–23). Amsterdam and Philadelphia, PA: John Benjamins.

de Vries, F. (2009) Error Analysis as a Tool for Measuring Second Language Development. Unpublished Master's degree, University of Groningen.

Grant, L. and Ginther, A. (2000) Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing* 9, 123–145.

Halliday, M.A.K., Matthiessen, C.M. and Yang, X. (1999) *Construing Experience Through Meaning: A Language-based Approach to Cognition*. London: Cassell.

Hood, G. (2004) *Poptools*. Canberra: Pest Animal Control Co-operative Research Center (CSIRO).

Jarvis, S., Grant, L., Bikowski, D. and Ferris, D. (2003) Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing* 12, 377–403.

Johansson, V. (2008) Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund Working Papers in Linguistics* 53, 61–79.

Leki, I., Cumming, A.H. and Silva, T. (2008) *A Synthesis of Research on Second Language Writing in English*. New York: Routledge.

MacWhinney, B. (2000) *The CHILDES Project: Tools for Analyzing Talk* (3rd edn). Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B. and Snow, C.E. (1990) The child language data exchange system: An update *Journal of Child Language* 17, 457–472.

McKee, G., Malvern, D. and Richards, B.J. (2000) Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing* 15, 323–383.

Miralpeix, I. (2006) Age and vocabulary acquisition in English as a foreign language (EFL). In C. Muñoz (ed.) *Age and the Rate of Foreign Language Learning* (pp. 89–106). Clevedon: Multilingual Matters.

Schmid, M.S., Verspoor, M.H. and MacWhinney, B. (2011) Coding and extracting data. In M.H. Verspoor, K. de Bot and W. Lowie (eds) *A Dynamic Approach to Second Language Development: Methods and Techniques* (pp. 39–54). Amsterdam and Philadelphia, PA: John Benjamins.

Spoelman, M. and Verspoor, M.H. (2010) Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics* 31, 532–553. doi: 10.1093/applin/amq001.

Systat Software (2015) *Table Curve 2D* (v5.01 edn). Richmond, CA: Systat Software Inc.

Tomasello, M. (2003) *Constructing a Language: A Usage-based Account of Language Acquisition*. Cambridge, MA: MIT Press.

van Dijk, M. and van Geert, P. (2007) Wobbles, humps and sudden jumps: A case study of continuity, discontinuity and variability in early language development. *Infant and Child Development* 16, 7–33.

van Dijk, M., Verspoor, M. and Lowie, W. (2011) Variability and DST. In M.H. Verspoor, K. de Bot and W. Lowie (eds) *A Dynamic Approach to Second Language Development: Methods and Techniques* (pp. 55–84). Amsterdam and Philadelphia, PA: John Benjamins.

van Geert, P. (2008) The dynamic systems approach in the study of L1 and L2 acquisition: An introduction. *Modern Language Journal* 25, 340–375.

van Geert, P. (2009) A comprehensive dynamic systems theory of language development. In K. de Bot and R.W. Schrauf (eds) *Language Development Over the Lifespan* (pp. 60–104). New York: Routledge.

Verspoor, M.H. and Sauter, K. (2000) *English Sentence Analysis: An Introductory Course*. Amsterdam and Philadelphia, PA: John Benjamins.

Verspoor, M.H. and Behrens, H. (2011) DST and a usage-based approach to SLD. In M.H. Verspoor, K. de Bot and W. Lowie (eds) *A Dynamic Approach to Second Language Development: Methods and Techniques* (pp. 25–38). Amsterdam and Philadelphia, PA: John Benjamins.

Verspoor, M.H., Lowie, W. and van Dijk, M. (2008) Variability in L2 development from a dynamic systems perspective. *Modern Language Journal* 92, 214–231.

Verspoor, M.H., Schmid, M.S. and Xu, X. (2012) A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing* 21, 239–263.

Wolfe-Quintero, K., Inagaki, S. and Kim, H.-Y. (1998) *Second Language Development in Writing: Measures of Fluency, Accuracy and Complexity*. Honolulu, HI: University of Hawaii, Second Language Teaching and Curriculum Center.

Appendix A: Definitions

Sentence complexity and clause types are as defined by Verspoor and Sauter (2000). Note that definitions are determined by whether the verb in a clause is finite or not.

A sentence was called:

(1) **Fragment**: a syntactically incomplete sentence, for example, when it has no finite verb or separate main clause.
(2) **Simple**: a sentence with one finite clause, which may or may not contain non-finite dependent clauses.
(3) **Compound**: a sentence that has two or more main clauses, which may or may not contain non-finite dependent clauses.
(4) **Complex**: a sentence with one main clause and one or more finite dependent clauses.
(5) **Compound-complex**: a sentence that has both compound and complex elements. It may contain two or more main clauses with at least one finite dependent clause in one of the elements. Otherwise, it may be a complex sentence with a main clause and coordinated finite dependent clauses.

Appendix B: Correlation Table of Syntactic Measures

Table 11.4 Correlation table of syntactic measures

		Average sentence length	Non-finite clauses	Finite dependent clauses	Combined fragments, simple and compound sentences	Combined compound-complex and complex sentences	Average noun phrase length	Simple sentences	Finite verb–token ratio
Average sentence length	r	1							
	sig.								
Non-finite clauses	r	-0.044	1						
	sig.	0.765							
Finite dependent clauses	r	0.160	-0.213	1					
	sig.	0.271	0.142						
Combined fragments, simple and compound sentences	r	-0.796**	0.080	-0.524**	1				
	sig.	0.000	0.585	0.000					
Combined compound-complex and complex sentences	r	-0.185	-0.222	0.751**	-0.326*	1			
	sig.	0.203	0.126	0.000	0.022				
Average noun phrase length	r	0.769**	-0.067	0.194	-0.617**	-0.035	1		
	sig.	0.000	0.649	0.182	0.000	0.810			
Simple sentences	r	-0.772**	0.074	-0.452**	0.918**	-0.241	-0.563**	1	
	sig.	0.000	0.614	0.001	0.000	0.095	0.000		
Finite verb–token ratio	r	0.611**	0.112	-0.414**	-0.393**	-0.536**	0.515**	-0.322*	1
	sig.	0.000	0.443	0.003	0.005	0.000	0.000	0.000	0.024

Notes: *Correlation is significant at the 0.05 level (two-tailed). **Correlation is significant at the 0.01 level (two-tailed).

Appendix C: Correlation Table of Lexical Measures

Table 11.5 Correlation table of lexical measures and finite verb–token ratio

		Average length of lexical words	Type–token ratio	Lexical diversity	Percentage of frequent lexical items	Percentage of unique lexical items	Proportion of tokens from academic word list	Finite verb–token ratio
Average length of lexical words	r	1						
	sig.							
Type–token ratio	r	0.251	1					
	sig.	0.082						
Lexical diversity	r	0.142	0.857**	1				
	sig.	0.331	0.000					
Percentage of frequent lexical items	r	-0.410**	-0.232	0.081	1			
	sig.	0.003	0.108	0.579				
Percentage of unique lexical items	r	0.149	0.182	-0.141	-0.673**	1		
	sig.	0.307	0.212	0.332	0.000			
Proportion of tokens from academic word list	r	0.849**	0.199	0.141	-0.326*	-0.040	1	
	sig.	0.000	0.170	0.335	0.022	0.784		
Finite verb–token ratio	r	0.537**	0.381**	0.294*	-0.531**	0.293*	0.497**	1
	sig.	0.000	0.007	0.040	0.000	0.041	0.000	0.000

Notes: *Correlation is significant at the 0.05 level (two-tailed). **Correlation is significant at the 0.01 level (two-tailed).

12 Lessons Learned from the Integration of Findings from Identical and Semi-Cognate Visual and Auditory Processing in Bilingual Cognate Studies: Implications for Future Studies

Kristina Cergol Kovačević

Introduction

This chapter presents a study of the influence of complete and partial orthographic overlap (identical and semi-cognates) on the processing of phonologically highly similar cognate equivalents. Result patterns obtained on lexical decision tasks (LDTs) in the visual and auditory modalities are compared to uncover differences in cognate processing between two languages of very different orthographic depths; while English is an orthographically deep language, Croatian is orthographically shallow in terms of letter–sound correspondence, which is believed to condition differences in the visual processing of the two and may contaminate the comparison of results between the given languages. To avoid the above-mentioned problem, in this paper the LDT results obtained in the visual modality are presented in relation to the LDT results presented in the auditory modality, and the effects obtained are discussed with reference to both modalities. Moreover, this chapter presents a longitudinal attempt to capture possible changes in the bilingual processing of the participants as the measurement is taken at two different points in time considered to be critical for English second language (L2) proficiency development in the studied speakers, brought about by the instruction and regular usage of the English language in the academic environment and the students' free time. This is also an attempt to create an (admittedly

very fragile) bridge between the currents of the Old and New Schools of psycholinguistic research (de Bot, 2010, 2015).

Cognates and Cognate Effects

Cognates have been defined in terms of form (phonological and orthographic) and meaning (semantic) overlap (e.g. Blumenfeld & Marian, 2005; Dijkstra, 2005; Sherkina-Lieber, 2004; Sunderman & Schwartz, 2008; Yule, 2006). Despite some orthographic and often more common phonological differences, speakers of two or more languages have repeatedly been shown to process cognate words with more ease than non-cognate words (e.g. Costa et al., 2000; Lemhöfer & Dijkstra, 2004; Roselli et al., 2014; Sherkina, 2003; Sherkina-Lieber, 2004). This cognate facilitation effect has also been found in multilingual processing (e.g. Lemhöfer et al., 2004; Szubko-Sitarek, 2011). An experimental psycholinguist may argue that the definition provided by Carroll (1992) may well encapsulate the experience of the participants involved in experimental studies. She proposes that cognates are lexical units from various languages for which bilingual (and multilingual) speakers feel that they are somehow 'the same thing' (Carroll, 1992: 94). Researchers use this fact and manipulate cognate features in terms of the extent of their similarity to establish the effects such manipulation has on cognate processing, reflecting the feeling of relatedness of cognate pairs between languages.

Owing to the features they share between languages, cognates used to be utilized in the investigations of language (non-)selectiveness of bilingual lexical access (e.g. Caramazza & Brones, 1979; Cergol Kovačević, 2012; Dijkstra et al., 1999; for an overview of cognate processing in bilingual and multilingual studies, see for example, de Groot, 2011). In other words, the idea was to research whether the two (or more) languages of a speaker are activated in parallel during lexical processing or not. Nowadays it is safe to say that researchers have reached a general consensus in terms of treating bilingual lexical access as parallel. This is reflected in a well-known and data-supported localist-connectionist model of visual bilingual lexical processing, the Bilingual Interactive Activation+ model (BIA+) (Dijkstra & van Heuven, 2002). In short, according to BIA+, upon the presentation of an input string, sublexical and lexical orthographic candidates from both languages of a bilingual are activated in case they share similarity with the input. Their activation depends on the resting level activation of individual items (due to their frequency, word familiarity, speaker L2 proficiency, etc.). The activated orthographic representations extend activation to the associated sublexical and lexical phonological and semantic representations. The activation between the matching levels of sublexical and lexical orthography and phonology is lateral. Both the word identification system and the task-decision system, which is affected by non-linguistic (contextual) information, are

involved in the processing. Cognate facilitation or inhibition effects, depending on the task involved, have been used to support this language-selective view of bilingual processing. For example, in an LDT, where participants are presented with words and pseudowords (strings of letters designed according to the phonotactic rules of the studied language(s) but which have no meaning) of one or more of their languages and need to respond as quickly and as accurately as possible to say if the stimulus is a word or not, cognate facilitation effect (faster reaction times (RTs) to cognates than to non-cognates) has repeatedly been found (e.g. Caramazza & Brones, 1979; Dijkstra et al., 1999, 2010; Lemhöfer & Dijkstra, 2004). BIA+ accounts for cognate facilitation in visual bilingual processing in terms of an orthographic-semantic priming effect, as overlapping shared orthographic and semantic representations from two (or more) languages become active when only one reading of the cognate pair is presented. Orthographic information activates shared semantic information of a cognate pair, which feeds back activation to both orthographic representations to strengthen them. Higher activation yields higher RTs to cognates as opposed to non-cognates. Even when orthographic overlap between cognate equivalents is incomplete, BIA+ predicts higher activation for such stimuli as the activation of both their partial differing representations increases overall activation. BIA+ also predicts the influence of L2 proficiency in terms of slower general activation for the less proficient L2, while cognate effect is expected to be higher in the less used language in comparison with L2 non-cognates, due to the similarity an L2 cognate shares with its quickly activated L1 equivalent. The mentioned orthographic depth effect could also be portrayed in the BIA+ model, as it can be predicted that the sublexical level may allow for faster visual processing of an orthographically shallow language (such as Croatian) as opposed to the orthographically deep language (such as English).

Cognate Types

Depending on the extent of the overlap in form, cognates can be divided into identical cognates (absolute overlap of form and meaning) (e.g. Croatian *problem* and English *problem*) and non-identical or semi-cognates (de Bot et al., 1995), which do not portray absolute form overlap but still share form and meaning in two languages (e.g. Cro. *spiker* and Eng. *speaker*). Achieving complete phonetic and phonological overlap between cognates may be difficult due to the language-specific phonology of each language (Dijkstra et al., 1999), especially if they are typologically fairly different, as is the case of Croatian (a Slavic language) and English (Germanic). If phonological differences between Croatian and English are considered, it is rather safe to state that phonologically (auditorily) absolutely identical cognates are rare, if not inexistent. However, speakers of the mentioned languages can still feel that

cognates are 'somehow the same thing' and the feeling of phonological similarity in cognate terms does not always correspond to the detailed phonetic description of the two cognate interpretations which may make them seem less similar than the speakers feel they are.

Cognate facilitation effect in LDT has been found for semi-cognates, exemplifying their processing advantage, especially for the speakers' L2, but it differed in intensity and stimulus list composition. Dijkstra et al. (1999) found an inhibitory effect of phonological (P) similarity and a facilitating effect of orthographic (O) and semantic (S) similarity. SOP overlap yielded facilitation which was accounted for due to the influence of the S and O similarity, which overrode the inhibitory P similarity. SP condition (in visual LDT) yielded no effect. Dijkstra et al. (2010) assessed four models of language processing to investigate how well these models accommodate the cognate effects they obtained in their research. They studied the processing of orthographically identical cognates and semi-cognates and found that RTs in an English (L2) LDT performed by Dutch-English speakers decreased with increasing similarity between cognate pairs. The phonological similarity condition increased the facilitation effect, but only in the case of identical cognates. No such effect was found for semi-cognates. The authors found the BIA+ model to be the most adequate model for the description of the obtained cognate effects, accounting for the LDT facilitation effects, which vary in strength depending on the degree of form overlap between the cognate pairs.

In a visual LDT with Spanish-Catalan bilinguals, Comesaña et al. (2015) found a facilitation effect for identical cognates but not for semi-cognates. The authors went on to explore how phonological similarities affected processing and found that participants responded more slowly to the stimuli with greater phonological overlap than when such overlap was smaller. In their second experiment, i.e. an LDT in which they used only semi-cognates as the cognate stimuli, they found cognate inhibition effects which were increased when there was greater mismatch between O and P overlap in the stimuli (semi-cognates with smaller O overlap and greater P overlap). They concluded that the stimulus list composition affected the processing, since they found faster RTs to cognate than to non-cognate words in the presence of identical cognates (facilitation effect), and slower RTs between the two stimulus types when identical cognates were absent (inhibition effect), also affecting the effect of phonological overlap. They interpreted this as the top-down effect of the language context on the relative activation of words from the two languages. They suggested that both similar but not identical orthographic and phonological representations of a semi-cognate get activated and cause lateral inhibition, while the two readings are modulated in a top-down manner by language context.

More cognate inhibition effects were found in studies in which activation of the phonological forms belonging to the cognate equivalents from the two languages was assumed. For example, Schwartz et al. (2007)

obtained longer naming latencies for cognates than non-cognates in a word-naming task carried out by English L1 speakers proficient in Spanish L2. The two languages differed in orthographic depth, as Spanish is an orthographically shallow language, although the authors made no note of that. They noticed that, when orthographically highly similar words mapped onto two more distinct phonological realizations, RTs were slowed. On the other hand, when the orthographic realizations were more distinct between the semi-cognate equivalents, no significant differences in the latencies associated with differences in phonological realizations were found. The authors interpreted this in terms of feed-forward activation (predicted in BIA+), where highly similar orthographic input crosslinguistically activates two different phonological forms and the increase in lateral competition causes a delay in naming.

Previous studies focused more on the influence of cognate stimuli in L2 processing and somewhat less in L1 processing. In the studies that did consider both languages of the participants, higher cognate facilitation effects were found in the non-dominant rather than the dominant language (see, for example, the picture-naming task performed by Costa et al., 2000) and with participants of lower L2 proficiency (Roselli et al., 2014). In other words, cognate form similarity seems to be more beneficial for L2 than for L1 processing.

In short, cognate effects vary depending (at least) on the S, O and P overlap between the cognate equivalents in different languages, stimulus frequency, stimulus list composition, participant L2 proficiency and task requirements. Various studies reported on inhibitory influence of phonological similarity (in various combinations with orthographic overlap between the cognate pairs) in cognate processing. This effect was interpreted in terms of lateral inhibition between the two realizations of the cognate equivalents in certain contexts.

Old and New Ways in Psycholinguistic Research

Recent changes in psycholinguistics point in a direction that is in line with the dynamic consideration of language processing, the Dynamic Systems Theory (DST; de Bot, 2012; de Bot; Penris & Verspoor; Török & Jessner, all this volume). The core idea of DST is that languages are complex systems which interact with their environment and change over time (e.g. de Bot, 2010). The present contribution takes a tiny step in this dynamic direction by looking at change (in cognate processing) over time, i.e. eight months of participants' intense study of EFL in combination with heavy exposure to the English language by means of media, brought about by the accumulation of experience in terms of language use and language proficiency. The idea of change over time influenced by the environment

conforms to the main pillars of DST; note, however, that this contribution takes a hybrid viewpoint in terms of recognizing the DST propositions as milestones in future studies of language processing while still measuring RTs in participants' responses to isolated stimuli at different points in time and trying to capture a change in the processing. DST proponents do not support the study of isolated elements, as any element in language is in interaction with other elements in its environment. Measuring RTs at different points in time to measure change is not supported either, since it has been shown that the stability of representations is a fleeting phenomenon (de Bot & Lowie, 2010). However, proponents of this view do admit that definite DST solutions for language processing research have not been proposed yet (Spivey, 2007, quoted in de Bot, 2010) which is why this contribution makes use of traditional RT measurement methodology with the above-described twists in terms of comparing visual and auditory processing in a longitudinal study design. Despite the elusive nature of the stability of representations, it seems logical that, if we find a constant effect in a longitudinal study that is repeated over a sequence of measurements, this result may be interpreted as a static moment situation capturing change over time. In this respect Duyck et al. (2007) have found facilitating cognate effects not only for isolated words but also for cognates used at the ends of sentences. It seems that cognate effects are rather consistent, and studying them in isolation is not overly obstructive for their interpretation despite their susceptibility to high-constraint contextual influences (van Hell, 2005). While awaiting new methods which would be in line with the propositions of DST, this chapter offers a modest contribution to the discussion of change in bilingual language processing based on the results of isolated cognate and non-cognate processing.

Cognate Effects and Speaker Second Language Proficiency

Cognate facilitation effects have often been found with participants who were described variously by the authors as bilinguals, non-balanced bilinguals, dominant bilinguals, etc., which makes comparison of results difficult between studies. This is why in this study the participants' L2 knowledge was tested and defined in terms of the levels provided by the Common European Reference for Languages (CEFR; Council of Europe, 2001). It is suggested that a shift in the processing could take place with a change in proficiency where the participants would react differently to cognate stimuli at different stages of their English language proficiency. De Bot et al. (1995) studied semantic and repetition priming effects on cognate processing in visual and auditory LDTs and found differences between the speaker groups they described as near-native and intermediate. The authors interpreted this

observation as evidence of a 'radical change in lexical organization' (de Bot et al., 1995: 7) in their participants. While cognate facilitation effects were found in the visual LDT in the intermediate group, the near-natives did not seem to be affected. In the auditory LDT they found a between-language repetition effect for cognates (and translation equivalents) and suggested that there was a strong and persisting effect of meaning activation in the auditory modality. The authors urged future studies to involve both visual and auditory tasks into their design.

To check the influence of a potential effect of change of L2 proficiency on cognate processing, a longitudinal study was envisaged predicting measurement at two different critical time points in the speakers' L2 development (Time 1 and Time 2). Time 1 was determined in such a way as to capture the moment before participants had intensive contact with the English language, i.e. when they had been merely exposed to limited classroom input in high school and before they were immersed in their English language major studies pursued at university level. The second data collection time (Time 2) took place after eight months of intensive work on the participants' English language skills at university. Should the result patterns change between these two measurement points, this change could be interpreted as the result of increased L2 proficiency, as described above. Due to the well-known instability of lexical representations in the bilingual lexicon, key to the effects studied in this paper is the investigation of the result patterns, rather than the comparison of RTs between the first and second measurement.

Research

In this study we set out to investigate: (1) if the ease of L1 and L2 processing differs in cognate equivalents as a function of high frequency and/or high phonological overlap with complete (identical cognates) and partial (semi-cognates) orthographic overlap (comparison of L1 and L2 processing); (2) how the absence of orthographic overlap (where only the inhibiting influence of phonological similarity is expected) influences RTs in auditory LDT (with reference to the differences in the orthographic depth between the studied languages); and (3) if cognate result patterns change with an increase in the participants' L2 proficiency.

If orthographic semi-cognates activate two similar phonological representations causing lateral inhibition, this may be reflected in slower RTs to such stimuli compared to non-cognates. Alternatively, the strength of the orthographic similarity of semi-cognates might override the inhibiting effect of similar phonology, yielding no cognate effect. Literature predicts facilitating effects of identical orthography on the processing of identical cognates (despite the high phonological overlap) in relation to L2 (English) non-cognates.

In order to avoid conclusions based only on the basis of visual processing where Croatian and English languages differ in orthographic depth, a fact which is expected to speed up the processing of orthographically shallow Croatian, auditory LDT, free of the orthographic depth effects, was also implemented. The auditory LDT was expected to yield results based only on semantic and potentially inhibiting phonological similarities, while there was no presentation of orthographic form that could facilitate the processing. If phonological similarity has an inhibitory effect on the processing, it should be clearly observed in the auditory LDT.

Finally, change in L2 cognate processing between Time 1 and Time 2 is predicted, as the similarity of form is expected to facilitate visual processing at Time 1, before the participants get immersed into their English major studies, while this effect may be lost at Time 2 with the increase in the participants' L2 proficiency resulting from eight months of intensive work on their English language skills at university level.

Methodology

Participants

Thirty Croatian university students of English who were studying to become elementary school teachers of English were recruited (28 female and two male; mean age: 19.8). The participants had all started learning English before the age of 11 and had not spent more than four weeks in an English-speaking country. When the first testing (Time 1) was carried out, the participants had just enrolled in higher (tertiary) education (study of English major at university level); they were only in their first week of classes and at a point when they were not yet as exposed to the English language as they would be in the next five years of their higher education. They were tested with the Oxford Placement Test (Allan, 2004) in the first week of classes and were found to be at least at level B2 of English language proficiency. By the second testing (Time 2) they had had eight months of intensive English language studies, in which they attended courses that were taught through English. They had eight hours of formal instruction a week in line with the following curriculum: 1st semester – Communicative Grammar in English 1 (2 ECTS), English Phonetics and Phonology (3 ECTS), Developing Skills in English 1 (2 ECTS); 2nd semester – Communicative Grammar in English 2 (4 ECTS), Developing Skills in English 2 (3 ECTS), Creative Teaching Activities in English 1 (2 ECTS). The numbers of ECTS points are provided to allow insight into the required amount of work invested into the courses. It is also important to mention that Croats are very much exposed to the English language in their environment, through media (TV programmes which are not dubbed, music in English, the internet, etc.). In order to find

out how much the participants used their English language for academic purposes as well as in their free time, they were asked to estimate to what extent they felt their English and Croatian (and any other language) were activated per week and to express this ratio in percentages; e.g. 50% of Croatian-language activation and 50% of English-language activation would mean that the participants felt that both of their languages were equally activated. The average values of their answers amounted to 32.3% of English-language activation time and 66.5% of Croatian-language activation time (1.2% pertained to the activation of some other language), showing that the students felt that English was activated one-third of their time as opposed to Croatian, which they felt was activated two-thirds of their time. They reported having normal or corrected-to-normal vision and hearing.

In addition to the 30 participants who carried out the LDTs, 160 students studying English as a major at the same institution were recruited to carry out a preparatory word familiarity test (see below).

Stimulus

The Croatian and English stimulus consisted of nouns comprised of 2–3 syllables and 5–8 letters, presented in their nominative singular forms, as nominative is known to be the most easily processed case in Croatian (Erdeljac & Horga, 2003). The frequency of each word was 50 or more tokens in a million (Leech et al., 2001; Moguš et al., 1999). However, frequency lists were only used for the initial choosing of the stimuli, and a word familiarity test (described in the Word familiarity test section below) was used as a more reliable control. Approximately half of the nouns in each stimulus group were concrete nouns, while the other half were abstract nouns. Orthographic and phonological neighbourhoods were controlled as well. None of the stimuli had stress on the final syllable because such a stress pattern is not common in the Croatian language. It was vital to pay special attention to this because of the known influence of phonetic processing in visual tasks.

Five visual word stimulus categories were formed, each comprising 20 stimulus items: Croatian non-cognates (e.g. Cro. *prozor* (Eng. *window*)), Croatian-based semi-cognates (e.g. Cro. *telefon* (Eng. *telephone*)), orthographically identical cognates (e.g. *student*), English-based semi-cognates (e.g. Eng. *project* (Cro. *projekt*)), and English non-cognates (e.g. Eng. *husband* (Cro. *muž*)). The Croatian- and English-based semi-cognate groups were formed out of different vocabulary items, so as to avoid priming in the LDT. Thus if, for instance, the English semi-cognate *paper* was used in the English semi-cognate stimulus group, its Croatian semi-cognate equivalent *papir* was not used in the Croatian semi-cognate stimulus and vice-versa. The orthographically identical cognates were visually the same in both languages (e.g. Croatian and English word *student*). No item bearing language-specific cues (in Engl. double letters, and *q, x, y, w*, which are non-existent in the Croatian alphabet; in Cro. ć, č, d,

š, ž, and the bigrams *dž, lj, nj*, which are not found in English) were allowed in the stimuli so as to avoid directing the processing in any way. American English spelling was used (*program* as opposed to *programme*), as it allowed for greater crosslinguistic orthographic similarity between languages.

An equal number of pseudowords per category was created out of stimulus words by substituting the initial, middle or final phoneme (the number of substitutions of each phoneme position was controlled) with a phoneme that differed from the original one in four distinctive features. This difference allowed for a design of pseudowords which fit the phonotactic and orthographic requirements of both languages and was perceptively (auditorily) large enough for experimental research. The total number of stimuli used in the visual modality was 200 (100 words and 100 pseudowords).

No identical cognate stimulus group could be formed for the auditory stimulus, so it was comprised of four stimulus groups: Croatian non-cognates, Croatian-based semi-cognates, English-based semi-cognates and English non-cognates. Matching pseudowords were created. Again, the stimuli comprised 20 vocabulary items per group.

American English pronunciation was used, since this is the variety the participants were most familiar with mostly due to media influence and self-reported preferences. The speaker who recorded the stimulus was a balanced bilingual raised bilingually by an American mother and a Croatian father. The speaker was an English-teaching employee at an institution of higher education, had a PhD in English, was trained in English phonetics and had taught English phonetics at university level. To ensure equal phonetic correctness in the pronunciation of the Croatian stimuli, a university teacher of Croatian with a PhD in Croatian helped model the prosody of the Croatian language where necessary. The recordings were examined and repeated until the desired correctness of the material was achieved. The acoustic length of the auditory stimuli was controlled. The total number of stimuli used in the auditory modality was 160 (80 words and 80 pseudowords).

Word familiarity test

An internet application connected to MySQL database (Kovačević, 2014) was designed, into which a word familiarity questionnaire was fed. It consisted of 100 vocabulary items, which were chosen according to the previously described requirements of stimulus selection; 50 words were added to serve as fillers. The participants ($N = 160$), of similar language profile to the participants of the main study, carried out this preparatory task. They were required to read each word and indicate on a 4-degree Likert-type scale how familiar the stimulus was to them ('click on one of the four radio buttons offered next to each word: very familiar (1), familiar (2), less familiar (3), completely unfamiliar (4)'). The mean answer for each word was calculated. The smaller the mean, the more familiar the word was. For example, if a word was very

familiar to all participants, its mean was 1. If it was completely unfamiliar to all of them, its mean was 4. Only words with a mean of 1.5 or lower were used. The familiarity test confirmed that the 100 words chosen for the stimulus were truly found to be the most familiar words by the participants.

Form overlap

The analysis of the orthographic and phonological overlap in the semi-cognate pairs was carried out to make sure that orthographic and phonological similarity between cognate pairs used in the stimulus was high (high phonological similarity in all stimulus groups was the precondition in this study so that the identical cognate or high orthographic similarity semi-cognate effects could be studied). Calculation of normalized Levenshtein distance (nLD) was performed. The Levenshtein algorithm is a measure of formal (orthographic or phonological) distance between words of different languages. It computes a minimum number of edit operations which are employed to map a string of letters or phonemes onto another one. There are three possible operations: substitution, deletion and insertion of an element of the string. Deletions and insertions are referred to as indels. Each string-changing operation has its cost: each indel is assigned a 0.5 penalty point; substitution, as a more complex operation involving the deletion of one element and insertion of another, is assigned 1 penalty point. To normalize the distance adjusting it for word length, the total cost is divided by the number of alignments. While calculating nLD for the orthographic representation of words is a rather straightforward process, certain issues involving the choice of transcription and elements involved in edit operations need to be resolved. Sometimes usage of broad transcription is not detailed enough to provide precise calculations of nLD between cognate pairs. Gooskens *et al.* (2015: 279) support the usage of narrow transcription in calculating nLD between languages, yet warn against using it in cases when high intelligibility is achieved as it may yield results which do not fit speaker intuition. To investigate what might count as related phonology between Croatian and English, Josipović Smojver's (2010) description of the Croatian-based lingua franca she refers to as 'Croglish' was used. The author presents segmental (and suprasegmental) differences between Croatian and English involved in L1 → L2 transfer, which often results in incorrect yet highly intelligible pronunciation of English by Croatian speakers. This is probably due to the fact that they create acoustic images of the L2 phonemes by mapping them onto the closest phoneme in their L1. Some examples involve producing /d/ and /t/ dentally (the Croatian way) as opposed to their alveolar English pronunciation, realizing the English phoneme /v/ as a labiodental approximant (to match the corresponding Croatian realization), etc. This corresponds to a Croatian speaker's intuition of the elements in the pronunciation of Croatian and English which are felt as somehow being 'the same thing', to use Carroll's

Table 12.1 Average normalized Levenshtein distance per cognate stimulus group

	Croatian semi-cognates	Identical cognates	English semi-cognates
Orthographic nLD of visually presented stimulus	0.24	0	0.28
Phonological nLD of visually presented stimulus	0.38	0.25	0.41
Phonological nLD of auditorily presented stimulus	0.31	x	0.28

(1992) words. For this study, this feeling was checked and confirmed with regard to native speaker intuition. Such elements were not penalized when Levenshtein distance was calculated, as it was believed that such a calculation would yield too large nLDs, which in reality would not affect the perception of the participants accordingly. The results of the analysis show low nLD measures and high orthographic and phonological similarity between cognate pairs. Table 12.1 shows the average nLD per cognate stimulus group.

Procedure

The participants carried out generalized visual and auditory LDTs. They were each seated in front of a computer screen on which the task was run. First the instructions were presented in both English and Croatian to ensure the bilingual set (Grosjean, 1998). After the presentation of instructions, the participants carried out a practice session and proceeded to the experimental session. They were instructed to look at the visually presented strings of letters (words and pseudowords in the visual modality) preceded by the fixation cross which was displayed in the middle of the screen for 1000 msec, or listen to the auditorily presented strings of sounds (words and pseudowords in the auditory modality) and decide if the stimulus was a word in either Croatian or English or not. Approximately half of the participants were instructed to press key 1 when they identified the presented stimuli as a word in one of their languages, and key 2 if they identified a pseudoword. For the other half of the participants the order of the keys was reversed. The stimuli were presented in both languages in the same experimental block in random order (Research Randomizer Tool was used). One part of the participants first carried out a visual LDT after which they were required to take a break and proceed to the auditory LDT; for the other part of the participants this order was reversed. The experiment was presented and RTs measured by means of the E-prime psychology software tool. The participants carried out the task at two different time points: Time 1, in October 2014, when they had just enrolled in their higher education English studies; and Time 2, in May 2015, after eight months of intensive work on their English language skills.

Results

Results of the visual lexical decision task

RTs of the correct responses to the stimulus were used in the analysis. RTs that were more than 2.5 standard deviations below or above the mean obtained for each stimulus type group were excluded from the analysis. A Kolmogorov–Smirnov test was applied on the trimmed data to confirm the normality of the distribution of the remaining results per stimulus type category. The distribution was found to be normal.

An analysis of variance (5×2 design) was carried out. A main effect of stimulus type was found ($F(4,116) = 61.048$; $p < 0.001$). A main effect of measurement time was found as well ($F(1,29) = 23.066$; $p < 0.001$). The RTs measured at Time 2 were significantly faster than the RTs measured at Time 1. Interaction between stimulus type and time point was found ($F(4,116) = 5.495$; $p < 0.001$). While at Time 1 the RTs to English semi-cognates were faster than the RTs to English non-cognates, at Time 2 this pattern was reversed, as the RTs to English non-cognates were faster than the RTs to English semi-cognates. The change in the result pattern is marked in Table 12.2.

As post hoc analysis, separate analyses of variance were carried out, one with the results obtained at Time 1 and the other with the results obtained at Time 2. The results of the first measurement have shown the effect of stimulus type ($F(4,116) = 45.335$; $p < 0.001$), and so have the results of the second measurement ($F(4,116)$ 27.956; $p < 0.001$). Post hoc tests showed no statistically significant differences between the RTs to Croatian non-cognates and Croatian semi-cognates, and the RTs to English non-cognates and English semi-cognates did not differ significantly either. The RTs to Croatian stimuli (non-cognates and semi-cognates) differed significantly from the RTs to English stimuli (non-cognates and semi-cognates). Curiously, the RTs to the orthographically identical cognates did not differ significantly from the RTs to Croatian semi-cognates but did differ from the RTs to Croatian non-cognates. The RTs to the identical cognates differed significantly from the RTs to the English stimuli (both semi-cognates and non-cognates). The results are presented in Figure 12.1 and Table 12.2.

Results of the auditory lexical decision task

The auditory LDT data were trimmed in the same way as the visual LDT data, and normality of distribution was confirmed. An analysis of variance (4×2 design) was carried out to find a main effect of time point ($F(1,29) = 11.184$; $p = 0.002$) and stimulus type ($F(3,87) = 64.929$; $p < 0.001$). The RTs measured at Time 2 were significantly faster than the RTs measured at Time 1. No interaction between stimulus type and time point was found ($F(3,87) = 920.076$; $p = 0.302$).
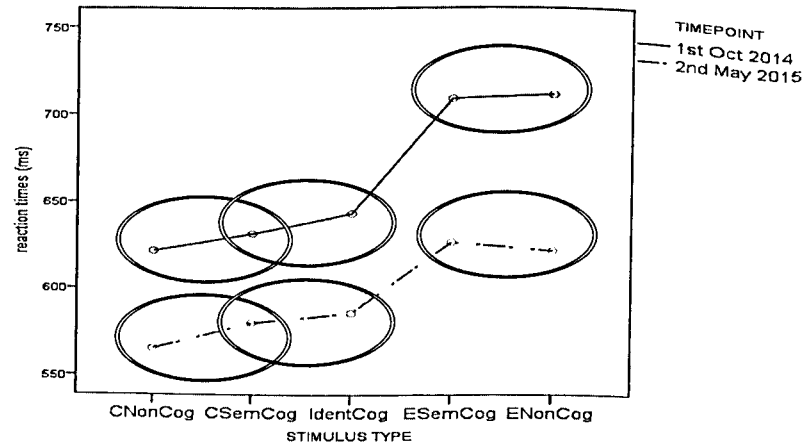
Figure 12.1 Reaction times to five categories of stimulus type: Visual word stimulus presented at Time 1 and Time 2
Notes: Circles mark values which do not differ significantly.
Legend (categories of stimulus type): CNonCog = Croatian non-cognates; CSemCog = Croatian semi-cognates; IdentCog = identical cognates; ESemCog = English semi-cognates; ENonCog = nglish non-cognates.

Table 12.2 Reaction times to five categories of stimulus type: Visual word stimulus presented at Time 1 and Time 2

Stimulus type	Time 1		Time 2	
	M	SD	M	SD
Croatian non-cognates	621.436	78.907	565.228	55.097
Croatian semi-cognates	631.541	82.683	579.888	59.153
Identical cognates	643.416	92.303	585.767	62.725
English semi-cognates	709.054	106.831	624.604	70.752
English non-cognates	711.389	117.444	621.556	63.537

As post hoc analysis, separate analyses of variance were carried out, one with the results obtained at Time 1 and the second with the results obtained at Time 2. The results of the first measurement showed an effect of stimulus type ($F(3,87) = 32.699; p < 0.001$), and so have the results of the second measurement ($F(3,87)$ 41.393; $p < 0.001$). Post hoc tests performed with the results obtained at Time 1 and Time 2 showed the same result pattern; the RTs to Croatian and English non-cognate stimuli did not differ significantly. The RTs to non-cognate stimuli were faster than the RTs to semi-cognate stimuli in both languages. The RTs to the English-based semi-cognate stimuli were significantly slower than the RTs to the Croatian-based semi-cognate stimuli. The results are presented in Figure 12.2 and Table 12.3.
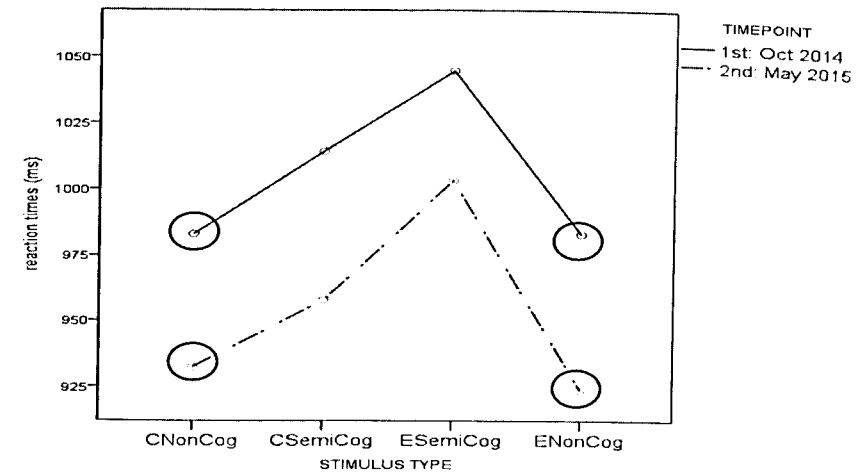
Figure 12.2 Reaction times to five categories of stimulus type: Auditory word stimulus presented at Time 1 and Time 2
Notes: Circles mark values which do not differ significantly.
Legend (categories of stimulus type): CNonCog = Croatian non-cognates; CSemCog = Croatian semi-cognates; IdentCog = identical cognates; ESemCog = English semi-cognates; ENonCog = English non-cognates.

Table 12.3 Reaction times to four categories of stimulus type: Auditory word stimulus presented at Time 1 and Time 2

Stimulus type	Time 1		Time 2	
	M	SD	M	SD
Croatian non-cognates	982.950	98.629	932.352	60.966
Croatian semi-cognates	1014.304	109.997	957.922	62.194
English non-cognates	1044.583	102.683	1003.131	70.98
English semi-cognates	982.944	98.970	923.759	68.39

Discussion

Visual versus auditory LDT results: Non-cognates

The analysis showed that the result patterns of the visual LDT may be misleading if interpreted without the consideration of the result patterns of the auditory LDT. Looking only at the non-cognate stimuli, one sees that in the visual modality, the RTs to Croatian non-cognates are much faster than the RTs to the English non-cognates, and this difference is statistically significant. Without the consideration of the results obtained in the auditory

modality, the researcher might be led to believe that this is simply due to L2 proficiency and the fact that Croatian is the participants' L1 and, naturally, their more proficient language. However, when the results of the auditory modality are taken into consideration, one sees no such pattern, as the RTs to non-cognate stimuli in Croatian and English do not differ. Despite the fact that English was the participants' L2 and their less proficient language, similar auditory Croatian-English result patterns are not surprising as the reader will recall that only very frequent and most familiar words (familiarity test) were used as stimuli. The ease of processing of such words caused by their familiarity is likely to surpass the limitations of language proficiency, and this was the idea behind using such stimuli. The conclusion which offers itself is the possibility of the languages' varied orthographic depths influencing the results obtained in the visual modality. The essential idea of the dual-route hypothesis of reading orthographically different languages in terms of orthographic depth is that 'it is easier to derive phonological codes in some orthographies than in others' and this variable may 'influence the extent to which phonological codes activate meaning' (Seidenberg, 1992: 87).

Seidenberg goes on to propose that, all other conditions in stimulus lists being equal (a precondition that is difficult to achieve), an orthographically shallow language may be easier to process than an orthographically deep language. This suggests that, when studying visual processing of two languages which are both written in the same (Latin) script yet possess such great difference in the orthographic depth as Croatian and English do, the result patterns of the visual data should be checked against the result patterns of the auditory data as the latter will presumably not be susceptible to the processes conditioned by the differences in the orthographic depth. In modelling, one may assume that faster processing of Croatian visual stimuli is possible, which may be to a great extent realized already on the basis of the prelexical phonological processes, put in terms of the BIA+ model, while the processing of the English stimuli may rely more on the lexical processes. This does not mean that both languages do not utilize both types of processes, only that each of them is more reliant on one rather than the other type. Since there is no difference in the RTs between the Croatian and English non-cognate stimuli in the auditory LDT, while this difference is significant in the visual LDT, this seems to be a reasonable explanation which one should bear in mind while interpreting the other results of this study and similar studies carried out with languages that greatly differ in orthographic depth.

Cognate processing

Considering the result patterns of the visually presented stimuli, one obvious difference between the RTs to the identical cognates and non-cognates in the two languages becomes evident: the RTs to identical cognates are, as expected, significantly faster than the RTs to English non-cognates

(cognate facilitation effect) and, unexpectedly, significantly slower than the RTs to Croatian non-cognates (cognate inhibition effect). A facilitating identical cognate effect in the processing of L2 has been repeatedly found in LDT. This effect is interpreted in terms of semantic co-activation of the more proficient (and thus quickly activated) L1 with that of the L2 of the cognate pair mapping onto the identical orthographic form, the effect of which is facilitating, and resulting in higher overall activation yielding faster RTs.

However, the interpretation of the cognate inhibition effect between visually presented Croatian non-cognates and identical cognates is problematic and is not in line with the above proposed interpretation as the facilitation of semantic co-activation between cognate pairs should work both ways. Mulder et al. (2015) report that when their participants were required to perform language decisions upon encountering an identical cognate, they were faster in choosing English as the language of their choice and slower to choose their L1 Dutch, showing a bias towards L2 English. Such an L2 bias intertwined with the orthographic depth effect could have caused facilitation of identical cognates in comparison with English non-cognates as opposed to the inhibition of identical cognate processing in comparison with Croatian non-cognates. Such a result can be interpreted in relation with Meuter and Allport's (1999) proposal according to which inhibition of L1, needed for the processing of L1, may persist upon subsequent processing of L1, causing its inhibition.

It is curious to note the interaction of RTs between stimulus type and time point in visual LDT, which potentially captures a change in the processing of English non-cognates and semi-cognates. While at Time 1 the RTs to English semi-cognates were faster than the RTs to English non-cognates, at Time 2 this pattern was reversed: the RTs to English non-cognates were faster than the RTs to English semi-cognates. One may suggest that a change in the processing has been witnessed which was brought upon by the increase in L2 proficiency. Recall that BIA+ predicts higher cognate facilitation effect in a less proficient L2 in relation to L2 non-cognate processing. It is possible that co-activation of different phonological forms linked onto similar orthography caused (more) inhibition in more proficient L2 speakers.

A semi-cognate inhibition effect was found in the RTs to the LDT in the auditory modality, and no semi-cognate effect was found in the visual modality. Phonological similarity is known to have an inhibitory effect on the processing and all the cognate stimuli in this study were purposefully chosen to have high phonological similarity. Without the facilitating effect of the orthographic similarity, the auditorily presented semi-cognate processing was slowed down, once again confirming the inhibiting effect of phonological similarity. In the visual LDT the facilitating effect of the orthographic similarity and the inhibiting effect of the phonological similarity are believed to have cancelled each other out, resulting in no cognate effect.

Measurement point effects: Some words of caution

Although a main effect of time was found and the obvious interpretation of faster RTs at Time 2 in terms of increased L2 proficiency comes to mind, this result should be treated with caution, considering that this effect, in post hoc testing, was found to be true both for English ($F(1,29) = 24.188$; $p < 0.001$) and Croatian ($F(1,29) = 18.816$; $p < 0.001$) in the visual modality, and for English ($F(1,29) = 10.409$; $p = 0.003$) and Croatian ($F(1,29) = 10.841$; $p = 0.003$) in the auditory modality. Practice effect with an eight-month interval between the measurements is hardly likely, and the only logical explanation seems to be the one where students felt more comfortable carrying out the experiment after one year of working at the institution where the testing was performed (Time 2) than in their first week of classes (Time 1). The reader will recall that this early measurement was performed so as to capture the participants' processing at the point when they had just started their English studies at university level. However, the main effect of measurement point has to be interpreted with caution. This is why further research should be conducted to follow the participants' L2 development throughout their university education. However, this does not prevent one from studying result patterns, which are the main interest of this study. Authors of similar studies in the future might consider carrying out the first testing of the participants after they have settled into the new environment. This could help prevent the potential feeling of unease the participants may feel when being tested in a new environment.

Attempt at explaining the inhibitory effect of cognate phonological similarity

Why does co-activation of two phonological interpretations have an inhibitory effect on bilingual cognate processing? The answer to this question may be provided by the exemplar theory. The exemplar theory was originally introduced in psychology as a model of perception and categorization (Pierrehumbert, 2001) to be subsequently used as a model of vowel categorization, as it allowed the interpretation of human capability to categorize various phonetic elements. According to the exemplar theory, each category is represented in the memory by a cloud of remembered tokens of that category (exemplars), where highly similar instances are closer, and dissimilar ones are far apart. Stored exemplars present the variation that occurs in a category. This is how, for example, the interpretation of allophones as different variations that gravitate towards the phoneme they belong to can be explained. The same tokens may be categorized into more than one category at the same time. The tokens are organized in the imaginary space on the basis of similarity with other tokens as well as their frequency and recency of usage. In the attempt to interpret cognate effects in terms of the exemplar

theory, the following interpretation is proposed: upon perceiving an auditorily presented semi-cognate, its interpretation may tend to gravitate towards both L1 and L2 phonological representations due to their shared features. As an allophone gravitates towards its phoneme category, the cognate may gravitate towards both cognate interpretations based on the similarity of form; the cognate equivalent of one language may behave, so to speak, as an 'allolexeme' of the other language's cognate (Cergol Kovačević, 2012). The gravitation force depends on L2 proficiency, higher proficiency yielding higher gravitation force towards both cognate equivalents, while lesser L2 proficiency causes the L2 cognate representation to gravitate towards the L1 interpretation. Notice that the gravitation-towards-two-options should be enhanced with higher phonological overlap between cognate equivalents, and reduced when the overlap is smaller, as research has confirmed. In other words, phonological similarity induces processing inhibition in more proficient L2 speakers as the interpretation gravitates towards two possible phonological interpretations.

Conclusions and Suggestions for Further Research

The study described in this chapter set out to investigate the influence of orthographic overlap in Croatian-English cognate processing by means of a comparison of orthographically identical cognates and semi-cognates in terms of the ease of their processing. The study was longitudinal in nature as one of the aims was to investigate the influence of change (in terms of improvement in the English language knowledge) in time on the processing. Finally, as Croatian is an orthographically shallow language as opposed to English which is an orthographically deep language, the results of the visual and auditory processing tasks were combined to exclude the potential influence of the greater ease of the visual processing of the Croatian language on the conclusions. The influence of high phonological overlap between cognate equivalents in the (semi-)cognate stimuli was found to be inhibiting in the auditory LDT, while this effect was inexistent in the visual LDT, presumably due to the facilitating influence of the orthographic similarity between Croatian and English cognate equivalents. The facilitating effect of orthographic similarity was confirmed in the case of identical cognates, but only in comparison with L2 non-cognate processing. The inhibition of identical cognates in comparison with L1 non-cognates remains difficult to interpret, but is believed to be influenced by the L2 bias in bilingual processing and differences in the orthographic depth between the studied languages. The inhibition found in the auditory modality could be tentatively explained by means of an exemplar theory postulate – the effect of a time-costly gravitation of an auditorily presented semi-cognate equivalent towards the cognate equivalents of both languages. Such gravitation occurs due to the features that both cognate equivalents share with

the input. Change in L2 proficiency is thought to influence change in cognate processing in the visual modality as English-based semi-cognates were responded to faster at Time 1 than at Time 2 when there was an increase in the participants' L2 proficiency. Higher L2 proficiency seems to render the speakers more susceptible to the inhibiting cognate effects.

Further studies could aim at capturing occasions of remodelling in the L1/L2 processing resulting from the increased – or, in the case of language attrition, decreased – proficiency in the L2, which would be in line with the future research envisioned by the pillars of the New School in psycholinguistics. Checking visual data against auditory data when comparing bilingual processing of languages with different, but also related, orthographic depths should become a regular feature in this discipline, since such a hybrid research methodology represents an attempt to bridge the gap between the old and new ways in psycholinguistic research (de Bot, 2010, 2015). It is hybrid in terms of merging the old ways in psycholinguistic research, where experimental studies focusing on isolated words were used to capture the specificities of language-processing mechanisms, with the new ways of the dynamic considerations of language processing (as outlined in Penris & Verspoor, this volume), where longitudinal measurement of language processing at different points in time may hopefully capture the change in the processing conditioned by a change in the dynamic language systems.

Acknowledgements

References

Allan, D. (2004) *Oxford Placement Test 2*. Oxford: Oxford University Press.
Blumenfeld, H. and Marian, V. (2005) Covert bilingual language activation through cognate processing: An eye-tracking study. *Proceedings of the Twenty-Seventh Annual Meeting of the Cognitive Science Society* (pp. 286–291). Mahwah, NJ: Lawrence Erlbaum.
Caramazza, A. and Brones, I. (1979) Lexical access in bilinguals. *Bulletin of the Psychonomic Society* 13 (4), 212–214.
Carroll, S.E. (1992) On cognates. *Second Language Research* 8 (2), 93–119.
Cergol Kovačević, K. (2012) Slušno procesiranje u govornika globalnog engleskog. *Govor* 1, 45–64.
Comesaña, M., Soares, A.P., Ferré, P., Romero, J., Guasch, M. and García-Chico, T. (2015) Facilitative effect of cognate words vanishes when reducing the orthographic overlap: The role of stimuli list composition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41 (3), 614–635.
Costa, A., Caramazza, A. and Sebastian-Galles, N. (2000) The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26 (5), 1283–1296.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
de Bot, K. (2010) Cognitive processing in bilinguals: From static to dynamic models. In R.B. Kaplan (ed.) *The Oxford Handbook of Applied Linguistics* (2nd edn) (pp. 335–348). Oxford: Oxford University Press.
de Bot, K. (2012) Rethinking multilingual processing: From a static to a dynamic approach. In S. Flynn, J. Rothman and J. Cabreli Amara (eds) *Third Language Acquisition in Adulthood* (pp. 79–93). Amsterdam and Philadelphia, PA: John Benjamins.
de Bot, K. (2015) From old school to new school research in psycholinguistics. Plenary talk presented at the 29th International Conference of the Croatian Applied Linguistics Society, 'Applied Linguistic Research and Methodology'.
de Bot, K. and Lowie, W. (2010) On the stability of representations in the multilingual lexicon. In M. Pütz and L. Sicola (eds) *Cognitive Processing in Second Language Acquisition: Inside the Learner's Mind* (pp. 117–133). Amsterdam and Philadelphia, PA: John Benjamins.
de Bot, K., Cox, A., Ralston, S., Schaufeli, A. and Weltens, B. (1995) Lexical processing in bilinguals. *Second Language Research* 11 (1), 1–19.
de Groot, A.M.B. (2011) *Language and Cognition in Bilinguals and Multilinguals: An Introduction*. New York and Hove: Taylor & Francis.
Dijkstra, T. (2005) Bilingual visual word recognition and lexical access. In J.F. Kroll and A.M.B. de Groot (eds) *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 179–201). Oxford: Oxford University Press.
Dijkstra, T. and van Heuven, W.J.B. (2002) The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition* 5 (3), 175–197.
Dijkstra, T., Grainger, J. and van Heuven, W.J.B. (1999) Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language* 41, 496–518.
Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M. and Baayen, H. (2010) How cross language similarity and task demands affect cognate recognition. *Journal of Memory and Language* 62, 284–301.
Duyck, W., van Assche, E., Drieghe, D. and Hartsuiker, R.J. (2007) Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33 (4), 663–679.
Erdeljac, V. and Horga, D. (2003) Mentalno procesiranje morfoloških jedinica u govorenoj riječi. In D. Stolac, N. Ivanetić and B. Pritchard (eds) *Psiholingvistika i kognitivna znanost u hrvatskoj primijenjenoj lingvistici* (pp. 239–246). Zagreb and Rijeka: HDPL.
Gooskens, C., van Bezooijen, R. and van Heuven, V. (2015) Mutual intelligibility of Dutch German cognates by children: The devil is in the detail. *Linguistics* 53 (2), 255–283.
Grosjean, F. (1998) Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1, 131–149.
Josipović Smojver, V. (2010) Foreign accent and levels of analysis: Interference between English and Croatian. In E. Waniek-Klimczak (ed.) *Issues in Accents of English 2: Variability and Norm* (pp. 23–35). Newcastle upon Tyne: Cambridge Scholars.
Kovačević, M. (2014) Word Familiarity Test. Internet application designed for the purposes of the study.
Leech, G., Rayson, P. and Wilson, A. (2001) *Word Frequencies in Written and Spoken English (Based on the British National Corpus)*. London: Longman.
Lemhöfer, K. and Dijkstra, T. (2004) Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition* 32 (4), 533–550.

Lemhöfer, K., Dijkstra, T. and Michel, M.C. (2004) Three languages, one ECHO: Cognate effects in trilingual word recognition. *Language and Cognitive Processes* 19 (5), 585–612.

Meuter, R.F.I. and Allport, A. (1999) Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language* 40, 25–40.

Moguš, M., Bratanić, M. and Tadić, M. (1999) *Hrvatski čestotni rječnik*. Zagreb: Zavod za lingvistiku Filozofskog fakluteta u Zagrebu i Školska knjiga.

Mulder, K., Dijkstra, T. and Baayen, H. (2015) Cross-language activation of morphological relatives in cognates: The role of orthographic overlap and task-related processing. *Frontiers in Human Neuroscience* 9: 16. doi:10.3389/fnhum.2015.00016.

Pierrehumbert, J.B. (2001) Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee and P. Hopper (eds) *Frequency and the Emergence of Linguistic Structure* (pp. 137–158). Amsterdam and Philadelphia, PA: John Benjamins.

Roselli, M., Ardila, A., Jurado, M.B. and Salvatierra, J.L. (2014) Cognate facilitation effect in balanced and non-balanced Spanish-English bilinguals using the Boston Naming Test. *International Journal of Bilingualism* 18 (6), 649–662.

Schwartz, A.I., Kroll, J.F. and Diaz, M. (2007) Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes* 22 (1), 106–129.

Seidenberg, M. (1992) Beyond orthographic depth in reading: Equitable division of labor. In R. Frost and M. Katz (eds) *Orthography, Phonology, Morphology, and Meaning* (pp. 85–118). Amsterdam: Elsevier Science.

Sherkina, M. (2003) The cognate facilitation effect in bilingual speech processing. *Toronto Working Papers in Linguistics* 21, 135–151.

Sherkina-Lieber, M. (2004) The cognate facilitation effect in bilingual speech processing: The case of Russian-English bilingualism. *Cahiers linguistiques d'Ottawa* 32, 108–121.

Sunderman, G. and Schwartz, A. (2008) Using cognates to investigate cross-language competition in second language processing. *TESOL Quarterly* 42, 527–536.

Szubko-Sitarek, W. (2011) Cognate facilitation effects in trilingual word recognition. *Studies in Second Language* 1 (2), 189–208.

van Hell, J.G. (2005) The influence of sentence context constraint on cognate effects in lexical decision and translation. In J. Cohen, K.T. McAlister, K. Rolstad and J. MacSwan (eds) *Proceedings of the 4th International Symposium on Bilingualism* (pp. 2297–2309). Somerville, MA: Cascadilla Press.

Yule, G. (2006) *The Study of Language* (3rd edn). Cambridge: Cambridge University Press.

13 The Impact of Attitudes on Language Retention of Russian as a Foreign Language in Hungary: Some Lessons to be Learnt from Attrition Studies

Szilvia Bátyi

Introduction

In this chapter we report on a study on the loss and retention of Russian as a foreign language in Hungary, with a special focus on the attitudes towards language learning, Russian people, teacher(s) of Russian, and the Russian language of learners of Russian in school. Our participants are 39 Hungarian learners who started learning Russian as a compulsory subject in Hungary between 1958 and 1988, and who had very little contact with Russian in the period of non-use (the incubation period) thereafter. Starting from the assumption that positive attitudes towards Russians and learning Russian would lead to higher retention scores than negative attitudes, the results indeed indicate that attitudes can have an effect on vocabulary reten-tion (including tip-of-the-tongue states), which is a novel finding in the field of language attrition. The study is also unique in the sense that the language under investigation is not a 'neutral' foreign language, such as French or German in Western European education, but a language loaded with political and societal 'weight'. From a Western perspective, the Russian influence was perceived as an occupation, and learning Russian was somehow equated with the learning of the language of a political enemy. Interestingly, our data show that this was not necessarily the perspective of Hungarian learners in the 20th century. Quite a few of our participants reported that they used to

appreciate Russian because of the beauty of the language and the rich culture that it represented. Our findings thus show that stereotypical views of the learning of languages with negative political connotations (e.g. 'everybody hated the Russians and no one learned Russian for pleasure') may be wrong and need to be re-evaluated. Another lesson is that social-psychological factors can have an effect over a much longer period of time than the two to four years typically reported in attrition studies.

The chapter is organized as follows. First, some methodological aspects of language attrition research will be discussed, before the analysis of three types of data will be presented, viz. quantitative attrition data, qualitative data from interviews, and data including tip-of-the-tongue phenomena as part of the attrition data. The reader should bear in mind that the retrospective nature of this study will not allow us to zoom in on the exact degree of attrition due to a lack of information about the initial conditions (for a detailed discussion of this, see Verspoor, 2015), including the language proficiency of our participants at the time of the onset of attrition, i.e. the start of the incubation period. Rather, the goals of this research are to study: (1) the amount of lexical retention (as far as possible); (2) the influence of extra-linguistic factors (see definition below); and (3) the strategies employed in lexical retrieval.

Literature Review

First, second and foreign language attrition

One of the major challenges of researching language attrition is the question of how to measure it. Although systematic research on the topic has been conducted since 1982 (the year Lambert and Freed published their seminal book on *The Loss of Language Skills*), there are still numerous methodological questions that remain unanswered. Several issues need addressing in the future in order to gain a deeper insight into attrition phenomena, the first and most crucial one being concerned with the kind of attrition under investigation: first language (L1) attrition, second language (L2) attrition, or foreign language (FL) attrition. With respect to the differences between L1 and L2 attrition, Bardovi-Harlig and Stringer (2010) suggest that while there are many similarities in the study of the two phenomena, in the latter, additional variables have to be considered, in particular the degree of success in acquisition and crosslinguistic interference. A further distinction should be made between L2 and FL attrition. In this respect, Schmid and Mehotcheva (2012) have argued that there are substantial differences between languages that were acquired naturally and languages learned through instruction. Language exposure, input and use are considered to be the main differentiating factors, although the distinction itself is not without problems. To give

an example: is learning a language through watching a movie in L2 with L1 subtitles considered natural or instruction based?

Foreign language attrition is defined in this study as the non-pathological decrease of language skills in a language that had previously been learnt through instruction by an individual (a modified definition of Köpke & Schmid, 2004). A crucial aspect in attrition research is the choice of the point of reference: how do we compare the two language proficiency levels (attrited and non-attrited) – one of the biggest dilemmas of cross-sectional studies. Ideally we should have longitudinal data from the same individuals, but such studies are rare and often unpractical. A related question is concerned with the 'symptoms' in the individual's performance that we take as evidence of language attrition. In her comprehensive article on L1 attrition, Schmid (2013) notes that various types of cross-linguistic influence (CLI) (borrowing, restructuring, convergence and shift, according to Pavlenko, 2004) are most visible on the surface, which is why they are often taken as evidence of language attrition. It is obvious that CLI does not necessarily indicate the loss of underlying knowledge, but shows that entities provided by the other language system play a role; that is, the symptoms of attrition may be caused by the failure to inhibit the L2. Accordingly, the problem of the executive control may serve as an explanation for the poorer performance in the language that is thought to be in the phase of attrition rather than the loss of previously acquired knowledge. On the other hand, it could be argued that more CLI in itself is a sign of attrition: apparently the L2 knowledge structures have become weaker and more amenable to influences from other languages (Olshtain, 1989).

Language attrition can be studied focusing on the process or on the outcome (or both). How and what we measure depends on whether we are interested in the process or the product of attrition. Schmid (2008) connects the two by stating that:

> (...) there are two stages of linguistic knowledge: a pre-attrition stage A and an attrited stage B, and the apparent difference between A and B, caused by the process of attrition, is then the phenomenon of attrition. (Schmid, 2008: 10)

This generates at least two questions: is there a moment from which we can call a person an attriter? In a way, this is a trivial question when one takes a dynamic perspective on language development (de Bot, 2008, this volume; de Bot *et al.*, 2007; Penris & Verspoor, this volume; Török & Jessner, this volume). Language skills are constantly changing, and there is no end point or ultimate attainment. Reduced accessibility of language components (words, rules) is a normal and effective strategy of the cognitive system to use resources sparingly: why keep a language completely active when it is not used? The second question that arises is: is it possible to define a point of

reference to which the attrited state is compared, in order to determine the peak attainment? So far none of the attempts has been successful at giving a satisfying answer to these questions. Or in Schmid and Mehotcheva's (2012: 116) words: 'It is often very difficult (or impossible) to determine ultimate attainment (i.e. level of proficiency before the onset of attrition) for such populations.' The question of ultimate attainment seems to lose its importance even more if one takes those documented cases into account in which *residual learning* and improvement in language skills took place after disuse (Cohen, 1975, 1989; Weltens & Cohen, 1989). As Weltens *et al.* (1986: 10) suggest: '(...) a period of non-use may just provide the learner with the opportunity to properly "digest" – whatever that may be – all the foreign-language material that he has been exposed to during instruction, so that some sort of "unlearning of incorrect patterns" can take place'.

De Bot and Larsen-Freeman (2011: 6) define language development by including language acquisition and language attrition as 'relevant outcomes of development processes'. From a dynamic systems theory (DST) perspective, language is characterized by constant change, non-linearity and inter-connectedness (among other features). Language development itself then can be positive or negative; however, given the above listed characteristics of the system, predictions about the learning outcomes are very limited and highly dependent on initial conditions (Penris & Verspoor, this volume; Verspoor, 2015). If the system rests in the so-called 'attractor state', external forces (e.g. the influence of the teacher, contact with a native speaker, travelling, etc.) may create a positive (or negative) attitude which moves the learner from the attractor state to one of the directions (improvement or decline).

Finally, according to Schmid and Mehotcheva's (2012) classification, the following variables count as extralinguistic aspects of language attrition/retention: age, age at the onset of attrition, attained proficiency, attitude and motivation, aptitude, time since onset of attrition, language contact and use, length of exposure to the language and typological proximity. With the exception of attained proficiency, aptitude and typological proximity, all the factors were included in our study with a special emphasis on attitude. In the following, I will have a closer look at these factors.

Attitude and motivation: Crucial extralinguistic factors?

Schmid and Mehotcheva (2012: 111) point out that when studying FL attrition, the motivation involved is different to some extent from L1/L2 attrition due to the classroom environment of language learning, where students are placed in artificial situations and 'the learner's willingness to communicate (WTC) derives more from the syllabus than from within her-/himself and may be thus less susceptible to individual differences than in naturalistic acquisition'. Early research on language attrition suggests a strong link between attrition and attitude. Gardner *et al.* (1985) analyzed

self-reports of students who attended a six-week intensive French course and were asked to fill in a questionnaire six months thereafter. Their results revealed that those participants with less positive attitudes showed significant loss, while the participants with more positive attitudes showed no substantial change. Similar results evolved from the study by Gardner *et al.* (1987).

It needs to be mentioned, however, that in more recent studies (Mehotcheva, 2010; Opitz, 2011; Varga, 2012; Xu, 2010), no clear connection between attitude and attrition was found. In her research with Romanian migrants living in Canada, Cherciov (2011) found some effect of attitudes on attrition. She states that levels of L1 attrition were most strongly dependent on negative attitudes towards the L1, combined with a younger age at the time of emigration and more education in the L2 country. To sum up, the results of studies on the interaction between affective variables and attrition are inconclusive, which could either stem from the dynamic nature of atti-tude and motivation or from methodological problems (Schmid & Mehotcheva, 2012).

Attrition and tip of the tongue

Reduced access to language components can be very temporary or more long term. Tip-of-the-tongue (TOT) phenomena are an example of the former. With non-use, access will become more and more difficult until ele-ments 'sink beyond reach' (Weltens & Grendel, 1993). Attrition is a process of gradual decline, rather than a sudden loss of a part of the language – although the network-based modelling of language attrition by Meara (2004) suggests that there may be a non-linear relation between non-use and decline. In some cases massive attrition hardly affects the language as a system, whereas in other cases even minimal attrition can lead to massive decline over time. TOTs are in a way the early warning signals of attrition, reflecting the instability of the language system: words can be accessible at one moment in time and inaccessible the next. TOT data can thus inform us about that gradual process, showing what part of the (lexical) knowledge is most vulnerable or resistant. In our data, a fairly large number of TOTs were found, which is not surprising given the language situation of the partici-pants: they were forced to spontaneously recall words they had not been using for decades (see below for more information).

The study of TOTs in the L2 or FL may give us information about a posi-tive process, that is, the retrieval path through which the individual tries to resolve the TOT experience, i.e. it provides us with information about how to reactivate the lexicon of a language that was reported by the individual as forgotten. Several studies claim that eliciting TOT phenomena is a way of exploring the retrieval process. Gollan and Brown (2006), for example, argue that TOTs should not be perceived as a retrieval failure but as a partially

successful retrieval. Ecke (2009) supports the relevance of the translation equivalent method to elicit TOT phenomena, that is, to ask participants to translate the stimulus word into their other language. When in the TOT state, individuals produce words or chunks of words while trying to recall the target word. TOTs with related words or word associates take longer to resolve than TOTs without associates, which raises the question as to whether they play any role in the resolution and whether they give any information about lexical organization (Burke *et al.*, 1991; Ecke, 2009). So far findings have shown that associates not only share the syntactic class with the target word in both L1 and L2, but they are also semantically related to the target (bilingual word association tests provide us with similar results; see Navracsics, 2007). Furthermore, associates can be similar to the target in sound and/or meaning. In the case of phonological similarity, the associates usually match the target on the initial sound or letter.

Learning conditions and language attrition

Ecke (2004) points out that attrition can be studied at the intersection of psycholinguistics and sociolinguistics, i.e. language attrition at the individual level may have a contribution to language change or even shift at the community or global level. However, languages gaining influence by globalization or migration may also lead to language attrition in the individual speaker. In addition to the psycho- and sociolinguistic nature, language attrition has another very important component, that is, a pedagogical orientation, or the *language teaching and learning context dimension*. As outlined above, when studying language attrition, the way the language (s) of the individual were acquired or learned has to be taken into account. In the case of FLs it is clear that, for the most part, language learning takes place in a formal, instructional setting. The methods of language teaching as well as the circumstances of language learning have changed tremendously throughout the last few decades. Schmid and Mehotcheva (2012) note in connection with the generalizability of language attrition studies from the 1980s and 1990s that, in the school system of many European countries, language use and language input were limited almost exclusively to the class without any real-life application. Nowadays, besides the wide range of teaching methods, language learning is fuelled by an abundance of opportunities to use and improve on the knowledge acquired in school (e.g. there are many websites where students can sign up and practise almost any foreign language with native speakers). What is more, technological developments and globalization make it possible not only to gain knowledge from the school, but also from social media, movies, computer games, music, etc., which also creates new challenges for the teacher. This suggests that the context and conditions of language teaching (e.g. teaching aids, opportunities to use the language outside the classroom, methods, etc.) should always be taken into account

when studying language attrition. In the case of the compulsory teaching and learning of Russian in Hungary we cover a past period when foreign language teaching was different in many respects from current circumstances; for example, teachers had to face very poor conditions in the classroom (including their lack of expertise in the language). The next section discusses that period in more detail.

A short history of Russian teaching in Hungary

Russian teaching had no history in Hungary before 1944/1945, which made the implementation rather difficult given the lack of trained teachers, teaching materials, textbooks, and even the lack of Cyrillic letters in the press. Research on language teaching and learning from this period is very limited; we can only rely on the comprehensive study by Szilágyiné Hodossy (2006) and the interviews with our participants. In order to have a better understanding of the context and circumstances of Russian teaching and learning in Hungary, the period of 1944–1989 will be briefly discussed here.

We know from the history of language teaching in different countries that it is politics that defines what languages are to be taught in the schools. This was also the case in Hungary in 1944 when the Russian occupation started. Without any precedent, it was very problematic to introduce Russian to the schools; there was no Russian teacher training in the country. According to Szilágyiné Hodossy (2006), the period can be divided into three parts:

- 1944–1949: 'pluralism'
- 1949–1956: 'centralization'
- 1957–1988: 'gradual liberalization'

Szilágyiné Hodossy (2006) notes that, besides the problems with the conditions and context of teaching, the content and methods were also inadequate. Students' age was not taken into account, which was mostly reflected in the topic of the texts used in the classroom (in the first period, socialism, political issues, etc.; later, Soviet culture) and the vocabulary. Regarding the methodology, the grammar-translation method combined with the direct method was used throughout the three periods (Szilágyiné Hodossy, 2006).

Another element of foreign language learning is the attitude towards the target language group which is very much influenced by the contact with its members. Gardner *et al.* (1987), in their early work on the role of motivation and use in second language attrition, stress the importance of the sociocultural context. They say that the availability or lack of the target language group and intergroup interaction could have an influence on attrition. In the period discussed here, Russian soldiers were stationed in the territory of Hungary, although they lived quite isolated in the society and limited their

interaction to the necessary conversations. Even though the target group was present in Hungary, students had no opportunity to form their attitude based on contact with the target group, apart from pen pals or in some cases trips to the Soviet Union.

The Study

The main goal of the study is to analyze the extent to which attitudinal factors have an effect on language attrition and/or the retention of Hungarian learners of Russian as a compulsory subject in Hungary between 1958 and 1988. As mentioned above, this study has a highly retrospective nature. Following Cherciov (2013) we tried to combine quantitative and qualitative approaches. In a very recent discussion of mixed methodology in second language acquisition (SLA), Singleton and Pfenninger (2015) conclude that '[e]ven where the qualitative dimension is in the nature of "add-on", of "putting flesh on the quantitative bones", insights may emerge which go well beyond what is discoverable via a quantitative investigative *modus operandi* alone'. Based on the five broad rationales (triangulation, complementarity, development, initiation and expansion) mentioned in Singleton and Pfenninger (2015), this study has a rationale of complementarity, with the qualitative part, the interview, serving to illustrate and clarify the quantitative data. While following Ivankova *et al.* (2006) the study has a mixed-methods sequential explanatory design, 'which implies collecting and analyzing quantitative and then qualitative data in two consecutive phases within one study'.

Our hypotheses are that: (1) participants will recall fewer words than they can recognize; and (2) positive attitudes will lead to better retention.

Participants

As mentioned above, all of our 39 participants (29 females and 10 males) learned Russian between 1958 and 1988, i.e. the period known as 'gradual liberalization'. In order to exclude age effects, that is, any decline in the knowledge due to age, the age range of our participants was restricted to 40–56. Participants were recruited and selected by the snowball approach (Goodman, 1961), which is a non-probability technique where existing participants help to recruit future subjects from among their acquaintances. The main criterion for selection was having had Russian as a mandatory subject before the dissolution of the Soviet Union (before the 1990s) (see Bátyi, 2015).

The disadvantage of a two-phase research design with multiple measurements over time is always that participants may drop out. This was the case in our experiment too, when the online questionnaire was filled out by many people but for several reasons they were not available for the next session. Altogether 39 participants completed the three parts of the study and were divided into three age categories: 40–45 (nine participants); 46–50 (14 participants); and 51–56 (16 participants).

Instruments and procedure

For the quantitative analysis we administered a lexical test and a questionnaire. The lexical test measured the level of retention of the vocabulary once acquired in school. Two (hierarchical) levels of the lexical knowledge were tested: recall and recognition (see also de Bot & Stoessel, 2000). The highest or deepest level is recall, where participants have to find the FL (in our case Russian) equivalent of the word they are presented with in their L1 (Hungarian). It is an active process and requires input or stimulus from the outside environment. The next level is recognition, or the ability to (passively) recognize words by translating FL (Russian) words into L1 (Hungarian). According to Ecke and Hall (2013) using translation equivalents is a good way to elicit TOT phenomena. Participants were presented with 50 words per language in both levels with no time constraints. Besides getting information about the level of vocabulary retention, this task provided us with data about the word searching process the participants went through. The lexical session was audio-recorded and the interviewees often thought aloud while looking for the target word in their mental lexicon (see Bátyi, 2015).

The stimulus words were selected from books that were used at the time of the language learning, for example: *Orosz nyelvkönyv* [*Russian Textbook*] (Kecskés, 1986); *Шаг вперед* [*Step Forward*] (Szilágyi & Oszipova, 1989); *Orosz nyelvkönyv* [*Russian Textbook*] (Hlavács & Rhédey, 1969). Only nouns were selected and categorized into the following groups: persons (e.g. *boy, grandmother*); places (e.g. *house, shop*); school-related (e.g. *pen, desk*); animals (e.g. *dog, elephant*); food and drink (e.g. *milk, bread*); body (e.g. *hair, head*); nature (e.g. *air, tree*); abstract (e.g. *peace, song*). Each level (recall and recognition) contained the same number of items from each category in a randomized order (to avoid a trigger effect). A complete list of items can be found in the Appendix.

Prior to the administration of the lexical test, the participants were asked to fill in an adapted online version of the Attitude and Motivation Test Battery (AMTB), which was originally developed by Gardner (1985). In the questionnaire, which was sent as an online form to the participants, they had to indicate on a 7-point Likert scale to which extent they agreed with the statements. There were seven randomized clusters:

(1) Attitude towards learning foreign languages (e.g. *If I were visiting a foreign country I would like to be able to speak the language of the people*).
(2) Attitudes towards Russians (e.g. *Russians are a very sociable, warm-hearted and creative people*).

(3) Integrative orientation (e.g. *Studying Russian was important for me because it enabled me to better understand and appreciate Russian art and literature*).

(4) Instrumental orientation (e.g. *Studying Russian was important to me because I thought it would someday be useful in getting a good job*).

(5) Attitudes towards learning Russian (positively worded items and negatively worded items) (e.g. *Russian was an important part of the school programme; learning Russian was a waste of time*).

(6) Anxiety in Russian class (e.g. *I always felt that the other students spoke Russian better than I did*).

(7) Attitudes towards the teacher, which includes positive and negative features (e.g. friendly–unfriendly, organized–disorganized); unlike other parts of the questionnaire, participants had to indicate on a 6-point Likert scale to what extent the positive or negative adjective characterizes their teacher.

The third part of the data collection, the interview, which included almost the same questions as in the questionnaire, was administered before the lexical test. While in the questionnaire attitude was operationalized by closed statements, the interview with the same questions elicited more detailed answers which gave us an opportunity for triangulation.

The three parts were administered in two sessions. In a first step, the online questionnaire was filled in by each participant, after which the experimenter contacted the participant and the interview was administered followed by the two-level lexical test. The interview method was used as a complementary method to the questionnaire in order to obtain a better picture of the participants' attitudes and motivation to learn Russian, the two constructs that are difficult to operationalize, especially in a retrospective study such as this one. The interviews were transcribed and imported to Maxqda and the text was coded. Participants were asked about their attitude towards learning Russian, Russian people and towards their teacher(s). Attitudes were coded as negative, positive and neutral.

For the analyses, SPSS and Maxqda software were used. The reliability of the AMTB was checked by computing Cronbach's alpha internal consistency reliability coefficients (see Table 13.1). All measures show an acceptable reliability coefficient except for instrumental motivation.

In Figure 13.1 the means are displayed for each cluster. On the group level it seems that attitudes towards learning foreign languages yielded the highest values, followed by attitudes towards learning Russian and Russian people, which reached quite high values on the 7-point Likert scale. Anxiety in the Russian class does not seem to have been a problem (with values under 3 on the 7-point scale), and integrative or instrumental orientations received medium scores.

Recall and recognition suggest two different processes; in the former the translation requires an impulse from within the system, while in the latter

Table 13.1 Reliability of the Attitude and Motivation Test Battery

Dimensions	Cronbach's α
Attitude towards Russians	0.928
Attitude towards learning foreign languages	0.773
Positively worded items	0.898
Negatively worded items	0.914
Integrative motivation	0.887
Instrumental motivation	0.571
Anxiety in the Russian class	0.925

stimulation is needed from the outside (de Bot & Stoessel, 2000). We hypothesized higher scores for recognition than for recall which was verified by the results (recall level: $M = 18.74$; recognition level: $M = 26.26$; $t(8) = -9.664$, $p < 0.001$).

A multivariate regression analysis was conducted to reveal the predictor variables, that is, to what extent the independent variables predict language attrition and/or retention. From the dependent variables (recall and recognition) a composite variable was created (their correlation was very high: $r = 0.955$, $p < 0.001$). For this part of the analysis, the results of the AMTB were used as independent variables. Traditionally, the *Contact with and use of the language* since the onset of attrition is taken into account; however, in our case, only a few participants mentioned that once or twice they had given directions in Russian, had had a small conversation in the language or had read Russian on signs, which is why this dimension could not be included in
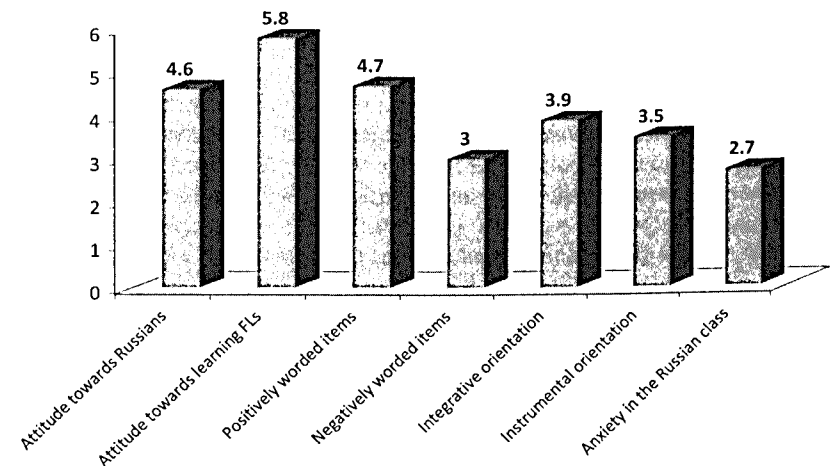

Figure 13.1 Average values of the Attitude and Motivation Test Battery clusters

the analyses. The following variables (extralinguistic aspects) were chosen: *age* (3 levels, see above); *sex, language exposure* (the number of years of studying Russian); *Attitude towards Russian people* (ATRP); *Attitude towards learning Russian* (ATLR) and *Attitude towards the teacher* (ATT). In the case of ATLR and ATT we wanted to create composite variables as both had positively worded and negatively worded items as well. The correlation of the positive and negative components of the ATLR ($r = -781$, $N = 39$, $p < 0.001$) let us create the composite variable (after the item reversal). For the teacher questionnaire the correlation between the two components was low ($r = -0.386$, $p < 0.005$, $N = 39$), so we decided to include them separately: negative attitude towards the teacher (ATTneg), positive attitude to the teacher (ATTpos). Participants also had the chance to rate more than one teacher in the questionnaire (it was likely that Russian was taught by several teachers). Only a few of them rated other teachers than the 'significant' one (to be discussed later) and the difference between the two ratings was not significant.

Results and Discussion

In the first part of the analyses, the results of the AMTB and the lexical test will be described briefly, followed by a discussion of the predictor variables in two quantitative analyses and, finally, the interviews will be used to support and enrich the quantitative analyses.

Altogether one dependent and seven independent variables were included in the multivariate regression analyses. The results from the stepwise analysis are illustrated in Table 13.2. Predictor variables were ATLR, language exposure and sex, explaining 61.3% of the variance. Surprisingly, the attitude towards the teacher had no predictive power on the level of retention. The interviews revealed that 34 out of the 39 participants (87.18%) had had a positive attitude towards their teacher, and they emphasized that the problem had not been the teacher but the fact that Russian had been a mandatory subject. After coding, a multivariate regression analysis was conducted with the same variables, but this time based on the result of the interviews. The results were very similar to what we got from the questionnaire (see Table 13.3).

Table 13.2 Multiple stepwise regression analyses (questionnaire)

Variable	R	R²	Adjusted R²
Attitude towards learning Russian	0.691	0.478	0.464
Attitude towards learning Russian/language exposure	0.773	0.598	0.576
Attitude towards learning Russian/language exposure/sex	0.802	0.643	0.612

Table 13.3 Multiple stepwise regression analyses (interviews)

Variable	R	R²	Adjusted R²
Language exposure	0.669	0.447	0.432
Language exposure/Attitude towards Russian language	0.732	0.536	0.510

This all shows that the merging of qualitative and quantitative approaches on the level of data collection and data analysis provides us with results that are supported by two different yet overlapping results. So, besides quantifying the interview results, we got useful information about the Russian learning context (methods, materials, teacher, motivation) and about the attitude towards the language.

Let us now discuss the Russian learning context through the participants' recollections. The AMTB contains a cluster of statements where attitude towards the target language group is concerned. We must realize that the statements take on other meanings when a person is immersed in the target language group or when the language is learned without any real contact with native speakers. Because of political, historical and ideological factors that were partly present in the schools and partly in the society or in the smaller environment of the students, i.e. in their families, the situation for Russian and attitudes towards the Russian people were more complex.

Example 1 exemplifies the isolation of the Russians in this period:

(1) *Once in Eger, in the disco I ran into Russians and there I had to talk ... but it was a unique experience that I could use the language with living Russian persons. Especially at that time when it wasn't so normal to meet a native speaker. (...) Even if you lived in their neighbourhood they had their own schools, shops and they were not encouraged to mix.*

Thus, the Russians were not part of the 'host' society, and intergroup contact was unlikely to occur.

Another dimension of the attitude towards Russian people is that our participants were students without clear views on the whole Russian invasion, as Example 2 illustrates:

(2) *I was curious how they lived and what kind of people they were, but thinking back, I had no such thoughts like they were different from us. As a child you accept other people easier.*

The general opinion about the Russians seems to be quite positive among the participants (see Example 3–6 below), although the attitudes may be negatively influenced by the experiences of family members. This discrepancy

between the influence of the school and the influence of the family can be explained in terms of the different experiences the parents of the participants undergone with Russians: after WWII, i.e. until 1956, there was great tension between the invaders and the invaded population which resulted in a revolution, and it generated negative feelings towards Russians in that generation. The historical memory in Hungary is still processing the horrors of the first years of the invasion (1944/1945) (see Pető, 2014), although humane acts of the Russians are also mentioned in the narrations. As Ungváry (2002) notes, on the individual level, there were people for whom the appearance of Soviet soldiers meant life, and others for whom it meant captivity, rape and property uncertainty.

(3) *I cannot (say) good or bad. There must be good and bad people among them too. (...) I am not even sure they were happy to be here. (...) People all over the world have the same feelings, only the environment and traditions are different and their life took a different path.*

(4) *They have a very deep spiritual life, despite the fact they suffered the years of communism, no-one could take away the deep faith in God from them. I am not sure our teachers worded the whole thing like this, but the Russian literature is also very religious, if you think of the works of Dostoievskij or Tolstoy. The educated and the simple Russian people are also very likeable, warm-hearted and hospitable.*

(5) *I think they are better people than the English. The Slavic is ... their heart is bigger.*

(6) *It was two different things what we learned at school and what we learned from our parents. In the school they tried to teach us the Russian culture, e.g. the Tretyakov Gallery, the monuments and famous buildings of Moscow, generally the culture. What I heard at home, that when they trooped into the town my parents found their behaviour very strange and unusual, and their thinking was so different from ours. So, for me it was hard to compare these two things, you know, what I heard from my parents and their experience and at school. ... It was because at that time it was hard-core socialism. And it was strengthened in the education too, the big Soviet brothers, their culture, and ... that we should be like them. That was in fashion to say so.*

It is clear from Examples 4 and 5 that the motivation for learning the language was not influenced by the attitude towards its speakers, as there was hardly any interaction. When asked about their attitudes towards learning Russian, the participants listed such aspects as language aptitude (see Example 7), the obligatory nature of the subject (Example 8), and the teacher in (Examples 9 and 10):

(7) *I didn't like it. Well ... simply ... how to say. I think the case is that I have no language aptitude. ... Maths and Physics are quite alright, but language was always a torture for me because I had to sit down and learn things by heart.*

(8) *We didn't like Russian ... not because of the country (SU) but because ... It was like every other subjects: you had to sit down and learn.*

(9) *Yes! What I learned in school was interesting. Maybe it was the teacher ... the way she treated people.*

(10) *Our form-master taught Russian who always wanted to be a French teacher, so there was a total lack of enthusiasm.*

Research on the effects of language aptitude on attrition (as described in Example 7) is very limited. A study by Bylund *et al.* (2010) suggests that higher language aptitude in L1 helps to prevent it from attrition, while Cherciov (2011) found that attrition is more dependent on attitudinal factors rather than language aptitude. However, it is important to note that both studies focused on L1 attrition.

In each interview, the respondents were asked about their Russian teacher (s), and it turned out – in line with previous research on language teaching (e.g. Cotteral, 1999) – that the teacher plays a key role in shaping students' attitude towards learning the language (see Examples 11 and 12). However, as it appeared from the quantitative analyses, the attitudes towards the teacher had no effect on the extent of language retention.

(11) *At the secondary school we had Russian besides English and both teachers were funny. He (the Russian teacher) was a terrific phenomenon. He looked like the teachers in the textbook with the typical glasses, shirt buttoned up to the neck, blade lips. Then the next teacher looked like a Natasha, like an old Russian lady. She never tried to make us love Russian, we loved to sing Russian songs and we learned a lot of them. I really regret I couldn't learn it back then because I still like it.*

(12) *My perspective was shaped by the personality of the teacher. I had several teachers, I liked the one in the elementary class, but then the vice-director taught it and I had to memorize grammar, which I didn't like at all.*

The responses reveal that the teaching methodology back in the day was the grammar-translation method (Examples 13–16), and in those rare situations when conversation-based tasks were used, students were placed into unnatural and unlikely situations.

(13) *We translated texts, there were no dialogues and we had to learn the words. And we also had to write coherent texts.*

(14) *We started with introduction, travelling, once at the exam I got the situation to buy a present for my wife, then I said what and where I bought. We translated a lot, this is how you teach the language, right?*

(15) *We learned very unlikely situations. For example Sasha invited István and his father to the Soviet Union to show them the Swan Lake in Bolshoj. Our speaking skill was zero, we translated sentences.*

(16) *There was no practical part at all. We couldn't ask for a glass of water. But we could find everything in Moscow without maps because we knew the places. And of course the characters of the Swan Lake, which I don't think I will need for anything in the future.*

Both the quantitative and qualitative results suggest that language attrition and/or retention are mostly influenced by the attitude towards learning Russian. Students in the period in question had had very few or no opportunities to have contact with the target language community, which means that their attitude towards learning the language could only be shaped in the school by the teacher and the applied methods, which is also described by the following quote from Littlewood (1984, quoted in Cook & Singleton, 2014: 93):

> There are some learning situations where many learners have not had sufficient experience of the second language community to have attitude for or against it. ... (M)any learners of French and German in Great Britain provide an example. In cases such as this, it is probable that attitudes relate more directly to learning as it is experienced in the classroom.

Many of the participants expressed their wish to reactivate their Russian and regretted that at school they had not invested enough energy to master it. Attrition studies at this point cannot answer the question of how to reactivate a forgotten language; more research is needed to explore the retrieval process in several types of language attrition (L1, L2, FL). However, our research design enabled us to assess retrieval paths through which the participants tried to reach the target word and these processes appeared to be similar to what happens in TOT states.

In our study participants showed signs of TOTs in 12.18% of all the word searching on the recall level and in 15% on the recognition level (cf. Gósy, 2004, who found 3.64% of TOT states with the definition method, including pauses, and Ecke, 2001, who reports 30.9% with the translation task, pauses not included). From the 249 TOT states on the recall level, 64 were resolved, while on the recognition level, 84 were resolved from the total of 324 TOT states (see Bátyi, 2015, for a detailed analysis).

Based on our data, 5 + 1 categories were created to classify word associates according to their relatedness to the target word (Table 13.4). The governing principle between associates and targets is meaning relatedness, e.g. TOLL (pen) – *карандаш* ('pencil') (7), *перо* ('feather') (in Hungarian the word *toll* is a homonym with two meanings: 'pen' and 'feather'), *бумаг* ('paper'). Previous research suggests that phonological relatedness is more typical of TOTs with L2 words even with newly learnt words (Ecke, 1997). In our case 21% of the associates are related phonologically to the FL target, while 30.8% were related to the L1 target. Table 13.3 clearly shows that phonologically related associates were activated more by the FL stimulus when searching for

Table 13.4 The relationship of word-associates with the target words

	Meaning-related	Sound-related	Mixed (meaning + sound related)	Cross-lexical	Context-related	Other
FL target	40%	16%	5%	19%	11%	2.3%
L1 target	42%	27%	3.8%	4.5%	1.4%	19.9%

Notes: See Bátyi (2015).

the L1 target than vice versa. If we take a closer look at the phonologically related associates on the recognition level, we can easily recognize that the relatedness often roots from the sounding of the stimulus, a phenomenon that has not been mentioned in the literature on bilingual TOTs (e.g. помидор /pomidor/ ('tomato') – (9) *folyosó* ('corridor'), (5) коридор /koridor/).

The context of vocabulary learning seems to be a factor in reactivating the target words; 11% of the TOTs on the recall level were based on remembering phrases and chunks of texts which helped the participant to find the target word (e.g. MESE ('tale') – *сказке конец иди работай* ('the story is over, go to work'), *ворона и рак* ('the crow and the crab by Tolstoi'). These findings also emphasize the importance of learning and teaching context when studying language attrition and/or retention. The analyses of the TOT states and FL attrition provide us with information about the retrieval process and strategies of accessing lexical information that has not been used for decades, besides revealing the connections in the mental lexicon.

Concluding Remarks

The main goal of this mixed-methods study was to analyze the extent to which positive or negative attitudes have an effect on vocabulary retention. To this end we looked into Hungarian learners' (former) attitudes towards Russian people and the Russian language, while testing their lexical skills in order to get an impression of the degree of lexical attrition. Given the dynamic nature of attitude, interviews were conducted in addition to a written questionnaire in order to get a clearer and more comprehensive view of the attitudes towards Russians, the Russian teacher(s) and learning Russian in a difficult time in history. As it was outlined in this chapter, the situation of Russian in Hungary in the 20th century was special in more than one sense. For instance, teaching conditions were not sufficient to guarantee the success of language learning due to the lack of trained teachers, books, dictionaries and teaching aids. The quantitative analyses showed that the extent of vocabulary loss is predicted by the learners' attitudes towards language learning as well as years of instruction. On the other hand, through the interviews with

the participants it was discovered why other factors, such as attitudes towards the teacher and towards Russians, do not show an effect on attrition. The isolation of the Russian soldiers in the country prevented the students from having any contact with the Russians and, given their (young) age, their attitude towards Russian people and towards the language was shaped by their teacher, the teaching materials and their experiences in the classroom.

The methodology of data collection, i.e. translation equivalents, led to a large number of TOT states which inform us about how our participants retrieve target words. Our data show that the lexical reactivation path is mainly semantic, while an important outcome of the study is that the context of teaching and learning plays an important role in recalling a language; i.e. finding the right cues may help in remembering. The latter findings are in line with results from psychology where several studies proved the importance of context-based cues for retrieval (e.g. Tulving, 1974; Tulving & Pearlstone, 1966) but until now it has not been applied in attrition studies.

Additionally, it should be kept in mind that, since attitudes are dynamic in nature (Dörnyei & Ryan, 2015), it might well be the case that the attitudes of the participants may have shifted over the several decades of non-use of Russian due to political and/or private changes and experiences. Hence the attitudes measured here could either reflect the sum of attitudes over time or merely the participants' current attitudes. The dynamic nature of attrition calls for more data points over time, with the risk that the testing itself may interact with the attrition process.

The present study has several implications for future research in the field of applied linguistics. One is that the assumed direct relation between attitudes and development is much more complicated than the literature suggests. It turns out that positive attitudes do not necessarily lead to better learning and more retention. The other is that the retrieval of a forgotten language seems to be dependent on context-based cues, which is in line with the idea of situated cognition in psychology, from which applied linguistics may also benefit. Thirdly, our data support the idea that both acquisition and attrition are dynamic processes and thus require a dynamic approach. Finally, a mixed methods approach appeared to be a useful method in the study of attrition; i.e. besides the quantitative data the interviews with former students of Russian provided us with rich and useful information on all the variables. Therefore 'first person accounts' should be considered as sources of relevant information in the study of language attrition and acquisition.

References

Bardovi-Harlig, K. and Stringer, D. (2010) Variables in second language attrition: Advancing the state of the art. *Studies in Second Language Acquisition* 32, 1–45.

Bátyi, Sz. (2015) Foreign language attrition: Elicited TOT phenomena. In J. Navracsics and Sz. Bátyi (eds) *First and Second Language: Interdisciplinary Approaches*. Studies in Psycholinguistics No. 6 (pp. 47–60). Budapest: Tinta Könyvkiadó.

Burke, D.M., Mackay, D.G., Worthley, J.S. and Wade, E. (1991) On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language* 30, 542–579.

Bylund, E., Abrahamsson, N. and Hyltenstam, K. (2010) The role of language aptitude in first language attrition: The case of pre-pubescent attriters. *Applied Linguistics* 31 (3), 443–464.

Cherciov, M. (2011) Between acquisition and attrition: The dynamics between two languages in adult migrants. PhD dissertation, University of Toronto.

Cherciov, M. (2013) Investigating the impact of attitude on first language attrition and second language acquisition from a dynamic systems theory perspective. *International Journal of Bilingualism* 17 (6), 716–733.

Cohen, A.D. (1975) Forgetting a second language. *Language Learning* 25, 127–138.

Cohen, A.D. (1989) Attrition in the productive lexicon of two Portuguese third language speakers. *Studies in Second Language Acquisition (SSLA)* 11 (2), 135–150.

Cook, V. and Singleton, D. (2014) *Key Topics in Second Language Acquisition*. Bristol: Multilingual Matters.

Cotterall, S. (1999) Key variables in language learning: What do learners believe about them? *System* 27 (4), 493–513.

de Bot, K. (2008) Introduction: Second language development as a dynamic process. *Modern Language Journal* 92 (2), 166–178.

de Bot, K. and Larsen-Freeman, D. (2011) Researching second language development from a dynamic systems perspective. In M. Verspoor, K. de Bot and W. Lowie (eds) *A Dynamic Approach to Second Language Development: Methods and Techniques* (pp. 5–23). Amsterdam: John Benjamins.

de Bot, K. and Stoessel, S. (2000) In search of yesterday's words: Reactivating a long forgotten language. *Applied Linguistics* 21, 364–384.

de Bot, K., Verspoor, M. and Lowie, W. (2007) A dynamic systems theory approach to second language acquisition. *Bilingualism, Language and Cognition* 10 (1), 7–21.

Dörnyei. Z. and Ryan, S. (2015) *The Psychology of the Language Learner Revisited*. New York: Routledge.

Ecke, P. (1997) Tip of the tongue states in first and foreign languages: Similarities and differences of lexical retrieval failures. In L. Díaz and C. Pérez (eds) *Proceedings of the EUROSLA 7 Conference* (pp. 505–514). Barcelona: Universitat Pompeu Fabra.

Ecke, P. (2001) Lexical retrieval in a third language: Evidence from errors and tip-of-the-tongue states. In J. Cenoz, B. Hufeisen and U. Jessner (eds) *Cross-linguistic Aspects of L3 Acquisition: Psycholinguistic Perspectives* (pp. 90–114). Multilingual Matters.

Ecke, P. (2004) Language attrition and theories of forgetting: A cross-disciplinary review. *International Journal of Bilingualism* 8 (3), 321–352.

Ecke, P. (2009) The tip-of-the-tongue phenomenon as a window on (bilingual) lexical retrieval. In A. Pavlenko (ed.) *The Bilingual Mental Lexicon: Interdisciplinary Approaches* (pp. 185–208). Bristol: Multilingual Matters.

Ecke, P. and Hall, C.J. (2013) Tracking tip-of-the-tongue states in a multilingual speaker: Evidence of attrition or instability in lexical systems? *International Journal of Bilingualism* 17 (6), 734–751.

Gardner, R.C. (1985) *The Attitude Motivation Test Battery Manual*. See http://publish.uwo.ca/~gardner/docs/AMTBmanual.pdf (accessed 23 February 2008).

Gardner, R.C., Lalonde, R.N. and MacPherson, J. (1985) Social factors in second language attrition. *Language Learning* 35 (4), 519–540.

Gardner, R.C., Lalonde, R.N, Moorcroft, R. and Evers, F.T. (1987) Second language attrition: The role of motivation and use. *Journal of Language and Social Psychology* 6 (1), 29–47.

Gollan, T. and Brown, A. (2006) From tip-of-the-tongue data to theoretical implications in two steps: When more TOTs mean better retrieval. *Journal of Experimental Psychology: General* 135 (3), 462–484.

14 Concluding Thoughts: A Road Map for Future Research in Applied Linguistics

Simone E. Pfenninger and Judit Navracsics

These are exciting times for applied linguists. The dynamism of applied linguistics that has been described throughout this volume has made it clear yet again that this is a multidisciplinary field of study, rather than an isolated and unconnected one. Seeing the discipline from the perspective of language problems highlights the fact that no single academic discipline can be sufficient to meet the needs of people facing multilingual/multicultural settings, bi/multilingualism and FL learning (Kaplan, 2010). Applied linguistics is said to deal with language-related problems that arise in 'real world' domains, but these are infinitely many and diverse (Widdowson, 2005). However, the diversity and the broad scope of the field should not be considered a weakness; according to Widdowson (2005: 23), they are the 'most commendable features' of applied linguistics. The chapters in this collection reflect this viewpoint and represent a wide variety of academic perspectives, from neurolinguistics to foreign language policy and second language acquisition (SLA). The volume has aimed to draw attention to key theories and methodologies in applied linguistics and, we believe, lays foundations for those new directions in this branch of linguistics which will move our subject forward. We have tried to find a balance among the theoretical, empirical and practice-oriented dimensions of applied linguistics, while showing the complexity and the multifaceted nature of the area. Although there was no space here for an exhaustive examination of all facets of applied linguistics, the issues singled out all focus on relatively new areas of study in applied linguistics and language studies methods. Accordingly, in this final chapter we take stock and identify our view of some of these key implications for applied linguistics as well as issues that will shape research agendas in the decades to come, informed by experience in allied disciplines.

What Are These 'Implications for the Future'?

We have identified three essential ingredients of future implications for applied linguistics as we currently see them – ingredients that are evident in every chapter of the book. We outline these in the following.

(1) As applied linguists we need to know about language myths and beliefs, as well as the ideologies that have been constructed, if we are to be able to present alternative solutions.

The contributions presented here directly confront the sometimes uncomfortable questions surrounding issues in multilingualism and SLA. The questions that arise in the conversation about issues in bi- and multilingualism (de Bot; Cergol Kovačević; Csépe; Navracsics & Sáry; van Heuven, all this volume) are complex yet fundamental to applied linguistics concerns in the light of the fact that the majority of people in the world are to some extent bi-/multilingual, and this bi-/multilingualism is associated with the need to negotiate life situations with other cultural and language groups. One of the hot topics in research on multilingualism is what has been labelled the 'bilingual advantage' (BA). De Bot (this volume) argues that following an initial phase in which the first excited reactions to findings on cognitive advantages in children and elderly people, and a consecutive countermovement that problematized some of the issues involved, we now find ourselves in a situation in which different types of advantages for different populations need to be the object of research on the BA. Thus, rather than being part of a camp that is for or against the bilingual advantage hypothesis, de Bot suggests that it is now time to investigate what benefits bilingualism may bring for different populations and tasks. According to Csépe (this volume), the main question to be answered in the near future is whether bilingualism relies on a common underlying proficiency composed of long-term representations with different access via the languages learned and maintained. From this point of view the development of working memory and the executive system may be more crucial than previously assumed, which can be investigated in depth with modern brain imaging techniques. In addition to these implications, these chapters also exemplify the interest of applied linguistics in cases where language goes wrong (see also Schmitt & Celce-Murcia, 2010). For instance, researchers may work on the impact of cognitive activity on Parkinson's disease (as mentioned in de Bot's chapter) in the belief that we can better understand how the brain functions when we analyze what happens when the speaker's language system breaks down or does not function properly (see also Bátyi's discussion of slips of the tongue, this volume).

The authors exploring the age factor in SLA and FL learning (Navracsics & Sáry; Pfenninger & Singleton; van Heuven) emphasize that age does matter in language learning, but the analysis of age effects is not a straightforward

enterprise. This is an area of SLA where lay people, educators, policy makers and researchers alike have been a little too cavalier in the generalizations made about age effects, particularly the co-called 'the earlier the better' assumption, which demonstrates the need for more nuanced analyses and more sophisticated methods (see below). As Hall *et al.* (2011: 18) rightly point out, it is part of the work of applied linguists to expose and explore the ideological and epistemological commitments and positions inherent in all our language-related beliefs, states of mind and social activity. According to Widdowson (2001: 795), applied linguistics, as a discipline that mediates between linguistics and language teaching, 'is of its nature a critical enterprise'. Gee (2008) claims that applied linguists have an ethical obligation to base their practices on what they believe is the best available evidence, as opposed to mandates from above, such as legislation and other voter initiatives proposed and imposed by non-linguists. He also points out that viable recommendations for best practice are often formed through a consideration of conflicting evidence and alternative viewpoints.

(2) Emphases on innovative research methodologies will play an important role in reshaping applied linguistics.

An ongoing challenge that all applied linguists face is to judge how well certain methods and approaches work in their own context, for their own participants. All contributors in this book would agree, we believe, that at its core, applied linguistics needs a coherent conception of language, whether this comes from linguistics or from some other discipline (see also Cook, 2009). One lesson that looms large in this volume is that the learning context should not – cannot – be removed from context and isolated from other systems. Dynamic systems theory (DST) has the potential to unite the major streams of research in the field of second language development, 'bringing together an understanding of learning and learner' (Larsen-Freeman, 2015: 12). This means that applied linguists ought to use research methods that honour this lesson, as de Bot (this volume) poignantly shows. The dynamic perspective presented in Penris and Verspoor's (this volume) linguistic complexity research yields a unique way of conducting a case study within the CDST framework. Working with dense data, they provide a framework in which development can be traced dynamically with adequate complexity measures, which can be applied to future studies for writers with different first languages. Even their research questions are substantively different from the majority of prior literature on complexity; rather than focusing on 'endpoints' or linear cause–effect relationships, they zoom in on different linguistic subsystems as they emerge and develop. The concept of DST is also picked up in further chapters by Jessner and Török (this volume), where the authors take into consideration the dynamics of the linguistic and other systems of the multilingual learners involved, thus proposing an alternative in the methodological approach to research on language learning strategies and multilingual benefits, and by

Bátyi (this volume), who discusses change and emergence as central to any understanding of the process of attrition. These chapters highlight the fact that CDST has fundamentally challenged our goal for applied linguistics research and our way of conducting it.

Several chapters in this volume also seek to demonstrate how research on psycholinguistics and multilingualism has impacted research methodology in applied linguistics, mostly by expanding it beyond traditional quantitative methods, implementing mixed methodology and/or developing an interdisciplinary perspective. Nowadays applied linguists need to be versed not only in interdisciplinary areas such as SLA, psycholinguistics and cross-cultural communication, but also in different types of data collection, particularly longitudinal data, data on variation, dense data gathering and analysis, and the integration of qualitative and quantitative methods in a single study. However, mixed method research is not yet commonplace in applied linguistics research, despite the immense benefits of strategically mixing or combining qualitative and quantitative methods, approaches and concepts in a way that produces an overall design with complementary strength and non-overlapping weaknesses (see also Johnson & Christensen, 2004; Singleton & Pfenninger, 2015). Where the two methodological dimensions are combined in a truly interactive, mutually informative fashion, the findings can be both satisfyingly rigorous and extremely rich.

Lazaraton (2005: 209) laments the fact that the whole discipline of applied linguistics seems to be struggling with a redefinition of its research goals, methods and paradigms, as can be seen in the growing acceptance of qualitative methods, increasingly pointed questions about the significance of research in this field, and continuing explorations of alternatives. However, despite the more widespread use of qualitative research (see, for example, Gabryś-Barker, in prep.), quantitative approaches will remain embedded in our daily lives, not least because one of the main goals of applied linguistics is to attach meaning to certain outcomes by generalizing to other individuals or groups of individuals (see Pfenninger & Singleton, accepted). In Gass's (2006: 213) words, 'we must have the capacity to generalize, for, if not, research remains at the "that's interesting" stage rather than moving the field along in any theoretically serious way'. We have argued in this book that we need ever deeper statistics training for new generations of applied linguists, of the kind suggested in the chapter by Pfenninger and Singleton. Several other chapters advocate innovative research agendas and methodologies. For instance, Navracsics and Sáry (this volume) call attention to the difficulties of compiling valid bilingual test material for phonological and semantic rhyming tests to measure the linguistic awareness of bilinguals, owing to differences in the depth and extent in which the available corpus linguistic databases present their data in the relevant languages. This ties in with the discussion by de Bot, who also examines the issue of the ecological validity of tests, discussing the compatibility and validity of different tasks that are supposed to establish the impact of bilingualism and

pondering the (im)possibility of isolating bilingualism as a factor from other co-occurring socio-affective and cognitive variables. Referring to de Bot (2010, 2015), Cergol Kovačević (this volume) recommends that checking visual data against auditory data when comparing the bilingual processing of languages with different, but also related, orthographic depths 'should become a regular feature in this discipline, since such a hybrid research methodology represents an attempt to bridge the gap between the old and new ways in psycholinguistic research'. Finally, the consistency analysis that van Heuven (this volume) advocates may reveal itself to be a useful tool to analyze the stability and well-definedness of sound categories in monolingual and bilingual language users, as he has found evidence that the consistency index may serve as a diagnostic of language dominance in (early) bilinguals.

Thus, one of the main implications for applied linguistics that results from the trends identified and documented in these chapters is the following: while applied linguists have more or less successfully included the learner as an active participant in recent decades by focusing on how the individual characteristics of each learner affects their learning (see Schmitt & Celce-Murica, 2010), we now ought to move away from the tendency to operate within 'decontextualized contexts' (Pennycook, 2005: 796); that is, we need to look at the learner *within context*. In particular, experimental SLA laboratory studies have been criticized for their lack of ecological validity because the language(s), contexts and activities do not represent those ordinarily encountered by language learners and users (Duff, 2010). However, in those cases where language learning is explored in, and in relation to, learning contexts and social contexts, the conceptualization of context is frequently neglected in the actual analyses. It is not enough, therefore, merely to draw connections between language and context, but context needs to be granted appropriate weight in the analyses.

(3) Even though the applied linguist is not an enforcer, applied linguists should contribute their own interpretation and advice in language policy and education.

Because our work deals directly with people facing different language-related problems in diverse communities and learning contexts, applied linguists are witnesses to how language shapes lives, as well as basic services such as education and legal protections (Hall *et al.*, 2011: 339). Cook (2009) warns that it is ultimately the client – and not the applied linguist – who has to balance all other factors and decide on the solutions. However, he concedes that the applied linguist is there to serve teachers' needs, to play a part in EU projects on translation and on linguistic diversity, to design syllabi, and to devise orthographies for languages that have no written form, among many other tasks. The 'centrality of client needs', as Hall *et al.* (2011: 17) refer to individuals' and groups' language-related needs, is at the centre of many chapters in this collection, e.g. in the discussion of critical thinking as a crucial factor in teacher education and teacher assessment (Fábián, this volume), and Bátyi's

(this volume) study of foreign language vocabulary retention as a function of learners' attitudes. The authors of these chapters have not just contented themselves with a description or empirical investigation of the role of language in the real-world problems in question, but great efforts have also been made to plan, test and evaluate potential solutions. After all, applied linguistics is a discipline concerned with the role that languages play in perceived problems of communication, social identity, education, politics and justice, *and* 'in the development of ways to remediate or resolve these problems' (Hall *et al.*, 2011: 15). FL learning in the conventional sense cannot be, in and of itself, sufficient because it is not merely a matter of learning phonology, vocabulary, and grammar but rather implicates behaviour in multicultural environments and all the communicative skills implicit in such activity (Kaplan, 2010: 575).

In a similar vein, politically motivated language policies might be grounded in language ideologies rather than language realities, as Csernicskó (this volume) illustrates in his description of the linguistic landscape of Ukraine, where the main political goal is 'to defuse the tension between the *de jure* situation (according to which Ukraine is a monolingual country) and the *de facto* situation (the reality where most of the population is multilingual)'. The discrepancy between politically motivated language changes and language realities is also reflected in Bülow and Harnisch's (this volume) analysis of the process of 'undoing gender' in German-speaking countries. Their plea for the use of the generic masculine as the most economical way to avoid gender – rather than the stylistically more cumbersome nominalized participles stipulated by top-down language policy efforts – demonstrates how language planning theory ought to explore closer conceptual links with sociolinguistics and the various branches of discourse studies. It seems that language ideologies occasionally elevate unproblematic language to the status of a 'problem' requiring treatment in overt processes of policy. However, any language planning activity must centrally derive its authority from the community(ies) of speakers (Kaplan & Baldauf, 1997) and, more importantly, include the totality of the ecosystem – notably the context – within which that activity is embedded (Eggington, 2005). This chapter also exemplifies how applied linguistics knowledge may be utilized in practical ways in non-educational areas.

These, then, are just some of the issues and challenges that applied linguists must address in the future from different perspectives and using a variety of approaches. The rapid development of applied linguistics and the array of available tools and resources applied linguists have available to them means that we will have to be vigilant in order to capture the new theoretical and methodological directions that are now coming to the fore.

References

Cook, V. (2009) Introduction: Applied linguistics and language teaching in the 21st century. In V.J. Cook and Li Wei (eds) *Contemporary Applied Linguistics. Vol. 1: Language Teaching*

and Learning. London: Continuum. See http://www.viviancook.uk/Writings/Papers/IntroAL08.htm (accessed 30 November 2015).

de Bot, K. (2010) Cognitive processing in bilinguals: From static to dynamic models. In R.B. Kaplan (ed.) *The Oxford Handbook of Applied Linguistics* (2nd edn) (pp. 335–348). Oxford: Oxford University Press.

de Bot, K. (2015) From old school to new school research in psycholinguistics. Plenary talk presented at the 29th International Conference of the Croatian Applied Linguistics Society, 'Applied Linguistic Research and Methodology'.

Duff, P. (2010) Research approaches in applied linguistics. In R.B. Kaplan (ed.) *The Oxford Handbook of Applied Linguistics* (2nd edn) (pp. 45–62). Oxford: Oxford University Press.

Eggington, W.G. (2005) Introduction. In P. Bruthiaux, D. Atkinson, W.G. Eggington, W. Grabe and V. Ramanathan (eds) *Directions in Applied Linguistics* (pp. 223–226). Clevedon: Multilingual Matters.

Gabryś-Barker, D. (in prep.) Applied linguistics and multilingualism. In D. Singleton and L. Aronin (eds) *Twelve Lectures on Multilingualism*. Bristol: Multilingual Matters.

Gass, S.M. (2006) Generalizability: What are we generalizing anyway? In M. Chalhoub-Deville, C. Chapelle and P. Duff (eds) *Inference and Generalizability in Applied Linguistics: Multiple Perspectives* (pp. 209–220). Amsterdam: John Benjamins.

Gee, J.P. (2008) *Social Linguistics and Literacies: Ideology in Discourses* (3rd edn). London: Routledge.

Hall, C.J., Smith, P.H. and Wicaksono, R. (2011) *Mapping Applied Linguistics. A Guide for Students and Practitioners*. London: Routledge.

Johnson, R.B. and Christensen, L.B. (2004) *Educational Research: Quantitative, Qualitative and Mixed Approaches*. Boston, MA: Allyn & Bacon.

Kaplan, R.B. (2010) Where to from here? In R.B. Kaplan (ed.) *The Oxford Handbook of Applied Linguistics* (2nd edn) (pp. 571–576). Oxford: Oxford University Press.

Kaplan, R.B. and Baldauf, R.B. (1997) *Language Planning from Practice to Theory*. Clevedon: Multilingual Matters.

Larsen-Freeman, D. (2015) Ten 'lessons' from complex dynamic systems theory: What is on offer? In Z. Dörnyei, P.D. MacIntyre and A. Henry (eds) *Motivational Dynamics in Language Learning* (pp. 11–19). Bristol: Multilingual Matters.

Lazaraton, A. (2005) Quantitative research methods. In E. Hinkel (ed.) *Handbook of Research in Second Language Teaching and Learning* (pp. 209–224). Mahwah, NJ: Lawrence Erlbaum Associates.

Pennycook, A. (2005) Critical applied linguistics. In L. Davis and C. Elder (eds) *The Handbook of Applied Linguistics* (pp. 783–807). Oxford: Blackwell.

Pfenninger, S.E. and Singleton, D. (accepted) Second language acquisition and age effects: The research conundrum. In K. Cergol-Kovačević and S.L. Udier (eds) *Applied Linguistic Research and Methodology: Proceedings from the CALS Conference 2015*. Frankfurt am Main: Peter Lang.

Schmitt, N. and Celce-Murcia, M. (2010) An overview of applied linguistics. In N. Schmitt (ed.) *An Introduction to Applied Linguistics* (2nd edn) (pp. 1–26). New York: Routledge.

Singleton, D. and Pfenninger, S.E. (2015) Insights from a mixed methods approach with respect to age and long-term instructed language learning. In J. Navracsics and S. Bátyi (eds) *First and Second Language: Interdisciplinary Approaches*. Studies in Psycholinguistics No. 6 (pp. 11–21). Budapest: Tinta Könyvkiadó.

Widdowson, H.G. (2001) Coming to terms with reality: Applied linguistics in perspective. In D. Graddol (ed.) *Applied Linguistics for the 21st Century. AILA Review* 14, 2–17.

Widdowson, H.G. (2005) Applied linguistics, interdisciplinarity, and disparate realities. In P. Bruthiaux, D. Atkinson, W.G. Eggington, W. Grabe and V. Ramanathan (eds) *Directions in Applied Linguistics* (pp. 12–25). Clevedon: Multilingual Matters.

Index

Note: Page numbers **in bold type** refer to main entries and definitions.