

# **Learner Corpora and English Acquisition**

A Collection of Studies

Edited by

Monika Černá, Jaroslava Ivanová and Šárka Ježková

---

Univerzita Pardubice 2015

Univerzita Pardubice

**Learner Corpora and English Acquisition.  
A Collection of Studies.**

Edited by

Monika Černá, Jaroslava Ivanová, Šárka Ježková

Pardubice 2015

The book was published with the support of the Czech Science Foundation project *Aspects of English Language Acquisition of Czech Students at the Onset of Teacher Education* (GA ČR 13-25982S).

© Univerzita Pardubice

© Lucie Betáková, Monika Černá, Bohát Róbert, Petr Dvořák, Christoph Haase, Libuše Hornová, Petra Huschová, Nina Horáková, Jaroslava Ivanová, Šárka Ježková, Jana Kořínková, Gabriela Lojová, John McKenny, Irena Reimannová, Maria Luisa Roca-Varela, Beata Rödlingová, František Tůma, Silvie Válková

Reviewed by: Doc. Mgr. Světlana Hanušová, Ph.D.  
Doc. PhDr. Eva Homolová, Ph.D.

**ISBN 978-80-7395-946-3**

# CONTENTS

LIST OF CONTRIBUTORS .....	7
LIST OF ABBREVIATIONS .....	9
PREFACE .....	12
INTRODUCTION .....	13
<b>I ANALYSIS OF L2 CORPORA .....</b>	<b>21</b>
CHAPTER 1 Some remarks on noun premodification in a Czech student corpus <i>Libuše Hornová</i> .....	25
CHAPTER 2 Diverse pronunciation results elicited by means of three different test types <i>Jaroslava Ivanová</i> .....	37
CHAPTER 3 Hesitation in learner's dialogues: distribution of individual means <i>Šárka Ježková</i> .....	47
CHAPTER 4 Academic writing: modal verbs in the role of hedges <i>Petra Huschová &amp; Irena Reimannová</i> .....	57
CHAPTER 5 Building and analysing an advanced English learners' written corpus <i>Silvie Váľková &amp; Jana Kořínková</i> .....	67
<b>II USE OF L2 CORPORA FOR L2 LEARNING .....</b>	<b>77</b>
CHAPTER 6 Strata of academic English in corpora: a parallel-genre approach <i>Christoph Haase</i> .....	81
CHAPTER 7 Building a Corpus of High School Academic Texts (COHAT) for academic English: Discovery learning in language acquisition <i>Róbert Bohát, Nina Horáková &amp; Beata Rödlingová</i> .....	91
CHAPTER 8 How a large corpus of dissertations can help teachers and learners of EAP writing <i>John McKenny</i> .....	105

CHAPTER 9	Unveiling vocabulary errors in non-native English	
	<i>Maria Luisa Roca-Varela</i> .....	117
<b>III</b>	<b>INDIVIDUAL DIFFERENCES IN L2 LEARNERS</b> .....	127
CHAPTER 10	An exploration of internal factors in foreign language learning	
	<i>Gabriela Lojová</i> .....	129
CHAPTER 11	Motivation in language learning: Focus on individual learners' stories	
	<i>Monika Černá</i> .....	143
<b>IV</b>	<b>CLASSROOM INTERACTION AND L2 ACQUISITION</b> .....	161
CHAPTER 12	Using conversation analysis in classroom interaction research	
	<i>František Tůma</i> .....	163
CHAPTER 13	Supporting language acquisition through teacher questioning	
	<i>Lucie Betáková</i> .....	173
CHAPTER 14	Supporting English acquisition through the teacher's social skills	
	<i>Petr Dvořák</i> .....	187
RESUMÉ	.....	199
REFERENCES	.....	203
INDEX	.....	223

## CONTRIBUTORS

Lucie Betáková, Faculty of Education, University of South Bohemia, České Budějovice, Czech Republic, [betakova@pf.jcu.cz](mailto:betakova@pf.jcu.cz)

Monika Černá, Faculty of Arts and Philosophy, University of Pardubice, Pardubice, Czech Republic, [monika.cerna@upce.cz](mailto:monika.cerna@upce.cz)

Bohát Róbert, International School of Prague, Praha, Czech Republic, [rbohat@isp.cz](mailto:rbohat@isp.cz);

Petr Dvořák, Faculty of Education, University of South Bohemia, České Budějovice, Czech Republic, [pdvorak@pf.jcu.cz](mailto:pdvorak@pf.jcu.cz)

Christoph Haase, Faculty of Education, University of J. E. Purkyně, Ústí nad Labem, Czech Republic, [christoph.haase@ujep.cz](mailto:christoph.haase@ujep.cz)

Libuše Hornová, Faculty of Arts and Philosophy, University of Pardubice, Pardubice, Czech Republic, [libuse.hornova@upce.cz](mailto:libuse.hornova@upce.cz)

Petra Huschová, Faculty of Arts and Philosophy, University of Pardubice, Pardubice, Czech Republic, [petra.huschova@upce.cz](mailto:petra.huschova@upce.cz)

Nina Horáková, International School of Prague, Praha, Czech Republic, [nhorakova@isp.cz](mailto:nhorakova@isp.cz)

Jaroslava Ivanová, Faculty of Arts and Philosophy, University of Pardubice, Pardubice, Czech Republic, [jaroslava.ivanova@upce.cz](mailto:jaroslava.ivanova@upce.cz)

Šárka Ježková, Faculty of Arts and Philosophy, University of Pardubice, Pardubice, Czech Republic, [sarka.jezkova@upce.cz](mailto:sarka.jezkova@upce.cz)

Jana Kořínková, Faculty of Education, Palacký University, Olomouc, Czech Republic, [jana.korinkova@upol.cz](mailto:jana.korinkova@upol.cz)

Gabriela Lojová, Faculty of Education, Comenius University, Bratislava, Slovakia, [lojova@fedu.uniba.sk](mailto:lojova@fedu.uniba.sk)

John McKenny, British University in Dubai, United Arab Emirates, [john.mckenny@buid.ac.ae](mailto:john.mckenny@buid.ac.ae)

Irena Reimannová, Faculty of Arts and Philosophy, University of Pardubice, Pardubice, Czech Republic, [irena.reimannova@upce.cz](mailto:irena.reimannova@upce.cz)

Maria Luisa Roca-Varela, University of Santiago de Compostela, Spain, [luisarocavarela@edu.xunta.es](mailto:luisarocavarela@edu.xunta.es)

Beata Rödlingová, International School of Prague, Praha, Czech Republic,  
brodtingova@isp.cz

František Tůma, Institute for Research in School Education, Faculty of Education,  
Masaryk University, Brno, Czech Republic, tuma@ped.muni.cz

Silvie Válková, Faculty of Education, Palacký University, Olomouc, Czech  
Republic, silvie.valkova@upol.cz

## ABBREVIATIONS AND ABBREVIATED WORDS

AF	absolute frequency
ALP	Applied Linguistics Project
AntConc	a freeware concordance programme by Lawrence Anthony
ASCII	American Standard Code for Information Interchange
AWL	the Academic Word List
B1	threshold, independent user according to CEFR
B2	vantage, independent user according to CEFR
BAWE	the British Academic Written English corpus
BNC	the British National Corpus
BUiD	the British University in Dubai
Buidcorp	the Corpus of Written Academic English (of the British University in Dubai)
C1	effective operational proficiency, proficient user according to CEFR
C2	mastery, proficient user according to CEFR
CA	conversation analysis
CAE	Cambridge Advanced Exam
CDO	Cambridge Dictionaries Online
CEFR	the Common European Framework of Reference for Language
CES	Czech speakers of English
CLAWS	the Constituent Likelihood Automatic Word-tagging System
CLC	computer learner corpora
CNC	the Czech National Corpus
COHAT	the Corpus of High School Academic Texts
concgram	all of the permutations of constituency variation and positional variation generated by the association of two or more words
CPH	critical period hypotheses
CSSE	the Czech Students' Spoken English (corpus)
EAL	English as an additional language
EAP	English for academic purposes
Edcorp	the Education corpus
EE	extended essay
EFL	English as a foreign language
ELF	English as a lingua franca
ELT	English language teaching
ESS	Environmental Systems and Societies
EU	the European Union

FCE	the First Certificate in English
FL	foreign language
FLA	foreign language acquisition
FLL	foreign language learning
FLT	foreign language teaching
GA ČR	the Czech Science Foundation (Grantová agentura České republiky)
GSL	the General Service List
IB	the International Baccalaureate
ICLE	the International Corpus of Learner English
ICT	information communication technology
InterCorp	a parallel multilingual corpus (a part of the CNC project)
IRE	initiation, response, evaluation
IRF	initiation, response/reaction, feedback/follow-up
ISP	the International School of Prague
IT	information technology
K	keyness
KWL	key word list
L1	mother tongue
L2	second language (sometimes including a foreign language)
LA	language aptitude
LGSWE	the Longman Grammar of Spoken and Written English
LPD	the Longman Pronunciation Dictionary
MICASE	the Michigan Corpus of Academic Spoken English
MICUSP	the Michigan Corpus of Upper-level Student Papers
MLAT	the Modern Language Aptitude Test
MT	mother tongue
MWEs	multiword expressions
n.d.	no date
NES	native English speakers
Ngram	a contiguous sequence of n items from a given sequence of text or speech
NNS	non-native speakers (of English)
NS	New Scientist; native speakers
PDF	Portable Document Format
pi	difficulty index
PLoS	the Public Library of Science
PNAS	Proceedings of the National Academy of Sciences of the USA
POS	part of speech
PPP	the present-practice-produce approach
ppt	parts per thousand
QUAL	qualitative
QUAN	quantitative
RF	relative frequency

rp	the Pearson correlation coefficient
SCT	sociocultural theory
SDG	Strategic Decision Group
SLA	second language acquisition
SPACE	the Scientific and Popular Academic Corpus of English
TEFL	teaching English as a foreign language
TEI	the Text Encoding Initiative
TESMC	teaching ESL students in mainstream classrooms
TESOL	Teachers of English to Speakers of Other Languages
TIA	typological interaction analysis
TL	target language
TOK	theory of knowledge
UAE	the United Arab Emirates
UCREL	the University Centre for Computer Corpus Research on Language, Lancaster
UJEP	Jan Evangelista Purkyně University
USA	the United States of America
USAS	the UCREL Semantic Analysis System
w1	words which are in the top 1,000 most frequent written English words
w2	words which are in the top 2,000 most frequent written English words
Z99	code indicating non-standard usage in a corpus

## PREFACE

The book is one of the outcomes of the project titled *Aspects of English Language Acquisition of Czech Students at the Onset of Teacher Education* and supported by the Czech Science Foundation. Within the framework of the project, the research team from the University of Pardubice organised the *English Acquisition and Corpora Building* conference, which took place at the University of Pardubice on June 4<sup>th</sup> 2015. Researchers from the Czech Republic and from abroad gathered to share the findings of their research into second language acquisition and corpora building. The papers that were presented in the aforementioned areas introduced a variety of theoretical approaches. On top of that, in some sections, the results that were presented revealed a high level of convergence and/or their topic focuses complemented each other to constitute a mosaic. In order to communicate such invaluable outcomes to a wider professional community, most presenters at the conference volunteered to contribute to this volume. The book you have just opened is its final product.

## INTRODUCTION

Second language acquisition (SLA) is an interdisciplinary field of study. Ellis (1997) defines SLA as “the way in which people learn a language other than their mother tongue, inside or outside of a classroom, and second language acquisition (SLA) as the study of this” (p. 3). Ortega (2009) expands the definition further in several aspects: she specifies the developmental phases of learners under investigation, i.e. late childhood, adolescence, and adulthood, and emphasises that “it [SLA] encompasses the study of naturalistic and formal language acquisition in second, foreign and heritage learning contexts” (p. 10). Our standpoint corresponds with that of Ortega, who views a foreign language as a subcomponent of SLA, which is a traditional umbrella term for the field of study. We are, however, aware of the difference between SLA and foreign language learning (FLL) in concord with Gass and Selinker (2008, p. 7); they propose that second language acquisition refers to the learning of a non-native language in the environment in which that language is spoken; the learning may or may not take place in the classroom setting. Foreign language learning, which most commonly happens in the classroom, refers to the learning of a non-native language in the environment of one’s native language (Gass & Selinker, 2008). The significant difference between the two types of learning is the access to the community of the target language users, which is substantial in the second language context. Contrary to that, in the foreign language context opportunities to interact with speakers of the target language are very limited unless they are actively searched for. They can be direct or mediated by means of various media and ICT. This dissimilarity of the two contexts, i.e. the naturalistic and formal ones, leads some scholars to emphasise that we should carefully consider the relevance of SLA research findings to FLL before making inappropriate overgeneralisations (Lojová, 2015, this volume). If viewed from the perspective of the proposed distinction, the individual studies included in the book fall into the area of FLL. Thus the book enlarges the body of FLL research, i.e. a less researched area within SLA.

There is a consensus in the SLA literature that most of the research in the field has been targeted at the English language as acquired by adult learners. Moreover, those learners often represent larger communities of English users, such as the Chinese, Japanese, and Poles, for instance. Therefore, the volume enriches the field by uncovering some aspects of the English acquisition of, especially, Czech learners, but also those from Spain and the UAE. Regarding the age of the learners, the studies in this book investigated almost exclusively adult learners; one exception is the study by Bohát et al., who focus on adolescents.

The goals of SLA involve both the description of processes and various ways of explaining them, i.e. the identification of internal and external factors that account for ways in which SLA can be realised (Ellis, 1997, p. 4). When searching for a framework to structure the content of the book, the editors found *A framework for investigating L2 acquisition* (Ellis, 2008, p. 34) relevant. The framework presented in Table 1 divides the “amorphous field” (Ellis, 2008, p. 3) of SLA into seven areas which fall into three broad subfields: *description of learner language*, *explanation of learning*, and *instructed SLA* (Ellis, 2008, p. 34). The chapters collated in this book target all three subfields, though not all seven areas. Five chapters belong to the first subfield, *description of learner language*, which equals area 1: *Characteristics of learner language*. The second subfield, *explanation of learning*, includes areas 2-5; it is area 4: *Inter-learner variability*, which is the focus of two authors. The third subfield, *instructed SLA*, involves areas 6-7. Three chapters aim at area 6: *Inside the “black box”*, i.e. classroom interaction and L2 acquisition, and four chapters at area 7: *Intervening directly in interlanguage*. We are fully aware that this is a crude categorisation, as we can identify a number of overlaps in some studies.

Table 1  
*A framework for investigating L2 acquisition (Ellis, 2008, p. 34)*

A framework for investigating L2 acquisition						
	General SLA					Instructed SLA
Description of learner language	Evaluation of learning					
Area 1	Area 2	Area 3	Area 4	Area 5	Area 6	Area 7
Characteristics of learner language	Learner external factors	Psycholinguistic processes	Inter-learner variability	The brain and L2 acquisition	Inside the ‘black box’	Intervening directly in interlanguage
Errors	Input and interaction	L1 transfer	Individual differences in L2 learners	Neurolinguistic accounts of L2 acquisition	Classroom interaction and L2 acquisition	Form-focused instruction and L2 acquisition
Acquisition order and developmental sequences	Social accounts of L2 learning	Cognitive accounts of L2 acquisition				
Variability		Cognitive accounts of L2 use				
Pragmatic features of interlanguage		Sociocultural accounts of L2 acquisition				
		Linguistic accounts of L2 acquisition				

The distribution of the chapters reflects the body of research in the respective areas. The research in the area of individual differences in L2 learners is still marginalised in SLA because of the prevailing concern with more universalistic aspects of acquisition (Ellis, 2008, p. 643) or, in the case of motivational research, because of the fact that its outcomes are not related concretely and in detail to the processes of second language development (Cook & Singleton, 2014, p. 102).

Obviously, the book also acknowledges the contribution of corpus linguistics to SLA since Sections 1 and 2 group together nine corpus-based studies where either the research is focused on learner language or the corpus analysis serves as a starting point for the design of data-driven learning strategies, or the authors combine both approaches. Various language corpora benefit from advances in computer technology (Hummel, 2014, p. 159), which enable researchers to approach learner language as an invaluable source from different perspectives and for an array of purposes. Over the last few decades, research on learner corpora has focused mostly on five major areas: “the nature of L2 knowledge and cognition, the nature of learner language, the role of L1, contribution of the linguistic environment and the role of instruction” (Chau, 2012, p. 192).

L2 language learner databanks can be viewed in terms of specific information about the learner and the communication situation, and as a comprehensive source of information examining the processes of L2 learning (Hummel, 2014, p. 159). With respect to specific information about the learner, the following aspects are taken on board: *language background* (L1=Czech for the majority of the learners, Arabic, Spanish, English native-speakers); *proficiency of FL learners* ranging from B1 to C2; *age*: learners older than 15 years of age; *sex*, and *exposure to TL*, i.e. English is used as a medium of education in other content subjects by native or non-native speakers of English or it is used “merely” as an FL.

With regard to the communication situation, data collection conditions such as “how, when, and where the data were collected” (Hummel, 2014, p. 159) play a crucial role in all corpora studies, which is further specified in the individual chapters with respect to university or secondary school settings. Needless to state, however, the prevailing majority of the learners are university students majoring in English in the FL environment in the Czech Republic. In this volume there are examples of both written, mostly academic writings (the chapters by Válková and Kořínková, Huschová and Reimannová, Hasse, and McKenny), or partially academic writings (the chapter by Bohát et al.) and spoken corpora (the chapters by Ivanová, Ježková, Hornová, and Bohát et al.). Various computational tools were used to analyse the data, e.g. Coh-Metrix 3.0 (the chapter by Válková and Kořínková), Wordsmith Tools 6 (the chapters by McKenny, Hornová, and Ježková), Wmatrix3 (the chapter by McKenny), and AntConc – version 3.4.3 (the chapter by Bohát et al.).

The data for the analysis vary from larger international corpora (argumentative essays from the Spanish subcomponent of the International Corpus of Learner English,

representing the B2, C1, and C2 levels according to the Common European Framework of Reference (CEFR, 2001), by Roca-Valera; the Corpus of Scientific and Popular Academic English by Hasse) to smaller specialised ones which include both written (the set of writings of C1 university students by Válková and Kořínková; the sample of bachelor theses by Huschová and Reimannová; the Corpus of High School Academic Texts by Bohát et al.; the Corpus of Written Academic English and British Academic Written English at C1 and C2 levels from the British University of Dubai by McKenny) and spoken language (the Czech Students' Spoken English (CSSE) corpus by Ivanová, Ježková, and Hornová).

A non-native learner corpus is usually analysed on the basis of comparison with a reference corpus (Ellis, 2008), be it, for example, a dictionary (the Longman Pronunciation Dictionary – Ivanová's study), a grammar book (the Longman Grammar of Spoken and Written English – Hornová's and Ježková's research), writing sections from coursebooks (New Headway Advanced, Landmark Advanced, Fast Track to CAE – Válková and Kořínková's project), or British Academic Writing English (the study by McKenny).

In relation to corpus data as a source of information, the chapters are clustered around two major areas: (a) learner language research – in many cases contrasted to native language research – mostly examining the frequency and distribution of various language structures; (b) corpus findings being used as a source of implications for teaching. The former area is represented by five studies analysing the language material from various viewpoints (the use of lexico-grammatical structure, pronunciation, discourse management, and pragmatics). The latter aspect is reflected in four chapters in which, on the basis of corpus analysis, the authors suggest strategies for form-focused instruction. Logically, there are numerous overlaps and combinations of both approaches at many points.

All the chapters in this book are divided into four sections according to the predominant topic, starting from the learner corpora research, mostly using interlanguage contrastive analysis, through the studies of corpora which propose various implications for teaching, proceeding to individual factors influencing the processes of SLA, and finishing with cases of exploration of classroom interaction.

Section One presents studies of characteristics of learner language and illustrates “an important link between the two previously disparate fields of corpus linguistics and foreign/second language research” (Granger, 2002, p. 4). In accordance with the claim by Farr (2010, p. 624), all the chapters attempt to contribute to the pedagogic applications of corpus linguistics, mainly a contrastive linguistic examination, as corpus-based techniques not only help us identify semantic and syntactic errors, but also reveal more and less frequent usages compared to NS ones. All the contributions in this section investigate L2 acquisition by collecting and describing samples of learner language. Ellis (1997) proposes three main ways of describing learner language:

(a) “focus on the kinds of errors learners make”, (b) the identification of “developmental patterns by describing the stages in the acquisition of particular grammatical features”, (c) the examination of “variability ... in learner language” (p. 15).

In the chapter *Some remarks on noun premodification in a Czech student corpus* Libuše Hornová investigates university students’ spoken performance at B2 level from the lexical point of view, that is, their ability to use primarily premodifying elements, i.e. adjectives, nouns, -ing and -ed particles, and gerunds in complex noun phrases, while postmodifying elements are of marginal interest. Jaroslava Ivanová’s chapter *Diverse pronunciation results elicited by means of three different test types* is devoted to an inspection of a very narrow field of pronunciation acquisition: the appropriate and inappropriate use of seven segmental features by B2-level Czech university students. The chapter by Šárka Ježková, called *Hesitation in learner’s dialogues: distribution of individual means*, is concerned with specific features of spoken learners’ performance, specifically the use of various strategies for gaining time while hesitating in conversation. She analyses the same language material as Hornová and, partially, Ivanová, i.e. the first half of the CSSE corpus, which represents transcribed recordings of 112 first-year university students, including monologues and dialogues. In their chapter *Academic writing: modal verbs in the role of hedges* Petra Huschová and Irena Reimannová discuss the role of modal verbs as hedges and university learners’ ability to identify modal verbs as hedges and use them accordingly in academic texts. The following chapter, by Silvie Válková and Jana Kořínková, entitled *Building and analysing an advanced English learners’ written corpus*, constitutes a corpus-based study focusing on certain aspects of learner production. It deals with different aspects of *grammatical and lexical complexity*, i.e. the usage of syntactic complexity (sentence length and complexity, syntactic similarity, and the occurrence of connectives), finite and non-finite verb forms, and the noun phrase in a written corpus.

Section Two groups together the chapters whose authors utilised the results of their corpus-based studies and suggested specific interventions in interlanguage development through form-focused instruction. The interventions include proposals for: (a) an approach to teaching some pragmatic aspects of English, such as academic versus popular academic genre (Hasse); (b) specific learning activities aimed at selected aspects of learners’ communicative competence (Bohát et al.); (c) a two-fold focus of using the Corpus of Written Academic English – first, by advanced learners in writing their dissertations, and second, by FL learners, here called language detectives, learning to use corpus tools (Wmatrix3, Wordsmith Tools 6) in order to explore corpora (McKenny), and (d) a type of lexis-learning-oriented activity (Roca-Varela). The suggested interventions undoubtedly have the potential to enrich language pedagogy; nevertheless, their effectiveness in the classroom has not yet been investigated.

The chapter *Strata of academic English in corpora: a parallel-genre approach* by Christoph Haase concentrates on the comparative analysis of two TL genres: *academic* versus *popular academic texts* from the point of view of linguistic parameters, specifically idiomatics and hedging. The study of the Corpus of High School Academic Texts by Róbert Bohát, Nina Horáková and Beata Rödlingová described in the chapter with the title *Building a corpus of high school academic texts (COHAT) for academic English: discovery learning in language acquisition* takes advantage of the fact that it comprises both spoken and written English produced by native and non-native speakers of English. It shows how it can be used not only for the identification of errors in interlanguage, but also for task-based learning aiming at the development of students' critical thinking and metacognitive skills, e.g. dealing with the occurrences of the personal pronoun "I" in written and spoken discourse. In his chapter *How a large corpus of dissertations can help teachers and learners of EAP writing* John McKenny examines lexical aspects in two types of academic corpora from Dubai University, the most frequent content words, the top keywords, and word bundles (e.g. *a lot of, the gifted and talented*), and multiword expressions. The chapter by Maria Luisa Roca-Valera, *Unveiling vocabulary errors in non-native English*, has a much narrower focus on lexical fields and inspects both the appropriate and inappropriate use of high-frequency English false friends by Spanish learners.

Section Three involves two chapters dealing with individual differences in L2 learners. After the sections uncovering the characteristics of learner language and proposing interventions in interlanguage development, this section contributes to the explanation of learning. In her chapter *An exploration of internal factors in foreign language learning*, Gabriela Lojová reviews major determinants of the effectiveness of foreign language learning. In the following chapter, *Motivation in language learning: Focus on individual learners' stories*, Monika Černá concentrates on one of the core factors which account for individual learner differences. More specifically, she explores the dynamic nature and temporal variation of English language learners' motivation.

Section Four contains three chapters dealing with classroom interaction and L2 acquisition, which is perceived by Ellis (2008) as an area within instructed SLA; he perceives interaction as processes which might either facilitate or limit various opportunities for language learning. Therefore, he views interaction in language classrooms as a "black box" (p. 35). By focusing on selected aspects of classroom interaction the three authors in the section throw some beams of light into the darkness of the "black box".

In the chapter *Using conversation analysis in classroom interaction research*, František Tůma approaches the topic from the theoretical perspective, outlining a research design grounded in dialogist approaches with conversation analysis as a research method. The other two chapters introduce outcomes of research into the

field. In her chapter titled *Supporting language acquisition through teacher questioning*, Lucie Betáková explores a question-answer mode of classroom interaction in relation to student language acquisition. *Supporting English acquisition through the teacher's social skills* is the title of the chapter by Petr Dvořák, who also examines teacher-student interaction but from a different point of view. He investigates the impact of teachers' social skills on L2 acquisition.



## Section One

# ANALYSIS OF L2 CORPORA

All five chapters in this section are focused on research into English learners' language. Although the approaches of the individual authors differ and they analyse corpora from different points of view (i.e. pronunciation, lexico-grammatical complexity, pragmatics, discourse analysis, and syntax), all the studies have a common intended aim: to modify the curricula of TEFL study programmes on the basis of the findings and thus to help students develop their communicative competence both in writing and speech.

The first chapter, by Libuše Hornová, is concerned with the assessment of the variety of lexical items in the adjective position in complex noun phrases. The research is based on a relatively small corpus of language produced by first-year students in university TEFL programmes (i.e. the first half of the CSSE corpus, which is one of the outcomes of a bigger multidisciplinary project called "Aspects of the English Language Acquisition of Czech Students at the Onset of Teacher Education"). It revisits some conclusions drawn previously from the analysis of this corpus, where one of the outcomes was an indication that the students' vocabulary was on an unsatisfactory level. She mainly focuses her attention on the premodifying elements of English head nouns in complex noun phrases – adjectives, nouns, and -ing and -ed forms. The results provide partial information about their occurrence in the complement, too. The frequency of individual lexical items in the corpus is compared with the corpus findings presented in the Longman Grammar of Spoken and Written English (Biber et al., 1999). Postmodification of nouns and the complexity of the noun phrases that were collected are considered only superficially, but can contribute to the intended practical goal of the analysis – to improve the vocabulary and structure of English noun phrases in the English discourse of Czech students.

In the following chapter Jaroslava Ivanová analyses the oral performances of the same group of students as mentioned above, though from a different aspect. The research question is whether there are any significant differences in the pronunciation of selected segmental features of 112 Czech foreign language learners elicited via three different subtests: a read-aloud 153-word text, a read-aloud list of 24 separate

words, and a sample of spontaneous speech. The English segmental features focused on are the front open vowel ash (*apple*), the weak central mid vowel the schwa (*again*), the voiced and voiceless dental fricatives (*then*, *threw*), the bilabial approximant (*away*), the velar nasal (*long*), and the pronunciation of word-final voiced consonants (*bag*) and (*bed*). The difficulty index ( $p_i$ ), which tells us the percentage of test takers whose answers are correct (Bachman, 2004, p. 140), is calculated. The overall average difficulty indices for three different tests are correlated (Pearson correlation coefficient). The results suggest that Czech learners receive lower success scores for a read continuous text than for a carefully pronounced list of words, and the least successful score for spontaneous speech. This is confirmed by other research studies with respect to Czech learners, although there might be differences in the case of individual phonemes, e.g. the word-final /d/.

As Šárka Ježková states in the third chapter of this section, the structure of spontaneous spoken discourse is significantly influenced by the principles of online production, which means that participants are concurrently required to keep talking, to plan ahead, and to consider what has already been said. Such a situation is reflected in a certain degree of dysfluency, which has an impact on specific features of the grammar of speech. Similarly, this chapter presents partial results of the same research project and works with the same corpus as that analysed by Hornová and partially by Ivanová. The study is focused on the distribution of various strategies that students use when they hesitate about what to say next. The categorisation of individual inserts and hesitators is based on the findings of Biber et al. (1999), who analysed native speakers' (NS) speech. The figures of the analysis reveal that in some situations Czech learners of English apply rather different strategies and do not use those more or less formulaic structures observable in NS conversation. It may be concluded that the distribution of individual hesitators could be influenced by the level of students' communicative competence in EFL, by the lack of instruction clarifying the difference between speech and writing, and in some cases also by the influence of L1.

After the first three chapters analysing students' production of spoken language, the picture is extended by the next two chapters, which are focused on analyses of written texts. That by Petra Huschová and Irena Reimannová is concerned with the use of modal verbs as hedging devices in bachelor's theses written by Czech students of English. Since modality is viewed as "the most important concept that cuts across the area of hedges" (Markkanen & Schröder, 1997, p. 6), i.e. expressions qualifying and toning down utterances "in order to reduce the riskiness of what one says" (Wales, 2001, p. 185), the chapter investigates whether and how students apply their explicit knowledge (Gass & Selinker, 2008; Ellis, 2005) concerning English modal verbs in the role of hedges in academic writing. Thus, the aim is to explore the acquisition and application of the particular pragmatic knowledge in the context of formal education. For the purposes of their study they used questionnaires to analyse

the processes of explicit knowledge formation, while in order to examine students' implicit knowledge they analysed the occurrence and use of modals in the discussion sections of selected final papers. The results indicate that the learners do not seem to fully exploit the pragmatic potential of modals in the role of hedges when presenting and interpreting the findings of their studies.

The last chapter in this section, by Silvie Válková and Jana Kořínková, sums up the preliminary results of a long-term project aiming at determining specific needs in teaching advanced English students at the Faculty of Education, Palacký University. At higher levels of proficiency, the progress of foreign language students is believed to be slower and difficult to measure. There is a tendency to make “quantitative” mistakes, i.e. to use certain foreign language items with a different frequency than that found in authentic natural texts. Through a series of primarily quantitative analyses of various features of advanced English students' writing, the authors aspire to make their teaching more focused and hence more efficient. In order to write smoothly flowing coherent texts, advanced students are expected to demonstrate both a wide range of vocabulary and efficient control of complex structures with clear and logical clause relations. The students' writing, however, often fails to manifest these qualities; students favour simple structures, with a strong preference for coordinated clauses and finite verb forms. The text parameters explored so far include syntactic complexity, the usage of finite and non-finite verb forms, and noun phrase complexity. The chapter explains the nature of their corpus together with the methods they use for its analysis, including the Coh-Metrix 3.0 computational tool.



## Chapter 1

# SOME REMARKS ON NOUN PREMODIFICATION IN A CZECH STUDENT CORPUS

*Libuše Hornová*

## 1 Introduction

Most researchers (e.g. Aijmer, 2009; Ellis & Barkhuizen, 2005; Granger, 2009) agree in their publications that learner corpora contribute to the improvement of methods in second language acquisition (SLA), bring a lot of information about typical structures and errors in the spoken and written language of foreign learners and also provide authentic material where the influence of the L1 can be identified.

This chapter tries to contribute to the improvement of the students' vocabulary level in their process of language acquisition through some research results based on the analysis of a learner corpus. The research team of the Department of English and American Studies at the University of Pardubice, with the help of colleagues from other universities, compiled a corpus of spoken English language of first-year university students (CSSE), which is the source material for various types of discourse analysis. I used this corpus before (Hornová, 2015) to find and evaluate the usage of syntactic functions of non-finite verb forms. Having spent a lot of time on a detailed analysis of the texts, I had the impression that the variety of lexical items in some structures is rather poor. Therefore, I decided to use the corpus again and concentrate on the students' vocabulary, more specifically, on the structure of complex noun phrases.

I compiled a survey of different premodifiers of head nouns in noun phrases, which includes adjectives, nouns, -ing and -ed participles and gerunds. The analysis incorporates some comments on the complexity of the noun phrases in the corpus, which might help introduce some innovations in practical language classes of university students majoring in English.

## 2 Description of the corpus

The learner corpus analysed here consists of a transcription of the spoken English discourse of 112 first-year students in the TEFL programmes of three Czech universities,

whose knowledge should correspond to the B2 level of the CEFR (2001). Their command of the English language was not influenced by their university study programme; it only reveals the knowledge and skills they acquired during their pre-university studies. The entire body of the corpus material was recorded in the autumn of 2013 and was later transcribed by the students participating in the recordings. The corpus comprises about 75,000 words and includes three parts: (a) a monologue where the students introduce themselves; (b) two-way dialogues between pairs of students who need some information that is provided by their partners; (c) a discussion between the same pairs of students on given topics. This corpus is not very large, but I agree that "... the SLA specialist attaches more importance to control over the many variables that affect learner production than to sheer size" (Granger, 2009, p. 17).

### **3 Aims and methods**

The first research task was to find if the students' vocabulary in the corpus is appropriate for the level of their English. This chapter does not try to perform a complete lexical analysis of the corpus, but only aims at presenting some partial findings, focusing on the structure of noun phrases and partly the complement. "Nouns are of pivotal importance in languages like English. It has been observed that along with verbs they are a dominant part of speech, and that the semantic content of sentences is borne mostly by nouns" (Algeo, 1995, p. 203).

In the syntactic positions of the premodification of nouns and in the complement, adjectives, nouns, participles (-ing and -ed) and gerunds appear repeatedly. Apart from that, they can also be found as non-finite condensed clauses or prepositional phrases in the postmodification of head nouns. "Adjectives and nouns are very common in all registers and certainly add to the informational density of the text" (Biber et al., 1999, p. 504).

Considering these well-known grammatical facts and the limited space available in this chapter, I decided to focus on the premodifying elements of head nouns in noun phrases (partly complement and postmodification) and consider the frequency and lexical variety of adjectives, nouns and participles and gerunds in such syntactic positions so as to be able to at least superficially assess the vocabulary level of the students participating in the corpus of spoken English that was used. The second aim of this chapter is to compare the structure of the noun phrases in this corpus with the corpus findings presented in the Longman Grammar of Spoken and Written English (LGSWE, Biber et al., 1999).

Finally, the research results are meant as a contribution to possible additions to the process of language acquisition in the advanced study of the English language. The lexical items found in the noun premodification in the corpus were compiled and counted manually and the figures are presented in the tables below. The occurrence

of individual premodifying elements was then checked with the programme WordSmith Tools (Scott, 2012). In order to evaluate the students' ability to form complex noun phrases, some examples of long noun phrases were collected.

#### 4 Adjectives in the premodification of head nouns (and complement)

“The whole object and purpose of language is to be meaningful ... It is because the needs of human communication are so various and so multifarious that the study of meaning is probably the most difficult and baffling part of the serious study of language” (Crystal, 2015).

Adjectives (and other lexical items used with that function) in various syntactic positions within the sentence structure certainly represent a word class that tells us a lot about the level of the students' lexical acquisition of the given foreign language. Therefore their occurrence within noun phrases and complements is analysed.

Table 1.1

*The occurrence of adjectives within noun phrases and complements*

	Total in corpus	Premodification
good (job, way, time, flat, salary)	257	87
great (site, experience, memories, etc.)	96	35
new (student, life, experience, etc.)	86	62
small (village, city, town, child, etc.)	69	66
big (family, problem, opportunity, etc.)	53	41
same (thing, course, school, etc.)	35	22
social (network, life, internet, etc.)	31	31
bad (idea, thing, habit, etc.)	28	15
international (course, stay, student, etc.)	25	19
difficult (grammar, version, examination)	19	3
foreign (students, education, course, etc.)	16	16
special (offer, time books, etc.)	15	11
young (people, learners, children, etc.)	14	14
possible (problem, work)	30	2
educational (institution, etc.)	11	4
Total	785	428
quiet, cold, religious, different, main, positive, low	only occasional usage	

Adjectives can be divided into descriptors and classifiers. Descriptors include those connected with colour, quantity and size (*big, small, heavy, etc.*), time (*young, new, recent, etc.*), evaluation (*good, great, lovely, nice, etc.*) and other ways of describing a given thing, idea or person (*positive, practical, free, strange, etc.*).

Classifiers usually specify the relation, classification or affiliation of the noun being modified to other referents (*additional, previous, similar, American, etc.*). They can also relate the noun to a certain topic (*commercial, political, phonetic, international,*

etc.). According to the context and the syntactic position of adjectives, the same lexical items can be sometimes used as descriptors and at other times as classifiers.

Descriptors are found in all four registers (conversation, fiction, news and academic texts) to a certain extent. Classifiers prevail in informational written registers.

In conversation descriptors have been found to be more common than classifiers in the attributive position, characterising size, time and evaluation, e.g. *big, little, new, old, good, nice* (over 200 per million words). The classifier *same* shows a similar frequency (Biber et al., 1999, pp. 508-513; Huddleston & Pullum, 2002).

Before I start evaluating the frequency of the premodifying adjectives in the present corpus, the conversational topics of the students should be mentioned. The students involved in the conversation were seeking some information, e.g. part-time job opportunities in their university town or abroad, sports facilities in the town, ISIC cards and their advantages, study stays abroad, the advantages and disadvantages of studying in their home town or far from home, plans for their future professional life, etc. Such dialogues are usually more transactional than interactional (McCarthy, 1991) and, therefore, it can be expected that the choice of lexical items used in such a conversation would be heavily influenced by the topic.

In spite of that, the first positions in the statistical counting correspond to the LGSWE corpus findings of descriptors and classifiers. The frequency of *good* is even far higher than that, which cannot be considered a positive quality of the students' spoken discourse. The adjective *good* is always a quick choice if other synonymous expressions of evaluation do not come to mind. The same is true of *great*, which functions as a descriptor of size as well as of evaluation in the corpus. With regard to the meaning of these two English adjectives, there is a wide choice of others which could have replaced *great* or *good* in that context.

The statistical figures summarised in Table 1.1 also give evidence of the occurrence of these adjectives in the predicative position. If we deduct the number of adjectives in premodification from their total occurrence, the rest is their frequency in the complement. The most numerous adjectives, *good* and *great*, are more commonly found in the complement, almost always following the verb *to be*, and only occasionally the verb *sound*. This fact supports the negative assessment of their frequency, because the sentences express the simplest way of communicating something about the subject.

Topical adjectives (*international, foreign, social, etc.*) occur repeatedly in the corpus, giving evidence of the students' ability to discuss specialised subjects/topics without problems or mistakes. Some adjectives, however, do not collocate with the head nouns, e.g. *foreign education, international student, etc.* The overall rating of the adjective frequency is not very positive, even if the numbers are close to those in the British corpus. The relatively high figures give us information about the students' lack of active knowledge of the variety of synonyms.

## 5 Nouns as premodifiers in noun phrases

Nouns as premodifying elements are rather common in the structure of English noun phrases, but mostly in written discourse, especially in news. In conversation they are rare. The LGSWE corpus findings introduce only four nouns in this register which are relatively productive: *car*, *Christmas*, *school* and *water* (Biber et al., 1999, p. 592).

Table 1.2

*The occurrence of nouns in premodification of head nouns and other syntactic functions*

	Total in corpus	Premodification
school (leaving exam)	246	16
university (halls, students, etc.)	209	38
language (exam, school, barrier, etc.)	99	46
student (sale, loan, agency, etc.)	77	38
sport(s) (centre, day, school, event, etc.)	48	28
part-time (job)	38	12
semester (course)	34	22
office (administrator)	25	10
summer (camp, course, etc.)	24	13
street (promoter, work)	22	11
Total	822	234
home, seminar, level, weekend, etc.	rare in premodification	

The majority of the nouns in the premodification of head nouns that were collected from this corpus are clearly influenced by the conversational topics. The noun *school* as a premodifier is not common here (unlike in the LGSWE findings), but in other positions it appears the most frequently of all the nouns in the whole text. Its major syntactic position is as the head noun of noun phrases with premodifiers such as *basic*, *secondary*, *grammar*, *high*, *primary*, etc. Sometimes it is used only with articles. The same is true of the noun *university* and *language*. Other syntactic functions of the nouns that were collected are typically objects, subjects and complements or parts of prepositional phrases used as adverbials.

In premodification, there is the noun *university*, which, in this context, can be expected and is used more often than *school*. The frequency of the two nouns taken together in both premodification and other positions is quite high. Almost all the nouns occurring in premodification and other syntactic positions in the corpus, e.g. *school*, *university*, *language*, *student*, *sport*, *semester*, whose frequency within the 75,000 words is rather high, fall into the same lexical field. Even if the variety of the premodifying nouns is not exactly outstanding, we should evaluate such structures in a very positive way, because nouns are not used in the premodification of other nouns in the students' L1. Thus, we can say that the students managed this feature of the structure of the English noun phrase quite well. As for the lexical

variety of nouns – both in premodification and in the other syntactic positions mentioned above – it is not very satisfactory.

## 6 Gerund in the premodification of nouns

The gerund as one of the non-finite verb forms does not exist in the students' L1 and, moreover, its position in the premodification of nouns is considered relatively unusual (Biber et al., 1999) and thus it is taught to students as occurring less frequently in such a position than the -ing participle.

Table 1.3

*The occurrence of gerunds in the pre- and postmodification of head nouns*

	Total in corpus	Premodification
working (hours, time, etc.)	62	23
teaching (assistant)	61	18
learning (material, etc.)	61	9
swimming (pool, course, etc.)	51	14
leaving (examination)	15	14
dining (hall)	11	11
shopping (centre, mall, etc.)	11	7
driving (licence)	5	5
housing (requirement, etc.)	4	4
Total	281	105
dancing, opening, diving, reading	less than four in the corpus	

Its high frequency of occurrence in premodification in the corpus examples is rather surprising unless the topic of conversation is taken into consideration. The result is the necessity of using the -ing forms that clarify the purpose or aim of the noun that is used (*swimming pool, leaving examination*) rather than any other meaning connected with the -ing forms.

When using -ing participles, the speakers inform the listeners about something that the head noun is doing (what is happening with it). In other words, noun phrases with -ing participles before the head noun are more common in a descriptive discourse. The students' ability to distinguish between the two functions of -ing forms should be acknowledged, because it provides information not only about the lexical items that have been acquired, but also about the students' knowledge connected with the grammatical features of the gerund and the -ing participle and their syntactic functions.

On the other hand, the very low total numbers of examples of the usage of -ing forms (both gerunds and participles), e.g. *working, learning, teaching*, is a significant sign of rather a low vocabulary level in the context of the corpus. The

figures that were collected in the computer review of the text actually cover the total usage of the given lexical items, including verb phrases expressing progressive aspect (-ing participles) and objects, subjects and complements (gerunds) and therefore the resulting numbers are not positive at all.

This partial conclusion should encourage teachers to place emphasis on the improvement of the students' skills in this field. It might also be valuable to repeat the analysis of the -ing forms as lexical items in different corpora with varied topics.

Gerunds in this corpus were also found in the postmodification of nouns (48 instances), following prepositions – *studying, teaching, working, travelling, buying, visiting*, and not many others. Apart from *teaching, studying* and *working*, their usage is only individual, e.g.:

- 1) ... *did you mentioned some possibility of travelling?* (p. 120)
- 2) ... *so I am really really ... afraid of working with ...* (p. 126)

Postmodifying participles, according to the LGSWE findings (Biber et al., 1999, pp. 631–632), are most common in academic prose with the -ed participle prevailing. In the corpus that was analysed there were about 72 postmodifying participles (both -ing and -ed), e.g.:

- 3) *I would have to have er somebody living with me, ...* (p. 6)
- 4) *I have a friend named, er, John Smith ...* (p. 70)

The examples of -ed participles are more numerous than those of -ing participles (-ed, 56, -ing, 16) and they are represented by lexical items such as *called* (22), *based* (11), *planned* (6) and a few others (*offered, filled with, related*, etc.) occurring only randomly. The number of postmodifying -ing participles is 16, covering items as *depending, living, concerning, including* and *telling*, whose incidence is 3 (the maximum number) to 1.

## 7 -ing and -ed participles in the premodification of nouns

The -ing and -ed participles mentioned in Tables 1.4 and 1.5 (below) do not only stand in the adjective position, but many of them display the grammatical characteristics adjectives. I did not consider this fact important for my research aim and thus they are presented here as a group of lexical items with the -ing or -ed suffixes in an adjective position, which usually either involves the premodification of nouns or is in the complement.

Apart from the two most frequently used participles (*interesting, interested*), the majority of the others in premodification are topic-related, e.g. *speaking* or *instructing* and *advanced* or *accepted*.

Table 1.4

*The frequency of -ing participles in the corpus*

	Total in corpus	Premodification
interesting (experience, course, etc.)	58	14
speaking (lesson, exercise, etc.)	37	12
entertaining (events, people, etc.)	7	4
exciting (work, etc.)	4	3
instructing (course)	2	2
following (week, year)	2	2
Total	110	37
outgoing, upcoming	insignificant occurrence	

Table 1.5

*The frequency of -ed participles in the corpus*

	Total in corpus	Premodification
advanced (students, exam, user, etc.)	19	6
accepted (card)	9	5
written (test, part, etc.)	9	4
required (documentation)	7	5
underprivileged (children)	7	7
limited (access)	6	4
spoken (language, material)	3	3
filled (form)	3	2
licensed (course)	2	2
Total	65	38
selected, motivated, prepared, complicated, equipped	less than two in the whole corpus	

The -ing and -ed participles found in the corpus occur more frequently in the subject complement position than in the pre- or postmodification of head nouns. The number of lexical items expressed by -ed participles in this syntactic position is slightly over 100 examples (apart from *interested* – 24), with a variety of other expressions, e.g. *welcome* (15), *located* (7), *equipped* (6) and *situated* (7) and many others with a very low occurrence (1-3), e.g. *advanced*, *satisfied*, *limited*, *prepared*, *reconstructed*, *well-known*, *exhausted*, *misused*, which illustrates the usage of a few individual students, not the vocabulary level of the whole group.

The number of -ing participles in the complement is low (42) and the participle/adjective *interesting* (22) is persistent. Other participles with sporadic usage are *annoying*, *motivating*, *entertaining*, *tiring*, *exciting* and a few others. Thus, the desired lexical diversity of the collected examples of all the -ing and -ed forms as lexical items in pre- and postmodification, as well as in the complement, cannot be considered satisfactory.

## 8 Complex noun phrases in the corpus

According to the LGSWE, long noun phrases with more than one premodification and complex postmodification is very rare in conversation (Biber et al., 1999, pp 597-606). In particular, multiple premodification is not frequent, because it is difficult for the listener to identify clearly the relations between the individual premodifying elements.

The typical postmodifying elements of head nouns are:

- a) prepositional phrases, sometimes in combination;
- b) postmodifying infinitive and participle (-ing , -ed) clauses;
- c) finite relative clauses;
- d) finite/non-finite appositive clauses.

These types of postmodification are not particularly common in any register and even less frequent in spoken discourse.

Surprisingly, in the present corpus very complex noun phrases could be found, with several items in both pre- and postmodification. Here are some examples, which are presented without any corrections:

- 5) *two original identification documents to prove your address* (p. 43)
- 6) *this opportunity to eat there for some couple of coins* (p. 48)
- 7) *high level English language skills in academic and professional environment* (p. 59)
- 8) *ladies street dance for girls and women in Dance Studio 341 environment* (p. 60)
- 9) *international children choir competition* (p. 62)
- 10) *your current situation with money* (p. 64)
- 11) *your opinion on professional agencies which offer ready-made seminar papers for students* (p. 67)
- 12) *an opportunity (here) to give you two rooms with kitchen for this amount of money* (p. 71)
- 13) *anybody telling you what to do with money* (p. 75)
- 14) *some possibility to write down some sport courses at the university* (p. 95)
- 15) *your services offered by your camp in USA* (p. 96)
- 16) *somebody who's not really rich* (p. 108)
- 17) *some information on the internet about your camp council program* (p. 118)
- 18) *ten percent discount on all tours organized by travel tours agency* (p. 133)
- 19) *the role of internet social network in our life* (p. 134)
- 20) *any possibilities in XX how to accommodate* (p. 135)
- 21) *Cambridge Advanced certificate which will be quite useful for you in your future* (p. 137)

- 22) *the main reason why you have chosen to be a teacher* (p. 138)
- 23) *a very nice job for a mother with two children or even more* (p. 139)
- 24) *a proof of high level English language skills* (p. 145)

The complexity of some of the noun phrases presented here is really quite remarkable, especially the length of the postmodification, which includes all the above-mentioned types: prepositional phrases, often several of them combined; non-finite clauses (infinitive, -ing and -ed participles); finite/non-finite relative and appositive clauses, sometimes in combination with prepositional phrases. Some of the quoted noun phrases (e.g. examples 11, 12, 18 and 23) include another postmodified head noun, and thus the relations among the individual parts of the noun phrase are rather complicated (Barlow, 2002).

How should we evaluate these structurally demanding pieces of the students' spoken discourse? I am afraid that structuring such noun phrases is only marginally connected with the students' knowledge about complex noun phrases and their skill in following the rules they have been given about pre- and postmodification of head nouns (Keizer, 2007). They simply use prepositional phrases or finite relative clauses in a way that is similar to what they would do in their L1. Non-finite postmodifying clauses may show a newly-acquired knowledge of the variety of English non-finite verb forms and their condensing power, but there is partial similarity to their L1, too. I am sure that they would be surprised if complimented on their usage of well-structured noun phrases.

Thus, the suggestion for further successful language acquisition aiming at the students' acquiring the ability to structure complex noun phrases, which are at the same time lexically rich, is to apply methods leading to the students' being aware of creating these very customary parts of English sentences. They give evidence about a condensed way of expressing ideas that is typical in the English language, and which is different in many ways from Czech discourse (Tárnyiková, 2007). This practice may influence the structuring of students' written texts more, but it will certainly enhance their language skills, leading to a type of discourse which is closer to that which is habitual for native speakers.

## **9 Conclusion**

The following outcomes and suggestions are the results of only superficial lexical analysis of the corpus of spoken English of university students. For a complete evaluation of the quality and quantity of the students' vocabulary in this corpus (and in general), a more detailed research must be performed. So far, I can offer the following conclusions:

- 1) the vocabulary of the students is not satisfactory – not so much in terms of statistical figures as in its richness;
- 2) the students' ability to use synonyms for ordinary lexical items is rather low, which is evident in the usage of most of the productive word classes – adjectives, nouns and verbs; participles and gerunds are derived from verbs and the statistical data provide a fair estimation of the lexical variety of all the verbs used in this corpus;
- 3) the students' use of nouns and gerunds in the premodification of head nouns is satisfactory, because of the structural dissimilarity to their L1;
- 4) the complex noun phrases included in the corpus are numerous and their structure is better than had been expected;
- 5) long complex noun phrases feature both pre- and postmodification, but it is doubtful that the students are aware of the complexity of these structures.

Some advice for the further development of the students' lexical competence in the process of language acquisition in advanced English language courses:

to focus on synonyms to all the productive word classes in practical language courses;

to explain clearly the structure of complex noun phrases (premodification + head + postmodification) and their frequency in different registers;

to practice the correct identification of complex noun phrases in texts of various styles;

to practice structuring noun phrases in different contexts and registers.

### **Acknowledgements**

This work was supported by a Czech Science Foundation Grant (“Aspects of English Language Acquisition of Czech Students at the Onset of Teacher Education” GA ČR 13-25982S).



## Chapter 2

# DIVERSE PRONUNCIATION RESULTS ELICITED BY MEANS OF THREE DIFFERENT TEST TYPES

*Jaroslava Ivanová*

## 1 Introduction

In the existing literature on diagnostic evaluation (Bachman & Palmer, 2010; Celce-Murcia, Brinton & Goodwin, 1996), little attention is paid to pre-course identification of university learners' pronunciation needs in their target language production by means of specific tasks. Without specifying a university or proficiency level of would-be students of English, it is proposed that two types of speech samples are obtained at the commencement of their being taught: "a standardised sample of the learner reading aloud" and "a sample of the learner free speech" (Celce-Murcia, Brinton & Goodwin, 1996). Both tasks can help the university teacher(s) identify areas for remedial pronunciation work, but they differ substantially in their potential to elicit the intended pronunciation features of English from all learners.

In assessing various features of English pronunciation, different task types can be used, with a focus on either perception or production (Celce-Murcia, Brinton & Goodwin, 1996; CEFR, 2001). In this article, only production tasks are discussed in detail, specifically reading aloud tasks and tasks involving semi-guided spontaneous speech in a "language teaching domain" (Bachman & Palmer, 2010, p. 68) in a foreign language setting.

Let us specify what is understood by language tasks in a language teaching domain. These aspects of tasks are outlined in line with applied linguists: they are: (i) "closely associated with, or situated in specific situations", (ii) "goal-oriented," and (iii) "involve the active participation of language users" (Bachman & Palmer, 2010, p. 59). In their definition of a "language use task" they emphasise that it is "an activity that involves individuals in using language for a purpose of achieving a particular goal or objective in a particular setting" (Bachman & Palmer, 2010, p. 68).

Unlike free speech tasks or open-ended tasks (Luoma, 2004, p. 50), reading aloud tasks do not conform to all the above mentioned specifications. They are highly structured speaking tasks since the examinees are provided with all the content they are going to convey (Luoma, 2004, p. 50). Reading aloud is mostly concerned with examinees' pronunciation and the extent of its comprehensibility. As a result, it lends

itself to comparative analysis of the pronunciation features under scrutiny. From the point of view of test-takers, not only do they limit their self-expression and involvement in a task, but they also lack pay-off value. In spite of the fact that reading aloud might be considered a closed-ended task as language material is provided by the teacher/interlocutor and not produced on the spur of the moment (Luoma, 2004, p. 50), we assume that in a way it manifests some features of open-endedness in view of the translation of the written form into its spoken one. It depends on: (i) the extent of the level of correspondence between written and spoken form of mother tongue (L1) in comparison with English (L2) (Crystal, 2010; Volín, 2010); (ii) the conditions under which it is conducted – whether the learner has been given enough time to prepare the reading aloud and if a model recording was available prior to the recording. As a result, reading aloud tasks, though generally used for testing or diagnosis of various aspects of pronunciation, they are not fully understood as a means of testing and that is why they deserve thorough discussion.

In comparison with silent reading for comprehension, reading aloud is viewed as one of the speech processes which, first, calls into play a person's analytical powers, immediately followed by synthesis (Hartl & Hartlová, 2000, p. 95). The process is stimulated by the graphical form of a word, which has to be mentally processed (Hartl & Hartlová, 2000, p. 95). Hendrich (1988, p. 191) perceives reading aloud as a combined language skill which makes the reader transform the written form into an oral one and the cognitive demands of this transformation grow with the level of relative ir-/regularity of the written-sound relationship of English (Crystal, 2010, p. 18). In this respect Crystal (2010, p. 224) asserts that about 75 per cent of English spelling-sound correspondence is rule-governed; however, there are about 400 English words, most of them the most frequent ones, whose written-spoken relationship is unpredictable. Moreover, Selikowitz (1998, p. 60) stated that the 26 letters and 44 phonemes of English face the reader of an English text with 577 realisations of letter-phoneme correspondence. On top of that, the level of comprehension in reading aloud affects the reader in terms of their ability to: (i) cope with some irregular sound-letter correspondence or consonant clusters that do not exist in their L1 (Richards, 1974), for example, thr- or -ght in *throw*, *taught*; (ii) retrieve the meaning of a less frequent lexical item from their mental lexicon (Selikowitz, 1998), and (iii) plan ahead further reading aloud. In other words, what is needed on the side of the reader according to the CEFR, is to activate receptive and productive strategies which enable them to: (i) plan (select a mental set and activate schemata); (ii) execute (locate cues and infer from them); (iii) evaluate (match the identified cues to schemata); and (iv) repair (remedy, if there is a mismatch between the identified cues and the activated schemata) (CEFR, 2001, p. 72).

All the likely difficulties mentioned above make us believe in line Richards (1985) that Czech readers are faced with negative transfer stemming from two facts. First, there is a consistently regular relationship between spoken and written Czech with

the exception of words of foreign origin (Volín, 2010, p. 35). Second, the pronunciation of the majority of Czech lexical items can be derived from a relatively limited set of rules (Holub, 2008). To sum up, negative transfer into Czech learners' interlanguage might stem from the irregularity of the spelling-sound correspondence and a relatively high number of rules governing the regular spelling-sound correspondence in English. This increases processing demands in reading aloud.

## 2 Experimental design

The study uses a mixed-methods design. The individual pronunciation results in three subtests are analysed, correlated, and qualitatively evaluated.

The statistical analysis of the categorical data (correct/incorrect pronunciation of segmental features in question) is applied to calculate (i) the classical difficulty index ( $p_i$ ), which tells us "how a given group of individuals performed on average on a particular item" (Bachman, 2004, p. 122); (ii) the average percentage success score per individual and task, and (iii) the Pearson correlation coefficients ( $r_p$ ) in order to calculate "an overall indication of how much the two variables covary" (Bachman, 2004, p. 86), here to indicate the strength of the correlation between different task types.

From the pronunciation point of view, these segmental features are the main focus of this study in all three subtests: the front open vowel *ash*, the weak central mid central vowel, i.e. the schwa, the voiced and voiceless dental fricatives, the bilabial approximant /w/, the velar nasal, and the pronunciation of word-final voiced consonants. These features were identified as problematic for Czech learners by Černá, Urbanová, and Vít (2011), Nádraská (2013) and Volín and Poesová (2008).

The application of a list of separate words and the enormous number of students ruled out a focus on prosodic features.

The British and American standards as presented in Wells' Longman Pronunciation Dictionary (2008) served as a point of reference.

### 2.1 Subjects

112 first-year university students participated in this study in autumn 2013. All of them were native speakers of Czech with B2-level of English according to CEFR (2001) and were recorded at the very beginning of their Phonetics and Phonology course or before its start. This is due to the differently structured curricula on teacher education courses in České Budějovice, Olomouc, and Pardubice.

The subjects agreed to be recorded and to cooperate throughout the project as they were asked to produce a transcript of their free speech (signed agreement). To

achieve good-quality recordings, the Sound Forge Pro 10 software (2003 – 2015), was purchased and used.

On the basis of the results of the 112 students reading aloud (subtests 1 and 2), 26 test takers' free productions (subtest 3) were selected. Each existing score range from 4 to 31 points (4.8-31.33) is represented by one speaker, including the highest and lowest five per cent of the cases. We came to this decision thanks to the robustness of the data obtained from the samples of free speech.

The subjects were recorded at their home universities by three university teachers.

## **2.2 Assessor**

The analysis was carried out using subjective auditory assessment because of the huge amount of the data and in order for it to resemble the real-life assessment of students' pronunciation of English phonemes by one university teacher, who listened to all the types of subtests twice to increase intra-rater reliability (Bachman, 2004).

## **2.3 Tasks**

Three data collection tools helped to elicit the pronunciation data. The subjects were requested to: (i) introduce themselves and answer two or three questions asked by an interlocutor (a university teacher), (ii) read a diagnostic passage aloud, and (iii) read a list of words aloud. The tasks were sequenced in this order.

Subtest 3 can be defined as semi-guided free speech (Chráska, 1999), guidance being provided by the interlocutors' topical questions, which were scripted: "Would you introduce yourselves? What about you and learning English? Did you attend any language courses? What about travelling in general and to English-speaking countries in particular?" The 26 speech samples differ in length in terms of words (mean = 151 words, max. 334, min. 77) and the number of tokens (mean = 83, max. 143, min. 53) and time.

Subtest 2, the diagnostic passage, consists of 153 words, out of which 98 are tokens, i.e. different words. From the point of view of word length in syllables, there are 103 one-syllable words (67.8%), 35 two-syllable words (23%), 11 three-syllable words (7.2%), and 3 four-syllable words (2%), and the average number of syllables per word is 1.43. The text characteristics of subtests 1 and 2 were obtained by means of *textalyser.net* (2004).

Subtest 1 contains a list of 24 words (19 lexical items are different) which are taken from the diagnostic passage. The reading aloud subtests were designed with the aim of enabling the comparative pronunciation analysis of the chosen words in the context of a text and in isolation.

Prior to being recorded, each reader was instructed to study the diagnostic passage and the wordlist for a minute, and to read the words on the list one by one, including the numbers at the beginning of each line.

## 2.4 Method

Each correctly pronounced feature was assigned one point, while each incorrectly enunciated feature was classified as zero out of the total of 36 points in subtests 1 and 2 respectively. The results were tabulated as item difficulty indices in order to achieve the total score for each of the 112 first-year university students and, simultaneously, to provide information on how the 112 students performed on average on a particular pronunciation feature (Bachman, 2004; Chráska, 2007).

In order to categorise the data, each individual pronunciation feature was treated separately. For example, first, all the words containing /æ/ (*lamp*, *bag*, *had*, *that*, *planned*) were put under one heading, then the total number of correct occurrences of the pronunciation feature being inspected was counted, and afterwards calculated as percentages, that is, difficulty indices  $p_i$ , as shown in Table 2.1.

Table 2.1  
*The front open vowel in subtest 1*

Phoneme	/æ/					
words	lamp	bag	bag	had	that	planned
acceptable pron.	læmp	bæg	bæg	hæd	ðæt	plænd
unacceptable pronunciations	*lemp	*bek	*bek	*het	*det	*plend
	*lem	*beg	*beg	*hed	*ðet	*plent
	*lɒŋ	*bʌk	*bʌk		*ded	*plent
	*lʌmp				*ðæt	*plɑ:nt
					*ðen	
sum /112	77	30	36	71	64	78
mean $p_i$	68.75	26.79	32.14	63.39	57.14	69.64

The word *bag* was read twice in the wordlist as it appeared twice in the diagnostic passage. Additionally, the word *bag* was examined later on as an example of a voiced ending together with *had* and *planned*. As a result, in the samples of free speech, in one word, such as *language*, the use of /æ/, /ŋ/, /w/, and /dʒ/.

Finally, the individual pronunciation results in the two above-mentioned subtests 1 and 2 (n=112) and subtest 3 of free speech were tabulated (n=26).

With respect to the transcripts of the free speech samples, they first had to be categorised with the help of textalyser.net (2004), which provides the list of words

with their frequencies. The individual words were then put under the appropriate heading of the pronunciation features being dealt with: the open front /æ/, the schwa in weak forms of function words, and in full-meaning words, voiced consonants at the end of weak forms of function words (/d/, /v/, /z/), and in full-meaning words (/d/, /v/, /g/, /dʒ/), the voiced and voiceless dental fricatives (/ð/, /θ/), the bilabial approximant /w/, and the velar nasal /ŋ/.

As the total number of occurrences of each pronunciation feature differs from one speech sample to another, for example, from three to 18 instances of /w/, the number of occurrences of correct pronunciations is expressed as a percentage. In order to be able to calculate comparable means (Table 2.2 below) based on other means, this equation was used (Volín, 2007, p. 50) to cope with different frequencies of a certain pronunciation feature in case of individual speakers:

$$(1) \quad \bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

Afterwards Pearson correlation coefficients ( $r_p$ ) were calculated separately for subtests 1 and 2 (n=112), taking into account the individual student's overall scores, likewise for subtests 1 and 3 (n=26), and finally for subtests 2 and 3 (n=26). The coefficients are intended to show to what extent better pronunciation results in one subtest might indicate better (or worse results) in another one, and the strength of this correlation (Bachman, 2004, p. 85; Chráska, 2007, p. 114).

When the pronunciation features in individual subtests were being assessed, a number of decisions had to be made in subtests 2 and 3 as they are context-dependent at word boundaries (Roach, 2009, p. 89).

### 3 Results

In comparison with faulty pronunciations of separate words in the wordlist and the diagnostic text, in the 26 free speech samples there is a substantial downtrend, showing an increasing level of difficulty in cases of the open front vowel /æ/ (53% → 16%), the schwa (63% → 42%), and the voiced dental fricative /ð/ (65% → 48%) and a gradual decrease in the instances of the voiceless dental fricative /θ/ (82.6% → 80.3%) and voiced bilabial approximant /w/ (69% → 63%) (see Table 2.2).

Table 2.2

*The relationships among subtests and the pronunciation features being examined*

Task types	Wordlist		Diagnostic text		Free speech	
	percentage	frequency	percentage	frequency	percentage	frequency
open front æ	52.98	672	45.08	672	16.41	179
schwa	63.39	560	32.19	672	41.83	776
final voiced consonants	51.50	896	75.95	1008	68.65	737
θ in <i>think</i>	82.59	448	80.89	448	80.26	37
ð in <i>the</i>	65.18	336	30.42	336	47.49	189
ŋ in <i>long</i>	32.37	448	54.42	448	41.40	206
bilabial w	69.42	448	57.35	448	63.15	216

The voiced alveolar plosive /d/ in both the verbs ending with -ed, (e.g. *waited*, *suspected*) and in the stems of words such as *had*, *and*, *bed*, and the voiced velar plosive /g/ in words such as *bag* appeared most problematic in Czech learners' enunciations in the wordlist (51.5%) and free speech (69%), while the mispronunciations of the velar nasal /ŋ/ achieve a fairly low difficulty index in the wordlist (32%), rising to a slightly higher level of achievement in free speech (41%).

Pearson correlation coefficients show that a higher proportion of correctly read separate words is likely to correspond with a higher number of correctly enunciated selected words in a continuous text ( $r_p=0.6817$ ) and chosen segmental features in spontaneous speech ( $r_p=0.6925$ ). The strongest  $r_p=0.7236$  was measured between subtests 2 and 3. We can accept  $H_A$ : the calculated coefficients show that there are medium ( $0.7>r\geq 0.4$ ) to strong ( $0.9>r\geq 0.7$ ) (Chráska, 2007, p. 105) positive relationships between the pronounced features in three subtasks at  $t_{0.05}(110)=1.984$  (subtests 1, 2) and  $t_{0.05}(24)=2.064$  (subtests 1, 3; 2, 3).

#### 4 Conclusion

Overall, these Czech learners received lower difficulty indices for a diagnostic text than for a carefully pronounced list of words which is confirmed by other research studies with respect to Czech learners of English (Volín & Poesová, 2008; Černá, Urbanová & Vít, 2011).

From prior experience of being exposed to the Czenglish pronunciation of university students at higher proficiency levels (B2-C1) for over 23 years, it was hypothesised

that the highest difficulty indices in pronunciation would be achieved in reading individual words aloud because the focus on separate words is less cognitively demanding, they are context-free, and there is more time to concentrate on pronunciation than in reading words within a diagnostic passage. The lowest difficulty indices were expected from the free speech samples as the students tried to get the message across, thus losing sight of language form, here specific pronunciation features. These assumptions were met in all but three categories: final voiced consonants, the velar nasal, and the schwa.

Because the pronunciation of the final voiced consonants in the grammatical ending -ed is the focus of all textbooks of English, we separated the individual subcategory of -ed, -d endings (e.g. *warned*, *needed*) to find the reason(s) why. To our surprise, their context-free pronunciations were rather poor in the case of individual words (see Table 2.3), but we had to take into account two influential factors having an impact on their pronunciations in a diagnostic passage (subtest 2) and free speech (subtest 3). The frequent occurrences of proper pronunciations of function words such as *and*, *had*, and *would* and the regressive assimilation of voice across word boundaries in the diagnostic passage and free speech subtasks contributed to the overall higher difficulty indices (see Table 2.3) than was assumed. This was so in spite of our expectations stemming from the problems Czech learners often experience when it comes to the pronunciation of a lenis consonant (i.e. voiced) at the very end of a word, e.g. *warned* is pronounced incorrectly with /t/. In the diagnostic text, however, the /t/ pronunciation was appropriate in the cases such as "... *had seen* ...", "... *had planned* ..." thanks to the regressive assimilation. The regressive assimilation affects mainly consonants across word boundaries in a manner that the lenis (voiced) consonant is enunciated as a fortis (voiceless) one being influenced by a fortis (voiceless) consonant at the beginning of the following word (Roach, 2009, pp. 110-111).

Table 2.3  
*Pronunciation of grammatical endings*

Task types	=-ed, -d /d/	Mean %	n
P <sub>i</sub> wordlist	<i>waited</i> , <i>suspected</i> (2x), <i>planned</i>	27.23	112
P <sub>i</sub> diagnostic passage	= 448 occurrences	52.96	112
P <sub>i</sub> free speech	= 51 occurrences	73.25	26

To diagnose the correct pronunciation of the velar nasal /ŋ/ was the most difficult nut to crack for the assessor. There were numerous crystal clear cases of its mispronunciations as /ŋk/, /ŋg/, or /n/ at the end of *long*, *everything*, and *locking*, but in reading the diagnostic passage aloud and the free speech sample its perception was affected by the speed of delivery of reading aloud, by the volume of the reader's

voice, and sometimes by a student's avoidance strategy, that is swallowing the ending of a word. As a result, these are the most dubious data obtained in this study.

The overall trend in the case of the schwa made us reinspect its pronunciation from the point of view of its distribution in the weak forms of function words and in the unstressed syllables of full-meaning words. It revealed the inability of Czech speakers to pronounce schwas appropriately in the weak forms of function words in free speech (25%, 482 occurrences), which became evident when we treated their pronunciations separately from those of full-meaning words (70%, 294 occurrences).

The greatest disadvantage of the wordlist is that it is context-free and requires the reading aloud of words such as *had* and *that* in their strong forms and, consequently, excludes the testing of weak forms of function words. In spite of the fact that reading the diagnostic passage aloud is not communicative by its nature and brings about errors caused by negative L1 transfer stemming from the differences between the sound-spelling correspondence in Czech and English, it permits multiple comparisons of context-bound pronunciation features and proves less time-demanding. The free speech subtask sheds light on the numerous facets of the pronunciation features that were being inspected in context (devoicing across word boundaries, weak forms of function words, different functions and meanings of words in context, e.g. different pronunciations of the word *that*: *that book* in contrast to *I didn't know that...*) without being affected by the way they are spelt in English.

As far as the limitations of this mixed-methods study are concerned, currently we are aware of some of them. Primarily, it is the number of the assessors, which limits the possibility to generalise the results for all first-year university students in teacher training programmes in the Czech Republic, even if the intra-rater reliability was rather high ( $\alpha=0.63$ ) (Bachman, 2004, p. 170). On top of that, we are faced with the question how far an inspected pronunciation feature can be perceived as a standard one and so appropriate in the pronunciation of a non-native speaker/reader of English. Second, the results are also influenced by the fact that the pronunciation focus in the diagnostic passage was limited to the words contained in the wordlist, which, on the other hand, made a comparative comparison feasible. Third, it might make a difference if the pronunciations of voiced consonants such as /v/, /z/ at the end of all the words were studied, compared with /d/ and /g/ pronunciations in the same positions in both the diagnostic passage and free speech. Fourth, although the overall results dealing with the pronunciation of the selected segmental features help us diagnose the pronunciation features in need of remedial teaching as tendencies to follow in teaching English segmentals, they do not shed light on an individual learner's variation. Fifth, the rather broad categories of the pronunciation features that were selected for analysis, e.g. final voiced consonants, schwa, have to be revisited in order to attend to the frequency and distribution of the troublesome pronunciation features for Czech university learners who would like to become teachers of English.

For example, most subjects in this study pronounced the schwa correctly in weak forms of function words such as *a*, *an*, *the*, if the last one was not followed by a word whose pronunciation started with a vowel sound, although they experienced enormous problems in the pronunciation of the remaining weak forms.

### **Acknowledgements**

This work was supported by a Czech Science Foundation Grant (“Aspects of English Language Acquisition of Czech Students at the Onset of Teacher Education” GA ČR 13-25982S).

## Chapter 3

# HESITATION IN LEARNER'S DIALOGUES: DISTRIBUTION OF INDIVIDUAL MEANS

*Šárka Ježková*

## 1 Introduction

There are obviously many differences that can be noted between written and spoken language. Unplanned and spontaneous speech is fundamentally affected by the fact that speakers produce the language on the spot. This chapter considers several aspects of spoken grammar from the point of view of second language acquisition. It presents an analysis of a learner corpus focused on selected structures typical of dialogue and spontaneous speech. The study explores the use and distribution of various types of hesitators and their effect on the flow of natural conversation. The findings are compared to the conclusions of studies of spoken discourse by English native speakers with an attempt at identifying the most salient differences and discussing the potential influence of the native language.

## 2 Principles of spoken discourse

Spoken discourse is characterised by several typical features which differentiate it from written texts, mostly resulting from a variety of social, psychological, and physical determinants. Participants in a conversation use the auditory channel when conveying a message (i.e. tone units, stress, pauses, voice quality, etc.), they share the context (i.e. the physical context of time and place, and also the broader context: social, cultural, institutional, etc.), and they usually avoid elaboration (i.e. there is a limited range of lexical and grammatical units). Other determinants influencing the character of a conversation are that it is interactive (i.e. there is an ongoing exchange between the participants), it takes place in real time (the participants produce utterances under pressure), and it expresses emotion, politeness, and attitude. Moreover, even though the roles of the speaker and listener change during a conversation, commonly there are quite few overlaps and remarkable silent pauses (Coulthard, 1985, p. 59).

As Biber et al. (1999) state, the grammar of speech is dynamic, i.e. constructed and interpreted under real-time pressure (p. 1066). As we cannot avoid linearity of

language production when we construct any type of a text, such principles of linear construction of a language are adapted to the particular purpose of speech, mainly to the principles of online production. The participants in a dialogue are supposed to adjust their behaviour to three main principles: (a) the *keep talking* principle, which makes both parties move the conversation on without significant interruptions, unnatural stops, or other breakdowns; (b) the *limited planning ahead* principle, which forces speakers to plan what they are going to say while they are already speaking and while they are limited by their human working memory; (c) the *qualification of what has been already said* principle, which pushes the participants to adapt and modify their planned contributions according to the previous utterances. The application of these three principles is reflected in the use of syntactic structures typical of the grammar of conversation (Biber et al., 1999, p. 1067).

Quite logically, unplanned, spontaneous talk results in looser organisation and a certain syntactic oddness (Hatch, 2001, p. 237). Thus Biber (1995, p. 43) infers that a certain degree of fragmentation in speech is considered undoubtedly acceptable since the information cannot be carefully incorporated into the text because of the time constraints. Finally, a minor but essential comment should be made: quite frequently terminological labels used for speech (including expressions such as *incomplete* or *non-fluent*) evoke negative connotations and bring negative evaluation (Ježková, 2012). Urbanová (2003), however, argues that “spoken language and written language constitute two different norms, which are not interchangeable” (p. 13).

### 3 Types of hesitators

The dominant interactional function of conversation affects many aspects of the grammar of speech. Schiffrin (1994) has shown that the interaction in conversation is structurally organised and that certain rules of turn exchange are crucial for successful conversational management (p. 238). Thus the mechanisms of adjacency pairs and turn-taking contribute significantly to the overall construction of a conversation, though the units through which speakers construct a turn may be various (sentential, clausal, phrasal, or lexical). Utterance launchers as special expressions placed at the beginning of a turn usually combine a variety of functions (e.g. taking a turn, performing interaction and cooperation, getting time for planning, etc.). Structurally, they may be categorised into three major groups: (a) fronting and left dislocation (with a primary linking function), (b) prefatory words (usually a one-word opening linking the utterance with a preceding turn, i.e. discourse markers, interjections, response forms, disjuncts, and conjuncts), and (c) overtures (i.e. longer expressions *I'll tell you what, the question is, etc.*, mostly giving speakers more time to plan ahead). The distribution of individual types of utterance launchers is influenced by the character of the conversation, the social and cultural context, and also by the relationship between the participants.

The studies on the relationship between syntactic structures and their discourse functions reveal that there is no one-to-one correspondence system, i.e. one syntactic construction may serve a variety of discourse functions and one discourse function may be accomplished by a variety of syntactic constructions (e.g. Schiffrin, 1994; Carter & McCarthy, 1995; Gee, 2005; Widdowson, 2007; Ward & Birner, 2008). Thus inserts, as units typically used in speech (i.e. single-word, non-clausal units, without syntactic relations with other structures), fulfil a wide range of discourse functions in conversation, although usually one function may be perceived as primary. The prevailing function of interjections is exclamatory; discourse markers mostly signal the transition of a turn but they may also serve as a means of interaction; response forms, as rather routinised responses to a previous turn, predominantly create an adjacency pair, and hesitators most importantly enable speakers to pause while signalling their wish to continue speaking. Within the category of hesitators we may differentiate between two major groups: repeats (i.e. one or more words which are repeated without any change) and filled pauses (i.e. special vowel sounds filling silence).

The use of all the types of inserts is influenced by many factors, including the three above-mentioned principles of conversation. The relatively high occurrence of inserts is a common feature of spoken language because conversation is not a prepared act and so a certain degree of dysfluency (which does not impede understanding) is considered to be a normal accompaniment of spontaneous speech. The use of all the types of hesitators serves as an irreplaceable means in strategies to gain time. Repeats are unplanned and involuntary; in contrast to deliberate repetition, they are used under online planning pressure, so they signal either that the speaker has not finished yet or that he/she would like to take a turn. It is also noticeable that hesitators have tendencies to co-occur.

#### **4 Analysis**

Spoken corpora are generally used not only to analyse phenomena typical of natural spoken language, such as discourse markers, hedges, tags, backchannels, ellipsis, etc., but also to understand information flow by focusing on the structure and functions of spoken interaction by analysing specific structural patterns such as turn-taking procedures, adjacency pairs, or various types of sequences (Aijmer & Stenström, 2005).

As Ellis and Barkhuizen (2005) suggest, the core interests of conversation analysis in second language acquisition are usually focused on the organisation of talk-in-interaction, including turn-taking, sequence organisation, and repair. In this study too the strategies of turn-taking are partially analysed, mainly in relation to the use and distribution of individual types of inserts and hesitators. The findings of the

analysis of the non-native speakers' (NNS) corpus described below are compared with the conclusions of studies researching the performance of English native speakers (NS). The questions of the research include: How far is the students' spoken language similar to the natural performance of NS? Where are the most significant differences? Why do or do not the interviewees sound "native-like"? Which features of spoken grammar should be focused on and practised in language courses? This chapter brings only a survey of the distribution of individual formal realisations of inserts with more detailed inspection of hesitators, exemplifying especially types of repeats, even though the author is aware of the fact that other features of the grammar of speech (e.g. repairs, non-canonical word order, discourse verbs, vagueness hedges, and incomplete and elliptical structures) are very closely interrelated. The use of other discourse management devices, however, is the subject matter of other studies.

#### **4.1 Motivation for the research**

This research is part of a bigger multidimensional project called "Aspects of the English Language Acquisition of Czech Students at the Onset of Teacher Education" (supported by the Czech Science Foundation) with the following overall objectives: (a) to compile a corpus of learner English used in spoken communication; (b) to perform an analysis of selected grammatical, discourse, and pronunciation features with conclusions for the second language acquisition processes, including the possible negative transfer of Czech grammar and pronunciation features into English; (c) to carry out an analysis of the data related to the learning histories; to identify critical incidents and people that influenced individual learning histories, in what way, and in what phase of the individuals' lives, and (d) to interrelate the above-mentioned analyses with the aim of modifying the contents of university courses in order to improve future English teachers' education.

The reasons why spoken performance is in focus are manifold. First, the research team believes that for future English language teachers the skill of speaking is crucial in their profession, while it seems that practising it has been rather neglected at primary and secondary schools, probably because of the fact that for teachers in schools it may be difficult to set transparent criteria for assessment. Moreover, similarly to the situation in other countries at all levels of the educational system, instruction is still mostly based on written language. Even though the studies of spoken corpora have revealed that there are significant differences between the grammar of speech and grammar of writing which are reflected in the discourse of speaking and writing, English language learners are seldom guided to realise such differences, let alone to acquire them actively. McCarthy and Carter (2001) point out that "there can be little hope for a natural spoken output on the part of language learners if the input is stubbornly rooted in models that owe their origin and shape

to the written language” (p. 51). Another motive for looking more closely at learners’ spoken performance is an apparent discrepancy between the CEFR level assessment criteria and corpus linguistics findings of NS conversation analysis. While the CEFR (2001) criteria for the C1 level state that a learner shows fluent spontaneous expression in clear, well-structured speech, and “can express him/herself fluently and spontaneously, almost effortlessly” (p. 28), Biber et al. (1999) argue that the grammar of speech is dynamic because it is constructed and interpreted under real-time pressure, and thus “correction and reformulation is possible only through hesitation, false starts and other dysfluencies” (p. 1066).

#### **4.2 Description of the learners’ corpus**

The Czech students’ spoken English (CSSE) corpus comprises about 150,000 words, i.e. transcribed recordings of 114 pairs of first-year students of TEFL Programme from three Czech universities at the beginning of their studies, approximately at the B2 level of the CEFR (2001). The interviews were recorded in autumn 2013 and 2014, transcribed by the students themselves, and later corrected thoroughly according to the recordings by the members of the research team. The process of compilation was carefully prepared and piloted beforehand since the process of data elicitation conditions the quality of the final product, i.e. the language corpus for further analysis. That is why the test was designed with the aim of obtaining language material that is as authentic as possible (Černá, 2014). For the process of transcription, the students were provided with detailed instructions concerning the whole process, with the emphasis put on exact wording, including all mistakes, false starts, non-word elements, etc. The subsequent editing by research team members covered checking, correcting, and standardising the transcription symbols.

Every conversation is structured and consists of three parts: (a) two monologues where the students introduce themselves and briefly describe their English language learning experience; (b) two dialogues where they have a task with an information gap, and (c) a discussion on a given topic, so there are 570 separate files in order to enable various contrastive studies. The personal profiles of the individual students are also added to the corpus (e.g. age, gender, number of years of studying English + where, staying in an English-speaking country) so that the interrelation between the level of their communicative competence and their learning history could be identified.

#### **4.3 Distribution of inserts**

As a starting point for the analysis, the findings of NS corpus (i.e. about 4 million words of British spoken English and 4 million words of American spoken English) research summarised by Biber et al. (1999) were used. Briefly, they claim that

repeats and filled pauses co-occur, the distribution of both categories is proportional, and the most frequent repeats are single functional words. Among them, the most frequent is the occurrence of the nominative of personal pronouns, possessive pronouns, the, and conjunctions, while the least frequent is the accusative of personal pronouns, prepositions, and verbs. Quite expectedly, the most frequent two-word repeats are subject pronouns + auxiliary contractions.

The figures presented here are based on the exploration of the first half of the CSSE corpus because at the point at which the quantitative (using WordSmith Tools) and qualitative analysis was carried out (the particular function of selected structures had to be judged individually), the final version of the second half was not available. Five types of inserts were investigated in relation to the structure of the interview and the overall strategy of conversation management.

Table 3.1  
*Overall distribution of inserts*

Type of insert	Proportional occurrence (%)
interjections	1
discourse markers	11
response forms	12
filled pauses	58
repeats	18

The studies of NS dialogues (e.g. Biber et al., 1999) show a rather balanced distribution of individual types. The figures in Table 3.1, however, prove the differences in the distribution of individual types of inserts, mainly the overuse of non-word filled pauses and underuse of interjections and discourse markers. On the basis of the analysis of individual students' dialogues, it can be generalised that the higher the level of a student's performance, the higher the frequency of all the types of inserts, including their complementary distribution and co-occurrence.

- 1) *Well, well, well, I can't because I still don't have a driving licence but I'm going to do it in, in [...] a very (er) near, near future ... (13017)*
- 2) *(er), well you get, (er), scholarship, and it depends on the host, host country and ... (13022)*
- 3) *(er), yeah, (er), maybe, and [...] have you [...] or [...] but [...] (+) Is it with accommodation or without? (13032)*

Examples (1) and (2) represent quite typical combinations of inserts students frequently use, i.e. filled pauses together with repeats and also discourse markers, even though the range is very limited – in most cases it is *well* or *so*. Such a performance is not very far from spontaneous NS talk, but unfortunately, in many interviews the natural distribution of inserts is not noted and instead of linguistic devices, we can see just filled and silent pauses, as in example (3).

#### 4.4 Distribution of hesitators

In a natural conversation, speakers use hesitators, which enable them to pause while signalling their wish to continue speaking. Generally, there are two main categories: repeats (i.e. exact repetition of one or more words) and filled pauses (i.e. a non-word filler: a special vowel sound filling silence). At the beginning of the research I intended to inspect only those fillers represented by words. But in the course of the analysis, it became necessary to look into both forms of hesitators and compare their distribution.

Table 3.2

##### *Distribution of hesitators*

Type of hesitator	Proportional occurrence (%)
filled pauses	69
repeats	31

Table 3.2 shows that both major groups of hesitators occur in conversations; however, it is apparent that the Czech students overused filled pauses. They also repeated non-functional words and longer structures (more than one word) much more often. The relatively high frequency of multiple-word repeats may be influenced by the character of the task; they mostly occur at points where students needed more time for thinking and planning the subsequent utterances in the second part of the dialogues, where they used information from a sheet that was provided. The following examples can serve as a comparison between two different strategies; in utterance (4) the student uses linguistic means to gain time, in (5) only non-word fillers.

4) *Yeah. Well, of course, but it is, yeah, of course you can have...* (13038)

5) *(Er), so I want to know, (er), (er), some, (er), the procedure recruit recruitments.* (13022)

#### 4.5 Distribution of repeats

Table 3.3

##### *Distribution of repeats*

Type of repeat	Proportional occurrence (%)
pronouns – nominative	24
subject + auxiliary (not only contractions)	15
other functional words (wide variety)	23
single-word lexical (variety of word classes)	13
multi-word (two or more words, many combinations)	12
partial (highly influenced by individual idiolect)	13

Surprisingly, in many utterances the Czech learners used repeats of other functional words (including demonstrative pronouns, conjunctions, and prepositions), single-word lexical repeats, mostly nouns, adjectives, verbs, adverbs, and multi-word repeats (e.g. *you don't have to, I don't know, I don't remember; to be, want to, my English*). These three categories of repeats represent almost half of all the occurrences.

- 6) *And you can probably better [...] you can better work with sources, sources because if you will find something in the book you will probably believe it's right.* (13039)

Example (6) illustrates such repeats, which are rather rare in NS dialogues, but which occurred relatively frequently in the interviews that were analysed. The question is how much this is influenced by the character of the interview or how much it is influenced by the mother tongue and the level of communicative competence in English.

## 5 Conclusion

Summarising the distribution of individual types of inserts, the underuse of interjections (i.e. 1%) may be caused both by the character of the task and the effect of negative transfer from L1. Even though it may seem peripheral in the system, the relatively high frequency of interjections in talk influences the overall impression of the students' performance very positively and brings their speech closer to a natural NS type. The use of response forms is very similar in both corpora (as regards their frequency and distribution) and thus it does not require any special attention. The strategy used for placing these words in dialogues is obviously very similar in both languages and that is why students do not have a problem in this area. It seems that the use of discourse markers may be influenced by L1 and thus this area should be concentrated on in the classroom. An appropriate range of discourse markers and common principles for their placement should be taught, explained, and trained in practical language courses.

The figures show that the inventory of devices is rather limited and that these structures are not always an inherent part of students' spoken performance. The distribution of hesitators differs rather significantly in the NS and NNS corpora, not only in that in the NNS corpus filled pauses occur more often than repeats, but also in that the repeated structures are different (in the NNS corpus repeats of multiple-word structures and non-functional words are much more frequent). In NS performances the vast majority of repeats are represented by the nominative of personal pronouns and contractions of personal pronouns together with auxiliary verbs, i.e. structures typically found at the very beginning of a sentence or clause. In the NNS corpus

such structures form only a smaller part and learners repeat lexical units and longer structures more frequently. One of the possible explanations may be the difference between the systems of the English and Czech languages. Since Czech, as a synthetic language, uses functional words (e.g. personal pronouns in subject position) far less often, learners do not use them in the target foreign language either. However, such an assumption would need further research aimed at the analysis of interviewees with various native languages to uncover the potential interrelation between the system of a native language and the oral performance in a foreign language. Apart from the dissimilarities in the use of various types of repeats, there is another case of an apparent effect of the students' native language, i.e. the considerable frequency of so functioning as a discourse marker, which could definitely be interpreted as a case of negative transfer because its Czech equivalent primarily has this function.

Since inserts play a very important role in the development of interaction, in conversation management, and in establishing the dynamic ongoing exchange, it is very important to pay attention to such structures in order to raise students' language proficiency (especially the use of discourse markers and both types of hesitators). McCarthy (1991) claims that these devices are essential in turn-taking and inevitable for overall conversation management, but since the strategies of their use vary from culture to culture, it may be difficult to apply them appropriately in foreign languages (p. 127). In accordance with the arguments presented by McCarthy and Carter (1994), it is absolutely necessary for learners to acquire discourse competence in order to be able to apply appropriate discourse strategies and to be able to communicate effectively in a foreign language. On the basis of the findings it can be concluded that the courses in the TEFL study programme should be modified with the aim of teaching students specific features of spoken discourse, bringing their attention to the whole inventory of structures, pointing out multiple functions of individual devices (e.g. discourse marker, filler, turn-taking signal, backchannel signal), and improving their communicative competence in speaking English.

### **Acknowledgements**

This work was supported by a Czech Science Foundation Grant ("Aspects of English Language Acquisition of Czech Students at the Onset of Teacher Education" GA ČR 13-25982S).



## Chapter 4

# ACADEMIC WRITING: MODAL VERBS IN THE ROLE OF HEDGES

*Petra Huschová, Irena Reimannová*

## 1 Introduction

In the context of the formal education system of the Czech Republic, opportunities for implicit acquisition from communicative contexts are quite limited and mastering English as a foreign language “usually requires additional resources of explicit learning” (Ellis, 2008, p. 1). That process of learning might be particularly difficult in terms of pragmatics, the study of people’s comprehension and production of linguistic action in context (Kasper & Blum-Kulka, 1993, p. 3), specifically in the area of hedging, which is considered an essential part of pragmatic competence. The area of hedging includes a wide range of lexical and grammatical devices; therefore, the focus of the study has been narrowed to modal verbs, which are usually listed as one of the most frequently used categories of hedges (Crompton, 1997; Hyland, 1994); in particular, epistemic modal verbs are generally viewed as prototypical hedging devices associated with degrees of uncertainty on the part of the author. Since hedges are also taken to be crucial in academic writing (Fraser, 2010; Hyland, 1999; Markkanen & Schröder, 1997), the major aim of the study is, accordingly, to investigate whether and how Czech students of English apply their pragmatic knowledge concerning the use of modals in their bachelor’s papers.

## 2 The interface affair: weak interface

Learning a foreign language, English, like any other cognitive task, involves conscious and unconscious processes that help to acquire and organise knowledge in the brain. Ellis (2008) claims that “human learning can take place implicitly, explicitly, or, because we can communicate using language it can be influenced by declarative statements of pedagogical rules (explicit instruction)” (p. 4). While we perceive implicit and explicit learning as the processes involved in such learning, knowledge is interpreted as the products of those processes (Ellis, 2009, p. 6), and explicit

instruction as “teaching about a language in classroom settings” (Schmidt, 1994, p. 20). The issue of consciousness and distinctions between implicit and explicit learning, knowledge, and instruction have been at the centre of research in cognitive psychology and foreign and second language acquisition (SLA) for many years.

According to Bachman and Palmer (1996), language knowledge includes organisational knowledge and pragmatic knowledge (pp. 67-70). Schmidt (1993) claims that if learning second or foreign language pragmatics is to happen, it is necessary to attend to those pragmalinguistic aspects which are to be acquired, i.e. “linguistic forms, functional meanings, and the relevant contextual features” (p. 35). To operationalise the construct of consciousness in the acquisition of pragmatics, we refer to Schmidt’s framework of awareness, where he explains noticing as “registering the simple occurrence of some event” (Schmidt, 1993, p. 26), while understanding “implies recognition of a general principle, rule, or pattern” (p. 26) at a higher level of awareness (Schmidt, 2001). Schmidt (2010) further proposes that understanding, or “knowledge of rules and metalinguistic awareness”, is “facilitative but not required” (p. 725), while noticing is believed to be necessary for learning a foreign language, as only the aspects that are noticed can serve as intake for their learning (Schmidt, 1993, p. 26).

On the basis of the evidence and findings from cognitive neuroscience, it can be inferred that noticing, or generalisations from examples, can be processed both implicitly, nonconsciously, *without any metalinguistic awareness*, and explicitly, consciously, intentionally, or deliberately, *with awareness* (e.g. Ellis, 2009; Schmidt 2010, 2001, 1993). Though processing different types of implicit and explicit knowledge happens distinctively, in separate cognitive systems, and is dissociable, those knowledge systems can interact and cooperate, especially in foreign language performance (Ellis, 2009, p. 15); their interface is “dynamic: It [*sic*] happens transiently during conscious processing, but the influence upon implicit cognition endures thereafter” (Ellis, 2005, p. 305). The distinctions between implicit and explicit knowledge are summarised in Table 4.1.

Table 4.1

*Distinctions between implicit and explicit knowledge (Ellis, 2009, pp. 11-16)*

Implicit knowledge	Explicit knowledge
tacit, intuitive	conscious
procedural	declarative
may or may not be target-like (developmental sequences, overgeneralisations)	imprecise, inaccurate
available through automatic processing	generally accessible only through controlled processing
fully internalised by the learner	a control “tool”
actual language use	verbalised (“nontechnical”) knowledge
limited access (age)	unlimited access (learnable)

The specific properties of the two types of knowledge were applied when evaluating learners' ability to describe and mainly to use hedges accurately and appropriately, or in a target-like manner, in the context of academic writing in an attempt to reveal whether pragmatic knowledge can be learned, i.e. whether foreign language learners notice and are aware of the role of the selected language devices in the given context.

### 3 The concept of a hedge

At first, it is necessary to delimit the concept of a hedge so that the modal verbs functioning as hedges can be identified. The term *hedge* as a linguistic concept was first used by Lakoff (1973), who referred to "words whose job it is to make things fuzzier or less fuzzy" (p. 471) and considered showing a lack of certainty on the part of the author as their main function. Lakoff's concept has been widened and the term has been associated with various pragmatic strategies, namely politeness, tentativeness, indirectness, mitigation, or vagueness. Hyland (1995, p. 34), for example, views hedges as devices representing an absence of certainty, employed to indicate a lack of commitment to the truth value of a proposition or a desire not to express commitment categorically. Similarly, according to Biber et al. (1999) or Wales (2001), hedges qualify the author's lack of knowledge about the truth of a proposition.

In summary, hedges can be described as modifiers of the speaker's commitment to the truth value of the proposition, which means that they typically qualify and tone down utterances to mitigate their strength. The definitions given above also imply that the prototypical hedging devices associated with various degrees of certainty are closely related to epistemic modality. Therefore, our study includes: the epistemic modal verbs *may* and *might* expressing uncertainty; *will* and *must* conveying logical deduction; the past tense forms *could*, *should* and *would* generally associated with tentativeness, and also *can*, which is not normally used epistemically, but is believed to appear in contexts where it can function as a hedge.

As one of our goals is to explore how students apply their knowledge concerning the use of hedges in their bachelor's papers, it is also important to introduce the function of hedges in written academic discourse. Generally speaking, it is difficult to find a consensus among researchers concerning the role of hedging devices in academic texts. Hyland (1994, p. 241) mentions two main reasons for using hedges, namely expressing claims with a certain degree of caution, modesty, and humility and diplomatic negotiation when referring to the work of colleagues and competitors. Salager-Meyer (1994) also discusses two uses of hedging, particularly making claims fuzzy/vague or alternatively "being more precise in reporting results" (p. 151), which enables authors to make the strongest claims possible. Similarly, Clemen (1997, p. 241) mentions increasing the credibility of propositions and protecting authors from

making possible false statements. In sum, hedges seem to be a significant resource for academic writers in that they help them produce credible and convincing claims. Therefore, learners should be aware of the fact that hedging devices allow them to express propositions with greater accuracy, to indicate subjective interpretations, or to soften categorical assertions.

In order to be able to identify modal verbs employed as hedges, we used Hyland's taxonomy (1999), which seems to represent the consensus among linguists; it is well-organised and easier to detect in comparison with other taxonomies. Nevertheless, only two subcategories, reliability hedges and writer-oriented hedges, will be briefly introduced in the following paragraph as these two are relevant for the purpose of our study.

According to Hyland (1999, pp. 166-169), reliability hedges, associated with the strategy of vagueness, typically express the author's tentativeness and assessment of the certainty of the truth of a proposition, which indicates that they are primarily expressed by epistemic modal devices allowing writers to present uncertain scientific claims, conclusions true under certain conditions, speculations, or alternative explanations. Therefore, this category seems to be crucial for our study. On the other hand, writer-oriented hedges, associated with the depersonalisation strategy, are employed, as Hyland claims (1999, pp. 170-177), to diminish the role of the author in the text and thus protect him/her from the consequences of being wrong. This strategy is based on using impersonal constructions or the passive voice and is important for saving the speaker's own face since it limits the damage which may result from categorical assertions. As the impersonal structures (especially passives) often co-occur with modal verbs, it is surmised that this strategy is important in our study as well.

## **4 Methods**

If we assume that learning pragmatics includes both implicit and explicit learning resulting in pragmatic knowledge in a foreign language, in our research we decided to examine the product of that learning, the knowledge attained in the context of formal education (see part 2). To evaluate learners' knowledge of and about hedges, we used questionnaires and analysed selected bachelor's papers.

### **4.1 Questionnaires: results and discussion**

The survey was conducted in the summer term of the 2014/2015 academic year. The instrument was created to evaluate declarative, precise, and verbalised knowledge about the selected pragmatic aspect and included contextualised tasks. The questionnaire

consists of one factual item identifying the phase of the bachelor's paper writing process, three open-ended items, and two closed items identifying learners' perception of hedging. The respective terminology (*hedging*) is not used in the questionnaire in order to reduce the danger of leading questions. The selected sample includes 11 first-year students of English Language Teacher Education who finished their final bachelor's degree qualification papers, and 18 third-year students of English for Business who have not submitted their final papers yet and are in different phases of writing their qualification papers. The questionnaires were administered in face-to-face lessons; the students were not allowed to use any additional materials, and the return of the filled-in questionnaires was 100 per cent.

As mentioned above, in the questionnaire the students were asked to demonstrate their explicit knowledge of language devices related to hedging. In the first part, consisting of open-ended items, the students declare that in academic writing they use the following language devices most frequently (according to their frequency of occurrence): passive structures (17 cases), complex sentences (14 cases), linking words (13 cases), modal verbs (seven cases), and hedging (four cases). Then, when being questioned about language devices toning down the strength of the statements in their final papers and explaining why they use them, the respondents list mainly modals (17 cases), hedges (12 cases), and adverbs (five cases), and state that these devices allow them to express possibility or uncertainty, hypothesise, avoid direct statements, speculate, and propose suggestions. The students' responses indicate that they seem to be aware of the function and uses of hedges in academic writing in that the reasons they mention are identical or similar to those discussed in the relevant linguistic literature (see part 3).

The answers to the last open-ended item, in which the learners were to identify the sections of their bachelor's papers which should be hedged most frequently, demonstrate that they mainly tend to hedge the analytical part (15 responses) and conclusion (11 responses), which is again in accordance with the literature. For example, Salager-Meyer (1994) and Hyland (1995) observe that hedging occurs predominantly in discussion sections, "where claims are made and the significance of results argued" (Hyland, 1994, p. 243). The students' responses also correspond with the results of the analysis, which illustrate that modals in the role of hedges prevail when discussing findings and interpreting examples in the practical parts of bachelor's papers (see part 4.2).

The two closed items of the questionnaire concentrated on the students' ability to distinguish between hedged and non-hedged statements and on identifying typical hedges in exemplary sentences occurring in bachelor's papers. In the first closed item the students were to recognise the sentences that are not hedged. The sample sentences used in this item are the following.

- a) *The answer to the question could be the following.*
- b) *This chapter will summarise some of the important film theories.*
- c) *An example of lexical repetition can be found in the following text.*
- d) *Their English may have progressed to the level at which it serves their particular international communicative purpose.*
- e) *The findings indicate that metaphorical expressions are more frequent.*
- f) *All evaluations must be done systematically and completely.*

The majority of the students correctly selected sentences (f) (23 instances) and (b) (18 instances) as non-hedged. However, as only seven (out of 29) students identified sentence (c) as non-hedged, we might assume that the remaining 22 students perceive the modal verb *can* in passive structures to be a typical hedge. Similarly, the findings of the analysis demonstrate that the modal verb *can* occurs most frequently in the excerpted bachelor's papers (209 instances out of 645) and the majority of its occurrences appear in passive structures (see part 4.2).

The last item of the questionnaire investigated the students' ability to recognise prototypical hedges in the following propositions.

- a) *The following words can function as both nouns and verbs.*
- b) *The item study can be replaced by the item research.*
- c) *The findings could lead to discussion as to how power was distributed.*
- d) *The same order of occurrence could be seen in both text types in the previous chapter.*
- e) *A possible explanation for this may be the gender issues.*
- f) *This might be considered the turning point in the novel.*

In this case the learners were able to identify prototypical hedged examples (sentences c, e, f) without problems: (c) was selected by 21 students, (e) by 24, and (f) by 21. The responses of the students show certain discrepancies in their interpretations/understanding of hedges. While in the previous item they probably considered *can* in passive structures to be a common hedge, in this case the majority of them do not seem to perceive the structure *can be replaced* as being a prototypical hedge (24 students), although the contexts are similar.

The questionnaire results indicate that the learners notice and are aware of the function and use of hedges, at least in terms of their declarative knowledge about hedges, because they are able to verbalise their knowledge accurately (see part 2) in all the questionnaire items except for one, in which the results imply the existence of a gap between the knowledge about hedges and the ability to recognise their functions in the given context appropriately. To question this finding from the perspective of actual language use, in the next step of our research we analysed selected final qualification papers.

## 4.2 Analysis of selected bachelor's papers: results and discussion

As mentioned in part 4.1, the studies by Hyland (1994, 1995) and Salager-Meyer (1994) suggest that it is mainly the discussion sections of research articles that are heavily hedged, and therefore our analysis focused on the occurrence of hedging devices, particularly modal verbs, in the practical parts of selected bachelor's papers. For the purposes of our research we selected 15 exemplary bachelor's papers that had been assessed as excellent in terms of their language proficiency. The example sentences in this section are not edited to illustrate the students' authentic use of English.

The size of the corpus is approximately 79,000 words and it includes 645 occurrences of the modal verbs that were analysed, i.e. *may*, *might*, *can*, *could*, *should*, *will*, *would*, and *must*. Out of the total number of tokens, 378 occurrences (59%) of the modal verbs are employed as hedging devices, whereas the remaining 267 occurrences appear in contexts where they do not seem to function as hedges, CAN being the most frequent (48%).

As the sentences in example (1) illustrate, students typically use the modal verb *can* (a) to refer to particular parts of their papers, (b) to convey ability, or (c) to express existential possibility in general statements.

- 1)
  - a) *An example of complex lexical repetition can be found in the following text.  
The self-designed observation sheet can be seen in Appendix 1.*
  - b) *Pupils can understand a lot more than they can say.*
  - c) *Relative clauses can serve as the means of condensation.*

In these sentences *can* is not interpreted as a hedge in that it conveys non-epistemic meanings and is predominantly found in rather general statements, or when referring to different parts of theses.

The fact that only 59 per cent of the modal verbs have been interpreted as hedges implies that, as Markkanen and Schröder (1997) argue, the modal verbs themselves are not "inherently hedgy but can acquire this quality depending on the communicative context or the co-text" (p. 6). In other words, particular words cannot automatically be identified as hedges in isolation as one form may have different functions, depending on the context. In addition, the amount of hedging seems to depend on the individual language users, which can be proved by the unbalanced use of modals in the papers that were analysed, e.g. the lowest incidence per paper was seven modals, whereas the highest incidence was 44 modals.

As for the modals under analysis that were employed as hedges (see Table 4.2), it is quite interesting that although it is epistemic modality that is primarily associated with hedging (Crompton, 1997; Hyland, 1999), the non-epistemic readings seem to

occur quite frequently in the role of hedges. Considering the frequency of occurrence, students tend to favour especially non-epistemic *can* (21.4%) in passive structures. Then they use the prototypical hedge *might* (17.2%) to express their uncertainty or hypothetical theoretical possibility in those contexts where it is usually interchangeable with *could* (13.2%). Epistemic *would* (16.7%) and *should* (13.5%) are associated with hypothesising, likelihood, or assessing probability based on facts. The prototypical epistemic modal verb *may*, expressing the speaker's attitude to the truth of the proposition, occurs rather infrequently (12.2 %) as the students seem to prefer *can*. However, *may* conveying epistemic possibility typically occurs in the contexts of subjective interpretation and evaluation, whereas non-epistemic *can* expressing possibility is used when discussing possible explanations based on reporting data. The epistemic modal verb *will* is used mainly when predicting outcomes, whereas *must* occurs infrequently as students do not usually express certainty in their theses.

Table 4.2  
*Modal verbs employed as hedges*

Modal verb	No. of occurrences (%)	Example structure
<i>can</i>	81 (21.4)	<i>can be concluded, can be explained</i>
<i>might</i>	65 (17.2)	<i>it might also imply certain beliefs</i>
<i>would</i>	63 (16.7)	<i>the transcript would be helpful</i>
<i>should</i>	51 (13.5)	<i>the meaning should be identical</i>
<i>could</i>	50 (13.2)	<i>could be considered, could be interpreted</i>
<i>may</i>	46 (12.2)	<i>this division may seem irrelevant</i>
<i>will</i>	19 (5.0)	<i>it is expected that the news will contain</i>
<i>must</i>	3 (0.8)	<i>there must be some factors</i>
Total	378 (100)	

As for the most frequently used modal, the findings indicate that *can* typically occurs in passive structures with implied agents, as in (2).

2)

*It can be also stated that the activities created in this thesis are beneficial for*

...

*It can be concluded that the use of the past progressive in the language of basketball commentary is rare due to ...*

Furthermore, this use of *can* may be in agreement with the findings discussed in part 4.1, where the questionnaire results indicate that to hedge their propositions, students tend to employ the modal verb *can* in passive structures. Such structures (e.g. *can be stated, can be concluded, can be explained*) are probably preferred to their active counterparts (e.g. *we can state, we can conclude, we can explain*) because of the students primarily attempting to avoid responsibility for their statements.

Although some authors consider non-epistemic *can* in passive structures to be a hedge in certain contexts (e.g. Fraser, 2010, who discusses so-called *hedged performatives*, which diminish the role of the author in the text), it is generally viewed as a marginal hedging device, particularly a writer-oriented hedge (Hyland, 1999).

As for the typical contexts of occurrence of hedges, example (3) illustrates that the students use the modal verbs when discussing particular examples (62%), when interpreting their findings, or when dealing with hypothetical situations (see Hyland, 1994; Salager-Meyer, 1994). These uses might be considered reliability hedges as they are associated with writers' uncertainty or lack of confidence (Hyland, 1999).

3)

*... which in this case might indicate the inevitability of the action.*

*The results suggest that the non-finite clause may be a possible alternative ...*

*In popular scientific style, the more appropriate paraphrase of this subheading would be ...*

*As the verb want is 'stative', this sentence could be re-phrased as ...*

When we compare these findings with the data obtained from the questionnaires, we can conclude that our students seem to be aware of the reasons why hedges should be used predominantly in the practical parts of their papers. The students who were questioned know that they need to be cautious about their claims when they are uncertain or lack confidence and realise that, by using hedges, they can communicate that their conclusions are tentative or that they convey subjective interpretations (see part 4.1). However, that overt knowledge appears to be mainly declarative since the analysis of the actual language use suggests that students tend to prefer primarily non-prototypical hedging devices; they seem to overgeneralise the use of modals in the role of hedges, and therefore they can fail to hedge their statements adequately and appropriately in the context of academic writing.

## 5 Conclusion

One cause of the inappropriate use of hedges might be country- and language-specific with regard to pragmatics and grammar and attitudes to them. This is also supported by Hyland's claim about the culture-specific features of academic writing (1994, p. 244). For example, Bloor and Bloor (1991) observe that "Czech research articles display a more direct, unhedged style" (in Hyland, 1995, p. 39). Those culture-, language-, and country-specific bounds appear to be crucial in the processes of second or foreign language acquisition in terms of learners' inability to be "ambiguity tolerant" (Lojová, 2015).

Other reasons might include the still preferred form-focused teacher-centred approach in instruction in which the awareness of grammatical accuracy in foreign

language learners seems to be overemphasised at the expense of providing opportunities for the systematic development of pragmatic awareness (Bardovi-Harling & Dörnyei, 1998; Lojová, 2015). Moreover, the use of modals might be determined not only by the particular contexts of occurrence and the writer's intention, but also by the subjectivity of the authors and individual styles in the use of hedges, indicating that higher pragmatic awareness does not necessarily result in "appropriate pragmatic production" (Bardovi-Harling & Dörnyei, 1998, p. 254). In conclusion, the appropriate use of hedges is clearly part of a language user's pragmatic competence, i.e. their ability to communicate the intended message effectively. The above-mentioned tentative conclusions open up space for further empirical research and implications in pedagogy in terms of the areas suggested above.

## Chapter 5

# BUILDING AND ANALYSING AN ADVANCED ENGLISH LEARNERS' WRITTEN CORPUS

*Silvie Válková, Jana Kořínková*

## 1 Introduction

As Soars, Soars, and Sayer (2003) state in their introduction to the New Headway Advanced Teacher's Book, "For some advanced students, their language production abilities have ossified. They might well acquire new vocabulary, but by and large they manage to avoid grammatical areas they are unsure of. By a process of circumlocution, they can restrict themselves to tried and tested phrases" (p. 4). In the course of our teaching practice, we have noticed that our advanced students' writing often fails to manifest qualities of "good writing" or "a neatly woven texture" (Halliday & Hasan, 1976) that is of smoothly flowing cohesive and coherent texts. Instead, the students tend to rely on rather simple familiar structures with a strong preference for finite verb forms, coordinated clauses, and a limited range of vocabulary.

The purpose of our long-term research project, whose partial results are presented in this paper, is to understand and determine the needs of advanced English language learners at the Department of English of the Faculty of Education, Palacký University in the field of developing their writing skills. Specifically, our research aims to compare frequencies of selected linguistic features in native and non-native English texts, that is, it explores the quantitative usage of some grammatical structures in the corpus texts written by Czech speakers of English (CES) with an equivalent corpus of native English speakers texts (NES). The results of the analysis should provide us with specific and detailed information about which features to target for in pedagogical intervention in teaching both the theory of the English language (various linguistic disciplines) and the practical language usage (language practice classes and language skills classes).

## 2 Building the advanced English learners' written corpus

A learner corpus is characterised as a corpus of a particular language produced by learners of that language (Baker, 2011, p.100). In 2013 we started building our own

learner corpus which would allow us to identify common errors and which would show us over- and underuses of lexis or grammatical structures when compared to an equivalent corpus of native speaker language. In our first quantitative analysis (Kořínková & Válková, 2013) we compared 50 pieces of formal and semi-formal writing focusing on syntactic complexity of native and non-native English writing, with the additional focus on the usage of finite and non-finite verb forms. The successive analysis focused on the complexity of the noun phrase, which was analysed in the corpus of 30 native and non-native texts. In our efforts, we were encouraged by the fact that “if a small difference or no difference is found then this is still a finding” (Baker, 2011, p. 102). The method of text analysis has been primarily based on manual counting of selected grammatical structures and items and categorising their functions and syntactic complexity.

### **3 Summary of our previous research**

#### **3.1 Syntactic complexity**

In this initial quantitative analysis 25 pieces of formal and semi-formal writing (essays, reports, and formal letters) produced by CES were compared to 25 texts by NES. The latter were sample texts and examples of good practice published in three advanced level coursebook writing sections: New Headway (Soars et al., 2003), Advanced, Landmark Advanced (Haines & Stewart, 2002, and Fast Track to CAE (Stanton & Morris, 1999). The CES texts were produced by Olomouc third year students of English under examination conditions as final essays of their three-year bachelor study programme. The two research samples proved to be very well comparable in size as both the number of words and sentences were similar.

The results showed that, on average, Czech students tend to use somewhat longer sentences (the average sentence length in words was 16.8 in NES and 18.2 in CES), although this may have been influenced by the fact that some Czech students used excessively long sentences which were sometimes too complicated and difficult to understand. The simple sentence was used more frequently by native speakers, although the difference was smaller than expected (33% of all sentences in NES and 28% in CES). In both research samples, about 50 per cent of all multiple sentences consisted of two clauses. In the CES sample, some instances were found of extremely long multiple sentences consisting of 7, 8, and even 9 clauses. As for the types of the multiple sentence, the most frequent type proved to be the complex multiple sentence in both research samples (72% of all multiple sentences in NES and 68% in CES). Czech students tend to use more compound sentences (14%) than native English speakers (9%), the remaining compound-complex type being represented almost equally (19% in NES and 18% in CES).

In texts written by Czech students, a significantly higher incidence of coordinators was found (the total of 60 in NES and 99 in CES), which is in accordance with the higher incidence of compound multiple sentences. As for subordinators, Czech students tend to rely on several basic subordinators (*if, when, because*) while the NES texts displayed a greater diversity in this respect.

### **3.2 Finite and non-finite verb forms**

Part of our analysis of syntactic complexity was a study of the typological and structural varieties of different clause types (Kořínková & Válková, 2013). In the same corpus material (see above) we found only moderate differences in the distribution of the three basic dependent clause types (i.e. nominal, relative, and adverbial) with relative clauses being slightly more common in NES (40% compared to 37% in CES) and nominal clauses slightly more common in CES (31% compared to 27% in NES). Greater differences were, however, identified in the incidence of the individual structural varieties of the dependent clause types: Czech students seem to prefer the finite varieties over the non-finite ones (i.e. infinitival and participial clauses). Non-finite clauses comprise 41 per cent of all dependent clauses in NES but only 23 per cent in CES. Substantial differences in the usage of the non-finite verb forms were also found in their syntactic position. In NES, 23 per cent of infinitives, 62 per cent of present participles, and 26 per cent of past participles were used adnominally (i.e. as modifiers in the noun phrase), whereas in CES the adnominal position was filled by only 5 per cent of infinitives, 38 per cent of present participles, and 7 per cent of past participles. This, in our opinion, demonstrates that for some reason Czech students of English use non-finite verb forms more comfortably in nominal and adverbial clauses than in relative clauses, where they form the postmodifying part of the noun phrase.

### **3.3 Complex noun phrase**

The results of our initial analysis, namely the usage of non-finite verb forms in adnominal position, inspired us to further analyse the structure of the complex noun phrase in NES and CES writing. The total of 30 texts was analysed, 15 written by advanced Czech English speakers and 15 by native English speakers, both samples being comparable in size. The latter were again selected from advanced English language coursebooks (see above). As for the genre of the texts, only reports and reviews were selected for the analysis, as these two genres are primarily descriptive and thus provide rich opportunities for using the complex noun phrase.

Our analysis focused on the overall incidence of the complex noun phrase in the two research samples, the length of the complex noun phrase, the proportion of

bare and complex noun phrases, the incidence and distribution of premodification and postmodification, and the incidence of selected types of premodification and postmodification. The results can be summarised as follows: in the Czech advanced students' texts the complex noun phrase tends to be less frequent (772 noun phrases were found in NES and 621 in CES), less complex (Czech students favour short one- and two-word noun phrases although some extra long noun phrases were also found), and the head nouns are more frequently premodified than postmodified. In premodification, Czech students use fewer adjectival nouns, adverbs, and participles; in postmodification they use fewer prepositional phrases and infinitives, and more finite relative clauses.

#### **4 Limitations of our corpus/research**

We are aware of the fact that our limited research data still lack some of the typical features of a corpus; namely our corpus is not fully representative of the variety under examination. Also, it has not existed in "machine readable" form, which required manual counting (McEnery & Wilson, 1996). Another problem we are aware of is the source of native English speakers' texts – for our past research we used texts published in coursebooks, which may have been edited and hence not fully authentic. We are still in the process of building a larger corpus of our advanced students' writing and we are searching for faster and more accurate techniques which would allow us to process larger amounts of data and identify patterns that we would not have noticed otherwise. One step in this direction was our decision to analyse part of our corpus with the computational tool Cohmetrix and compare the results with the outcome of our previous manual analyses.

#### **5 Using Coh-Metrix**

Coh-Metrix is an online computational tool developed by scholars at the University of Memphis, USA. It is freely available to all researchers whose aim is to analyse large stretches of language on various measures of textual cohesion, language features and readability in a short time and with minimum effort. It makes use of existing models on computational linguistics (Graesser et al., 2004, p. 201), such as parsers, lexicons, and word frequency databases. The current version, Coh-Metrix 3.0, calculates measures of 108 indices providing word information (word concreteness and word familiarity, average age of acquisition, part of speech incidence scores, type-token ratio), sentence information (sentence length, syntactic complexity, incidence of various syntactic structures, incidence of connectives), and other types of information about texts whose similarities and differences can then be exposed in a matter of minutes.

30 texts (book and film reviews) written by Czech advanced students of English were selected from our corpus to be compared with 30 authentic book and film reviews randomly downloaded from various British and American websites. First, five texts from both NES and CES were analysed with the aim to identify those Coh-Metrix indices which might potentially show different measurements in the two samples. Then, 41 indices in total (not all of them related to our previous research) were selected as the most promising ones and computed in all 60 texts.

### 5.1 Coh-Metrix indices relevant to previous syntactic complexity research

Index 6 DESSL shows the average number of words in each sentence in the analysed text. Index 7 DESSLd calculates the standard deviation of the measure for the mean length of sentences within the text, where higher values of standard deviation indicate larger variation of the sentence lengths within the analysed text (Cohmetrix version 3.0, 2014). Index 15 PCSYNp measures the syntactic simplicity of the text, taking into account the sentence length and complexity, as well as variability of syntactic patterns. The following Table 5.1 shows the minimum and the maximum value of indices 6, 7, and 15 in NES and CES, together with the percentage difference in the mean values of the indices in CES compared to NES.

Table 5.1  
*Sentence length and complexity*

Index	NES min-max	NES mean	CES min-max	CES mean	Difference in CES
6 DESSL	10.5-38.6	21.5	10.6-32.8	16	-25.4%
7 DESSLd	5.8-25.1	10.1	4.2-15.5	8.1	-21.2%
15 PCSYNp	1.1-65.5	18.4	2.3-85.5	41.1	+123%

The data seem to confirm our previous findings that the range of average sentence length in NES and CES is not remarkably different. What is different, though, is the mean value, which was 21.5 words in NES and 16 words in CES. It is also clear that within one text in NES the sentence length is far more variable, which suggests that native speakers use more diverse sentence types and structures. The most different feature in the compared research samples proved to be syntactic simplicity, which was remarkably higher in CES. The tendency to use shorter, less complex and less variable sentences is further illustrated by the output of Index 72 SYNSTRUTa and Index 73 SYNSTRUTt, which measure syntactic structure similarity of sentences in the analysed text. This is achieved through calculating the proportion of intersection tree nodes between all adjacent sentences and all sentences/across paragraphs respectively (Cohmetrix version 3.0, 2014). Table 5.2 below indicates that in texts

written by Czech advanced users of English, the syntactic similarity of adjacent sentences as well as of all sentences within the text is more prominent.

Table 5.2  
*Syntactic similarity*

Index	NES min-max	NES mean	CES min-max	CES mean	Difference in CES
72 SYNSTRUTa	0.03-0.19	0.068	0.05-0.15	0.086	+26.5%
73 SYNSTRUTt	0.03-0.14	0.067	0.05-0.15	0.081	+20.7%

Coh-Metrix 3.0 also includes indices that can be used to calculate the incidence of various categories of connectives per 1,000 words of the text. Here, our findings can be matched with our previous manual research just partially, since the tool identifies only general categories (logic operators, additive, adversative, causal, and temporal connectives) but not the representation of concrete individual connectives, which means that their variety cannot be assessed. The overall incidence of all connectives was found to be almost identical, the mean value in CES being only 4 per cent higher than in NES. What is interesting, though, is the higher incidence of logic operators (Index 52 CNClogic) and of all semantic categories of connectives (additive: Index 56 CNCAdd, adversative: Index 53 CNCADC, causal: Index 51 CNCCaus) except of the temporal ones (Index 54 CNCTemp). Although the afore mentioned preference of our students for coordination over subordination cannot be directly ascertained, we can conclude that their writing seems to be more explicit due to the higher incidence of conjunctions (logic operators) explicitly expressing relationships between clauses and sentences. The differences are shown in Table 5.3 below.

Table 5.3  
*Incidence of connectives*

Index	Mean NES	Mean CES	Difference in CES
52 CNClogic	33.1	39.4	+18.7%
56 CNCAdd	49.1	51.7	+ 5.2%
53 CNCADC	16.4	20.9	+27.9%
51 CNCCaus	20	26.9	+34.9%
54 CNCTemp	17.5	14	- 20.4%

## 5.2 Coh-Metrix indices relevant to previous verb form research

Index 83 WRDVERB calculates the incidence of verbs per 1,000 words and index 75 DRVP the incidence of verb phrases. In both cases the mean value of the computed output was higher in CES (by 11.9% and 6.6% respectively), which confirms our

previous conclusions that Czech advance students of English, perhaps influenced by their mother tongue, favour verbal structures over nominal ones.

Our previous results concerning the usage of finite and non-finite verb forms could not be fully confirmed, even though Coh-Metrix 3.0 includes two indices calculating their incidence. Index 81 DRINF measures the incidence of infinitives, while Index 80 DRGERUND claims to measure the incidence of gerunds. However, through referring back to the analysed texts, we found that any -ing form not categorised as an adjective had been classified as a gerund. Therefore, the output of Index 80 must be understood to represent the incidence of gerunds and present participle forms indiscriminately. We were much surprised by the higher incidence of infinitives in the CES sample, which seems to directly contradict the conclusions of our prior research. It is worth noticing (see Table 5.4 below), however, that the range of DRINF values is much wider in CES, which could suggest that the result may have been influenced by extreme values rather than the overall tendency. On the other hand, the incidence of verbal -ing forms was found lower in CES, which is in agreement with our previous findings.

Table 5.4

*Incidence of non-finite verb forms*

Index	NES Min-max	NES mean	CES Min-max	CES mean	Difference in CES
81 DRINF	3.7-29	13.2	0-41	15.6	+18.7%
80 DRGERUND	0-44.2	15.5	0-52	12.4	-20.2%

**5.3 Coh-Metrix indices relevant to previous noun phrase complexity research**

Index 82 WRDNOUN calculates the incidence of nouns per 1,000 words and index 74 DRNP the incidence of noun phrases. Surprisingly and contrary to our previous research, the incidence of noun phrases in NES and CES was determined as almost identical. The incidence of individual nouns was slightly lower in CES, which again seems to reflect the general preference of Czech users of English for verbal rather than nominal structures. Index 68 SYNNP measures the mean value of modifiers per noun phrase. Here, the Coh-Metrix output yet again supports our past findings suggesting our students' preference for shorter, less complex noun phrases.

Table 5.5

*Noun phrase incidence and complexity*

Index	NES min-max	NES mean	CES min-max	CES mean	Difference in CES
82 WRDNOUN	193-379	272	170-339	255	-6.2%
74 DRNP	292-427	362	302-417	360	-0.3%
68 SYNNP	0.6-1.3	1.1	0.6-1.3	0.9	-16.3%

## 6 Conclusion and future research

Although analysing texts with the help of Cohmetrix is far less time consuming than manual analysis, it has its limitations. It provides precise information on objective parameters of texts (sentence length, word length, lengths of paragraphs, etc.) but in other less objective aspects, the results may be less clear or more difficult to interpret. First of all, as authors of Cohmetrix confess, the calculated results “are often subject to the output of third party parsers, lexicons and word frequency databases, all of which are outside of the control of CohMetrix” (Cohmetrix version 3.0 indices, 2014). Another issue is linguistic terminology, which is far from being universal, therefore if one decides to follow e.g. the incidence of gerunds, logical operators or other language items which may be defined/labelled differently by various linguists, it is necessary to verify the meaning of the terms used in the context of Cohmetrix. Finally, Cohmetrix naturally identifies occurrences of language items and structures in general categories, which means it will not help to disclose differences in the usage of specific individual items, e.g. specific connectors, personal pronouns, prepositions.

It can be concluded that although our evidence of quantitative differences in the texts of Czech learners of English in comparison with the texts written by native speakers is not yet robust, we find the results a useful basis for our further research. In the future we plan to keep enlarging our advanced learners’ corpus both in size and different genres of writing. We would like to investigate other computational possibilities besides Cohmetrix, which would enable us to obtain fast and precise information on various parameters of texts. We would also like to focus on other areas of grammar that showed different frequency of usage in our so far researched samples (personal pronouns, the passive voice, adverbial phrases). Finally, it is important to emphasise that the ultimate goal of our research is practical – to apply the results in teaching English to Czech advanced students.

According to the CEFR (2001), a proficient user, i.e. on C1 level, “consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot” (p. 114). When we take this presupposition as a starting point, then following a fixed syllabus covering all areas of grammar with the same importance seems to lose its justification. We believe that it is more reasonable to focus on those areas of complex structures that proved to be problematic (avoided or overused) for Czech advanced learners in our research. In theoretical disciplines, discussion of various problematic forms and working with authentic and learner texts might be more relevant than the usual presentation and practice sequence. In practical disciplines, instead of completing course book tasks designed for students regardless of their nationality, specifically focused tasks may be more challenging and effective for our students. We hope that building and analysing the corpus of advanced Czech

students' English writing will ultimately lead to designing new teaching materials reflecting this approach.



## Section Two

# THE USE OF CORPORA FOR L2 LEARNING

Section Two comprises four chapters dealing with different ways of describing corpora which might, to a lesser or larger extent, become a tool for intervention into L2 learners' interlanguage. In this section you can find authors who not only describe the corpora they compiled themselves from different sources, whether written native sources of English (Hasse), written non-native sources of English (McKenny), or both written and spoken on one hand and native sources and non-native sources of English on the other hand (Bohát et al.), but also suggest specific interventions in L2 acquisition through form-focused instruction or strategy and learning skill training. Roca-Valera's concluding chapter utilises a sub-corpus of an official non-native corpus of Spanish learners to gain insights into the field of Spanish-English false friends and to suggest lexical strategy training.

In the first chapter Christoph Haase introduces a number of approaches to the quantitative evaluation and profiling of native-speaker academic written texts in English. He argues that a major problem for practitioners and students of academic writing is the appropriate balance between a subjective display of the individual research and an objective display of the results obtained. The problem is exacerbated by the fact that there are no unique and specific linguistic markers of this balance. Here he presents the results of a corpus-based research project in which specialised academic texts are juxtaposed with popular academic texts. The main difference measured is not in content but in markers of different levels of lexicalisation, conflation, subjectivity, and objectivity. To this end, texts from diverse natural sciences (physics and biosciences) have been compiled into a corpus (the Scientific and Popular Academic Corpus of English, i.e. the SPACE corpus). The research is based on the parallel structure of the corpus. His contribution includes a presentation of this corpus, its annotation, and how it was tagged for expressions of subjectivity. Finally, he makes suggestions on how to incorporate this theoretical insight into academic writing course modules and into the study of English for Academic Purposes (EAP).

In the second chapter Róbert Bohát, Nina Horáková and Beata Rödlingová present their learner corpus, called COHAT (Corpus of High School Academic Texts), which

is currently being developed at the International School of Prague (ISP). It is designed to be a bank of both spoken and written texts for data-driven and student-centred discovery learning with room for IT integration, critical thinking, and metacognition. While they appreciate the value of error detection and the study of interlanguage, they also try to use their learner corpus to detect what their students, native and non-native speakers of English, did well in their academic writing or presentations. In their study they present the first set of tested discovery learning activities and worksheets based on COHAT. The topics of research activities are subjectivity, hedging, keywords, collocations, and grammar. The worksheets or *language discovery lesson plans* are designed to help students perform the data collection in the corpus and then generalise their findings into trends or language rules. They argue that their experience shows that students in the lower grades of high school can exercise their critical thinking skills in learning academic English. At the same time, they envision an expansion of COHAT through teacher texts and primary sources of native English as well.

In the next chapter, John McKenny investigates how the compilation and interrogation of a specialised corpus of successful Master's dissertations might help advanced learners of academic English to write their own dissertations. The corpus in question is the British University in Dubai (BUiD) Corpus of Written Academic English (Buidcorp), which is a corpus of non-native writers of English. Over the past two decades various corpora of learner English have become available (e.g. the Cambridge Learner Corpus and the Chinese Learner English Corpus). Individual studies have focused particularly on dissertations/theses written by learners of English (e.g. Charles, 2006; Kwan, 2006; Thompson, 2005; Isik-Tas, 2008). There was a felt need for a focus on dissertations completed by research students in the Gulf region. The principal aim of the Buidcorp project was to meet this need and further develop the corpus as a useful teaching and research tool for FL learners. These learners are taught how to use different corpus tools (Wmatrix3, Wordsmith Tools 6) in order to explore various linguistic features of L2 interlanguage presented in Buidcorp in comparison with BAWE (British Academic Written English).

In the final chapter Maria Luisa Roca-Varela maintains that English people are used to hearing non-native speakers using English. It is not rare that second language learners mix words up in the foreign language, especially when their spelling in different languages is similar or identical, but their meaning is not. In most situations native speakers are fully aware of this fact. Native speakers often notice that non-native speakers fail to differentiate certain words in English from time to time; meanwhile learners strive to overcome their lexical difficulties to keep communication going. Learning a foreign language is a complex task which involves time and effort, trial and error, seeking, blundering, and learning from mistakes. The author tries to give us some clues to understand the peculiarities of learner vocabulary better. She examines the most remarkable lexical problems in learner language and also provides particular examples showing those problems. She extracted the data from the International

Corpus of Learner English (ICLE), a learner corpus which constitutes a reliable source for the study of non-native English and provides a number of examples of negative transfer in the field of false friends in English and Spanish. The analysis of this corpus reveals the recurrent misuse of some high-frequency English words which might be an important downside to foreign language acquisition and use on the part of Spanish learners. She finishes her contribution to corpus linguistics with some pedagogical considerations for the EFL classroom and pinpoints some issues for further research.



## Chapter 6

# STRATA OF ACADEMIC ENGLISH IN CORPORA: A PARALLEL-GENRE APPROACH

*Christoph Haase*

## 1 Introduction

This contribution will introduce a corpus project which has been pursued in the last few years and which is currently hosted at the English department at Jan Evangelista Purkyně University. It connects the study of English for Academic Purposes (EAP) to a cognitive linguistic perspective but it is also informed by methods from the natural sciences. Beginning with the design of a specialized corpus, it was a consequential development that a research interest arose in linguistic markers of academic language, especially in the language of the natural sciences. While the humanities and social sciences have different standards in academic writing, it is helpful to remind students of these subjects that the yardstick of academic work is in the hard sciences and the application of rigorous argumentation and logic enhances rather than impedes the “softer” humanities and social sciences. The question whether the actual research in the sciences is dependent and perhaps enhanced by the appropriate language used, is a long-standing one and one that has received a substantial amount of interest from linguists, especially in the last few years. From the side of corpus linguistics, a major influence is Biber’s work on registers (Biber, 2006) while the focus has been firmly on the lexico-grammatical features (Biber & Barbieri, 2007, p. 263) and the study of lexical bundles has led to a large number of academic styleguides and wordlists, sometimes of the very applied “for Dummies” type (e.g. Olsen, 2010), sometimes from the teaching perspective (Basturkmen, 2010) and often informed by linguistic studies (Ruiz-Garrido, Palmer-Silveira, & Fortanet-Gómez, 2010). The following study is located in this tradition but at the same time it owes much to the genre approach by Swales (Swales, 1993; Swales & Feak, 1994, 2009) in that it is not only interested in the academic styles as such but also in the different strata that exist within the genre. The strata obtained in the corpus analysis can then in an additional step be used to make linguistic strategies of the authors transparent for teaching. Therefore, in teaching academic writing, an appropriate starting point of any class on academic writing is a display of the actual practice of writing.

## 2 The characteristics of EAP and the implications for teaching

### 2.1 The critical perspective

In displaying texts of the following type, the student awareness can be raised by taking the detour of what can be considered a type of shock treatment in order to educate critical thinking about the genre and the “cultivation of an empirical perspective on texts” (Canagarajah, 2002, p. 3). The following example is a good indicator of how to cultivate this type of thinking by showing it in the EAP classroom:

*0104PN Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants*

Phylogenetic relationships within Plantaginaceae were determined from a 4,730-nt data set consisting of portions of four chloroplast regions (ndhF, rbcL, and intergenic spacers atpB rbcL and trnLtrnF). Relationships within *Plantago* subgenus *Plantago* were analyzed from a 9,845-nt data set containing two additional chloroplast regions (intergenic spacers psaAtrnS and trnCtrnD). Maximum likelihood (ML) trees were constructed with PAUP\* by using the general time-reversible model, a gamma distribution with four rate categories, and an estimate of the proportion of invariant sites. The rate matrix, base frequencies, shape of the gamma distribution, and proportion of invariant sites were estimated before the ML analysis from a neighbor-joining tree constructed from the data. Divergence times outside Plantaginaceae were taken from ref. 27. Those within the family were calculated by using a penalized likelihood approach (28) as implemented in the R8S program (29) and a time constraint of 48 million years. [...]

After giving the students a few seconds of processing of this text, the response is usually that such example practice creates considerable anxiety among the students as they assume this is the kind and caliber of text they have to produce, e.g. for their linguistics classes. In consequence, however, and after some detail discussion, we come to agree in the classroom that this is not the correct field of knowledge for the students, and, in a concrete application of genre theory (Swales, 1993) in which the reader is as much part of the creation process of an academic text as the author, it becomes clear that either the reader or the author is “in the wrong spot”. In the case of the text (which is taken from the field of genetics) it is somewhat natural that a student of linguistics (or the instructor, for that matter) does not understand it. This leaves to discuss if either the author has made a mistake in writing for the wrong readership – one that follows implicit and explicit expectations to the text type – be it topical or also within the tolerance span of expected and conventionalized lexicalization patterns – or the reader has made the mistake in ignoring clear markers of the text type (the journal in which it appears, the title that indicates who might be the target group, keywords or an abstract given as descriptors, etc.). It is thus an opportunity to remember what characterizes the different respective text types and it can lead to a classroom discussion why a reader does not automatically understand

any text given that it is written in a language he or she is competent in. So returning to the initial question, what is difficult about this text? After all it is written in English and a student readership should be able to process it. When we continue by looking at the lexical items, the students usually agree that it is the nominal expressions that create problems for students. This understanding is exacerbated by displaying a counterpart to this text which looks as follows:

*0104NS Plant DNA shows speedy changes*

The mitochondria of a group of nondescript flowering plants contain the fastest-evolving DNA yet known.

Until now, the mitochondrial genomes of plants were thought to evolve slowly. But when Jeffrey Palmer and colleagues at Indiana University in Bloomington compared mitochondrial DNA from nine species of plantain (members of the genus *Plantago*) and 41 other plants, they found that some *Plantago* sequences changed [...]

Obviously, this is a popular science text that corresponds in content to elements of the specialized text. And after that students tend to agree that the second text is understandable it is processable for people with perhaps some academic education but maybe not education in the special field of the first text.

## **2.2 A possible consequence of text-type specific characteristics**

Both texts come from the corpus introduced here – which is called SPACE (Scientific and Popular Academic Corpus of English) (see part 3). The corpus shows the common features of academic writing in which the nominal (i.e. complicated) style prevails. In direct comparison we can see that semantic information is densely packed in the first text. Processes and operations appear highly specialized and morphologically conflated. The text is written exclusively in the passive voice. Of course, all these are the known parameters of academic texts but at the same time it is interesting to see also that in the natural sciences especially the author/s is/are striving for objectivity. They are further very reluctant to include obvious markers of involvement or author commitment in the way a researcher commits to their research results:

*0082PN* GFP expression observed in the gustatory neurons of the labial palps and leg tarsal segments (Fig. 1 C and D) was suppressed by targeted GAL80 expression (data not shown), as expected from the previous observation that the 3.3-kb *Cha* regulatory DNA directs gene expression in most if not all chemosensory neurons in the peripheral nervous system (23, 41). Concomitant with the further restriction of the GAL4 activity in C309 by the *Cha*3.3kb-GAL80 construct, the temperature-induced courtship chain formation and head-to-head interactions were suppressed completely. [...]

As can be seen, there is little space for subjectivity in natural sciences and while this is intended to give an objective account of the author's involvement or participation

in the study, the author commitment is often stereotypically lexicalized by for example modal expressions like modal auxiliaries, modal adverbs, or factive verbs:

*AX0039* indicate a presence [...] well within [...] current observation bounds could cause early star formation at a level sufficient to explain the high reionization redshift [...]

As a result, we can determine the relevance of the genre parameter for academic writing: academic writing happens at different strata, the most obvious within the sciences are the specialized and the popular-scientific strata. Therefore, to juxtapose both strata was the initial idea of the compilation of the SPACE corpus.

### **3. The Scientific and Popular Academic Corpus of English (SPACE)**

#### **3.1 Structure and sections**

The present size of the corpus is approximately 1.5 million words distributed across its defining characteristic: the parallel structure of an academic section and a popular section which is superimposed on a dual structure of a “hard” science section (physics) and a “soft” science section (biosciences). The corpus has been described at length elsewhere (Haase, 2008, 2013, 2014) so the following will be a short summary for reasons of brevity.

The apparent relevance of strata in itself is not a novel idea, for example, the Michigan Corpus of Upper-Level Student Papers (MICUSP) collects student papers of the upper-intermediate stratum but the texts differ substantially depending on the skills and competences of the students (Römer & Swales, 2010). However, in SPACE, the two sections correspond to each other because they come from parallel publications or rather from an original publication and a subsequent popular-science summary of the same content. The origin of the original research papers was retrieved from preprint servers, for example *arxiv.org*, mainly for the physics texts, and from the Proceedings of the National Academy of Sciences (PNAS) as a source for the bioscience section. All popularized versions come from the UK-based popular-science journal *New Scientist* which routinely publishes the output of science journalists who write simplified versions of original research papers. The parallel popular papers were retrieved from the *New Scientist* website which was made possible via a regular subscription to the journal. The backbone of this parallel structure of the corpus in practice can be found in the following table.

Table 6.1  
*Domains and word counts in the SPACE*

Subcorpus	Descriptors	Word count
arXiv	physics, astrophysics, quantum mechanics	809,320
New Scientist – physics	physics, astrophysics, computer science, quantum mechanics	203,470
Proceedings of the National Academy of Science (PNAS)	biochemistry, genetics, genetic engineering, microbiology	267,105
New Scientist – biosciences	biochemistry, genetics, genetic engineering, microbiology	30,499
Public Library of Science – medicine (PLoS)	medicine, virology, clinical psychology, public health	217,254
New Scientist – medicine	medicine, virology, clinical psychology, public health	17,050
total		1,544,698

When comparing the domains, we can see that the *arXiv* physics component (corpus file code: AX) and the PNAS bioscience component (corpus file code: PN) are the largest components. In comparison, for example, this means that academic physics is represented with ca. 800,000 words while the popular physics component with texts from the *New Scientist* (corpus file code: NS) only has 200,000 words. Overall, the popular versions are always shorter by a factor of 0.2 which has to be taken into account in the data analysis which is based on normalized amounts. Further, for teaching it is important to note that the popular-academic versions resemble much more student writing than the specialized-academic texts.

A look at the structure shows the ontology of the biosciences in PNAS as an alphabetical, 1-tier list with 17 descriptors of biological sciences: agricultural sciences, biochemistry, cell biology, developmental biology, ecology, evolution, genetics, immunology, medical sciences, microbiology, neuroscience, pharmacology, plant biology, population biology, psychological and cognitive sciences, sustainability science, and systems biology. For the purposes of SPACE, it was decided on a much simpler ontology with three components (from the very small (microbiology) via molecular biology (genetics) to the super small (biochemistry)). The physics ontology at *arXiv* shows that the field of physics is extremely diversified and includes descriptors for 153 categories (not given here for reasons of brevity). In the corpus the decision was made to simply include three categories: quantum mechanics, particle physics, and cosmology.

### 3.2 A parallel look at popularization, conflation and lexicalization

With the parallel structure of the corpus it is possible to measure an information “trickle-down” from the academic to the popular register which very often involves (linguistically) complex strategies of simplification which we call popularization and which uses conflation of the overall text (as seen in part 3.1 where text length was compared) but also conflation of the basic idea or title into an attention-grabbing popular title.

Table 6.2

*Parallel titles in physics and biosciences (AX and PN for academic, NS for popular)*

Code	Title
physics	
0019AX	Does the Rapid Appearance of Life on Earth Suggest that Life is Common in the Universe?
0019NS	Alien life gets more probable every day
0020AX	Flat manifold leptogenesis in the supersymmetric standard model
0020NS	Quantum foot in the door
0022AX	Neutrino-induced Collapse of Bare Strange Stars Via TeV-scale Black Hole Seeding
0022NS	Mini black holes do the monster munch
0023AX	The Ultimate Fate of Life in an Accelerating Universe
0023NS	Never say die
0033AX	Measuring the Small-Scale Power Spectrum of Cosmic Density Fluctuations Through 21 cm Tomography Prior to the Epoch of Structure Formation
0033NS	Cosmic dark age found in shadows
0034AX	The last Gamma Ray Burst in our Galaxy? On the observed cosmic ray excess at particle energy 1018 eV
0034NS	Ancient explosions pelt earth with shrapnel
biosciences	
0049PN	Life extension in Drosophila by feeding a drug
0049NS	Keep young and beautiful
0051PN	Spirochete and protist symbionts of a termite ( <i>Mastotermes electrodominicus</i> ) in Miocene amber
0051NS	Tiny fossil has guts
0056PN	Hermaphroditic, demasculinized frogs after exposure to the herbicide atrazine at low ecologically relevant doses
0056NS	Thought of food
0062PN	Elevated mutation rates in the germ line of first- and second-generation offspring of irradiated male mice
0062NS	Suffer the children
0069PN	Biosilica formation in diatoms: Characterization of native silaffin-2 and its role in silica morphogenesis
0069NS	Natural glass

The mirroring effect is sometimes surprisingly close to the original, so, for example, the “last gamma ray burst” (0034AX) becomes an “ancient explosion” that “pelts

earth with shrapnel” (0034NS) (which, physically speaking, is a “cosmic ray excess” (0034AX). “Elevated mutation rates ... of ... offspring” (0069PN) is a more complicated way to say “Suffer the children” (0069NS). In the following comparison, the academic paper talks about “possible observational facts” and popular version says “we should be able to spot”, which is essentially the same meta-content. Therefore, if the original idea of the corpus was that the science journalists only process the content and then write a summary, it turns out they actually also mirror the syntax the description of procedures of the original paper, albeit in a shortened form and in a less formal register (in which “wormhole structures” become a “superfast transport network”). This can be seen in the following text sample:

*0008AX* This paper continues our study on the *possible observational effects* that struts of negative masses would produce if they are isolated in space. Since *wormhole structures* require the violation of some of the most sensitive energy conditions at the wormhole throat, wormholes are natural candidates –if they exist at all– for stellar size negative mass objects. [...]

*0008NS* If sophisticated aliens are commuting across the Galaxy using a *superfast transport network*, *we should be able to spot* the terminuses. A multinational team of physicists has shown that “wormholes”—gateways to distant regions of space—should stamp a coloured hallmark on light from distant stars as it travels past them on its way to Earth. [...]

Observations like this can be systematized throughout the corpus and therefore give a valuable aid for learners and practitioners of academic writing at student level as it raises the awareness of text-type specific registers which are independent of the factual content (and thus the linguistics of academic writing can be deduced from its purest form – in the natural sciences – and transferred to the social sciences and humanities, raising the standards there as well.

In the following Figure 6.1, the students were asked to find the described parallel relationships and provide some contextual markup.

#### Virtual Black Holes in Hyperbolic Metamaterials

Optical space in electromagnetic metamaterials **(a) may be engineered to emulate** various **(b) exotic space-time geometries.** However, these **(c) metamaterial models are limited in many respects.** It is believed **(d) that real physical space-time strongly fluctuates on the Planck scale.** These fluctuations are usually described as virtual black holes. Static metamaterial models introduced so far do not exhibit similar behavior. Here we demonstrate that thermal **(e) fluctuations of optical space** in hyperbolic metamaterials lead to creation of virtual electromagnetic black holes. This effect is very large if the dielectric component of the metamaterial exhibits critical opalescence.

Recent breakthroughs in nanofabrication enabled rapid progress in development of novel electromagnetic metamaterials [1], which gain their properties from structure rather than composition. Considerably extended range of material parameters (local dielectric permittivity  $\mu_{ik}$  and magnetic permeability  $k_{ik}$  tensors) in these media enabled novel approaches to electromagnetic design, which take advantage of the field [...]

How to **(a) cook up** **(b) 'foamy' space-time** in the lab

**(c) Can the way the universe behaves** at the tiniest scales **(c) be recreated in the laboratory?**

Physicist Igor Smolyaninov of the University of Maryland in College Park has a recipe for cooking up a lab-scale version of a “quantum foam”, the **(d) choppy substance that constitutes space-time in some theories of quantum gravity.** According to these theories, space-time may appear smooth and curved, but zoom in, and it is actually made of virtual black holes, each just 10-35 metres wide, **(d) which flit in and out of existence.**

To mimic this structure, Smolyaninov suggests exploiting “critical opalescence”. [...] Smolyaninov calculates that the differences could perhaps be made so great that **(e) incoming light would bounce repeatedly around the interior of the patches, becoming trapped like light in a black hole.** Dependent on fluctuations in heat, such “black holes” would be transient like those in a quantum foam. It’s not clear what this artificial foam could tell us about quantum gravity, but Smolyaninov says that, at the very least, it could allow physicists to “ask more intelligent questions about how to detect the real thing”.

Figure 6.1. Contextual markup in parallel samples from the SPACE (extended section)

Here, the difference in register but identity in content is even more striking: cf. (a) “may be engineered to emulate” becomes “cook up...in the lab”; cf. (d) “real physical space-time strongly fluctuates” becomes a “choppy substance that constitutes space-time”. More research into these strategies is needed that could at some point lead to computational-linguistics application that routinely simplifies texts.

### 3.3 Hedging in SPACE

Hedges, while being a Lakoffian idea from the early 70s that have more to do with formal logic and which are today summarized as locutions in evaluatory frameworks (Keefe, 2003, p. 181), they nevertheless create a gradient (Keefe, 2003, p. 94) of commitment to a truth value of an utterance. Further, they are “words used to modify the meaning of a statement by commenting on the uncertainty of the information or on the uncertainty of the writers” (Durik et al., 2008, p. 218). In academic writing, they have been an object of interest from the beginning (e.g. Hyland, 1994; Aijmer, 2005). The following is taken from a study on hedge expressions, exemplified here in an academic text from psychology which seems extremely hedged.

*0202PS Save the last dance.*

Whenever competing options are considered in sequence, their evaluations *may be affected* by order of appearance. Such serial position effects *would threaten* the fairness of competitions using jury evaluations. Randomization cannot reduce potential order effects, but it does give candidates an equal chance of being assigned to preferred serial positions. Whether, or what, serial position *effects emerge may depend* on the cognitive demands of the judgment task. In end-of-sequence procedures, final scores are not given until all candidates have performed, *possibly burdening judges' memory*. If judges' evaluations are based on how well they remember performances, serial position effects *may resemble* those found with free recall. *Candidates may also be evaluated step-by-step*, immediately after each performance. This procedure *should not burden memory*, though it *may produce* different serial position effects. Yet, this paper reports similar serial position effects [...]

This sample was selected from the psychology section of the SPACE and illustrates how in a social science the commitment to the cause-effect relationships is considerably lower than in the natural sciences where causation replaces the much weaker correlation, which is the best social science can hope for in their findings. Although the example is anecdotal, it points toward general trends that can be observed when the science cultures are taken into account, too. Cautiousness is a hallmark of the social sciences as they study the aggregated behaviour of humans. However, the disadvantage of a qualitative analysis is that it remains anecdotal and the disadvantage of a quantitative analysis of hedges is that there is no isolated identifiable class that represents a hedge and it is very difficult to automatically tag hedges. Further, the use of hedges does not always indicate imprecision, as the study by Jucker, Smith and Lüdge (2003)

shows. They note that vagueness, while “often seen as a deplorable deviation from precision and clarity” (Jucker, Smith & Lüdge, 2003, p. 1737), in conversation, vagueness can be more informative than precision.

As a taggable class, modal auxiliaries and adverbs may lead to results but what about lexical verbs like *suggest*, *doubt* or *deny*? In the SPACE, we tried to find a solution to this by manually classifying these expressions. In a very long and laborious process, a student assistant tagged all hedge expressions in the corpus and has assigned every hedge expression a value between 1 and 10 in which a 10 means certainty and a 1 means near impossibility.

We call these tags *propensity tags* after Popper and Eccles (1984), with propensity being the probability of something to be the case.

*0090PN Topical DNA oligonucleotide therapy reduces UV-induced mutations and photocarcinogenesis in hairless mice*

UV-induced DNA damage gives rise to mutations and skin cancer. We show that topical pTT pretreatment enhances the rate of DNA photoproduct removal, decreases UV-induced mutations, and reduces photocarcinogenesis in UV-irradiated hairless WT repair-proficient and Xpc heterozygous *partially\_AV0\_M\_6* repair-deficient mice, both transgenic for the lacZpUR288 mutation-indicator gene. These data support the existence of inducible mammalian DNA damage responses that increase DNA repair capacity after DNA damage and hence reduce the impact of future exposures to environmental carcinogens. The ability of topically applied pTT to induce protective physiologic responses that *normally\_AV0\_M\_8* result from DNA damage *suggests\_VV\_M\_8* a previously undescribed means of reducing skin cancer in high-risk individuals.

Skin cancer accounts for *at least\_AJS\_M\_9* 40% of all human malignancies, 1,000,000 cases annually in the U.S. (1, 2). Incidence is *clearly\_AV0\_M\_9* linked to UV exposure and increases exponentially with age (1, 3). Skin cancer risk is *greatly\_AV0\_M\_8* increased in the rare disease xeroderma pigmentosum (XP), because of mutation in one of several DNA repair enzymes responsible for nucleotide excision repair (NER) (4–6). Development of a hairless mouse model (7, 8) and more recently hairless XP gene knockout mice that mimic the human cancer susceptibility (4) has *greatly\_AV0\_M\_8* facilitated studies of photocarcinogenesis. In particular, lowdose daily UV irradiation of XP group C (Xpc) mice leads to the development of skin cancer with a short latency time (80–100 days) and 100% prevalence, and *partially\_AV0\_M\_6* repair deficient Xpc mice are also more prone to UV-induced skin cancer than their WT counterparts (9). Finally, to study DNA mutations induced by physical and chemical agents in tissues of higher eukaryotes, transgenic mice carrying multiple copies of a lacZpUR288 mutation-indicator reporter plasmid have been generated (10) and crossed into WT and repair-deficient mice strains (11). Our laboratory has shown that *many\_PNI\_M\_6* protective responses triggered by UV irradiation are duplicated by treatment of cells in vitro and in vivo with thymidine dinucleotide (pTT) (12, 13), originally selected for study because it is the obligate substrate for formation of *most\_PNI\_M\_8* [...]

Thus, for example in *clearly\_AV0\_M\_9* we have the expression and we have added the CLAWS (Constituent Likelihood Automatic Word-tagging System) tag (a recent overview can be found in Kennedy, 2014, p. 212) for the part of speech (AV0 for

adverb), the position (M for medial) and the value of propensity (9 for near certainty). Compare this with *partially\_AVO\_M\_6*.

Of course this process has its shortcomings because of its subjectivity but if the same person ranks the hedges then the errors or idiosyncrasies cancel each other out.

Table 6.3

*Mean propensity values for the different domains in the SPACE*

	Academic	Popular	Academic	Popular	Academic	Popular
biosciences	microbiology		genetics		biochemistry	
	5.07	5.51	5.07	5.67	5.09	5.93
physics	cosmology		particle physics		quantum mechanics	
	5.52	5.78	5.54	5.66	5.35	5.73

The tabulated results show a tendency towards higher propensity in the popular genres. This is the case for all investigated texts on average, the largest difference being the one between academic and popular texts in biochemistry (5.09 for academic in comparison to 5.93 for popular biochemistry texts). However, this could be a statistical effect due to the number of texts (which was the lowest in this section). Overall, the difference between 5.27 (all academic texts in the corpus in comparison to 5.65 (all popular texts in the corpus) can not be considered significant.

#### 4 Conclusion

The investigated linguistic parameters in the SPACE show that it is possible to attempt quantitative genre research in academic writing when parallel texts are concerned and systematically compared. In this contribution, a specific type of genre analysis could be demonstrated on the example of a novel corpus which achieves this comparison through its parallel structure. The integration of corpora in teaching academic writing and in the study of English for Academic Purposes is not new (e.g. Chambers, 2005) but involving students in active search for the appropriate text-type specific idiomatics still is underemployed in academic writing classes. It therefore turned out helpful to demonstrate to students that academic writing occurs on different levels or rather strata of proficiency and competence, thus ameliorating for the frustration effects experienced by students when confronted with authentic examples (see part 3.1). But in a second step, these perceived differences can be quantified as well if an isolated parameter is selected. In our example, however, the perceived difference between the “softer” and less precise popular-academic (thus more student-like) texts was not substantiated by statistics. The upshot of this is that on the one hand side, strata can be studied with corpus means but they may be more similar to each other than previously expected. For students, this is usually good news.

## Chapter 7

# **BUILDING A CORPUS OF HIGH SCHOOL ACADEMIC TEXTS (COHAT) FOR ACADEMIC ENGLISH: DISCOVERY LEARNING IN LANGUAGE ACQUISITION**

*Róbert Bohát, Nina Horáková, Beata Rödlingová*

### **1 Introduction: How it all started**

How can one make language learning more enjoyable? What is the importance of learning languages anyway? “Education is considered to be full if the pupil’s mind is brought to wisdom, the tongue to eloquence, and the hands to the activities that they are expected to perform in life,” said Jan Amos Comenius, a seventeenth century Czech intellectual, in the Preface to his *Orbis Sensualium Pictus* (Comenius, 1991, p. 21). In order to bring “the tongue to eloquence,” it was necessary “for the intellect and language to go always in parallel” so as to promote genuine understanding of language, not just rote learning and reciting decontextualized vocabulary items, for “repeating words without understanding belongs to parrots and not to people” (Komenský, 2003, p. 23). He also criticized some language textbooks for creating unnatural sentences, combining words that do not naturally collocate, “so that some of the words themselves would wonder how they got together” (Comenius, 1986, p. 265). For these – and other – reasons, Comenius (1986) suggested that language learning be based on “... words and phrases in one concise body [Lat. *corpus*] of all language” that would enable learners “in a short period of time” make their learning as an “easy, pleasant and safe transition to real authors” (p. 264) in the target language. The result was his world famous *Janua linguarum reserata* (Gate of Languages Unlocked), a textbook originally published in 1631, translated into many languages and reprinted for more than three centuries.

It is interesting that in an attempt to find a solution to some language teaching problems (such as the use of artificial sentences, unnatural collocations and vocabulary learning without context, as well as the challenges of polysemy and homonymy), Comenius (1986) proposed the creation of “*unum redactae corpus*” – “one concise corpus or body” (p. 264) of language; for him, this “corpus” was his *Janua*, a corpus of 1000 sentences organized thematically under 100 topics. However, more than 300 years later, Comenius’ call for a corpus of language is answered by corpus linguistics; its application to language acquisition and teaching allows teachers and

textbook writers to use real life sentences with natural collocations and contexts, exposing the learner to an authentic language experience. This, in turn, will allow learners to deduce a specific meaning of a polysemic word from the context. Naturally, the word *corpus* has a much wider meaning today – and a much wider application to language learning and teaching.

Inasmuch as rote learning and mechanical vocabulary memorization (often as fast as vocabulary forgetting) are sadly still a problem in many a twenty-first century classroom, a solution is called for. How could we teach words in their natural context and collocations? How could we make sure that *intellect* and *language* work together, using critical thinking and metacognition in high school academic English classrooms? And how can this be made enjoyable to the student? In an attempt to answer these questions, the English as an additional language (EAL) Department at the International School of Prague (ISP) designed the *Applied Linguistics Project* (ALP) – a student-led research into academic English and mother tongue (MT). Mimicking real research, grade 9 students formulated a research question, designed a method of data collection and analysis, and finally wrote an academic paper with an abstract in their MT and English. This was then presented at a student conference in front of a panel of experts at the Faculty of Education, Charles University, Prague.

Since the launch of ALP in 2013, several students have chosen to carry out studies that included the use of corpora. For example: “*How are ‘dobrý’ (good) and ‘velký’ (big) translated depending on the context?*” This was researched by a Czech native speaker using the multilingual corpus InterCorp (Cz-Eng), provided by the Czech National Corpus. Another topic, “*How are selected academic words used depending on their context?*”, was a quantitative study by a Korean speaker using the British National Corpus (BNC) to analyze the actual use and meanings of cognition related words such as *observe, conceive, assume, believe*, etc. A third example is “*Comparing Proverbs in the British National Corpus and Czech National Corpus*”, performed by a Russian speaker, using the BNC and Čermák’s (2003) paremiological minimum. Interestingly, this student – having experienced the huge difference between the implications of absolute and relative frequency while processing raw data – commented: “This is the first time in my life that I’ve seen mathematics being useful.” This was a beautiful example of metacognitive reflection and transdisciplinary skill transfer.

The insights gained during the research and the quality of learning achieved suggest that corpus linguistics may indeed be a valuable tool for student discovery learning *about, through* and *of* a language, and also provided an impetus for a more extensive incorporation of corpora into academic English teaching and learning at ISP. Yet, when it comes to academic corpora, most of them are at university level, beyond the reach of high school teenagers. Thus, to provide the students with texts that best suit their academic needs and language level, the idea to build a Corpus of High School Academic Texts (COHAT) inevitably followed.

## 2 COHAT and a wider definition of a learner

COHAT is a specialized academic English learner corpus of exemplary high school texts. As of June 2015, it is made up of 100 texts – written and spoken – with ca. 150,000 words in 20 different genres (text types). There is an important difference from other learner corpora, however: while most learner corpora define *learners* as students of English as a second language (L2), we define learners as *all students* of academic English at our school, including native speakers of English. After all, when reading academic texts, many native speakers feel that “it’s all Greek to them”, which is often almost literally true, given the Greek origin of many scholarly terms. Thus, the first subdivisions of COHAT are into the native speaker subcorpus (currently about 25% of the total) and the non-native subcorpus, currently ca. 75 per cent of the total. This 1:3 ratio reflects the makeup of ISP’s student body quite accurately, and this carries with it a great potential for a comparative study of native, near-native and non-native speaker production of academic writing.

Another important detail about COHAT structure is expressed by the word *exemplary*: by exemplary texts written by high school students we mean texts (written, written to be spoken, or spoken and then transcribed) that reached standard levels of 6 or 7 (7 = the best), roughly equivalent to the traditional A and B grades (or 1 and 2 in the Czech system). This approach was inspired by the BAWE Corpus (Alsop & Nesi, 2009). The meta-information includes: subjects, text types, gender, grade level, and mother tongue. Currently, plain text files of the corpus are processed by AntConc (Anthony, 2014), but we are negotiating the possibility of COHAT being hosted by the Czech National Corpus Institute on their online interface at <http://korpus.cz>.

The *target audience* of our corpus are primarily high school students (native and non-native English speakers), but also teachers, year 1 university students, teachers in training, researchers focusing on youth language, etc. Another unique feature of COHAT is its *purpose*: it aims to go beyond error detection and interlanguage studies, or even the fascinating field of L1 transfer into L2. All of the above are good and useful goals, but we hope not to stop there, but proceed towards the detection of *desirable language patterns for particular subjects and text types* – syntax, morphology, word choice – and to *empower our students* (as our school’s mission says) through *discovery learning*, using the corpus to study and reflect on what language features made their peers’ texts successful. This should be especially useful due to the fact that, linguistically speaking, the texts were produced by students approximately their age, so even a novice is likely to decipher another teenager’s language. Psychologically speaking, the fact that texts with the top two grades were included shows that a certain amount of mistakes in texts need not impede their success.

The teaching-learning cycle within the *Teaching ESL students in mainstream classrooms (TESMC)* program is a sequence of deconstruction of exemplars, joint

construction by students with teacher guidance, and independent construction (Jones, 2007). In this connection we see the use of COHAT also as a treasure trove of texts for deconstruction, especially in a genre-based approach to teaching academic writing (Derewianka, 2003). Deconstruction involves familiarizing students with characteristic features of a genre, using an *analytical approach and evaluation* of sample texts – a process rich in opportunities for critical thinking and reflection with respect to which features of language are successful in which context and why. This increases the students’ sense of security and satisfaction thanks to understanding what is required of them in any given genre, unlike traditional rubrics that are often misunderstood even if they are read. Also, COHAT and other learner corpora could be used in analyzing language through the lens of Halliday’s (1993) *functional approach* and Vygotsky’s *zone of proximal development* (Vygotsky, 1986).

Table 7.1  
*COHAT Structure and Categories*

COHAT: Number of texts by subject		
	Texts (number)	Texts (percentage)
Academic writing	1	0.8
English as an additional language (EAL)	24	19.5
English language and literature	19	15.4
History	15	12.2
Natural Sciences	15	12.2
Maths	4	3.3
Psychology	4	3.3
Social studies	17	13.8
Theatre	14	11.4
Theory of Knowledge (TOK)	10	8.1
COHAT: Number of texts by gender		
	Texts (number)	Texts (percentage)
Female	73	59
Male	48	39.4
Gender information not available	2	1.6
COHAT: Number of texts by first language		
	Texts (number)	Texts (percentage)
Native speakers of English	27	22
Non-native speakers of English	96	78
COHAT: Number of texts by grade level		
	Texts (number)	Texts (percentage)
Grade 9 (age 14-15)	65	52.8
Grade 10 (age 15-16)	20	16.3
Grade 11 (age 16-17)	18	14.6
Grade 12 (age 17-18)	20	16.3

### 3 Limitations of COHAT

The first limitation of our corpus is its relatively small size of just above 150,000 words. As Table 7.1 shows, another important limitation of COHAT is that it is not yet balanced and fully representative. In terms of grade levels, grade 9 seems to dominate, which can obviously skew some investigations of the whole corpus, and this bias towards grade 9 usage needs to be borne in mind when analyzing data. Gender balance is almost achieved, but there is a stronger presence of female authors in the current version of the corpus. As for subject representation, we have very few mathematical texts, as well as a very small number in the field of physics and chemistry, creating a strong bias towards the humanities (making up ca. 78% of the corpus).

The native/non-native ratio, though *representative* of our school, also needs to be balanced out. It is our goal to eliminate these disadvantages and irregularities, making our corpus balanced and representative; in the meantime, however, we need to take these limitations into consideration when analyzing data. Some of the bias in data analysis can be alleviated by paying close attention not only to the frequencies of linguistic phenomena or items, but also to their distribution across genres, grade levels, gender, or subjects, normalizing or relativizing the raw data accordingly. Another remedial step could be comparing COHAT data with an external corpus that is both balanced and representative, or with a specialized corpus for a given field. Comparisons of this kind may help identify at least some problematic areas due to the lack of balance and representativeness.

### 4 COHAT: Application in the Classroom

In spite of its relatively small size and limitations, COHAT has proved to be capable of yielding meaningful data during our preliminary research and case studies. The presence of student errors (the texts are unedited) makes it a useful tool for error or interlanguage diagnosis, helping us to identify the areas of weakness that need to be revealed and minimised. As stated above, however, that is only the first step; the most productive and constructive use of COHAT lies in the positive data it contains. Even in this constructive approach, teachers can benefit from identifying the underuse or overuse of certain language features, helping them in lesson planning for the topics their students really need, preventing the plague of teaching what our students already know.

Furthermore, we have devised a growing set of exercises as small research tasks for students to perform, to observe academic English in action and then generalize their observations and analysis into grammatical, syntactical, stylistic or other rules. Needless to say, a grammar rule discovered by the student himself or herself is more

likely to stay in active use and long-term memory than a memorized abstract rule without context. This is where learning can truly be a guided rediscovery of knowledge and the class can “become a community of inquiry” (Wells, 2001, p. 20).

Another benefit: the lack of absolutes in the data analysis helps students practice pattern and trend detection, and then understand most (if not all) rules as what they typically are: *trends, not hard and fast rules* “written in stone”. This approach engenders intellectual flexibility in reflecting language as it really is – largely uncertain and unpredictable, full of surprises and exceptions to the rules. This is especially the case in English where one could say that the only rule that has no exceptions is the rule that all rules have exceptions. Additionally, it can help EAL students cope with linguistic irregularities in English, seeing as they will find that there are many outliers in their data, as well as clear trends.

To illustrate the potential for the use of COHAT in teacher and student language research, we would like to present a few mini-research tasks, i.e. pre-tested (and some also solved) samples or exercises. These are designed as lesson plans that can be used as templates applicable to work with other corpora, too. For more examples, please see the ISP Corpus site (see the end of this paper) or write to the authors – an exchange of expertise and suggestions can only benefit all learners involved.

#### **4.1 Lesson plan 1: Subjectivity: in the *I* of the beholder**

Research task: How could we evaluate the degree of subjectivity in a text?

Learning goal: Facilitate students’ own analysis of subjectivity indicators and the general trend towards a more neutral, objective reporting in academic texts in comparison with spoken texts.

Introduction: One indicator of the degree of subjectivity is the relative frequency of the personal pronoun *I* in a text. (This can be presented as a hint, or it could be arrived at by a series of Socratic questions by the teacher to the students.)

Hypothesis: Using your experience with academic and social (everyday) language, hypothesize which subcorpus, written or spoken, will have a higher frequency of the personal pronoun *I*, and why.

Sample Hypothesis 1: It seems reasonable to hypothesize that in a more subjective text the speaker/writer will focus on himself or herself more and use this pronoun *I* more frequently (with a higher relative frequency) than in a more objective text.

Sample Hypothesis 2: The relative frequency of *I*, *me*, *my/mine* will be higher in the spoken corpus; academic writing will show a lower relative frequency due to the requirement of objectivity in reporting.

Test/Method: Find the frequency of the first person singular pronoun *I* in the Spoken Subcorpus and in the Written Subcorpus. Remember to eliminate false hits such as the Roman numeral *i.* or *I.*, etc.

Collect the raw data and record the absolute frequency (AF).

Data Processing: Normalize your results (and thus make them comparable) by calculating the relative frequency (RF) of hits per 1,000 words. [ $RF = AF / \# \text{ of thousands in subcorpus}$ ].

Results: Spoken Subcorpus of 7,697 words shows the absolute frequency of *I* as 434 hits. Interestingly, AntConc also shows a very high *keyness* (K) value for this item:  $K = 1246.6$ ; this could be another aspect of subjectivity evaluation to be researched.

Calculation:  $RF = AF / \# \text{ of thousands in subcorpus}$ ; therefore,  $RF = 434 / 7,697$ , i.e.  $RF = 56.3 \text{ ppt}$  (parts per thousand).

The Written Subcorpus of 104,000 words (as of May 2015) produced 159 hits for *I*; thus  $RF = 159 / 104 = 1.5 \text{ ppt}$ .

Conclusion: The first person personal pronoun *I* in the spoken subcorpus has a relative frequency of 56.3 ppt (plus a high *keyness*), while in the written corpus it has a relative frequency of 1.5 ppt. This matches the hypothetical prediction, showing that the RF of *I* in the spoken subcorpus is ca. 38 times higher than in the written subcorpus.

Further sub-tasks: What other indicators of subjectivity can you think of? List them and justify (explain why they indicate subjectivity). Choose one of these examples and test them in a corpus.

Reflection: What does this difference in frequency tell us about subjectivity in the academic writing style? Why is this important?

## 4.2 Lesson plan 2: Key word lists (KWL)

Research task: Investigate the similarities and differences in the use and meaning of selected subject-specific academic words.

Learning goal: Discovery of the differences in vocabulary frequencies and use in natural sciences vs. humanities, including the tendency to prefer different shades of meaning of the same words.

Introduction: Understanding similarities and differences in high-frequency words and keywords across subjects contribute to a more efficient learning of academic language and a deeper understanding of academic vocabulary. Corpora can be used

to generate keyword lists such as Table 7.2, which can serve as the basis for further vocabulary analysis.

Research sub-task 1: Sort the following keywords from COHAT into two categories – what categories can you think of?

Table 7.2

*Selected keywords from COHAT subcorpora (unsorted)*

aggression, become, civil, containers, data, different, enlightenment, experiment, food, participants, peasants, plastic, political, public, revolution, snails, use, value, violence, water

Research sub-task 2: Use Table 7.3 to estimate which words in the natural sciences column you could also easily find in the humanities column, and what the differences in their meanings are. Write down your estimates and test the frequency of at least two of these words in the humanities subcorpus.

Table 7.3

*Selected high-frequency keywords in natural sciences and humanities*

Natural sciences (biology, environmental studies and societies, chemistry)	Humanities (social studies, history)
data	revolution
water	aggression
food	political
different	civil
value	enlightenment
containers	violence
experiment	peasants
plastic	become
snails	public
use	participants

Research sub-task 3: Use COHAT to find the differences in the meaning of the words in these two different contexts and compare with their definitions in Cambridge Dictionaries Online (CDO).

Sample analysis for the word *value*: The definitions of the word *value* are given as follows: “noun: money (the amount of money that can be received for something), importance (the importance or worth of something for someone); verb: money (to give a judgment about how much money something might be sold for), importance (to consider something important)” (CDO).

The random sample of 25 hits in COHAT reveals that in the humanities subcorpus 88 per cent of the occurrences are a noun meaning *importance*, while in the natural sciences 90 per cent are a noun with the meaning of a numeral quantity, which is not

included in the dictionary. This suggests that the language of natural sciences tends to be more narrowly subject-specific and differs greatly from everyday use.

### 4.3 Lesson plan 3: Language of caution

Research task: What language features can we use to express the degree of (un) certainty of the speaker about what he or she says or writes?

Learning goal: Help students notice a pattern of academic caution through the use and frequency of modal verbs and selected lexical verbs typical for expressing uncertainty in evaluating data.

Introduction: The use of hedging, or language of caution, appears to be a typical feature of what is considered a correct academic expression, and one of the major differences between academic and social language. While in everyday conversations we tend to speak more often in an absolute way, saying that *things are this way* or *x causes y*, in academic English we are required to express any uncertainty possible, suggesting that *things seem to be this way* or *x may cause y*, etc. Language of caution typically includes the use of adverbs (*largely, reportedly*), lexical verbs (*tend, indicate*) and modal verbs (*may*), often in combination to express a varying degree of the speaker's commitment to the utterance.

Table 7.4

*Absolute frequency of modal verbs may and might in COHAT*

Verb / use (AF)	Lexical use ( <i>be allowed to</i> )	Hedging
may	19	146
might	0	58

Sample analysis: The use as a hedging marker appears to be prevalent; 88 per cent of the uses of the verb *may* and 100 per cent of the uses of the verb *might*. Furthermore, the distribution of *may* in its meaning *be allowed to* shows all the examples come from the same type of document (historical writing on the topic of the Treaty of Versailles), which further supports the claim that *may* in academic language is primarily used as a hedging marker.

Table 7.5

*Absolute frequency of the verbs appear, indicate, seem, suggest in COHAT*

Verb / use (AF)	Lexical	Hedging
appear	15	20
indicate	0	18
seem	0	50
suggest	2	26

Sample analysis: The use for hedging is prevalent, to a varying degree (57-100%), for all the four verbs. Upon closer inspection of individual examples, it is interesting to note that *indicate* is in 67 per cent of the examples further hedged; 22 per cent by a preceding verb (*may, seem to*), 33 per cent by a reference to the source of information (*other studies, estimates, another researcher's name*) and 5 per cent by an adjective (*largely*). This suggests the speaker's subjective perception of the verb is not sufficiently effective as a hedging marker and reveals the creative power of combining the various hedging devices to modify the degree of commitment.

Further sub-tasks: Use the collocation function to investigate adverbs that frequently modify the selected verbs.

Reflection (food for thought): Which use should be considered as primary when teaching or learning certain lexical and modal verbs?

#### 4.4 Lesson plan 4: The *iffy* nature of our *ifs*

Learning goal: Allow for an inquiry-based reflection on the use and meaning of various types of conditional sentences, as well as a comparison of conditional patterns in English with conditionals in the students' mother tongues.

Introduction: Why should we care about *if* sentences (also known as conditionals)? There are several basic types of conditionals in English and each of them has a specific meaning. If we use a type of conditional that does not "match" the meaning we want to make, we could very well end up saying what we did not mean. Matching our conditionals to the right situation or context will thus make our communication accurate and effective, which is important at school as well as in everyday life.

Research task A: Genre frequency and distribution – which text types have the highest frequency of conditional sentences and why? Determine the frequency and distribution of *if* sentences in all genres (text types).

Recommended method:

- i. Find all the occurrences of *if* with AntConc Concordancer in COHAT;
- ii. Identify how many times (absolute frequency) it is used in each subcorpus, subject and text type and in how many different texts (distribution);
- iii. Calculate the *relative frequency* (number of hits divided by the number of thousands in the corpus => hits per thousand) and *average per text* (number of hits divided by the number of different texts in which they appear).

Results: COHAT statistics: *if*: 187 hits (in 112,000 words [as of June 2015])

Table 7.6  
*Frequency and Distribution of “if” in COHAT*

	Absolute frequency (hits)	number of texts	Average per text	Subcorpus size (number of words)	Relative frequency (hits per thousand words)
Spoken subcorpus	23 hits	7 texts	ca. 3.3	7,697 words	2.9
Written subcorpus	164 hits	26 texts	ca. 6.3	104,303 words	1.57

Sample analysis: It is interesting that the relative frequency of *if* in the spoken subcorpus is almost twice as high as in the written corpus. This more frequent use of conditionals might be a sign of a more speculative character of the spoken language – a proposal that would need to be verified by analysing the actual content of all the conditional sentences. A second observation relates to the *distribution*: the *average per text*. In the spoken subcorpus, the overall relative frequency is almost the same as the frequency average per text, which makes them appear evenly spread across the corpus. In written texts, however, the *average per text* is 6.3 times per text, almost twice the average in the spoken subcorpus, and more than three times the overall relative frequency. In other words, *if* sentences in the written subcorpus appear less frequently but tend to be more concentrated (probably depending on the genre). What text types (genres) tend to have a higher concentration of conditionals?

Table 7.7  
*Conditional frequency and distribution by subject and text type (genre)*

Subject	Extended essay (EE)*	Short research paper	Lab report	Analytical essay
Biology	18 (1 EE text)	13 (3 texts)	x	x
Chemistry and physics	x	x	2 (1 text)	x
Academic English / writing	x	x	x	23 (7 texts)
Environmental systems and societies (ESS)	25 (1 text)	x	x	x
History	x	x	x	50 (9 texts)
Math	x	11 (1 text)	x	x
Psychology	14 (3 texts)	x	x	x

\*In the International Baccalaureate (IB) program, the EE is a type of longer research paper.

Sample analysis: Out of 164 occurrences, 57 appear in an EE research paper, amounting to almost 35 per cent of all the conditionals in the written subcorpus. Another 24 hits appear in a similar (albeit shorter) text type: a short research paper;

together, a research paper genre occupies a total of 81 out of 164 hits, a surprising 49.4 per cent of all. Does this indicate a more *iffy*, speculative tone in this text type? Could it be due to the frequent formulation of hypotheses and their evaluation in view of the data (which often leads to the formulation of new hypotheses, etc.)?

Research task B: Grammar – do academic texts use future forms of verbs after *if*? In many languages, *if* is often followed by a future form, such as in Czech: *Jestli přijdeš, budu rád* (literally: If you *will* come, I will be glad.) How is this idea expressed in English?

Recommended method:

- i. Find all the occurrences of *if* with AntConc Concordancer in COHAT;
- ii. Identify how many *ifs* are followed by a future tense form *will, shall*;
- iii. Record any other verb tenses that occur.

Results (partial): Out of the total of 187 *if* sentences, there were 0 results for *if* followed by a future tense form (in the *if*-clause). The verb forms following *if* in the whole corpus are: *the past simple tense, the present simple tense, or would*.

Extension sub-task 1: Get the statistics for all these occurrences and analyze the meanings expressed by each combination. Which tense is used in the type of sentence that corresponds to the Czech *\*If you will come, I will be glad?*

Extension sub-task 2: Organize all the conditional examples into *categories by verb form* used in the *if*-clause. What meanings are prevalent in the *if* + past tense? What meanings are associated with *if* + present tense? And what meanings go with *if* + *would*? What kind of situations, contexts and text types do the examples appear in? Are there any differences between texts by native speakers as compared to non-native speakers?

Teacher note: Extension 2 may lead to a rediscovery of the traditional conditional taxonomy: zero, first, second, third, or mixed conditionals. Students may be asked to give each category their own name before they are introduced to the traditional terminology; experience shows that students' own names for categories often help as simple definitions or memory aids for the standard nomenclature.

Extension sub-task 3: Compare COHAT results with corresponding text types in the BNC or another external corpus and analyse the similarities and differences.

## 5 COHAT: Next steps and future vision

As stated above, the immediate goal in the building of COHAT is to make it a referential, balanced, representative corpus of high school academic texts. The speed of this

process depends on several factors, the most important being the willingness (or lack thereof) of our students and colleagues to make their exemplary texts available to us, followed by the time needed to transcribe these and convert into an appropriate electronic format. In terms of size, we were able to gather above 100,000 words in approximately four months (from February to June 2015); hopefully, we might be able to reach the goal of 500,000 words by the end of the 2015-16 school year.

Apart from the quantitative growth, we would like to introduce tagging (parts of speech/morphological, syntactic, semantic), hopefully in cooperation with the Institute of the Czech National Corpus. Structurally, we would like to add texts produced by high school teachers themselves, both samples of what the assigned text type should look like, as well as a subcorpus of spoken “teacherese“ in lecture transcripts. This could be expanded by a written corpus of instructions, test questions, word problems, rubrics, handouts, narrative reports, etc. to enable us to study the similarities and differences between the academic language of students and teachers. Furthermore, we could see how (in)accessible teacher texts are to the students and their parents, especially the language of narrative reports.

An analysis of the linguistic gap between the two parts of the school community might help us identify ways to bridge the gap where and when necessary. In other words, how can we, teachers, improve in the clarity and unambiguous wording of our homework or lab instructions? What about the use of English idioms and cultural references in international schools with high EAL student populations? How can we keep the beauty of the teachers’ idiomatic language without losing the attention (and understanding) of our EAL students? What are the lexical density and lexical richness of teacher-produced texts and how do they compare with student texts? Can we use teacher samples as a support for the students’ journey into the *zone of proximal development*?

In time, we would like to make COHAT publicly available through the Czech National Corpus user-friendly interface, and also expand it with non-ISP academic texts from academic writing courses, international schools, bilingual schools, EFL high-proficiency students at monolingual schools, the IB program, etc. We welcome student and teacher texts and other ways of cooperation.

COHAT, as we envision it, will be accompanied by a comprehensive set of heuristic exercises for students to identify trends in language use and discover the rules themselves (with teacher assistance). In the field of grammar, this may improve the understanding of language in use, lead to better long-term retention and increased ability to cope with uncertainty and irregularities in English grammar.

A subcorpus encompassing frequently used primary sources (including fiction and poetry, historical documents and religious texts) will enable students to use concordances and keyword lists in literary analysis of motifs, symbols, etc. These linguistic features can also help them to detect the tone, perspective or bias in historical or political texts, in journalist reporting, as well as frequencies and distribution of

key terms in the traditional religious texts in social studies. This small learner corpus can also help in teaching students explicitly how to create their own specialized mini-corpora for specific projects and language learning self-reflection. Last but not least, working with COHAT will serve as training grounds for using bigger corpora.

Once balanced and representative, COHAT can also be used to create our school's own *general service list* and *academic word list*. These could be used in addition to the official lists to meet the immediate needs of especially the younger EAL students for their academic communication, giving them a vocabulary that they are likely to need in our school, with the possibility of making these lists grade- or subject-specific. That could help us be quite specific in answering the question of which words are general, i.e. used across all or most disciplines, and which words are highly specialized, i.e. used only in a limited number of subjects or genres. Such vocabulary teaching would need to be more than a mere list of words; these would need to be much like the keyword lists of our corpus (or one of its subcorpora), just “one click away” from the concordance with every word in a sentence with its context and appropriate text type, with the real collocations used in exemplary texts – just like Comenius envisioned it almost 400 years ago. Using learner and academic corpora, such as COHAT, might therefore be a step towards the creation and opening of the interactive and electronic *Ianua linguarum nova* (*A New Gate of Languages*).

Note: The International School of Prague (ISP) was founded in 1948 and thus is the oldest independent school in the Czech Republic. It serves more than 800 students from 60 different countries. ISP is accredited by the Council of International Schools and the New England Association of Schools and Colleges and offers the International Baccalaureate (IB) diploma program to students in grades 11 and 12. ISP is also authorized by the Czech Ministry of Education to educate foreign nationals and Czech citizens. ISP's high school students have been involved in corpus based data-driven language learning since 2013. More information about our school is available at [www.isp.cz](http://www.isp.cz).

## **Acknowledgements**

We would like to thank Lawrence Hrubes (ISP EAL Dept. Leader) for proofreading the paper, his moral support and many practical suggestions; ISP Administration for their moral and material support; PhDr. Petr Chalupský, Ph.D., PhDr. Bohuslav Dvořák, and Mgr. Karel Žďárek from the Faculty of Education at Charles University in Prague for making the ALP student conference possible; doc. Mgr. Václav Cvrček, Ph.D., Mgr. Michal Křen, Ph.D., and Mgr. Lucie Chlumská from the Institute of the Czech National Corpus for their valuable training and professional help, and all the enthusiastic ISP students who made this project such a thrilling adventure of learning. Róbert Bohát would further like to thank his wife and family for their support without which most of this work could not have been completed. *Soli Deo Gloria*.

## Chapter 8

# HOW A LARGE CORPUS OF DISSERTATIONS CAN HELP TEACHERS AND LEARNERS OF EAP WRITING

*John McKenny*

## 1 Introduction

This chapter investigates how the compilation and interrogation of a specialized corpus of successful Masters dissertations might help advanced learners of academic English to write their own dissertations. The corpus in question is the British University in Dubai (BUiD) Corpus of Written Academic English (Buidcorp). Over the past two decades various corpora of learner English have become available (e.g. Cambridge Learner Corpus, The Chinese Learner English Corpus). Individual studies have focused particularly on dissertations/theses written by learners of English (e.g. Charles, 2006; Isik-Tas, 2008; Kwan, 2006; Thompson, 2005). There was a felt need for a focus on dissertations completed by research students in the Gulf region. The principal aim of the Buidcorp project was to meet this need and further develop the corpus as a useful teaching and research tool.

## 2 The Buidcorp corpus of written academic English

In many ways the Buidcorp corpus can be characterized as opportunistic. The data is there so why not use it – in the spirit of Sir Edmund Hilary, who, on being asked why he climbed Mount Everest in 1953, with Sherpa Tenzing, replied: “Because it was there”.

Masters dissertations submitted by BUiD students in the Education, Business and Engineering Faculties since 2004 were available in machine-readable form. These were collected as PDFs and converted into text only (ASCII) format. Arabic is the L1 of the large majority of the students; all students on submission had agreed to their work being used for research. File names helped mark subject area, year of submission, but also served to protect the identity of the student authors. TEI (Text Encoding Initiative) metadata (headers, section divisions) were introduced to make *abstracts, introductions, literature reviews, methods, results, discussion, conclusion* and *references*

sections retrievable as more generically focused sub-genres of the macro-genre, *dissertation* (Paré et al., 2009). It is intended to carry out genre analysis on sub-corpora containing all the examples of the major sections of the dissertations (e.g. introductions, conclusions) possibly as optional student-led project work. Interestingly, as some Masters students now embark on research projects for their Masters dissertation based on the corpus, their work will soon form part of the corpus. Thus the corpus could be said to be self-propagating. The findings of these genre analyses will be published as reference material on the virtual learning environment platform of the University's Academic Success Unit.

Table 8.1  
*Makeup of the Buidcorp corpus*

Research fields	Number of words
91 M.Ed. dissertations TESOL and leadership	2.3 million
206 M.Sc. dissertations business	5.3 million
140 M.Sc. dissertations engineering and IT	2.4 million
approximate total	10 million

### 3 The haecceity of the Buidcorp corpus

Within the corpus linguistics paradigm perhaps the principal consideration in corpus design is the representativeness of the corpus. A corpus can be representative of, for example, a language variety, a dialect, a register such as legal English or business English or the academic writing of a particular group or level of students. What is being stressed is the typicality of the language and the repeated patterns found in the corpus. In the sub-discipline of applied linguistics designated *computer learner corpora* (Granger et al., 2002), there is also a focus on the distinctive features and even idiosyncrasy of each learner corpus. In the case of the Buidcorp, I will be exploring two distinct dimensions: the degree to which it is representative of L2 academic writing and, the obverse of this dimension, which I will call the *haecceity* of the corpus. The term haecceity was coined by the medieval philosopher John Duns Scotus. It derives from the Latin term, *haecceitas*, i.e. the *thisness* or distinct qualities, properties or characteristics of a thing which makes it a particular thing (individuation, identity).

The approach of *computer learner corpora* (Granger et al., 2002) tends to be diagnostic: in what ways does the language of the learners fail to correspond to standard English; which expressions have student writers overused or underused? I align myself with the *English as a lingua franca* (ELF) approach (Seidlhofer, 2011)

which sees learner Englishes as valid varieties of English worthy of study in their own right. I emphasize what the writers in the Buidcorp do well and remind my students that the writers featured in Buidcorp are all successful writers initiated and accepted into the academic community. We should remember that the original meaning of a Masters degree was a qualification to teach or practise a profession. Students of academic writing should not be afraid of picking up bad writing habits from past Masters graduates. Using a corpus approach they will see many successful realizations of the phraseology of academic English. The few lapses will be overwhelmingly outnumbered by the successful realizations of the phraseology of dissertations.

Wmatrix3 (Rayson, 2009) is a corpus analysis and comparison tool which was developed at Lancaster University in the Centre for Computer Corpus Research on Language (UCREL). The UCREL semantic analysis system (USAS) is a framework used by Wmatrix3 for the automatic semantic tagging of text assigning each word to a domain of discourse. Among the USAS tagset, the semantic tag, Z99, provides a rapid calibration of the non-standard usages in the corpus. Z99 is the tag that Wmatrix3 (Rayson, 2009) gives to language which is not contained in the USAS dictionary. There is no suggestion that it is wrong. It simply cannot be classified by Wmatrix3. It might be a compound or hyphenated word not usually written as the writer has written it. It could be a neologism or foreignism not yet found in standard dictionaries. Many words tagged Z99 might simply be typos and misspellings. Using USAS tag Z99, some of the Buidcorp writers' innovation and creativity can be revealed (as well as slips they made). The following are expressions from the Business sub-corpus which could not be categorized by USAS (labelled Z99): *CSR, PMO, transformational, 0.00%, HRM, PMOs, GCC, erp, ppm, emirati, BUiD, emiratisation, BCM, EFQM, IFRS, arbitral, id#, Emiratis, normalized, Hofstede, sukuk, SPSS, opm3, vol, 2-tailed, fidic, EOT, PMI, his/her, CCM, OCB, Etisalat, Cronbachs*.

#### **4 The hard and soft approaches to corpus linguistics**

Sinclair (2003), proposed a distinction between *corpus-based* and *corpus-driven* linguistics. Basically, according to the terminological distinction, corpus-based linguistics is traditional linguistics carried on with corpora while corpus-driven linguistics is a non-theory-laden approach to language without preconceived ideas which starts from scratch and follows where the data lead. Corpus-based and corpus-driven linguistics could be described respectively as “top-down” and “bottom-up” approaches to the analysis of lexical data. The corpus-driven paradigm invites the linguist to look afresh at words and their patterns in concordances and texts in order to devise new *a posteriori* explanatory models which not only cover the general patterns and tendencies which can be discerned but also attempt to encompass every irregularity or localized idiosyncrasy of language that is uncovered. Sinclair (2003)

reconciles corpus-based and corpus-driven linguistics in the introduction to his work *Reading Concordances*:

The amount of variation in actual usage makes accurate generalization rather difficult. The difference is often said to be between “top-down” and “bottom-up” approaches; starting from the “top” it is extremely difficult to arrive at a description that fits the facts of usage, while starting from the “bottom” it is not easy to formulate sufficiently general statements. However, the experience and intuition of the researcher are available in both approaches, and so the so-called “bottom-up” approach, properly conducted, is really a two-pronged attack on the data from the top and bottom simultaneously.

(Sinclair, 2003, p. x)

## 5 Data-driven learning

Tim Johns (1994) in his influential work on *data-driven learning* described his students as becoming “language detectives” as they manipulated and investigated the language data discovering the rules for themselves. This kind of learning is different from traditional approaches to language teaching in that it stresses the incorporation of real data into the language classroom. One of the traditional approaches to language teaching is the *present-practice-produce* (PPP) method by which students practice a previously presented grammatical or lexical construction and are then able to produce a sentence or text which exhibits the newly learned feature, paying special attention to the notion of *fluency*. This approach is mainly *product-based* in that it presents specific aspects of the language to the students. *Data-driven learning*, on the other hand, uses corpora and concordance software to teach language. It is *process-based* in nature which means that it encourages students to discover certain aspects of language themselves and experiment with the language. It assumes that grammar is a relatively flexible system and not a set of static rules.

Advantages of the product-based approach are obvious. It gives students a sense of direction and structure while learning many aspects of a language. However, a product-based approach shows only fragments of language or grammatical rules, which are based to a certain extent on the intuition of the textbook author. Recent corpus research shows that these rules are often inaccurate. In the *process-based* approach, students are encouraged to actively think about texts and to recognize language patterns.

Carter and McCarthy (1995) suggest supplementing the PPP approach with an *illustration-interaction-induction* approach. Illustration takes place via an examination of real data, while interaction with the text draws students’ attention to certain aspects of the data, such as interpersonal grammar. The induction stage signifies the students’ ability to draw conclusions about the interpersonal functions of lexicogrammatical constructions.

More research is necessary to determine how students learn from corpus data. For now it is clear that concordancing in the classroom is an extremely powerful hypothesis testing device which allows controlled speculation, makes hidden structures visible, enhances at the same time imagination and checks it by inductivity, thus making higher degrees of objectivity possible. Other corpus data, for example word and Ngram frequency lists can make students aware of what are the usual and preferred ways of saying things and what are the less usual ways.

We can see two main English for Academic Purposes (EAP) scenarios: the EAP practitioner relays or mediates corpus findings to writers who need to join a discourse community or who need to transfer/translate their L1 skills and knowledge to handle L2 discourse conventions. The second kind of EAP practice considered in this paper is the initiation of soon-to-be autonomous learners in the use of corpus methodology to become researchers of the language actually occurring in their discourse community. There is a growing interest beyond TESOL more widely within the social sciences, in the role played by language (Alvesson & Karreman, 2000). This interest in language has encouraged linguists and non-linguists alike to probe the potentialities of text and discourse analysis. Many teachers and students' disciplines outside TESOL are not necessarily interested in English *per se* but have instrumental motivation to understand the conventions of thought and communication which define their professional area and how they are textualized in English (Widdowson, 1998, p. 9).

In this chapter two different tools are described that *language detectives* can use in exploring corpora: the already mentioned Wmatrix3 (Rayson, 2003) and Wordsmith Tools 6 (Scott, 2012). Wmatrix3 automates a good deal of the corpus analysis that needs to be done. The user uploads corpora to the University Centre for Computer Corpus Research on Language (UCREL) and works mainly online. Wmatrix3 tags for *part of speech* (POS) and for semantic domain (USAS). It compares a corpus with a control corpus and produces frequency lists of words, parts of speech, and lists of multiword expressions. At all stages, concordances and lists can be produced. Wmatrix3 presents almost instantly to the investigator a great deal of information about the language of the corpus at many levels, giving an overview of the whole corpus. Such lavish and all-encompassing data might support the enquiries of a student who prefers a top-down approach.

Wordsmith Tools lends itself to working with a bottom-up approach. The corpus linguist can start with wordlists (lists providing words in descending order of frequency) and move on to keywords, concordances, clusters (Ngrams), and concgrams. It has many more functions than can be described here. Sometimes Wordsmith compliments Wmatrix3. If a corpus contains many files, these files have to be joined using Wordsmith's file utilities before they can be uploaded as one file to Wmatrix3. Overall, Wmatrix3 suits a researcher who prefers a top-down approach to investigating language while Wordsmith Tools suits a more bottom-up approach. Students can be initiated into the workings of both tools in a relatively short time by a teacher in

one-to-one or small group sessions. There are various tutorials within the system for users to deepen their knowledge of the tools’ potentialities.

## 6 The control corpus

The selection of a comparable corpus to use as a reference corpus or control corpus is an important step in corpus analysis. The choice is simple within Wmatrix 3. A number of corpora are provided to choose from or the corpus analyst could use one of the ones they have developed themselves within Wmatrix3. Permission to download the British Academic Written English (BAWE) corpus was obtained from the University of Oxford Text Archive (ota.ox.ac.uk). The Masters dissertations in arts and humanities and in the social sciences which obtained Merit or Distinction were extracted. This yielded a control corpus of expert user writing of just under one million words. The fact that my two sub-corpora of education and business dissertations were much larger did not render BAWE sub-corpus unfitting as a control corpus. This is because the measure of significance used in Wmatrix3 (i.e. log likelihood) already makes adjustments for differing corpus size (Rayson, 2003; Rayson, 2008; Dunning, 1993).

Our first approach to the education sub-corpus (ca. 2.5 million words) of Buidcorp corpus was via a straightforward word frequency list generated by Wordsmith Tools 6 (Scott, 2012). The 20 most frequent content words are shown below. The overwhelming majority of the 100 most frequent words in a corpus will be grammatical or synsemantic words. Table 8.2 below shows the first 20 content words in descending order of frequency after the functor words have been eliminated. An examination of Table 8.2 shows that most of the highly frequent content words therein are nouns. This “nouniness” suggests that our sub-corpus has a greater level of formality and the dissertations are informational (Biber, 1988).

Table 8.2

*20 most frequent content words in Edcorp (2.5 million word corpus of Education dissertations in Buidcorp)*

	Word	Frequency		Word	Frequency
1	students	18,202	11	English	5,301
2	school	10,714	12	teach	5,193
3	teachers	10,063	13	student	5,023
4	education	8,206	14	children	4,884
5	learning	7,599	15	about	4,873
6	research	7,102	16	time	4,624
7	language	6,842	17	teaching	4,224
8	teacher	6,746	18	uae	4,071
9	study	6,704	19	these	3,941
10	schools	5,788	20	needs	3,900

The next level of analysis is to find the keywords of the texts in our corpus. Keywords are words which are significantly more frequent in one corpus than others. The concept of keyness is now the preferred term. The idea is to identify words specific to a particular kind of language use. The formula calculates the proportional use of a word in the specialised corpus divided by the proportional use of the word in a general corpus. The identification of keywords can indicate what a corpus (or text) is about, i.e. its “aboutness” (Cheng, 2012; Phillips, 1985).

Keywords have been championed and explained in many publications by Mike Scott, designer of Wordsmith Tools, and his associates (e.g. Scott & Tribble, 2006). Table 8.3 lists the 50 top keywords in education sub-corpus. If this keyword table is examined, certain characteristics of the education corpus become apparent.

Table 8.3

*First 50 keywords from education sub-corpus of Buidcorp (in descending order of log likelihood values)*

	Word/phrase	Edcorp freq	%	Control corpus freq	%	+/-	Log likelihood
1	students	441	1.52	305	0.03	+	2096.54
2	technology	339	1.17	94	0.01	+	1925.73
3	teachers	267	0.92	71	0.01	+	1525.91
4	writing	257	0.89	115	0.01	+	1346.96
5	laptop	199	0.69	22	0	+	1251.14
6	initiative	164	0.57	43	0	+	939.14
7	computers	130	0.45	11	0	+	833.46
8	laptops	119	0.41	6	0	+	785.24
9	learning	159	0.55	106	0.01	+	762.83
10	United_Arab_Emirates	101	0.35	1	0	+	695.84
11	student	142	0.49	94	0.01	+	682.46
12	science	147	0.51	157	0.02	+	617.55
13	schools	140	0.48	132	0.01	+	611.28
14	to-	82	0.28	0	0	+	574.01
15	one-	82	0.28	1	0	+	563.25
16	classes	94	0.32	26	0	+	534.17
17	study	161	0.56	321	0.03	+	532.65
18	one-to-one	75	0.26	2	0	+	506.59
19	100010	72	0.25	0	0	+	504.01
20	id	84	0.29	32	0.01	+	453.33
21	teaching	86	0.3	66	0.01	+	397.99
22	classroom	58	0.2	18	0.01	+	323.91
23	education	92	0.32	189	0.02	+	300.24
24	curriculum	50	0.17	18	0.01	+	272.52
25	research	108	0.37	360	0.04	+	272.46
26	skills	71	0.25	105	0.01	+	266.07
27	EFL	37	0.13	0	0	+	259.01
28	computer	59	0.2	57	0.01	+	255.73

29	high	92	0.32	291	0.03	+	239.55
30	using	100	0.35	373	0.04	+	234.92
31	teacher	53	0.18	48	0.01	+	234.18
32	use	111	0.38	567	0.06	+	207.31
33	emirati	29	0.1	0	0	+	203.01
34	educational	50	0.17	67	0.01	+	194.4
35	in	836	2.89	15816	1.7	+	190.63
36	marks	27	0.09	0	0	+	189
37	classrooms	31	0.11	5	0	+	188.3
38	IAT	25	0.09	0	0	+	175
39	high_school	26	0.09	1	0	+	173.51
40	implementation	40	0.14	39	0	+	172.89
41	achievement	40	0.14	39	0	+	172.89
42	researcher	34	0.12	19	0	+	170
43	Internet	24	0.08	0	0	+	168
44	program	34	0.12	23	0	+	162.53
45	UAE	23	0.08	0	0	+	161
46	grade	32	0.11	18	0	+	159.77
47	write	45	0.16	80	0.01	+	156.56
48	integrate	26	0.09	5	0	+	154.92
49	professional	43	0.15	84	0.01	+	143.58
50	2009	26	0.09	9	0	+	142.65

Certain fairly predictable concepts emerge with high *log likelihood*, meaning their frequency of occurrence is much higher than could be predicted judging from the reference corpus (most notably *student(s)*, *teacher(s)*, *classroom(s)*, *technology*, *computer(s)*, *laptop(s)*, *writing*). The prominence of these last words points to the major revolution which has occurred in education in recent years in which the uses of technology in education has moved centre stage. It might be surmised that the occurrence of two verbs among the keywords, *write* and *integrate* demonstrate two major foci of the education dissertations, the teaching of writing and the integration of skills. Such speculation can be corroborated by follow-up concordancing.

The most frequent words in Wordsmith Tools and the keywords in Wmatrix3 can be used by language detectives as starting points for further investigation using the concordancer in each tool. Students can begin to investigate the collocations of the words and the phraseology surrounding them and the functions that words and phrases serve in the texts they are found in.

## 7 The supralexic level: phraseology

Writers of EAP need to develop *phraseological competence*, formerly referred to as *collocational competence* (Howarth, 1998). The greatest challenge faced by novice writers of academic prose is the production of idiomatic English with a phraseology

which expresses the writer's meaning well without drawing attention to itself through unnaturalness or strangeness. Many applied linguists have addressed this problem and the Longman Dictionary of Spoken and Written English (LGSWE) marked a milestone in the quest for ways of providing a taxonomy of this protean area of linguistic studies. Biber et al. (1999) in their monumental LGWSE started by coining the term *lexical bundle*. This term was very close in meaning to the established term *Ngram* except that Biber and his co-authors stipulated that to qualify as a lexical bundle the sequence of words would have to occur a certain number of pre-ordained times per million words. Thus lexical bundle is determined purely on statistical grounds. Using Wordsmith Tools Index, the 3-5 word lexical bundles were extracted from the education sub-corpus:

Table 8.4

*Most frequent 55 of the 3-5 word bundles from the corpus with original frequencies before filtering.*

	Bundle	Freq.	%	Texts
1	in the uae	1,704	0.07	72
2	in order to	1,510	0.06	89
4	gifted and talented	861	0.04	11
5	as well as	828	0.03	86
7	the use of	724	0.03	82
8	of the students	689	0.03	72
9	with special needs	642	0.03	33
10	of the study	626	0.03	78
11	united arab emirates	614	0.03	70
12	the importance of	607	0.03	82
13	in this study	580	0.02	77
14	on the other	543	0.02	75
15	in the classroom	534	0.02	69
16	the other hand	507	0.02	70
17	on the other hand	505	0.02	70
18	the impact of	501	0.02	61
19	of this study	492	0.02	77
20	the united arab	492	0.02	65
21	in terms of	491	0.02	80
22	ministry of education	489	0.02	64
23	the role of	486	0.02	77
24	the number of	484	0.02	80
26	the united arab emirates	483	0.02	65
27	the gifted and	468	0.02	9
28	the gifted and talented	454	0.02	8
29	in the school	448	0.02	51
30	do you think	444	0.02	79
31	a lot of	444	0.02	41

33	the implementation of	428	0.02	65
34	in the united	418	0.02	85
35	of the teachers	417	0.02	43
36	of the research	411	0.02	71
37	the ministry of	409	0.02	50
38	students with sen (special educational needs)	407	0.02	71
39	due to the	407	0.02	63
40	the end of	406	0.02	11
41	that there is	405	0.02	84
42	there is no	397	0.02	75
43	in addition to	393	0.02	82
44	of the school	393	0.02	77
46	the fact that	392	0.02	60
48	be able to	386	0.02	76
49	university in dubai	385	0.02	32
50	teaching and learning	384	0.02	79
53	the needs of	380	0.02	80
54	the purpose of	371	0.02	60
55	of students with	371	0.02	65

A cursory glance at this list of lexical bundles reveals many interesting features of the writing in the Buidcorp. The overuse of *a lot of* (444 occurrences in 41 dissertations) shows that students need to be advised that this expression is overly informal and could be replaced with more formal expressions such as *a great deal*. The expression *in terms of* used 491 times in 80 dissertations was probably overused. Students could be invited to examine a partial concordance of this term and decide whether its use was always necessary. Often it is used to sound formal but can be replaced by a much simpler preposition. The EAP teacher can derive a great deal of benefit from lists of lexical benefits as can the autonomous learner or language directive.

## 8 The use of Wmatrix3 to extract multiword expressions (MWEs)

Another approach at the phraseological level is the search for multiword expressions. MWEs are not lexical bundles nor formulaic sequences although there is some overlap. Wmatrix3 searches out MWEs using a two-pronged approach. First, it has a large lexicon of sayings, idioms, clichés, commonplaces and other fixed expressions for which it searches in the corpus. This is further supplemented by a search for variable phrases using a set of matrices and formulae with variable slots, for example:

1st, 2nd, 3rd, 4th ... of January/February/March etc. ;  
 first/second/third ... in line to the throne/team to win the World Cup etc. ;  
 the largest/second/fifth largest/richest ... city in the world ;  
 the longest/second longest ... river in UK/Europe/the world.

Table 8.5

*Wmatrix counts MWEs as one word using underscores*

	Occurrences in BAWE		Occurrences in Buidcorp (education)		+/-	Log likelihood
United_Arab_Emirates	1	0	201	0.04	-	435.55
private_schools	1	0	178	0.04	-	384.6
ministry_of_education	1	0	108	0.02	-	229.8
public_schools	2	0	89	0.02	-	180.45
high_school	6	0	97	0.02	-	174.91
human_rights	236	0.03	3	0	+	162.51
higher_education	15	0	106	0.02	-	157.18
United_States	210	0.02	1	0	+	156.87
a_lot	72	0.01	157	0.04	-	121.66
academic_year	4	0	57	0.01	-	100.52
Hong_Kong	148	0.02	3	0	+	95.16
special_needs	2	0	43	0.01	-	80.93
east_Asia	99	0.01	0	0	+	78.89
international_relations	126	0.01	3	0	+	78.58
english_speaking	2	0	40	0.01	-	74.54
North_Korea	90	0.01	0	0	+	71.72
I think	2	0	38	0.01	-	70.29
rather_than	411	0.04	81	0.02	+	67.67
developing_countries	115	0.01	4	0	+	65.53
education_system	8	0	47	0.01	-	65.36
mother_tongue	8	0	46	0.01	-	63.45
exchange_rate	75	0.01	0	0	+	59.76
going_to	46	0.01	86	0.02	-	57.39
primary_school	4	0	36	0.01	-	57.3
in_fact	249	0.03	39	0.01	+	56.8

## 9 Conclusion

The would-be language detective can pick up on interesting levels of Log Likelihood which its inventor calls a measure of “surprisingness” and follows through with concordances of the most interesting MWEs. Four entries from Table 8.5 will serve to illustrate the power of the concept of MWE as a tool for increasing language awareness. There is a startling contrast between the BAWE writers and the Buidcorp writers in relation to their interest in human rights (236 mentions in BAWE and 0 in Buidcorp). This could be said to be at the level of content analysis. *I think* is overused by the Buidcorp writers. This epistemic stance marker, usually assertive, sometimes a hedge, is more frequent in spoken English. The “good” writers of BAWE almost completely abstain from it in their dissertations. On the contrary, these high achieving writers use *in fact* much more often as a stance marker (249 vs. 39).



## Chapter 9

# UNVEILING VOCABULARY ERRORS IN NON-NATIVE ENGLISH

*Maria Luisa Roca-Varela*

### 1 Introduction

Learning a foreign language (L2) is not an easy task. It involves mastering many linguistic features such as pronunciation, vocabulary and grammar and being able to use the language communicatively. Among other things, in order to be communicatively competent in the L2, it is necessary to have a good command of the foreign language vocabulary and use it effectively. However, the linguistic performance of non-native speakers of English frequently reveals a distinctive use of the language and remarkable lexical problems. These lexical errors may be an obstacle to successful communication and may get in the way of fully understanding a normal conversation. Generally speaking, errors in vocabulary are seen as communication distracting elements (Agustín Llach, 2005) which may have a rather negative effect on message understanding (Hughes & Lascaratou, 1982).

Within the area of vocabulary errors, it has been widely acknowledged that there are certain words which are more problematic than others. False friends represent a peculiar group of words that are particularly challenging for language learners. False friends are words that resemble in form in different languages and have different meanings in each language (Brown, 2006). They are difficult because these words look very much alike in the L1 and in the L2 but contrary to what it looks like, they are not semantically the same (e.g. English *chef* “main cook” vs. Czech *šéf* “boss”, English *carpet* “rug” vs. Spanish *carpeta* “folder”). There are false friends in many different languages. False friends are fairly common in closely related languages (e.g. English *gift* “present” and German *gift* “poison”, Dutch *brand* “fire” and English *brand* “make”, Italian *burro* “butter” and Spanish *burro* “donkey”). Yet, this does not imply that false friends are only found among cognate languages which are etymologically related or descend from a common ancestor. As a matter of fact, there are false friends among unrelated languages, such as Polish and Spanish (e.g. Polish *spirytus* “alcoholic drinks” and Spanish *espíritu* “soul”), English and Russian (e.g. English *sympathetic* “compassionate” and Russian *simpatichtnyy* “good-looking”)

or Portuguese and Czech (e.g. Portuguese *bunda* “buttocks” and Czech *bunda* “coat”).

The existence of these misleading words is admitted to be a bidirectional problem for both native and non-native speakers. On the one hand, non-native speakers might confuse the real intention of an innocent word by a native speaker. Thus, Chamizo (1999) mentions that an American speaker was misunderstood by a Romance language speaker when he uttered the noun phrase “fastidious speech” to describe the talk of the non-native speaker. Apparently, the Romance speaker felt upset because the adjective *fastidious* means “boring” to Romance speakers of French, Italian or Spanish, not “comprehensive/exhaustive” as in English. On the other hand, native speakers might not understand a comment like *Eastwood’s last film was an \*exit* from a non-native speaker. An English speaker might have difficulties understanding this comment where the word *exit* is used in the sense of “box-office success”. In the light of these examples, it can be said that false friends may not only affect the accuracy of a message but also the communicative purpose of language. The communication of a given message can be even more difficult when there is no interaction, when we are in an unidirectional communication mode, i.e. a speech, a lecture, email, when there is no possibility of negotiating meaning on the spot. The presence of false friends in real learner language is the concern of this chapter.

## 2 False friends and learner language

As aforementioned, learning a foreign language is a complex task. It involves time and effort, and making mistakes on the way to proficiency. Apart from that, a good language learner is able to make the most of these mistakes, learn from them and use different strategies wisely to make an effective use of the foreign language. Nevertheless, casual observation of learner language shows the occurrence of some lexical errors which are recurrent in learner language and which might “fossilise” in the learners’ productive use of the foreign language (Selinker, 1994). Thus, when we examine learner language, we may find mistakes which somehow alter the intended message (e.g. *My brother was MOLESTING me yesterday when my mum came into the room.*) or funny slip-ups which make us laugh (e.g. *I like eating SOAP.*), both problems may have their origin in the misuse of a lexical item, such as a false friend. For this reason, the main research aim focuses on the peculiar use of some high-frequency English words which happen to be false friends between English and the learners’ native language that is Spanish in this case. In particular, this survey emerges from the need to identify and unveil the problem of false friends which really lies under the recurrent misuse of some everyday English words in the learners’ writings.

### 3 The study

#### 3.1 Study aims

The present study analyses learner language. Specifically, this study provides evidence of the problem of false friends in the learners' written language and illustrates the constant misuse of high-frequency English words in learner language with corpus data. The data provided in this paper have been extracted from the International Corpus of Learner English (ICLE), a learner corpus which contains argumentative essays written by advanced EFL learners with widely different L1 backgrounds. On the other hand, the lexical items examined are high-frequency English words, i.e. they are words which are listed at the top 2,000 most frequent English words in the Longman Communication 3,000 - a word list, based on the 390 million-word Longman Corpus Network, which illustrates the 3,000 most frequent words in both spoken (s) and written English (w). The last part of this chapter will finally reveal those false friends which continue to be challenging for Spanish learners at upper intermediate and advanced level of English.

#### 3.2 Corpus analysis

After an initial examination of the learners' texts in the Spanish subcomponent of ICLE, which contains 200,376 words and is made up of 500 to 1,000 word argumentative essays written by upper intermediate to advanced EFL learners whose L1 is Spanish, it has been observed that, alongside *actual* and *career* (also problematic for EFL learners with other L1 backgrounds), the English adverb *actually*, the noun *success* and the verb *support* are recurrently used by learners. All these words figure prominently among the 2,000 most frequent English words. Thus, according to the Longman Communication 3,000, *actual* and *career* are classified as w2 words which means that they are in the top 2,000 most frequent written English while *actually*, *success*, *support* are words which are in the top 1,000 most used written words in English (w1). Drawing on corpus data drawn from ICLE, the present study shows how frequent these five words are in learner language. The conclusion is that all of them occur over 13 times in the Spanish subcorpus and it is also remarkable that they show a "quite peculiar" use which differs from the native English use of these words. For instance, phrases like *\*actual society* and *university \*career* are really frequent. The examples below illustrate this.

(1) *ACTUAL society is extremely violent, television has undoubt[ed]ly influenced [in] this increase of violence. <ICLE-SP-UCM-0005.5>*

(2) *When you choose to study a university CAREER, you expect you may get a job within the branch you have chosen; but in the majority of the cases, that is not so. <ICLE-SP-UCM-0030.4>*

### 3.3 Results

The language in this corpus constitutes an important source of information to determine the specific problems with the aforementioned English words which are false friends between both English, the learners' foreign language, and Spanish, the students' mother tongue. The data show that there are some words classified as English "false friends" which are constantly used in the corpus. This corpus also provides us with evidence that these "repeatedly-used" words occur in strange phrases which sound entirely foreign and which show an incorrect word choice. The most outstanding results are illustrated in the table below. Table 1 shows the total number of occurrences of these common false friends, together with the percentages of inaccurate uses attested in the corpus. The columns labelled "raw frequency" show the frequency of occurrence of these words in the Spanish subcomponent of ICLE, and the columns following it illustrate the number of errors and the percentage of lexical mistakes (% of inaccuracy) found.

Table 9.1  
*Frequencies and error percentage in ICLE*

ICLE	Raw frequency	Number of errors	Error percentage
actual	19	17	89.5
actually	21	10	47.6
career	14	12	85.7
success	20	2	10.0
support	42	5	11.9

As illustrated in Table 9.1, there are five words which are quite common in the corpus (see first column) and all of them are misused in at least 10 per cent of the cases.

If we examine the occurrence of these words in learner language, we observe that the misuse of English false friends influences the quality of the learners' written performance negatively. The use of these words is shown in the qualitative analysis that follows. This analysis presents the false friends under investigation and contains information about part of speech (whether it is a verb, adverb, adjective, noun, etc.), frequency (if it belongs to the top 1,000 words –w1– or 2,000 words –w2–) and examples of use from the corpus. The examples presented in this chapter have not

been corrected or modified so the sample sentences below constitute a reliable copy of the learners' words in ICLE. If we focus on the learners' use of these words and the specific problems found, we should draw attention to the following issues.

ACTUAL (adjective, w2) is one of the most popular examples of the phenomenon of false friends. Data in ICLE show that there are nineteen examples which contain the adjective *actual*. Only two out of these nineteen examples show the correct use of the English adjective *actual*. These data indicate that even upper intermediate or advanced Spanish learners do not use this English item in a correct way. Thus, noun phrases, such as \*the *actual* government, \*the *actual* law, \**actual* life, \*the *actual* moment, \*the *actual* social situation, \*the *actual* society and \*the *actual* world, instead of using *present-day* or *current*, remind us of the use of Spanish *actual* "current, present-day".

- \* *Professional soldiers [...] would be a good alternative to our ACTUAL system of military service. <ICLE-SP-UCM-0016.3>*
- \* *The ACTUAL prison system should be \*renovated. <ICLE-SP-UCM-0049.3>*

Apart from that, the second example illustrates the misuse of another English verb "renovate" which would be inappropriate in this context.

ACTUALLY (adverb, w1) is also a well-documented example in the literature of false friends. This lexical item is a false friend between English and many different European languages, such as German, French or Italian. However, the word *actually* seems to be less problematic for learners than its corresponding adjective *actual* (previously analysed). The influence of the learners' mother tongue and the Spanish adverb *actualmente* "currently" is not so obviously perceived in the corpus. This may be due to the fact that both *actually* and *actualmente* may occupy similar sentence functions and can be used in initial position as sentential adverbs in both the learners' L1 and the L2. In most of the sample sentences, *actually* is used to add new information to what has been previously said. It is likely that any Spanish speaker could perceive the implicit meaning of "nowadays" in the students' use of this word. However, this is not clear and the semantic influence of Spanish on the use of this English word does not hinder the understanding of the message; therefore, utterances such as the one below are regarded as correct because it makes sense to any English person, and it is felicitous in terms of structure.

- ✓ *Future generations will suffer the consequences of this relationship between television and the disintegration of the family. Recent surveys in Great Britain indicate that very young children ACTUALLY spend with their parents less than ten minutes per day, which is, undoubtedly, a chilling average (...)*  
<ICLE-SP-UCM-0022.5>

CAREER (noun, w2). The presence and use of the English noun *career*, which is at the top 2,000 most frequent words in English (w2), reveals that learners tend to assign the Spanish sense of “studying a degree” to this word in the learners’ written productions. Thus, the original sense of this English word “job or profession” is completely missing as illustrated in the example below.

- ✘ *When you choose to study an university CAREER, you expect you may get a job within the branch you have chosen; but in the majority of the cases, that is not so. <ICLE-SP-UCM-0030.4>*

Problems with high-frequency English words such as *career* show that learners draw strong associations between the lexical stocks of their mother tongue and their second language. As a consequence, despite the semantic differences and the frequency of these words in each language, whenever there are similarities between both word stocks, the L1 meaning seems to prevail and is also attached to the English lookalike.

SUCCESS (noun, w1). Corpus data suggest that the English noun *success* is being influenced by its Spanish homograph and homophone *suceso* “event” on some occasions. The following example illustrates a different use of the English noun *success* which does not exist in English. The learner uses this word in its plural form \*successes to mean “events/happenings/circumstances”, which constitutes a clear example of transfer. This idea is stressed by the fact that this noun, commonly regarded as uncountable, occurs in its plural form.

- ✘ *Why do the machines exist and work but because a mad absent-minded scientist finds the way to improve our standard of living? Fortunately, this wise person has the wonderful attitude to create thanks to his or her dreamings and objectives, making a continuous mixture of real and imaginaries SUCCESSES intermingled with every new results of the investigations. <ICLE-SP-UCM-0050.3>*

Yet, most instances of this noun show that, in general, students know what *success* refers to in English. They use it in the sense of “accomplishment of something” as well as in well-known word combinations such as *great success*, *to become the key to success* or *with success*.

- ✓ *They didn't get a great SUCCESS. All the romantic poets Wordsworth, Coleridge, Byron... wrote romantic tragedies, they took place in extravagant and distant places. <ICLE-SP-UCM-0014.8>*

SUPPORT (verb, w2). When the verb *support* occurs in positive sentences, it is used in the sense of “enduring difficulties” under the influence of Spanish.

- ✘ *[...] responsibility causes depressions and nervous breakdowns due to the following problems: taxes, children, our jobs and the most modern of all*

*the fear of loosing our jobs, too many difficulties to be SUPPORTED by a person. <ICLE-SP-UCM-0029.3>*

However, there is a high percentage of instances where the verb *support* is correctly used. Whenever the English sense of the verb is kept in learner language, it frequently collocates with words such as *idea* or *argument*.

- ✓ *Let me give more examples to SUPPORT my idea. <ICLE-SP-UCM-0032.3>*
- ✓ *Finally some people SUPPORT the idea of an army formed totally by professional soldiers. <ICLE-SP-UCM-0040.3>*

The results of this study prove that these words occur in unusual contexts and in atypical collocations. There seem to be some factors that make these words special. In particular, the existence of similar looking words in the native language (Ringbom, 2007) and the overgeneralisation process (Scovel, 2001; Kaweera, 2013), strategy consisting of extending the use of certain words to contexts where they do not apply, constitute an important downside to language acquisition and use.

As for the overall results and taking the quantitative data into account, it is quite remarkable that 32 per cent of the examples found of these high-frequency English words are misused by learners, and of course, they may be problematic and lead to important misunderstandings. In fact, the results of the qualitative analysis reveal that the mishandling of these lexical items may generate confusion and misinterpretations. Most of these lexical problems reveal an evident influence of the L1 and the effects of negative transfer on the use of English false friends.

At any rate, the data in the corpora show that high-frequency English words which are false friends with words in the learners' mother tongue pose serious problems which should be addressed in EFL classrooms. Thus, high-frequency words such as *career* or *actual* (among others) are frequently misused by learners in their written performance. When analysing the lexical mistakes in the use of false friends, we observe that the learners' mother tongue exerts a powerful influence on the use of the L2 vocabulary; thus, the fact that there are similar words in the learners' mother tongue have pushed learners to use English false friends as if these words were fully translation equivalents in both the L1 and the L2. Therefore, negative transfer accounts for most of the lexical errors affecting English false friends found in the corpus. Consequently, these results point at the need for the teachers' action in the classroom.

#### **4 General conclusions**

This chapter was conceived to analyse some of the most remarkable problems in learner language. It focuses on the misuse of high-frequency English words which are also false friends between the L1 and the L2 in the written production of Spanish

learners. This study was carried out with the support of a learner corpus which contains samples of written texts produced by Spanish learners of English. One of the main aims of this corpus-based study was to examine the learners' problems with high-frequency English words which are common in learner language and challenging for learners so as to help teachers meet the needs concerning this lexical area.

It is worth noting here that this study shows with empirical data that mistakes with English false friends still persist in the production of upper-intermediate to advanced learners. In a general sense, the number of false friends which are accurately used is higher than the number of false friends which are incorrectly used. Nonetheless, the influence of the students' mother tongue is perceived in 32 per cent of the total. Therefore, the data in the learner corpora analysed indicate that Spanish students have problems with some of these words. Thus, Spanish learners resort to some English false friends and use them in a different way. A noun, such as *career*, which is frequently used in learner language, is persistently causing problems. Learners use it to mean "university course", which constitutes an obvious case of crosslinguistic transfer. In fact, students appear to be inevitably tempted to use this word in the wrong context when they are talking about their university studies (e.g. *Few years ago, the study of a CAREER was destined to the offsprings from wealthy families* <ICLE-SP-UCM-0001.3>). In this case, teachers must mention this problem explicitly and provide learners with some clues so as to avoid any mistakes that could arise from the misuse of this word.

Regarding the reasons accounting for the learners' problems with English false friends, the data gathered in the corpora reveal that Spanish learners of English resort to some English false friends and use them as if they were true translation equivalents for their Spanish quasi-homograph counterparts. Learners rely on their mother tongue excessively when using English; learners use Spanish as a reference point to write in English. Most of the mistakes found are very likely the result of the students' unawareness of these words. An increasing exposure of the learners to the English language, an early incorporation of these lexical items in language learning, the use of suitable techniques for the teaching and recycling of these peculiar words (e.g. meaningful examples, clear contexts of use) and the learners' effort to learn these words and to be accurate in their use of English vocabulary are needed for the correct acquisition and use of these lexical items. All these issues are connected with the following part which discusses the pedagogical implications of this study.

## **5 Pedagogical implications**

Learner corpora provide language experts with valuable information on the specific problems that learners might encounter. Those lexical mistakes involving the misuse of a false friend should not be overlooked since these lexical mistakes may hinder

communication (e.g. *Nowadays in the middle of this kind of life based in the COMODITY and the try to be in a high level of life[.] <ICLE-SP-UCM-0041.3>; In other places such as New York, where it never have snowed in this way, last week there were people incomunicated and they were SUPPORTING temperatures of fifteen degrees bellow zero. <ICLE-SP-UCM-0052.4>*).

As regards the teachers' action in the EFL classroom, language instructors should raise the learners' awareness of false friends; thus, explicit instruction is highly recommended (Chacón Beltrán, 2006) so that learners become aware of the semantic differences between similar items in the second language and their own language in a conscious way. Lengeling (1995) mentions some common vocabulary teaching practices which might be useful for an effective teaching of false friends. She urges teachers to use strategies, such as *explaining* "how these words are different and what the correct word is for the corresponding word in the target language", *collecting* "those FF that cause problems and incorporate their teaching in the classroom" and *recycling* those problematic items from time to time (p.5).

As is the case with any other vocabulary items, teachers should emphasise the role of context in vocabulary learning (Lewis, 1993) and should not approach lexis as a compilation of single words with fixed meanings. Providing illustrative definitions and suitable collocations might be useful to promote the students' suitable learning and correct use of English false friends. A useful task could be to make flower-like diagrams that help students illustrate the most basic semantic and pragmatic features of false friends (Figure 9.1).

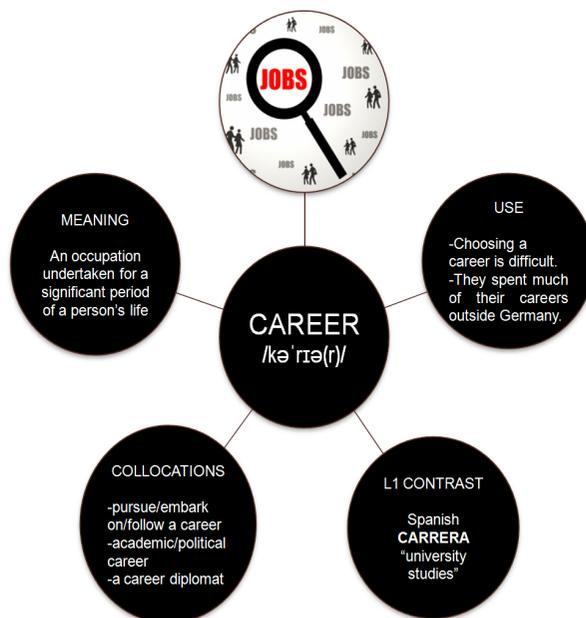


Figure 9.1. Example of a flower-like diagram

The aim would be to work on the different sides of vocabulary knowledge and keep vocabulary records well-organised. In this activity, students are asked to write down the information about the false friends they consider most important and distribute this information into five main sections: meaning, collocations, examples of use, a representative picture and main contrast with the mother tongue.

The false friend studied should figure prominently at the core of the flower with its corresponding pronunciation (see Figure 9.1); the petals are filled in with information about meaning and use.

Besides the use of a wide range of classroom tasks, the teachers' attitudes towards these words and their strategies to deal with them should be also examined. This issue opens a new interesting path for further research. Thus, research on the techniques and strategies used by language teachers to introduce, explain and practice these lexical items is needed in order to test the usefulness of these measures and the effectiveness of these strategies to reduce the lexical problem of false friends. In sum, teachers and learners must join forces and work together to overcome difficulties and make the most of the learning experience.

### Section three

## INDIVIDUAL DIFFERENCES IN L2 LEARNERS

SLA is an interdisciplinary field which involves the study of many areas, including the explanation of individual differences in SLA. Though individual learner differences, such as language aptitude, learning styles, motivation, and anxiety, are reported in the SLA literature as correlating strongly with L2 achievement, their investigation has been marginalised in the field of SLA.

In the first chapter of this section Gabriela Lojová offers an overview of selected internal factors determining the effectiveness of foreign language (FL) learning and contributing significantly to the successful implementation of some principles of humanistic education. The overview, based on state-of-the-art knowledge of applied psycholinguistics, aims to identify and briefly analyse some topical aspects of the internal factors in question. After providing a brief contextual framework, she puts the main emphasis on the impact of psycholinguistic insights on the effectiveness of FL learning in the Central European educational context and on the role of the teacher in acknowledging learners' internal factors and individual differences. Adopting this perspective, she briefly analyses the following factors: brain functioning and brain-based FL learning, age, FL aptitude, the impact of intelligence on FL learning, some learning styles, and the power of effective learning strategies. The importance of applying a learner-centred approach implicitly underlies all the ideas presented in the chapter, as its aim is to foster the humanisation of FL teaching by shifting the focus from teaching to learning. This approach can undoubtedly help teachers gradually replace enduring traditional teaching approaches with more effective ones. She believes that in today's context teachers clearly have to develop their pedagogical-psychological-linguistic thinking to enable them to flexibly adjust their teaching to learners' needs and changing pedagogical situations. Furthermore, encouraging introspection may help FL learners get to know themselves better, which is an essential precondition for autonomous lifelong FL learning. Therefore the importance of including these internal aspects into FL teachers' professional development is emphasised throughout the chapter, as is the urgent need to carry out more in-depth research in our educational context.

After an overview of several factors that account for differences between individual learners, in the following research-based chapter Monika Černá focuses on motivation. The perspectives and approaches that have been adopted to investigate motivation have changed since the introduction of Gardner and Lambert's social psychological construct of integrative motivation in the 1970s. This chapter presents a study which is a part of a larger research project. It focuses on two aspects of language learning motivation, its dynamic nature and temporal variation, while applying a process-oriented approach. The author aims to shed light on the temporal progression of language learner motivation over a long period of time. The respondents are Czech learners of English enrolled in English language teacher education programmes at three universities in the Czech Republic. On the basis of quantitative data obtained through a questionnaire, respondents for narrative interviews were selected. The interviews elicited qualitative data, i.e. the respondents' own accounts of their English language learning experience. The biographies of three selected learners are analysed with the aim of uncovering the dynamicity of individual learners' motivation and its fluctuations resulting from changing contexts.

# AN EXPLORATION OF INTERNAL FACTORS IN FOREIGN LANGUAGE LEARNING

*Gabriela Lojová*

## 1 Introduction

Despite many changes in foreign language (FL) teaching in Central European countries over the past three decades, experience still shows that students achieve a relatively high level of knowledge about the language they are learning, but their ability to use this knowledge in real life communication is often inferior, which signals that the methodologies used are not as effective as hoped. In an attempt to enhance the quality of language education, experts search for more effective approaches and models of FL teaching, providing teachers and FL methodologists with sound theoretical background underpinning various teaching methodologies. In this endeavour it is of vital importance to bear in mind not only linguistic and pedagogical aspects of FL teaching processes, as is common in traditional approaches, but also psychological (or psycholinguistic) aspects determining the process of FL learning. Researchers in this latter field are aiming to explain the neuropsychological and psycholinguistic foundations of learning, create models of learning/acquisition processes, argue their practical importance, and critically analyse various psychological and educational aspects, etc. In so doing, experts have been emphasizing the need to shift our main focus away from teaching and more towards the learner and learning processes. Nowadays, it is generally accepted that teaching approaches which respect the developmental and individual differences of learners are considered to be the most effective. This orientation is rooted in a social-constructivist theoretical framework emphasising the importance of not only cognitive processes but also other internal and external/social components determining the effectiveness of learning/teaching (Lojová, 2010). According to these learner-centred approaches and in accordance with fundamental humanistic and constructivist positions, it seems obvious that everything that is achieved in the classroom eventually depends upon what goes in the learners' minds. Whatever pedagogical decision we take, we have to keep this fact in mind.

Learning is an extremely complex and multifaceted process determined by numerous internal and external (environmental) factors which enter into variable relations in

dynamically changing pedagogical situations and which have a significant impact on the final effectiveness of teaching and learners' outcomes. Though being aware of a complex system of interrelated and co-functioning external determiners, such as social, cultural, linguistic environments, and educational conditions, the focus of this paper is on internal factors that are considered especially influential in the specific process of learning foreign languages.

However, before exploring these internal factors the primary issue to emphasise from the perspective of the co-functioning of internal and external factors is the difference between second language acquisition (SLA) and foreign language learning (FLL).

Despite long-running discussions among experts about the differences between these two processes, particularly among cognitive psychologists (e.g. Doughty & Long, 2003; Krashen, 1981; Lojová, 2011b; Sternberg, 2009) it seems striking how frequently the notions are used interchangeably and fundamental differences neglected or underestimated, even in professional contexts. It might be due to diverse research approaches, related topics being discussed, or differing educational or linguistic contexts. However, a fundamental difference which has not been resolved yet seems to be the dilemma about the nature of language, i.e. Chomsky versus Piaget. The relevance of this tension not only for SLA but also for FLL should be considered and researched more urgently, as well as the related nature-nurture questions and their practical implications.

Being aware of the plurality of opinions among experts, for our purposes the following definitions may be adopted: a *second language* is any language different from the first language/mother tongue a child is exposed to that is used in its environment as a natural means of communication; whereas a *foreign language* is any language used in the school context other than the first language, the second language or the language of instruction. *Acquisition* is a subconscious process whereby children acquire a language in an intuitive way in a natural environment or in non-formal settings; whereas *learning* is a goal-oriented conscious process whereby children realize that they are learning a language, usually in formal instructed settings. The psychological analyses of the phenomena are very complex and are discussed in many professional publications. However, from our pedagogical perspective the essential question is how learners perceive a language. This may be perceived as follows:

(1) a natural means of communication, with implicit intrinsic motivation, that they acquire in their daily environment (or informal settings) without realizing it. Here the social environment naturally stimulates their second language performance and provides them with immense language input, as well as with corrective feedback needed for subconscious testing of hypotheses about language functioning; or (2) as just one of many subjects at school occurring in formal settings, for which learners have to learn, do their homework, be evaluated and marked etc. Learners quite frequently

cannot see the purpose of learning a language as they do not live in a target language environment.

These fundamentally different subjective perceptions stimulate different learning mechanisms and affective variables, particularly motivational, which are of crucial importance.

Even if there is no clear-cut boundary between these two processes (learning and acquisition) in many real life situations or educational contexts, if one of the processes (SLA or FLL) is highly dominant, it is of crucial importance to take the differences into consideration. The distinction leads to diversity in education resulting in the existence of different models of language learning. It is obvious that in most mainstream schools in Central European countries, learning English as a foreign language is clearly dominant.

Another reason why this distinction is so important is that SLA has been the subject of more frequent research than FLL. How much can we draw on SLA research findings, how relevant are they for FLL? Can they be automatically adopted and adapted to FLL and teaching? Even if there are some obviously relatively universal phenomena, it is necessary to critically analyse and reflect and carry out research on the importance of diverse linguistic, socio-cultural, and educational environments: How powerful and how different is the impact of these factors upon the processes of SLA and FLL? We should be cautious about tendencies to overgeneralize on the basis of existing SLA research, which might lead to ineffective or inappropriate pedagogical decisions, or can even be counterproductive (e.g. the attempts to teach EFL in central European environment without any focus on form, which might work quite effectively in SLA environment; or the boom of introducing FLT to kindergartens because of the influence of the popular belief that *children learn languages more easily and faster*).

## **2 Internal factors: individual differences in FLL**

There is a broad range of differences in the level of FL proficiency individual learners attain. Some achieve a high level with apparent ease, whereas others have a hard time mastering even the basics of FL after the same amount of learning. Why do such considerable differences in learners' outcomes exist? Researchers have been examining a variety of internal factors or inner predispositions within learners' minds, namely linguistic, cognitive, and affective components, to explain why some learners are more successful than others (Ehrman, 1996; Dörnyei, 2005; Lojová, 2011b; Skehan, 1989). The following sections focus on a brief overview of the most relevant factors for FL learning.

## 2.1 The brain

FLL is an extremely complex neuropsychological process based on neurobiological functioning of the brain. Despite the recent rapid development of neurosciences and brain scanning technologies, research findings relevant for FLL are still very limited. However, the possibilities of the application of neuroscience's knowledge to the theory and practice of FL teaching are very promising as far as higher effectiveness is concerned. So far FLT recommendations have been often speculative, based on trial and error or on generalizing empirical experience without understanding the underlying neural mechanisms. Brain-based language learning, i.e. organizing FL teaching around the way the brain learns best, may be the simplest and most critical core reform to be initiated. Based on sound research findings, it provides an explanatory framework for redesigning FLT methodology with all its components (Jensen, 1996).

The main areas of inquiry in neurosciences relevant from our perspective are:

- 1) The localization of language in the brain, i.e. of L1, L2 and FL. The knowledge of the localization of different languages and their functions, their interconnection and mechanisms of co-functioning could shed more light on questions related to interlingual transfer and interference. It might help us decide when and how to use the mother tongue in the FL classroom effectively, give hints for more effective teaching and learning strategies.
- 2) Understanding the dominance, cooperation and complementary functioning of brain hemispheres processing speech stimuli differently (the left hemisphere "scientifically" whereas the right one "artistically") can shed more light on many aspects of FL learning and teaching (e.g. Blakemore & Frith, 2005; Jensen, 1996; Lojová, 2011b; Schumann et al., 2004). Nowadays it is obvious that in these processes numerous structures and functions of both hemispheres are actively involved. For effective learning both hemispheres should be stimulated, independently of which one is dominant, although learners' individual differences must be respected. However, traditional FL teaching, still dominating in many schools, activates the left hemisphere much more, overloading it with a lot of information to be learned. Learners are supposed to memorize large amounts of often unnecessary, irrelevant vocabulary and phrases, and boring grammar rules in a decontextualized way, which they do not find meaningful and are not able to use in real communication. Activating also the right hemisphere and stimulating its functions can undoubtedly enhance learning. Whole-brain learning enables learners to connect learned pieces of information in their own way, to create their own associations, to construct the inner representation of the FL system in their own unique way and ultimately to use learned FL in their real life communication.

As for brain functioning and FLL, the main tasks for teachers are:

- a) To use a wide range of learning activities, stimulating both hemispheres so that learners can subconsciously or consciously select the ones that enable them to learn better according to their dominance (For FL learning and teaching, the principles of brain-based learning are incorporated in numerous teaching materials available as “brain-friendly” publications on <http://www.brainfriendly.co.uk/shop/index.php>).
  - b) To help learners get to know themselves better, identify their dominance so that they can create more effective learning strategies and understand possible problems they may encounter in FLL.
  - c) Teachers should identify their own dominance to avoid the subconscious tendency to prefer teaching strategies and learning activities that are effective for themselves, rather than for their learners.
- 3) The neurophysiological roots of cognitive processes are a complex and multifaceted field of research, whose findings are of key importance for establishing effective FLT methodologies. Understanding how learners process FL (e.g. perceive FL visual and auditory stimuli, how they discriminate sounds, recognize and integrate language units, generalize and differentiate, how they create perceptual schemata, concepts and higher level structures; how FL messages are encoded, transmitted and decoded) should determine various pedagogical aspects, such as how to modify input, help learners to think in phrases, enhance linguistic tolerance, develop guessing and paraphrasing skills, understand the need for meaningfulness, etc. (Sternberg, 2009). Neuropsychology of memory functioning seems to be of core importance in explaining how to create more permanent and multifaceted associations, how the learned material is remembered, stored and retrieved from memory, how declarative and procedural knowledge is stored and interconnected, etc. The practical implications of this knowledge for FLT are immense (e.g. Lojová, 2011a; Lojová et al., 2011).
- 4) Nowadays it is obvious that equally important is the knowledge of the neurobiological base of numerous non-cognitive or affective factors (e.g. motivation, emotions, relations, attitudes, anxiety) whose impact upon the learning achievements might be of prime importance.

## 2.2 Age

The internal factor most frequently discussed among professionals as well as the general public is that of age and FLL. Generally speaking, there are two fundamental issues driving research on the significance of age: (1) what is the right age to start FLL? (2) differences in FLT methodologies applied to different age groups.

### 1) What is the right age to start FLL?

As for the best age, state-of-the-art research does not provide sound evidence to support any generally acceptable recommendation. Actually, research on age effects in FLL is still in its infancy. It seems that the fundamental question of the existence (or not) of a critical period has not been answered yet: Is there a biologically determined optimal, sensitive age when language functions can develop easily and more effectively? If yes, how can it be applied to FLL? (Lennerberg, 1967; Lojová, 2011b). Experts differ in their opinions considerably. There are lively arguments for and against early FL teaching supported by different theories, research findings and observations (De Keyser, 2000, 2012; Nikolov & Djigunovič, 2006).

Generally speaking, experts endorse three different opinions: *the sooner the better*, *later is enough*, and *it depends*.

#### a) *The sooner the better*

Experts taking this position support the Critical Period Hypothesis (CPH). They believe in brain plasticity at an early age which enables rapid subconscious learning and developing innate brain capacity. They also suggest that children's potential is wasted if it is not being developed sufficiently at an early age. Therefore this potential, and cognitive capacity in particular, must be stimulated by an enriched environment, so why not also by the opportunity to learn a foreign language.

#### b) *Later is enough*

Experts endorsing this opinion do not believe in the CPH. Generally they fight against overloading children at an early age, which is, they argue, predominantly the age of play. Children's potential should be developed and stimulated in a natural way and in a natural environment. However, a foreign language is not considered to be a natural environment, unlike the mother tongue or a second language. Furthermore, they warn that cognitive or language development should not be overestimated to the detriment of other components and functions, particularly emotional and personal. They also believe that children can learn a foreign language faster, better and more effectively later when the learning mechanisms that encounter to the conditions of instructed classroom language learning are developed (Doughty & Long, 2003; Hinkel, 2005; Krashen, 1982; Muñoz, 2008).

#### c) *It depends*

This is a kind of integrated position. Many experts claim that each age has some advantages and disadvantages that have to be seriously taken into consideration when deciding about a specific child or educational settings. In addition, they do not consider age to be the only powerful or decisive variable. They emphasise rather the importance of other variables, namely various inner predispositions, linguistic and socio-cultural environments, and educational conditions in particular. This position seems to be the most plausible also in current Central European context.

However, nowadays there is an overall tendency to lower the age of starting to learn and teach languages, supported also by EU language education policy (European Commission, 2002). The underlying theoretical framework, objectives and methodological guidelines are published in document *A Policy Handbook: Language Learning at Pre-primary School Level: Making it Efficient and Sustainable* (European Commission, 2011, p. 7). The general recommendation for Central European educational context could be concisely summarized as follows: it is not recommended to teach a foreign language in a kindergarten if there is not an appropriate educational context (qualified teachers in particular). It is crucial that children be exposed to more than one language, as a way of opening their minds to multilingualism and an awareness of diversity of languages and cultures. In doing so, children develop the sense of otherness and cognitive flexibility, as a preparatory stage for creating the inner representation of a foreign language that will be subsequently learnt in primary school as part of the formal curriculum. The basic principles and recommendations for FL teaching in pre-primary education in our context are analysed in (Lojová, 2005/2006a, 2005/2006b, 2012/13).

Another topical question is whether to introduce the first foreign language in the first year or in the third year of primary education (or even later). Developmental psychology claims that under normal conditions at the age of six cognitive capacity, structures, mental functions and mechanisms are mature enough to start creating the inner representation of another linguistic system (Farkašová, 1998, 2002; Moon & Nikolov, 2000; Nikolov & Djigunovic, 2006). However, adequate pedagogical conditions are crucial, namely qualified teachers and a FL teaching methodology adapted in such a way that the target language is spontaneously acquired rather than consciously learnt (Lojová & Kostelníková, 2012; Lojová & Straková, 2012).

Despite the considerable plurality of experts' opinions on the most suitable age to start to learn a FL, nowadays in professional discussions there is a growing consensus that, viewed from a long-term perspective, learners who start earlier are at an advantage (Harleyová & Wangová, 2004, in Štefánik, 2004; Marinova-Todd et al., 2000). It is suggested that if learners are expected to study for many years (through primary and secondary education), it is better to start earlier as they are likely to gradually achieve a higher level of FL proficiency and, what is of vital importance, their achievements will be more stable and longer-lasting. This hypothesis is supported also by neuroscience, mainly by research focussed on the CPH (e.g. Bialystok & Hakuta, 1994; DeKeyser, 2000; Lojová, 2011b; Skehan, 1998).

## 2) Differences in FLT methodologies applied to different age groups.

As for specific methodologies applied to various age groups, the knowledge of developmental characteristics relevant for FLL and their age specific differences is undoubtedly of paramount importance. Not only the theoretical knowledge but mainly the ability to apply it to a language classroom, to understand the impact that

developmental characteristics have on learners' behaviour and the process of learning. Generally speaking, when contrasting young and adult learners, the crucial differences are in their cognitive processes (thought, perception, memory functioning, attention, language) as well as in numerous affective components (motivation, needs, interests, anxiety, social relations, attitudes, etc.) (e.g. Krashen et al., 1982; Lojová, 2011b). Respecting these differences and adjusting all aspects of methodology accordingly is an essential prerequisite for effective FL teaching.

Even if nowadays there are numerous academic publications on developmental psychology available, there seems to be a need for a more practically oriented guidebook helping FL teachers to use the abstract knowledge directly in a FL classroom. Chapters on practical developmental psychology for FL primary teachers are available in Lojová and Straková (2012); for pre-school education in Lojová (2005/2006 a, b).

### **2.3 Aptitude**

Researchers have been looking for answers to numerous questions related to the existence, structure and impact of language aptitude (LA) such as: Is there such a thing as LA? What is its nature or substance? Do we need any specific aptitude for mother tongue acquisition? If yes, is it the same as the aptitude for learning a foreign language or is it different? Is a specific aptitude necessary for spontaneous language acquisition as well as for instructed FLL? What is the structure of LA? Can LA be measured and how? What is the predictive value of LA for successful FL learning?

Research on LA flourished in 1960-70s resulting in developing various psychometric tools. Leading experts Carroll and Sapon (1958) created the Modern Language Aptitude Test (MLAT), which became very popular and its numerous modifications spread quickly worldwide. The boom of LA testing reached also Slovakia in the 1980s. Adapted to our language and context it was used for the selection of pupils for "language classes" in primary education (Malíková, 1989). However, subsequently experts started to criticise it and doubted the validity of any aptitude test due to the lack of sound underlying theoretical knowledge (Dočkal & Schranzová, 1992), considering LA to be simply a hypothetical construct. As Skehan (1998) says, we cannot measure LA unless cognitive psychologists and neuropsychologists find out what is at the core of aptitude. Otherwise it is only speculative and, however sophisticated, tests cannot be valid. The research in the field has been at a standstill since.

As far as the structure of LA is concerned, i.e. what LA consists of, it has not been definitely specified yet. However, it seems obvious that LA is multicomponential. According to Skehan (1989) it consists of three important components that have been theoretically analysed: language analytic capacity, memory ability, and phonetic

coding ability. However, they are neither clearly defined nor are their roles in language learning satisfactorily specified. A relatively frequent hypothesis suggests that working memory may be a central component of LA (Carroll, 1965; Skehan, 1989). Its capacity may influence the ability to acquire native-like sensitivity to linguistic cues in second language and to comprehend structurally complex sentences correctly and efficiently (Miyake & Friedman, 1998). Of crucial importance for the effectiveness of FLL is the fact that LA components can compensate for one another. Therefore for language learners it may be useful to know the structure of their LA, i.e. their learner type, so that they could understand the sources of relevant learning problems and create more effective FLL strategies.

Another important question is the predictive value of LA, or in other words, how important it is for successful FLL. Some believe, that LA is one of the most important factors (e.g. Bialystok & Hakuta, 1994; Skehan, 1989; Sparks & Ganshow, 2001) as numerous psychometric studies have identified the learner's LA as an important determinant of FL proficiency (Miyake & Friedman, 1998). However, this hypothesis is also highly disputable and there is no clear research evidence to support it.

It is important to mention that nowadays there again seems to be an increasing number of primary schools in Slovakia trying to test LA as a tool for selecting pupils to some specialized schools focused on FLT. However, educational psychologists warn against such tendencies for several reasons: (1) the negative impact of "labelling"; (2) the validity of the used tools is not guaranteed; (3) primarily due to the very basic psychological equation:  $skill = aptitude + activity$ . Therefore if somebody insists on testing pupils, then a carefully selected set of tests should be used measuring not only aptitude but also other inner predispositions and personality traits that jointly contribute to final learning achievements.

Another closely related inner predisposition is *intelligence*, which experts tend to approach as a kind of aptitude. Are more intelligent people better FL learners? The answer again is more complex. The state-of-the-art knowledge could be summarized in a simplified way as follows:

- 1) Successful instructed FL learning may be enhanced by higher IQ. However, IQ does not seem that influential for natural acquisition of a language;
- 2) It depends on different language skills: the impact of IQ on the effectiveness of developing speaking and listening comprehension seems less significant than on academic skills such as reading and writing;
- 3) IQ seems crucial for achieving a higher level of proficiency based on complex inner representation of a target language and the higher level of cognitive processing;
- 4) From an educational perspective, the structure of IQ seems to be of fundamental importance rather than an overall IQ score. As Howard Gardner's theory of multiple intelligences claims: Everybody is good at something different. Let us learn English through our dominant IQ and we will learn more effectively

(Gardner 1993, 1999). As this theory is quite popular and widespread among educators, nowadays numerous teaching materials are available for EL teachers (e.g. Christinson, 1996; Rinvoluceri, 2005).

## **2.4 Learning styles (cognitive styles, learner types) and strategies**

*“If the child can’t learn the way we teach, maybe we should teach the way they learn.”* (Estrada, n.d.).

Each learner constructs his/her subjective picture of the reality, processes information and solves problems in his/her own unique way due to his/her unique learning style. Learning styles are generally considered to be important individual features determining the effectiveness of FL learning/teaching processes. An individual learning style is a complex characteristic consisting of a number of physiological, cognitive and affective features reflecting the complexity of mental functioning. It is a relatively stable, consistent way of functioning, reflecting underlying causes of behaviour. Generally, it is believed that learners differ considerably in their preferred ways of learning foreign languages and that the best results are achieved when learning/teaching processes can be adapted to the specific learning styles of any individual. Learning styles are being studied and researched intensely, not only by cognitive psychologists but also by experts in FL education. The most influential researchers in the field are R. Oxford, J. Sperry, R. Schmeck, M. E. Ehrman and J. M. Reid (e.g. Ehrman, 1996; Doughty & Long, 2003; Dörnyei, 2005; Hinkel, 2005; Reid, 1998; Skehan, 1998). They attempt to explain how learners differ in perceiving foreign language stimuli, in processing and storing knowledge about a language, creating the inner representation of a target language system, in responding to stimuli in varied conditions, and in their unique way of using learnt knowledge and skills in speech production and communication in diverse real life situations. So far, however, research findings have provided us with very few unequivocal conclusions. Therefore theoretically there are a lot of controversial claims leading to a plurality of opinions, different interpretations and taxonomies. In the specific processes of FL classroom learning/teaching, the most frequently discussed seem to be learning styles classified according to following criteria:

- Dominant perceptual channel (visual, auditory, kinaesthetic type);
- Dominant brain hemisphere (left vs. right hemisphere);
- Structure of multiple intelligence (Gardner’s theory);
- Structure of aptitude (memory vs. analytic type);
- Field dependency;
- Ambiguity tolerance;
- Reflexivity and impulsivity.

Field dependency seems to be the most frequently researched learning style in second language acquisition. All the listed learning styles are analysed in detail in professional literature (e.g. Ehrman, 1996; Lojová, 2011b; Lojová & Vlčková, 2011; Mareš, 1998).

*Learning strategies* are so closely related to learning styles that in reality they can hardly be separated. However, theoretically they differ significantly. Learning styles are considered to be consistent, enduring tendencies and more or less given predispositions to approach learning tasks in a specific way, i.e. to develop particular strategies, which vary interpersonally. On the other hand, learning strategies are developed behaviours or specific methods of approaching a learning problem or task, which might vary also intrapersonally.

Learning strategies were brought into research focus in the 1970s when educators realized that learners were not learning effectively. Due to the influx of information to be learnt, traditional teaching approaches were not effective anymore and the need for change in education became urgent. At the core of the change was the idea of educating autonomous (independent, self-directed) learners who are able to think, understand, analyse, evaluate, and consciously control their own learning process. The main task has been to teach learners how to learn effectively, i.e. to get them to think about how they learn, to analyse their needs and problems, to suggest possible solutions, and to self-evaluate their own progress; rather than to transmit knowledge and get learners to memorize facts and passively follow teachers' commands. Currently there are numerous publications available also in the field of foreign language teaching (Chamot et al., 1999; Lojová & Vlčková, 2011; O'Malley & Chamot, 1990; Oxford, 1990; Vlčková, 2010) providing theoretical analyses, taxonomies, research findings, as well as recommendations for classroom teaching. They analyse learning strategies for all language skills and means, phases of learning process, for various learning environments, etc. Understandably, there are numerous factors having an impact on the development of learning strategies. However, pedagogically speaking, the crucial issue seems to be training and intervention, and the role of a teacher in particular. The main idea is expressed in a motto: *I know that I can't teach my learners everything therefore I have to teach them how to learn on their own effectively* (n.d.).

### **3 Practical implications**

As far as FL teachers are concerned, there are four general implications for their pedagogy arising from the perspective discussed in this chapter:

- 1) They should develop their learners' metacognitive awareness, get them to know themselves better through introspection, help them identify their learning styles

and utilize the knowledge for enhancing their learning, i.e. consciously utilize their strong points, reduce their weaknesses and develop effective learning strategies based on their learning styles.

- 2) They should get their learners to think about how they learn and to analyse their learning strategies and possible problems. They should help them create more effective learning strategies either directly or indirectly through various elements of their methodology such as: evaluation, types of homework, discussions, using relevant classroom activities, etc. (Lojová & Vlčková, 2011; Long & Doughty, 2009). In so doing, they would educate more independent/autonomous FL learners. Nowadays in many textbooks there seem to be an increasing number of learning tasks and activities focussed on learning styles and strategies that can considerably contribute to fulfilling this objective. (The overview of the textbooks used most frequently is available in diploma thesis by Pastorková (2013)).
- 3) Each teacher should get to know his/her own learning styles and related strong and weak points. Subsequently they should consciously control and modify their classroom behaviour so that all the learners could benefit from their methodology. Teachers must consciously reduce the subconscious tendency to prefer and overestimate learning tasks and activities favouring their own learning styles and keep in mind that all learner types should benefit from their lessons equally.
- 4) Being aware of the characteristics analysed above and respecting them in their work, teachers can considerably enhance the effectiveness of foreign language teaching by using a wider range of learning tasks and activities so that each learner can consciously or subconsciously select the ones that match their learning styles.

#### **4 Conclusion**

Respecting internal factors and individual differences can be considered one of the most effective means of humanizing FLT. Nowadays it is obvious that knowledge about learners' individual differences opens up a new perspective and view on FL teaching. Understanding their nature may throw some light on how to maximize the outcomes of FL learning and teaching. It helps teachers understand better their learners – their behaviours, attitudes, achievements – and to see the roots of some problems that learners are faced with. It enables teachers to purposefully control classroom activities and facilitate learning process more effectively, to create more suitable conditions and a positive classroom atmosphere in which learners can learn a target language in accordance with their needs, interests and potentials. It provides them with the guidance on how to approach learners with different capabilities and develop their foreign language communicative competence to their own potential. It enables them to achieve the level of humanisation where learners can enjoy FLL and

develop their communicative competence optimally rather than be stressed, frustrated, overloaded and demotivated, as it is rather frequent in a traditional classroom.

As far as research is concerned, nowadays there is an urgent task to pay more attention to the field discussed. In the interdisciplinary field of FLL researchers do not have to start from scratch. They can turn not only to SLA research findings but also to other branches of study examining human being, adopt their findings, evaluate their relevance for FLL, adapt and apply them to these specific processes. Key areas include experimental, educational, social and developmental psychology, psycholinguistics, pragmalinguistics, neurolinguistics, neurodidactics, etc. Drawing on insights from these areas would undoubtedly lead to an increasing efficacy of FL teaching and to the gradual replacement of traditional teaching approaches with innovative ones.



## Chapter 11

# MOTIVATION IN LANGUAGE LEARNING: FOCUS ON INDIVIDUAL LEARNERS' STORIES

*Monika Černá*

## 1 Introduction

Motivation is a term with a meaning which is understood in various ways. It is sometimes used as a catch-all term for explaining the success or failure of any complex task when learners are simply described as *(un)motivated*. However, things are not simple at all. Though motivation is difficult to define, it plays a major role in second language acquisition (SLA). In awareness of the difference between second language acquisition and foreign language learning (Gass & Selinker, 2008, p. 7), the former is used in this text as a superordinate term to denote both processes (Ellis, 2008, p. 6) in situations when the distinction is not important. Contrary to that, the term foreign language learning is utilised to highlight the difference. The same principle guides the use of the term second/foreign language; L2 is used as a synonym for second language. The present study aims to explore how foreign language learners' motivation to learn English has changed in the course of their lives.

## 2 Motivation in SLA

Second language acquisition is an interdisciplinary field which involves the study of many areas, including the explanation of individual differences in the process of second language acquisition. According to Ellis (2008), this area of investigation has remained marginalised in SLA, but recently there has been research attempting to explain why some learners succeeded in learning a second language better than others (p. 643). To answer the question it is not enough to explore cognitive factors. Ortega (2009) emphasises that because of the nature of human behaviour it is volition and motivation, i.e. conative influences, that can make language learners succeed or fail (p. 168). In addition to language aptitude, personality, and anxiety, motivation is considered a “core factor” (Ellis, 2008, p. 644) which accounts for individual learner differences. The label implies that motivation is likely to influence the progression of second language learners, as well as their ultimate achievement, in a substantial

way. There have been attempts to discern the centrality of motivation in explaining the relative degree of success achieved by different second language learners. For example, Gass and Selinker (2008, p. 426) quote Skehan (1989), who claims that motivation appears to be the second strongest predictor of success, trailing only aptitude. Ortega (2009) analyses several studies and reports that, depending on the study being quoted, motivation explains up to 35 per cent of the variation shared between motivational quantity and second language achievement (p. 188). The conclusions confirm the central role of motivation in language learning; the psychology of motivation, however, does not explain how its findings relate, concretely and in detail, to the processes of second language development (Cook & Singleton, 2014, p. 102).

The current understanding of the concept of second language learning motivation is the result of decades of L2 motivation research. Throughout the years various schools of thought in psychology have contributed to the (re-)construction of the concept – moving on the path from Lambert and Gardner’s influential concept of *integrativeness*, introduced in the 1970s, toward the *L2 Motivational Self System* proposed by Dörnyei in the end of the first decade of the twenty-first century. (For an overview see e.g. Cook & Singleton, 2014; Ellis, 2008, 2015; Ortega, 2009; Williams & Burden, 1997). This chapter, however, focuses on the issues in motivational research that are relevant to the study.

In spite of the extensive research, L2 motivation has not yet been defined clearly. Doing that is challenging because the nature of motivation is not clear (Gass & Selinker, 2008, p. 426), and nothing about motivation is directly observable; its effects are susceptible to different interpretations (Cook & Singleton, 2014, p. 102). Moreover, “human behaviour is very complex and influenced by a great number of factors ranging from basic physical needs [...] to higher values and beliefs” (Dörnyei, 2001, p. 7). In spite of the obvious problems inherent in second language motivation research, recently there has been a consensus among scholars regarding the complexity of L2 motivation (e.g. Dörnyei, 2001; Ehrman et al., 2003; Ellis, 2008, 2015; Williams & Burden, 1997) and the understanding of motivation as a “desire to initiate L2 learning and the effort employed to sustain it” (Ortega, 2009, p. 168). Nowadays, motivation is conceived as a dynamic construct, with the emphasis being placed on temporal variation, context-related variation, and changes in learner behaviour.

The view of motivation as a dynamic construct appeared in the period around the turn of the millennium; since then motivation has not been perceived as a stable trait of a learner but as something that changes over time. Dörnyei and Ottó (1998) proposed a process model of language learning motivation. The model was based on the *Action Control Theory* of Heckhausen and Kuhl (1991, in Dörnyei & Ottó, 1998, p. 46), who believed that there were distinct temporally ordered phases in the motivational process, namely *predecisional* and *postdecisional* phases, which were energized and directed by largely different motives.

Dörnyei and Ottó’s model (1998) is considered “the fullest attempt to represent the complex, dynamic nature of motivation” (Ellis, 2015, p. 51). In spite of its limitations, specifically its linear view of motivation, the model is very useful because of its link to the L2 classroom (Dörnyei, 2001). The *Process Model of L2 Motivation* (Dörnyei, 2001) thus provides “a basis for identifying specific strategies that teachers can employ to help motivate learners” (Ellis, 2015, p. 51). The model (Dörnyei, 2001) consists of preactional, actional and postactional phases (see Figure 11.1).

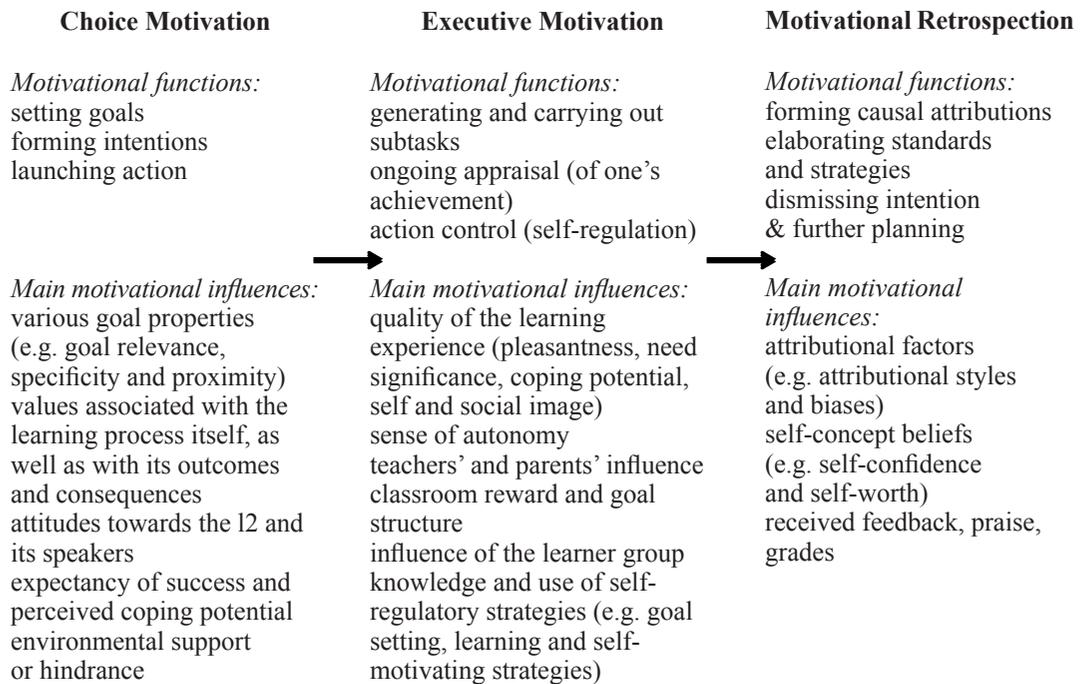


Figure 1: A process model of learning motivation in the L2 classroom (Dörnyei, 2001, p. 22)

In the preactional phase motivation needs to be generated to select the goal or task to pursue. Therefore, motivation in this phase is referred to as choice motivation. In the actional phase, “motivation needs to be actively maintained and protected while the action lasts” (Dörnyei, 2001, p. 21). Dörnyei labels this motivation executive and maintains that it is particularly relevant to classroom instruction, where learners are exposed to distracting influences. In the postactional phase learners go through retrospective evaluation after completing the action. That is why the phase is called motivational retrospection (Dörnyei, 2001). The distinctive nature of the three phases implies that the motives which influence learners’ behaviour are different in each of them. Figure 11.1 provides a list of the main motivational influences classified according to the phases. Viewing motivation as a process was a novel approach

which was in agreement with the findings of some other researchers. By the same token, applying a social constructivist approach Williams and Burden (1997) proposed a model of motivation involving the following three stages which affect each other: reasons for doing something → deciding to do something → sustaining the effort, or persisting. The first two stages are related to initiating motivation, the last one to sustaining motivation. Distinguishing the two types of motivation implies that decisions to act and the amount of effort invested to achieve the set goals are driven by different motives. Though Williams and Burden's model (1997) differed from that of Dörnyei and Ottó (1998) in other aspects, the basic assumption is the same, i.e. the existence of discrete temporally ordered phases in the motivational process energised by different motives.

### **3 Motivational transformation episodes**

Dörnyei's process model of language learning motivation was found relevant to the present study of foreign language learners' motivation for several reasons: first, because of the temporal dimension of motivation that the model acknowledged; second, for classifying motivational influences with respect to individual phases of the motivational process, and third, for its focus on motivation in the classroom setting. The model, together with Shoaib and Dörnyei's (2004) study, thus provided a theoretical framework for the analysis of learners' biographies. In their inspiring study Shoaib and Dörnyei (2004) analysed individual learner biographies and identified what they termed motivational transformation episodes, i.e. recurring patterns of motivational change present across varied learning situations which result in profound restructuring of the individuals' disposition:

- Maturation and gradually increasing interest.
- Stand-still period.
- Moving into a new life phase.
- Internalising external goals and "imported visions".
- Relationship with a "significant other".
- Time spent in the host environment.

(Shoaib & Dörnyei, 2004, p. 31)

The first type of episode is one of maturation and gradually increasing interest. It refers to those situations in people's lives when they gradually realised the importance of learning English, which they were not initially aware of for many reasons. The respondents in Shoaib and Dörnyei's study gave the following reasons: they were too young, learning English was not their choice, it was just another school subject, etc. (2004, p. 31). The stand-still period means that the people suspended language

learning for a period of time but started again after the conditions changed. The learning motivation remained at the same level but the process was interrupted. Moving into a new life phase is related to a change in learning goals. Regarding this type of transformation episode, Shoaib and Dörnyei also noticed what they called “sharpening of the focus in language learning orientations” (2004, pp. 32-33) – in the course of their lives the language learning goals of the respondents tended to become more specific. A very interesting episode is that of internalising external goals and *imported visions* because there were learners who were originally forced to learn a foreign language but who later became intrinsically motivated. This phenomenon of internalising initially extrinsic motives is called “identified regulation” (Ortega, 2009, p. 176). Another episode acknowledges the role of a significant other, namely a partner who is a native speaker. The last episode type is called time spent in the host environment. Shoaib and Dörnyei provided evidence that experiencing a stay in a host environment can have a significant impact, both in a positive and in a negative sense (2004, p. 34).

## **4 The study**

### **4.1 Methodology**

Building on the principles of mixed methods research, the presented study is the qualitative part of an explanatory study of a sequential design (Creswell & Plano Clark, 2011, pp. 217-218). In this type of study the quantitative data is elicited first, and then a research sample for the qualitative phase is selected (i.e. type QUAN → QUAL). Finally, the outcomes of both studies are interrelated. In the context of this study questionnaires and narrative interviews were used to obtain the data for the two analyses. The chapter focuses almost exclusively on the qualitative study, with the exception of the respondents’ profiles, which are grounded in the quantitative data.

In his review of (auto-)biographical research Benson stated that qualitative research in SLA was gaining ground, though it still remained a minority interest (2004, p. 17). A decade later, the situation has not changed significantly. The present study responds to the call for “greater emphasis on qualitative research directed at the holistic description of second language learning experiences and for greater emphasis on the social, affective and conceptual dimensions of the learning process” (Benson, 2004, p. 12). Its aim is to identify temporal fluctuations of motivation in learners’ stories of their experience, i.e. motivational transformation episodes with respective influences. This is in accord with the suggestion of Ehrman et al. (2003) who propose that “motivation is a very complex-compound factor in learning and must be considered in the light of non-affective variables” (p. 322). Among the

variables they list, the authors also include language learning history and beliefs (Ehrman et al., 2003, p. 323). The study is based on recollective data elicited by means of narrative interviews.

Narrative interviews are widely discussed in the literature on social and educational research. There is a lack of agreement among scholars regarding, among other issues, the number of interviews with each participant and the formulation of questions. Concerning the number of interviews, if one interview is to be used, then ninety minutes is the optimum length for a qualitative research interview, as several authors suggest (Elliot, 2012, p. 32). However, Seidman (1998, in Elliot, 2005, p. 32) argues that the most appropriate procedure involves three interviews with different goals and content. In spite of the recommendation, researchers use various procedures; frequently, two interviews are organised with each respondent. Rosenthal (1995, in Hendl, 2012, p. 177) proposes that researchers should first elicit a narrative covering the respondents' whole life story, and only then should they focus on the selected phase of their lives. Contrary to that, to avoid obtaining extensive amounts of data of low relevance Hollway and Jefferson (2000, in Elliott, 2005) argue that "the best questions for narrative interviews invite the interviewee to talk about *specific times and situations* [my emphasis], rather than asking about the respondent's life over a long period of time" (p. 30). This suggestion was followed when formulating the questions.

The interview started by referring to my previous encounter with the respondents in the earlier part of the research project. Then, questions focused on current issues were posed first to prepare the informants to think and talk about the past. The questions used to elicit the main narrative were directed to specific times, i.e. pre-school, elementary school, secondary school, and post-secondary periods, and to specific situations, i.e. learning English in various contexts. Moreover, items of various types for the questioning phase were also prepared. In the pilot phase it appeared to be the most suitable procedure to combine the main narrative phase with the questioning phase within one interview. To maintain the momentum of the interviews I asked types of questions suggested for the questioning phase, especially specific questions about a situation already mentioned by the respondent or questions aimed at clarifying what the person had said before (Hendl, 2012, p. 177). Furthermore, I also provided carry-on and reinforcement feedback and encouraged elaboration through the use of various probes (Dörnyei, 2007, pp. 142-143).

The interviews were conducted in June and July 2014; all of them were audio recorded. The researcher decided to interview the students in the Czech language to minimise the potential negative impact of the informants' ability to produce a narrative in English. Subsequently, the recordings were transcribed word by word using the technique of commented transcription (Hendl, 2012, p. 208). The comments that were inserted indicated emphasis and laughter occurring during the interview and also provided extra-linguistic information regarding the use of pauses and time-

gaining devices. Otherwise, the transcriptions were not edited, in order to preserve authenticity. The selected biographies were subsequently analysed using the Atlas.ti7 software (1993 – 2015). With reference to Mishler's framework, the analysis was focused on the content of the interviews; attention was paid both to its descriptive and evaluative function (1996, p. 38). The researcher analysed the three respondents' life stories of learning English and identified motivational influences, referring to the process model of language learning motivation (Dörnyei, 2001), thus applying the model at the macro level. After the motives had been detected, patterns of motivational change were further explored using the framework of motivational transformation episodes described above. However, the analysis was not restricted by the framework to searching for patterns not identified by Shoaib and Dörnyei (2004).

## **4.2 Research sample**

The present study is a part of a larger project investigating aspects of the second language acquisition of Czech learners at the beginning of tertiary education. The research sample of the quantitative study included all those Czech students in English language teacher education programmes at the Faculty of Education in České Budějovice, the Faculty of Education in Olomouc, and the Faculty of Arts and Philosophy in Pardubice who started their first year in autumn 2013 and who met the criteria for being included in the sample (n=112). For the purpose of the qualitative analysis nine students were interviewed and out of them three respondents with different profiles were selected for this study. The profiles are based on the questionnaire data collected in October and November 2013. They provide basic information regarding: the students' age when they were first exposed to English; the location of the schools they attended; the learning contexts; attitudes to the English language and to English as a school subject; their experience with native speaker teachers, and stays in English-speaking countries. Regarding the learning contexts, the distinction between formal, informal, and non-formal contexts of learning (Kotásek, 2002) is used in the study. A formal context refers to compulsory education; an informal context denotes learning in courses organised by various other institutions, including one-to-one tuition. Non-formal learning is non-institutionalised and is a part of everyday life. In order to preserve the privacy of the selected students, fictitious names – Alice, Cindy and George – are used in the chapter to refer to their life stories. The extracts from the interviews quoted in this study were translated into English by the researcher.

## **4.3 Findings**

### **4.3.1 Alice**

#### **4.3.1.1 Profile**

Alice was ten years old when she started to learn English as a compulsory subject in grade four of her elementary school (In the Czech Republic elementary schools provide primary and lower-secondary education). The school was in the village where she lived with her family. At the age of fifteen Alice started to commute to a small town nearby to attend a secondary grammar school, i.e. to start upper-secondary education. Her experience with learning English up to the age of eighteen was limited to formal learning, with the exception of a short language school course taken before the school-leaving examination. Her attitudes to both the English language and English as a school subject were prevailingly positive, although her attitudes to the school subject turned rather negative during her first years at secondary school. Throughout her school-based English language education she experienced neither a native speaker teacher nor a stay in an English-speaking country. While her contact with English outside the school was limited to listening to songs while at elementary school, it intensified at secondary school because of increased exposure – browsing the internet, listening to songs, and watching films and TV series were Alice's regular free time activities at that time.

Alice's biography was chosen for analysis because her learning of English was bound solely to the formal educational context and outside the school English did not play a significant role in her life. The analysis uncovered two motivational transformation episodes.

#### **4.3.1.2 Outcomes of the qualitative analysis**

The first episode was quite a long period of maturation and increasing interest; it lasted from the compulsory start of English language education in grade four of elementary school till the end of secondary school. Because it was compulsory for Alice to start learning English, choice motivation influences, i.e. those related to goal setting, forming intentions and launching action, were not available for her to draw on in the preactional phase of the motivational process.

During the actional phase both positive and negative influences were traced. At the beginning English was just another school subject, which Alice perceived in a neutral way; compared to maths and physics it was relatively pleasant, and compared to German it was easier. While at elementary school, learning English was something she was able to cope with effortlessly. She used to get the best marks, i.e. ones on

the scale from one to five: “*During the elementary school ... my best friend and I were stars, we had only ones*” (1:59). She did not believe she was that good: “*I had the best marks but I think they did not match the quality of my performance*” (1:19) since she did not have an opportunity to compare her level with learners from other schools: “*... I was considered someone who is good at it [English], ... but we did not participate in any competitions ... never ... there was no chance to show off with your English*” (1:62). Elementary school English lessons were viewed rather negatively, because for her the lessons were too liberal and chaotic, homework was ignored, and some tasks were perceived as irrelevant: “*... if young children are expected to talk for half an hour about what they did at the weekend, it was not good ...*” (1:64). Overall, the experience with learning English towards the end of elementary school was positive, though the mosaic of influences consisted of positive, neutral, and negative ones. Alice confirmed this by saying that “*... I knew I would definitely continue studying English [at secondary school]*” (1:4) even if she had not had to do so.

After passing a placement test on entering secondary school, Alice was placed in the average group, which was taught by a teacher with a liberal educational style and an interest in music. From her perspective the teacher focused on songs and translation but did not pay attention to other aspects of English: “*What I missed was practising vocabulary ... he could have placed more emphasis on vocabulary learning*” (1:18) and used speaking tasks which lacked variety “*... the only type of lesson was: talk about what you did at the weekend, just the same thing all the time*” (1:21). Although he was perceived as an educated and experienced person, in her opinion, he did not know how to teach well “*we were not sure how to say basic things but we had to learn complex phrases ... his English was good ... he had all the knowledge but failed to teach it*” (1:35). Unfortunately, the students stayed in the same groups for four years, which Alice did not like: “*... we came with what we learnt at the elementary school and we couldn't influence it; we got into the bad, average group, where we stayed and there was no chance to change groups or to progress ... where we started, there we finished ... and we, the students, couldn't influence it*” (1:36). She regretted not having a chance to get into the best group. She would have preferred to have a more demanding teacher, the one teaching the top group of students, who were enthusiastic about his lessons. She was convinced that pressure from the teacher was a necessary motivational influence: “*If there is no pressure from the teacher, you do nothing*” (1:34). In the first two years of secondary school her marks were average, which reflected little invested effort: “*I was average ... I was happy with threes ... There was a turning point after elementary school ... we left with best marks and we didn't have to do much, but at the secondary grammar school you had to do something and I was rather sloppy ...*” (1:60). While elementary school ended on a positive note, in the first half of her secondary school career negative influences, whether the quality of the learning experience or the teacher,

prevailed. The formerly strong influences, the coping potential and marks, lost their power to maintain the qualitative balance.

The process of growing interest in English intensified considerably with the introduction of an optional English seminar with a different teacher in the third year of secondary school. The number of lessons increased and she started to find the learning experience more challenging, since learning activities were oriented towards a specific goal, i.e. passing the school-leaving (“maturita”) exam. She started to be more successful in learning English, which resulted in increased enjoyment, as well as commitment. *“If you are successful in something, you start to enjoy it and you start to be interested in it”* (1:48). The student began to acknowledge the quality of the learning experience, which, up to that moment, had exerted rather a hindering influence for the variety of reasons mentioned above and which took a turn for the better towards the end of secondary school.

To summarise, during the episode of maturation and increasing interest the executive motivation of the student fluctuated in response to the changing contexts. In the long run the main motivational influences while at elementary school appeared to be the coping potential and ongoing positive appraisal, while during the second half of her secondary school career it was the quality of the learning experience and the impact of the teacher.

The second motivational transformation episode identified in Alice’s history is type four, internalising external goals and *imported visions*, described in Shoaib and Dörnyei’s study (2004). Regarding choice motivation, values associated with learning and its outcomes became important for Alice. She found mastering English useful and personally relevant: *“English is useful nowadays ... you can use it in the future ... use it for travel”* (1:50); therefore, she decided to study English at university. She had also considered her potential and concluded that languages, i.e. English, were the right choice: *“I started to make up my mind ... I realised that I will not be a physicist, but I’ll study the language [English]”*(1:30). Although Alice claimed to have selected her study programme autonomously, her parents obviously influenced the decision. Especially her father’s opinion regarding the usefulness of English nowadays seems to have been influential since it was mentioned several times during the interviews: *“My father uses English at work ... he supported me ... he kept saying that English is important and in demand”*(1:55). The teachers also contributed to the decision by placing an emphasis on passing the “maturita” exam and directing the students’ thinking towards their future: *“They [teachers] reminded us repeatedly of the maturita exam, and that we have to make our living somehow”* (1:29). They thus energised the students’ executive motivation. The motivational influences related to the second episode converged with those associated with episode one towards the end of secondary school. Apart from her increased interest in learning English, Alice also internalised the external goals and *imported visions* of her parents, specifically her father, and also of her teachers.

## 4.3.2 Cindy

### 4.3.2.1 Profile

Cindy also lived in a small town. Similarly to Alice, she had no experience of learning English prior to the compulsory beginning of English language education in grade four of elementary school. She was ten years old when she started to learn English as a foreign language. While at elementary school her attitudes to English, both to the language and the school subject, were reported to be negative, which was, however, challenged by the qualitative analysis. The situation changed after she enrolled in an eight-year programme at a secondary grammar school in the same town. Throughout her years of formal education in English, none of her teachers was a native speaker. She also experienced informal learning to a considerable extent since her learning experience at school ran parallel with learning in private lessons. Otherwise, the role of English in her everyday life was not very important up to the age of fifteen. Her contact with English was restricted to listening to music and browsing the internet on a weekly basis. Her exposure to English increased in the following years because of songs, films, videos, and TV series. She started to use internet social networks for communication in the target language; the only opportunity to communicate with native speakers in English face to face was available during summer camps, which she went to repeatedly. Moreover, she briefly visited Great Britain and the United States of America before finishing secondary school.

Cindy's story was selected for analysis because of the richness of the contexts in which she learnt English in the course of her life. The L2 motivation in her life was much more dynamic in comparison with Alice; seven motivational transition periods were identified.

### 4.3.2.2 Outcomes of the qualitative analysis

The first episode in her life was also maturation and increasing interest. Similarly to Alice, she also started to learn English as an obligatory subject in grade four of elementary school. It was not her decision; therefore, considering the choice motivation in this episode is not relevant and, also, no influences corresponding to this type of motivation were identified in Cindy's biography.

In terms of executive motivation, it was especially the quality of the learning experience, specifically pleasantness and coping potential: *"I remember that we spent most of the time in the middle of the carpet playing games ..."* (3:18), which fuelled her interest in learning English. Cindy perceived the experience as very pleasant. She enjoyed the lessons because of their game-like nature and because the lessons were within her coping potential. Her comment that *"the teacher did not impose any*

*requirements*” (3:2) implies that she was able to meet the given requirements without any problems. Furthermore, obtaining excellent marks was also a powerful influence together with the influence of the teacher, whose lessons she liked. After one school year Cindy was a pupil who enjoyed learning English. The process of gradual maturation and gaining interest was interrupted by two episodes.

The second motivational transformation episode is about losing motivation as a consequence of the harmful effects of the teacher or the learning experience. Interestingly, this type of episode was not defined by Shoaib and Dörnyei (2004), although fifteen respondents out of twenty-five in their study reported a harmful role of teachers and methodology (2004, p. 30). Probably, the experience did not lead to profound restructuring of the respondents’ disposition. This was also the case with Alice who went through a period of motivational decline, though without deep changes in her disposition. Cindy, however, did undergo such fluctuations. The loss of motivation was a negative consequence of a sudden change of teacher in grade five when the classes were reorganised. The energy of former motivational influences was lost. The quality of the learning experience altered considerably; the atmosphere in class was not very friendly, in contrast to the classes of the previous teacher. Neither did the new teacher respect the learners’ prior learning experience and knowledge: “... *a different teacher, different learning materials and ... what the classmates learnt in the previous year I have never seen before ... she expected us to know it as well and she did not respect us*” (3:5). It was something Cindy was not able to cope with. As a result she obtained average grades in her mid-term report in grade five, which affected her self-image. Regarding her motivation for learning English, at the end of the episode she was the exact opposite of the person she had been half a year before.

The third motivational transformation episode in Cindy’s life, lasting for about half a year, was a reaction to the previous demotivation phase. It may be called regaining interest. Similarly to the second episode, it was not identified in Shoaib and Dörnyei’s study (2004). The characteristic feature of the pattern is that in reaction to the loss of motivation new motivational influences are found to replace those which diminished. Its potential is to counterbalance the loss of motivation, if a learner has choice motivation and decides to act. Goal properties and environmental support were the influences that motivated Cindy’s action. Building on the support of her mother, who organised private lessons, she found the goal of private tuition, i.e. to improve her average grades, personally relevant: “*He [the private teacher] mainly tried to prepare me for the tests to achieve better marks ... to improve*” (3:26). Furthermore, the role of private lessons in improving her English was significant: “*Without the extra lessons ... I wouldn’t have been able to make progress*” (3:22). Therefore, she maintained action; more specifically, she invested effort into learning in one-to-one lessons which helped her to develop her English. She attributed the results to her effort: “*I think I tried hard to get to their level*”

(3:28). Cindy regained interest in learning English and gradual maturation was able to continue.

However, the fourth episode is again about losing motivation as a consequence of the harmful effects of the teacher or the learning experience. The quality of the learning experience changed with a new teacher in the first year of secondary school: *"I didn't look forward to the subject [English] at all, I was stressed ... when getting things for the upcoming day I thought 'Oh, dear, English again' ... it was because of the teacher"* (3:53). Lessons of English turned into a traumatic experience for several reasons: the way the teacher taught the class did not help her learn since she needed more time; in her opinion the teacher prioritised certain learners or groups of learners (girls over boys), used labelling, and ignored a handicapped classmate: *"She told him: I don't know what to do with you. Just listen"* (3:80). On top of that, before the "maturita" exam, there was a critical incident – the teacher ridiculed Cindy in front of the class and made critical evaluation of her appearance and non-verbal signals rather than her spoken performance: *"During the revision of the maturita topics I was asked several questions in front of the blackboard. I answered and I was quite happy with my performance and that day there were two classes joined together, so about thirty people, and suddenly she started to mime me and told me I looked terrible, my face was terrible ..."* (3:12). Furthermore, she was advised not to sit for the exam unless she changed her appearance and body language. As a result of this unprofessional behaviour of the teacher her self-image was destroyed and her self-confidence before the "maturita" exam lost. It was again her mother who provided the necessary support and encouragement needed for passing the exam.

Episode number five may be termed regaining interest, similarly to episode three. The main motivational influences relate to the quality of the learning experience again, though outside the school. While at secondary school she participated in international English camps, which were based on cooperation with an American church. This allowed young people to meet, learn, and work together. Mostly, the camps were held in the Czech Republic but Cindy also participated in a camp which took place in the USA. Thanks to her participation in the camps, which she enjoyed, her ability to speak English developed. She also became more self-confident and willing to learn: *"there I realized what English is ... at school I didn't like it at all ... [in the camp] I tried to learn something"* (3:47). She appreciated real-life opportunities to communicate in English, which is something she missed in classes since they were prevalingly targeted at completing exercises in textbooks. The quality of the learning experience in the camps influenced her decision to study English considerably: *"If it hadn't been for the English camps, I wouldn't have decided to study English"* (3:46).

The last two motivational transformation episodes, numbers six and seven, are related to moving into a new life phase, i.e. type four in the study by Shoib and Dörnyei (2004). The sixth episode represents the decision to take a one-year intensive

course in English after not being accepted to university. Improving her level of English was perceived as being worth investing in. Her resolution was driven by choice motivation – the goal and outcomes of learning were perceived as relevant and valued. *“I wanted to learn what nobody taught me at school, I wanted to be able to say one day that I really know English well”* (3:65). During the course of action, her self-confidence increased thanks to passing an international exam in English; furthermore, she gained a more positive self-image thanks to the comparison of her performance with that of her peers. The last episode is of the same kind as the previous one. The new life phase was her becoming a student of English at the university. The preactional phase was quite long since a decision was forming gradually in her mind. On completion of the one-year intensive course Cindy was about to choose the field of her future university studies. She was considering informatics or English and finally decided on the former. She also signed up for an English course. Unfortunately, its level was too low for her and she did not make any progress. After one term, she started to miss studying English. Coupled with a certain dissatisfaction with her selected programme, it was the reason for changing her field of study: *“During the year I realised that I missed English ... that I liked the language and that I should study at a different faculty”* (3:93). The main motivational influences included goal relevance, and, specifically, the value of learning and its outcomes.

### **4.3.3 George**

#### **4.3.3.1 Profile**

George, living with his family in a small town, started to learn English as a pre-schooler on a language school course. At the moment he began to learn English at the primary stage of elementary school George had four years of experience of foreign language instruction. His attitudes to English, both to the language and the school subject, were positive throughout his life. After primary school he studied at a local secondary grammar school with an eight-year programme. He experienced five English teachers in those eight years but no native speaker teacher. Even after the beginning of formal English education in grade four of elementary school he continued learning informally on various courses. Equally, non-formal learning was also important in his life. Even before the age of fifteen he spent time every day listening to songs, playing computer games, and browsing the internet. Less frequently, once a week, he used internet social networks to communicate in English. Before reaching the age of eighteen he also visited Great Britain and the United States of America.

The last biography analysed is that of George, which differs from the previous stories in the age at which he started learning English and in the level of learner autonomy.

#### 4.3.3.2 Outcomes of the qualitative analysis

The first motivational transformation episode identified in George's story of learning English was maturation and increasing interest. This episode started in the pre-school period, when his parents made the decision to register him for a language course. Since they were very active themselves they enrolled their son in numerous courses and afternoon clubs. While he refused to attend some of them, learning English was an activity he initially enjoyed because of its game-like nature. The pleasantness of the learning experience and his parents' support were the main motivational influences in the period preceding the compulsory start of learning English at school. From when he was a pre-schooler learning English in an informal context ran in parallel with second language education at school. George liked attending language school courses, but, on the other hand, he had negative attitudes to school. It was because George was an introverted child who felt isolated in the class. He was also influenced by his parents' strong negative opinion regarding the alternative vision of education implemented by the school. In spite of that, his attitudes to English remained positive mainly because of the previous knowledge he had gained on the courses he had attended before the compulsory beginning of English instruction. "*This [learning English] was something I was good at ... I was better than the others... it motivated me*" (7:5). Thus the quality of the learning experience, specifically the coping potential and self-image, energised his executive motivation: consequently, his interest in learning English increased.

The characteristics of the second episode did not match any of the types described by Shoaib and Dörnyei (2004). Drawing on the analysis of motivational influences, the episode has been termed an autonomous action which is intensive and focused. At the age of fourteen George came across a method of learning languages which advertised promising outcomes, conditional only on a proper implementation of the method. He decided to apply it and, consequently, organised his everyday learning in accordance with the guidelines and recommendations. He was self-disciplined and was able to sustain his efforts for about a year. Regarding choice motivation, the motivational influences were multiple. The first was goal relevance, specificity, and proximity, since improving English through the method was a relevant goal which was specific enough and achievable. Second, the learning outcome was perceived as valuable: "*I could see that I would at least learn this [English] since I'm not good at anything else*" (7:16). Third, expectancy value and perceived coping potential were obviously strong influences: "*This [applying the method] was something that ... that one could find the time for... that I believed I could manage ... I could succeed*" (7:19). In the course of the action the pleasantness of the learning experience energised executive motivation, as well as self-assessment based on both individual and social progress norms: "*... you can compare yourself with your classmates ... and I could see that I started to make progress*" (7:13). Furthermore, George needed

more autonomy in learning, which he lacked in classes at school: “[*When learning*] *I need my own pace. This [applying the method] was something I could control myself. Not just ‘do this, do this, do this, and if you don’t do it, there will be sanctions’*” (7:56). Being an autonomous learner, he was aware of self-regulatory strategies and was able to use them. He set the goal he would pursue; he was able to allocate several periods of time to learning every day. Since he believed in success he managed to maintain this action for a year. In retrospect he realised the benefits of the learning outcomes, which he attributed to the effort he had invested. Further on in his life he was quite critical of the teachers who taught him at school but, at the same time, he was able to meet the study requirements rather effortlessly; outside school he enjoyed autonomous learning, drawing on a variety of motivational influences.

## 5 Conclusion

The study of the learners’ histories uncovered individual patterns of motivational dynamicity; there were differences among the respondents in terms of the number of motivational transformation episodes and their nature. If compared to the study of Shoaib and Dörnyei (2004), the results converge to a considerable extent. Regarding the types of episodes, their first type, maturation and gradually increasing interest, was identified in all three biographies, albeit with varying length. Moving into a new life phase, the third type, was recognised in Cindy’s story only because her path to the point of entry to studying English at university was not straightforward and took much longer. The fourth episode, called internalising external goals and *imported visions*, was identified once, although, in the lives of the other two learners external goals also undoubtedly played a role. In a way there is an overlap with type one, since all the learners who do not start to learn English of their free will are expected to internalise external goals. However, the findings diverge in that type two, the stand-still period, was not observed in the learners’ stories that were analysed. Similarly, episodes of types five and six, i.e. a relationship with a “significant other” or time spent in the host environment, did not appear in the biographies. Furthermore, three other motivational transformation episodes were identified in the present study. The first was an episode labelled demotivation, two occurrences of which were observed in Cindy’s biography. Similarly, the second one, termed regaining interest, appeared in her story twice, both times as a reaction to the demotivation phase. The third new episode was called an autonomous, intensive, and focused action, which was observed in one of the biographies. However, further analysis of a larger research sample would be needed to confirm the significance of these episodes.

The study also confirmed the added value of biographical research, i.e. learning about the meaning which people attach to events in their life. Alice’s story implies how significant the effects of external differentiation may be. Cindy’s story shows

that the impact of the teacher's behaviour on the learner may be immense. George's story provides evidence of the power of learner autonomy. By the same token, potentially strong motivational influences, native speakers and visits to English-speaking countries, may fail to be influential.

The results clearly indicate that the quality of the learning experience and the teacher are major motivational influences, especially in those situations where choice motivation does not apply. Moreover, the findings suggest that motivational influences across contexts are complementary; under favourable circumstances influences that are missing from the formal setting may be replaced by those originating in informal and/or non-formal environments.

What are the implications of these findings for foreign language teaching and learning? First of all, not surprisingly, the study provides evidence that motivation is a powerful affective factor and that motivational patterns are of an individual nature. These patterns are strongly influenced by two factors, the quality of the learning experience and the teacher, which are very much interrelated. The two factors are essential, especially in the foreign language context, for example in the Czech Republic, where children either start to learn English as pre-schoolers because their parents want them to do so or they start at elementary school because it is a compulsory subject. They do not have a choice. Therefore, it is the teacher's responsibility to help them internalise initially external goals – those of their parents or curricular objectives. In the early stages of foreign language learning the pleasantness of the learning experience appears to be crucial, which confirms the priority of affective aims in this phase of learning. In order to help the learners mature and increase their interest in learning English teachers have to respect learners' needs. Considering the three biographies, whenever the individual learner's needs were not respected, there was a more or less profound motivational decline. Conversely, opportunities for being autonomous in contexts outside the school brought about an increase in motivation. Thus, building on the results of the study, the right strategy towards a motivated language learner seems to be making English lessons more learner-centred. This brings us back to the teacher who is able to teach English with respect to the differences between individual learners.

The stories of the three learners ended well in spite of several periods of prevailing negative influences. Who would have guessed that Cindy, a learner with average marks in grade five, would one day decide to study to become an English teacher? This story may encourage teachers to think of their learners in the long-term perspective, beyond the "here and now" reality.

## **Acknowledgements**

This work was supported by a Czech Science Foundation Grant (“Aspects of English Language Acquisition of Czech Students at the Onset of Teacher Education” GA ČR 13-25982S).

## Section Four

# CLASSROOM INTERACTION AND L2 ACQUISITION

The authors of the chapters approach classroom interaction and L2 acquisition from a different standpoint.

In his chapter František Tůma argues that dialogist approaches in general and conversation analysis (CA) in particular have not yet been commonly used in Czech educational research on classroom interaction. However, CA conducted on video recordings of classroom teaching makes it possible to consider the local and dynamic nature of interaction. The aim of his theoretical-methodological chapter is twofold. First, it introduces the dialogist framework and related views of (language) learning, as well as CA as a research method. Second, it presents the research design of a research project entitled “Classroom interaction in English language teaching in higher education”. He pays special attention to the methodological aspects of an analysis of classroom interaction in the context of the language classroom.

The chapter by Lucie Betáková deals with classroom discourse analysis. The main aim of her study is to show that within classroom interaction teachers’ verbal activities both within the initiation and feedback moves support the language acquisition of the learners. She concentrates mainly on the initiation move, in which elicitation by the teacher is very common. That is why she has studied the types of questions teachers ask and how the types of questions influence learner language in the response move. It is generally believed that open questions are of higher quality as they give the learners a chance to express themselves, and in addition they promote thinking and elicit more language. Betáková shows that asking referential, i.e. real questions, in comparison with display questions, is very important to develop classroom interaction and thus promote language acquisition. She proves that open questions, though, are not the only key to success. Even teachers who do ask open questions may elicit very little language from the learners if other aspects of classroom interaction, i.e. psychological aspects, such as a positive classroom atmosphere or thinking time for the learners, are neglected.

Petr Dvořák adds a social dimension to the study of foreign language classroom interaction which is highly specific in that the target language is both the tool and

the ultimate end-goal of instruction. He views the processes applied during those interactions as closely linked to social functions and social milieux, in this case within the classroom environment. Social skills are an important component of the English language teacher's professional competence as they are instrumental in effectively and creatively realising the processes of foreign language acquisition. The author presents research findings focused on analysing the role of social skills within foreign language interactions and discourse, as implemented by secondary school teachers of English in the Czech Republic. The researcher applied a mixed-methods research design using both qualitative and quantitative data acquired with the help of a questionnaire and structured observations. First, the teacher-student interactions were analysed on the basis of two dimensions reflecting social skills that form part of the teacher's educational style: the *educational management* and the *educational relationship* between the teachers and the students. In the second part of his research Dvořák concentrates on verbal communication and discourse within selected segments of interactions. The results show that individual teachers' social skills and educational styles have a direct impact on the target language interaction, communication, and discourse. Further, they influence the effectiveness of the learning and language acquisition processes.

## Chapter 12

# USING CONVERSATION ANALYSIS IN CLASSROOM INTERACTION RESEARCH

*František Tůma*

## 1 Introduction

The shared activity and mutual influence between teachers and learners constitute a fundamental part of the teaching and learning processes in the classroom. It is therefore vital that educational research describe and understand the nature of classroom interaction.

Due to the grounding of this research in dialogism (and relatedly in ethnomethodology and conversation analysis) we use the term classroom interaction, which we understand in line with the concept of dialogue, i.e. “any dyadic or polyadic interaction between individuals who are mutually co-present to each other and who interact through language (or some other symbolic means)” (Linell, 1998, p. 9). Classroom interaction therefore denotes the participants’ mutual influencing and reacting realized primarily by means of spoken language in the classroom.

The phenomenon of classroom interaction has been studied from a number of perspectives, including quantitative observation methods, ethnographic research, linguistic approaches, sociocultural theory, and ethnomethodological conversation analysis (Mercer, 2010; Mitchell, 2009; Rampton, Roberts, Leung, & Harris, 2002). In this chapter we adopt dialogism as an epistemological framework, within which sociocultural theory can cast light on the nature of mediated action and learning. Ethnomethodological conversation analysis presents a tool for analyzing classroom interaction.

As will be shown in part 4, contemporary Czech educational research on classroom interaction does not reflect the dialogist framework to a greater extent. The aim of this chapter is therefore to introduce the theoretical background and methodological aspects of an analysis of classroom interaction which reflects the dialogist framework. At the same time the theoretical and methodological considerations underlie the research project “Classroom interaction in English language teaching in higher education”, whose research design is presented in part 5.

## 2 Theoretical background

The framework of dialogism has been introduced in more detail elsewhere (e.g. Linell, 1998, 2009; Marková, 1982; Rommetveit, 1988; Tůma, 2014a, pp. 878-883). The crucial point is that dialogism views interaction as a collective process in which the participants mutually influence each other and in which any utterance “makes response to something and is calculated to be responded to in turn” (Vološinov, 1973, p. 72). Therefore interaction is not understood as information transfer from the sender to the receiver, which is typical of monologism. It follows that dialogist approaches to interaction take into consideration context and the local and dynamic nature of interaction. Conversation analysis (CA) can be seen as a research method which reflects these assumptions, despite some reservations regarding the direct compatibility of dialogism and CA (e.g. O’Connell & Kowal, 2003, pp. 201-203).

As far as learning through interaction is concerned, sociocultural theory (SCT), which builds on the works of Vygotsky and his co-workers, can explain how human mental processes function with regard to their cultural, historical and institutional settings (Wertsch, 1991, p. 6). In SCT, mediated (inter)action that involves the use of language is seen as central and is related to the functioning and development of higher mental functions, of which speech is of our interest (Vygotsky, 1978, pp. 24-26, 52-57; Wertsch, 1991, pp. 8-13, 28-43). According to the general genetic law of cultural development, every function appears on two planes: first on the social (or interpsychological) plane, and then it is reconstructed onto the intrapsychological plane (Vygotsky, 1978, pp. 56-57, 1981). This complex reconstruction can be called internalization and is possible only if the function lies within the zone of proximal development of the individual (Vygotsky, 1978, pp. 55-57, 84-91; Vygotsky, 1981; Wertsch, 1991, pp. 19-28), which presupposes intersubjectivity between the learner and the more knowledgeable other.

The concepts of mediation, zone of proximal development and internalization have been used in sociocultural second language acquisition (SLA) theory (e.g. Lantolf & Thorne, 2006). This way language learning is conceptualized in relation to interaction. There exists a body of research in which the conversation analysis methodology is combined with (mainly) sociocultural theory (e.g. Markee, 2007). More generally, the view of learning and interaction in unity, i.e. learning through interacting, rather than seeing learning and interaction as separate processes, is inherent to social SLA theories (Firth & Wagner, 2007) and is reflected in related sociological and linguistic schools of thought (for more details see Auer, 2014).

As far as the nature of language is concerned, dialogism generally views language as a form of social action (for other views of language see Cook, 2010). This view of language is also relevant to conversation analysis (Schegloff, Koshik, Jacoby, & Olsher, 2002, pp. 4-6), in which interaction realized by means of language is referred to as “talk-in-interaction” or “institutional talk” (Drew & Heritage, 1992).

### 3 Conversation analysis

CA aims at understanding social order through analysis of talk-in-interaction. CA builds on ethnomethodology (e.g. Auer, 2014, pp. 120-139; Goodwin & Heritage, 1990, pp. 284-287), which is compatible with dialogism.

The underlying assumption is that talk-in-interaction is important in social life in terms of both everyday concerns and society at large (ten Have, 2007, p. 10). Relatedly, CA deals with naturally occurring talk-in-interaction, which is the object of the analysis. Other defining positions of CA include an emic (i.e. participants') orientation (ten Have, 2007, pp. 34-35), the assumptions that "talk-in-interaction is systematically organized and deeply ordered" and that "the production of talk-in-interaction is methodic" (Hutchby & Wooffitt, 1998, p. 23; Goodwin & Heritage, 1990, pp. 287-288; Seedhouse, 2004, pp. 14-16). Context in CA is viewed in a dynamic way: it is presumed that each action is both context-shaped (i.e. it emerges from the context) and context-renewing (i.e. it becomes a part of the context), which is compatible with dialogism (Goodwin & Heritage, 1990, p. 289; Schegloff, 1992; Seedhouse, 2004, pp. 42-45).

It has been pointed out that the view of language adopted in this research is that of a form of action, on which CA adopts a complex view. Connections are explored between "the particularities in the details of human action and the generalities of shared organizational problems and resources" (ten Have, 2007, p. 10). Therefore, CA aims at describing the specific features of individual phenomena and at the same time it brings those specifics "under the umbrella of a generalized account of some sequential pattern or interactional device" (Hutchby & Wooffitt, 1998, p. 95), which is reflected in the analytic procedures that we outline below.

The participants' actions and sequences of actions comprise the phenomena of interest. More specifically, the core research areas in CA, as mentioned by Schegloff (1989, in ten Have, 2007, p. 120) and Auer (2014, pp. 135-138), comprise the organization of turn taking (Sacks, Schegloff, & Jefferson, 1974), sequence (Schegloff, 2007) and repair (Schegloff, Jefferson, & Sacks, 1977). In addition, CA has often been used in research on bilingual interaction (including foreign language classroom interaction) in order to cast light on the sequential organization of code-switching (Auer, 1995; for reviews see Martin-Jones, 1995, 2000).

CA can be characterized as an inductive and data-driven analysis. An important part of the analysis is transcription, which includes details of the talk-in-interaction under investigation. During the analysis, attention is paid both to the original recording, which is seen as a reproduction of the event, and the transcript, which represents the event through detailed transcription (Hutchby & Wooffitt, 1998, pp. 73-92).

The analytic procedure can be outlined in three steps. First, the analyst starts by identifying relevant phenomena and related passages, which is often referred to as "unmotivated looking" (ten Have, 2007, pp. 120-121; Seedhouse, 2004, p. 38). Second,

the analyst carries out a micro-analysis of the passage in which the phenomenon of interest occurred. As a result, a provisional analytic scheme is formulated. It should be pointed out that the participants', i.e. emic, perspective is adopted; mechanistic treatment of the data is generally rejected in CA (Mori & Zuengler, 2008, p. 17; Schegloff et al., 2002, p. 18). Third, other instances are to be found in the corpus in order to refine the provisional analytic scheme. This can be achieved by locating similar cases or deviant cases (ten Have, 2007, pp. 147, 153). Steps 2 and 3 require a systematic and thorough analysis (ten Have, 2007, p. 162).

The findings related to the organization of turn-taking, sequence and repair are generated by the refining of the provisional analytic scheme. Starting in a unique (micro)context, similarities are sought with other instances, thus working with a more abstract contextual level (Schegloff, 1987; Seedhouse, 2004, pp. 208-215). The robustness of the analysis can be achieved by comprehensive data treatment, which covers all the relevant data in the corpus, including deviant cases (ten Have, 2007, pp. 47-149).

#### **4 A review of literature**

In this part we will outline the state of the art of research on classroom interaction. First, we briefly review the research conducted abroad. Second, we outline the state of the art in the Czech Republic.

A number of reviews of literature exist regarding classroom research and classroom interaction (e.g. Ellis, 2008, pp. 775-835; Hall & Walsh, 2002; Nunan, 2005; Thoms, 2012). These reviews reflect the methodological and theoretical plurality of the contemporary classroom research on classroom interaction. For the purposes of this chapter we will focus mainly on the use of CA in classroom research.

Recent CA research on classroom interaction is reviewed by Mori and Zuengler (2008), according to whom the research has addressed the details of turn-taking, repair, various types of speech exchange systems, non-pedagogical talk in the classroom and peer tutoring. They also add that close single case analyses are conducted. As far as research on foreign language classroom interaction is concerned, Seedhouse (2004, pp. 86-88) reviews a number of relevant studies and introduces the outcomes of his analysis of classroom interaction in EFL classes based on a large corpus of data from different educational settings, focusing on the organization of turn-taking, repair and sequence (Seedhouse, 2004). As mentioned above, CA is also deployed in research on code-switching in the classroom (for reviews see Martin-Jones, 1995, 2000).

The above outline confirms that CA is a common research method in research on classroom interaction. The reviews (Martin-Jones, 2000; Mori & Zuengler, 2008; Seedhouse, 2004, pp. 86-88) also suggest that language teaching and learning

processes in secondary and higher education are commonly researched. However, this does not seem to be the case regarding the educational research on classroom interaction conducted in the Czech Republic. From three recent reviews available (Mareš, 2009; Tůma, 2014a, 2014b) it follows that relatively little attention has been paid to classroom interaction in foreign language teaching. Moreover, whereas lower-secondary educational contexts have been addressed by a considerable number of studies of classroom interaction in the last 20 years, studies from the area of higher education seem to be lacking. The reviews also suggest that dialogism has not been reflected in Czech research on classroom interaction to a greater extent yet; monologist research seems to predominate. From the articles reviewed by Tůma (2014a, 2014b) it also follows that conversation analysis has not been used in Czech educational research on classroom interaction at all. In the following part, a research project that attempts to fill this gap in Czech educational research is introduced. Special attention is paid to the methodological aspects.

## **5 Research design of the project**

The aim of the research project “Classroom interaction in English language teaching in higher education” is, by means of conversation analysis conducted on the empirical data (procedures of data collection are introduced below), to uncover some of the mechanisms underlying classroom interaction, namely in the areas of the organization of turn-taking, sequence and repair, including the sequential organization of code-switching.

The research will be conducted as a case study, since it allows for investigating “a contemporary phenomenon in depth and within its real-life context” where “the boundaries between phenomenon and context are not clearly evident”, and it can also rely on multiple sources of evidence (Yin, 2008, p. 18).

The case will be a group within an EFL course taught at the Language Centre at the Faculty of Education, Masaryk University. The course is designed for prospective teachers (typically bachelor second year students) with different specializations. Although the course content includes elements of English for academic purposes, its general aim for the students is to reach the B1 level (CEFR, 2001). The lecturer holds a Master’s degree in EFL teaching and has been teaching English for more than four years at the Centre, which is expected to guarantee relatively stable performance on the part of the lecturer. The learners in the course will have satisfied the requirements from the preceding part of the course. One group usually comprises between 15 and 20 students.

Since students are requested to do a foreign language course at the B1 or B2 level (CEFR, 2001) at the majority of faculties of education in the Czech Republic, we assume that at least some of the phenomena discovered and described within

the case study will be relevant to similar educational settings elsewhere (we discuss the generalizability of the results at the end of this part).

During the autumn semester of the academic year 2015/16 it is expected to record at least 10 lessons (i.e. 15 hours of recording) in the EFL course. In addition to video recordings, the data collected in the research will include the researcher's fieldnotes, collected classroom materials, teacher's lesson plans etc., to which we refer as "ethnographic data". From the perspective of CA, such ethnographic data can complement the analysis (Mori & Zuengler, 2008, pp. 23-24; Seedhouse, 2004, pp. 88-93), which we discuss in more detail below.

### **5.1 Specific aspects of the analysis with regard to English language teaching**

In CA it is generally held that the analyst pays attention primarily to the talk-in-interaction, other data sources (e.g. participants' or institutional background) being irrelevant unless the participants refer to these sources in their interaction (Markee, 2000, p. 22; Schegloff, 1992, p. 197; Seedhouse, 2004, pp. 15-16, 42-46, 91). However, additional data can be included to complement the recordings, depending on the focus of the analysis and on the nature of the activities under investigation (ten Have, 2007, p. 78). Since L2 classroom interaction is influenced by pedagogy, classroom interaction reflects the constant tension between the intended (i.e. planned) and actual pedagogy (Seedhouse, 2004, p. 85). Therefore Seedhouse (2004) maintains that ethnographic data can provide "a link between the micro and macro levels" (p. 92). In addition, Mori and Zuengler (2008) hold that "the participants share a physical space and use nonvocal resources along with their talk to accomplish the sense-making processes" (p. 23), it is therefore vital that the analysts take these resources into account. Thus the data sources in this project will also include data of ethnographic nature.

From the dual role of target language in foreign language teaching (it is both a medium of instruction and its goal) it follows that language teachers have to deal with the following mutually related concerns: (1) their plans (aims) and the content to be taught, (2) the learners' ideas, (3) the learners' accuracy (e.g. Brown, 2002; CEFR, 2001). As a result, a number of different types of activities occur in the classroom, which involve a number of different situations of language use (for a review see Ellis, 2008, pp. 788-790). Relatedly, Seedhouse (2004) distinguishes the following four contexts in his CA study: form-and-accuracy contexts, meaning-and-fluency contexts, task-oriented contexts, and procedural contexts, in each of which the interactional mechanisms differ.

It follows that the interaction recorded in this project will include a number of different contexts. In line with Seedhouse (2004, pp. 196-197), the following types of evidence will be used when determining the type of context for interaction: (1) the

teacher's statement of the intended pedagogical focus of an activity, (2) ethnographic data (lesson aims, classroom materials, etc.), (3) the participants' orientations to the context evident from the details of interaction. Clearly, the third source of data is the most relevant to CA, yet the other sources will also be taken into consideration during the analytical processes to guarantee the sensitivity of the analysis and its findings in relation to the actual interaction.

## **5.2 On generalizability**

It should be emphasized that CA does not aim at empirical quantitative or statistical generalizations (ten Have, 2007, p. 149). Concepts such as "representative samples", "relations between variables" or "patterns of conditions and consequences" are relevant neither to CA, nor to dialogism. Instead, CA aims at "theoretical grasp of interactions' underlying 'rules' and 'principles'" (ten Have, 2007, p. 150; see also Seedhouse, 2004, pp. 253-261), which can be achieved through analytic induction as characterized in part 3. The notion of analytic induction is not dissimilar to analytic generalization (Yin, 2008), which is inherent to case study research.

As far as the sample size is concerned, a minimum of 15 hours of video recording (i.e. 10 lessons) will be used as a basis for analysis. Relatedly, Seedhouse (2004) maintains that "recent classroom research into communication in both L1 and L2 classrooms has considered between 5 and 10 lessons a reasonable database from which to generalize and draw conclusions" (p. 87). Within the case, interaction will occur in different types of activities and contexts. It is therefore assumed that sufficient amount of data will be collected with regard to the purpose, nature and outcomes of the analysis.

The reliability and validity of the outcomes will be achieved by (1) collecting natural classroom data, (2) comprehensive data treatment (see part 3), (3) discussions and consultations with experts and presenting the outcomes at conferences, (4) including transcripts in publications (ten Have, 2007, pp. 65-90, 140-142; Seedhouse, 2004, pp. 253-261).

## **6 Concluding remarks**

As suggested in part 4, Czech educational research tends to focus on basic (especially lower-secondary) education rather than on higher education. The research seems to reveal little about the dynamic and situated nature of classroom interaction, since the majority of analyses are conducted in the monologist tradition (Tůma, 2014a, 2014b). These two areas appear to be a "gap" in Czech educational research which the research project presented in part 5 aims to fill.

The outcomes of this project, i.e. the CA findings related to the organization of sequence, turn-taking and repair (including the sequential nature of code-switching) underlying classroom interaction, will reflect not only the social actions, but also the participants' mutual understanding of the social actions. These organizations comprise a component of "the architecture of intersubjectivity" (Seedhouse, 2004, p. 237; Markee, 2000, pp. 84-96). Thus the outcomes can contribute to our understanding of socially shared cognition (Schegloff, 1991).

As far as the relevance of this project to the field of foreign language teaching/didactics is concerned, it should be pointed out that one of the features of the learner-teacher interaction is the fact that the teacher often not only provides a corrected version of the learners' utterances (Seedhouse, 2004, p. 238), but also uses various questioning techniques and clarification requests to maintain interaction. These and other interactional mechanisms will comprise the project outcomes. In the interaction between the teacher and the learner(s) we can observe (1) what the learners can say or do with the foreign language and (2) how the teacher models what can be said (or done), creating the potential for learner development. Relatedly, the actual state development and the zone of proximal development can be referred to, respectively (Vygotsky, 1978, pp. 55-57, 84-91), hence the relevance to sociocultural theory of second language acquisition (Lantolf & Thorne, 2006, pp. 263-290). It is then assumed that the students learn the foreign language through their participating in classroom interaction. Thus the project outcomes can also contribute to our understanding of how the interactional mechanisms facilitate intersubjectivity and foreign language learning, which is in line with the social view on SLA proposed by, for example, Firth and Wagner (2007).

In relation to the outcomes of CA, ten Have (2007) holds that "the core phenomena have been identified, but they can be explored further and there exists an enormous variety of settings, conditions, and languages for which the local organization of talk-in-interactions can be fruitfully studied" (p. 11), the same is acknowledged by Mori & Zuengler (2008, p. 16). Schegloff et al. (2002, p. 14) add that it is desirable to investigate how classroom talk differs from ordinary conversation and from talk in other educational contexts. In line with this, the outcomes of the presented research project will contribute to the body of CA research on classroom interaction by analyzing the underlying interactional mechanisms in EFL classes in Czech higher education and by comparing these with general findings from CA and with findings from other studies conducted abroad in EFL classes and higher education.

Theoretically and methodologically, the presented research project intends to introduce dialogism and conversation analysis in Czech educational research, which was the aim of this chapter.

## **Acknowledgements**

This work was supported by a Czech Science Foundation Grant (“Classroom interaction in English language teaching in higher education”, GA ČR 15-08857S).



## Chapter 13

# SUPPORTING LANGUAGE ACQUISITION THROUGH TEACHER QUESTIONING

*Lucie Betáková*

## 1 Introduction

Discourse analysis has a relatively long tradition. Various authors since 1960's have concentrated on discourse types which are easily recognizable. For example Gumperz (1977) notes that:

Members of all societies recognize certain communication routines which they view as distinct, separate from other types of discourse, characterized by special rules of speech and nonverbal behavior, and often distinguished by clearly recognized openings and closings. (p. 17)

He gives an example of a joke, recipe, argument etc. A school lesson would be another example.

From the sociolinguistic perspective, lessons are seen as products of the interaction among participants, i.e. teacher, students, texts/materials (Mehan, 1985; Green & Harker, 1988; Cazden, 2001). Lessons are viewed as reciprocal, collaborative affairs involving both teachers and students (Bellack et. al., 1966; Malamah-Thomas, 1987). Interaction is regarded as being central to language acquisition (Malamah-Thomas, 1987; Ellis, 1994; Johnson, 1995; Walsh, 2006).

Classroom interaction serves first of all the pedagogic purpose (or cognitive purpose in Cazden's terms) but there is also a social purpose in the classroom communication. In the language classroom, Malamah-Thomas (1987) argues, "where the pedagogic purpose of communication is inherently linguistic in nature, the social and pedagogic purposes are often conflated" (p. 18).

Also Walsh draws (2006) a distinction between content-based subjects and language classrooms pointing out that "the linguistic forms used are often simultaneously the aim of a lesson and the means of achieving those aims. Language is both the focus of activity, the central objective of the lesson, as well as the instrument for achieving it" (p. 3).

How students talk and act in classrooms greatly influences what they learn or in our terms, acquire. Full participation in classroom activities requires competence in both the social and interactional aspects of classroom language – in other words,

*classroom communicative competence*. (Wilkinson, 1982; Johnson, 1995). Johnson (1995) points out:

Classroom communicative competence represents students' knowledge of and competence in the structural, functional, social and interactional norms that govern classroom communication. Without such competence, second language students may learn little from their classroom experiences. Just as communicative competence is considered to be essential for second language learners to participate in the target language culture, classroom communicative competence is essential for second language students to participate in and learn from their second language classroom experiences. (p. 6)

For second language students, classroom communicative competence means not only successfully participating in classroom activities, but also becoming communicatively competent in the second language. The patterns of communication in classrooms represent a crucial aspect in the learning process. A lot of teachers would agree with the idea that classroom communicative competence leads to communicative competence outside classroom. There are many countries, including the Czech Republic, where the classroom represents the only encounter with English, i.e. the target language, and the classroom context has to compensate for natural interaction outside classroom. That is why it is so important for the teacher to support language acquisition of the learners through his/her verbal activities.

## **2 Typical structure of classroom discourse**

If we study the structure of classroom discourse we may find out that it is very similar in different parts of the world (Mehan, 1985). He claims that classroom events have internal constituent structures composed of the interactional work of teachers with students. These sequences have three interconnected parts: an initiation act, a reply act and an evaluation act. Mehan (1985) considers the three part sequence "fundamental to educational discourse" (p. 126).

This underlying structure of classroom lessons is usually referred to as *IRE* (*initiation-response-evaluation*). It is important to note that a similar three-part exchange was introduced by Bellack and his co-workers in 1966. Bellack et al. (1966) use *solicit, respond and react*, nowadays they are more commonly described as initiation, response and feedback, but even today, this type of exchange is regarded "the very fabric of classroom interaction" (Walsh, 2006, p. 41). For the IRF structure Walsh uses the term *the essential teaching exchange*.

Buzzelli and Johnston (2002) point out that "certain forms and patterns of classroom discourse promote certain types of learning, and thus engage children as active learners" (p. 23). We would like to argue that the types of questions teachers ask considerably influence language acquisition of their learners. Questions, though, represent only

one type of initiation moves as introduced by Sinclair and Brazil (1982) in their book called *Teacher Talk*.

### 3 Initiation categories

Sinclair and Brazil (1982) state that initiation of language interchanges by the teacher is the main instrument of education. The authors argue that by asking questions, giving instructions and giving information, the teacher guides and controls his class. We can see from the quote that in 1982 the focus was on guiding and controlling the class, whereas the focus of 2015 has shifted to the learners and their language acquisition. We would like to argue here that teachers' verbal activities within initiation move should develop classroom interaction and thus promote language acquisition of learners.

Sinclair and Brazil (1982) deal with initiation in a considerable detail. They describe the move of initiation in terms of individual acts. As a starting point, the authors discriminate between initiations "according to what would be a fitting response for each" (1982, p. 55). They abandon the terms statement, question and command because they can refer to either form or function. Instead, they use declarative, interrogative and imperative forms which correspond to the functions of: information, elicitation, and direction.

#### 1) informing

Informing requires the receiver only to acknowledge.

#### 2) eliciting

Eliciting anticipates some contribution of information from the addressee. By eliciting we can seek *decisions*, *agreement* or *content*.

##### a) decisions

Elicitation is in the form of yes/no questions. The minimal response is either *yes* or *no*.

##### b) agreement

In this case the speaker seeks agreement with what he/she is saying. He/she indicates their own position more or less clearly, e.g. *Isn't the article interesting?* (example by Betáková, 2010)

##### c) content

The syntactic structure of this type of elicitation is a wh-question.

There is an interesting type of elicitation in teacher talk where the initiation and response are members of the same syntactic structure:

Teacher: *So the capital of Scotland is...*

Pupil: *Edinburgh.* (example by Sinclair & Brazil, 1982)

The teacher requires the addressee to finish the sentence. The intonation at the end of the utterance is usually level, indicating that more is to come in the pupil's turn (Sinclair & Brazil, 1982).

It is interesting to note, as far as elicitation is concerned, that in the language classroom more than just minimal response is usually required, even though such an exchange would be very unnatural in everyday conversation. Teachers very often require whole sentence responses in both mother tongue and foreign language classes. The aim, obviously, is to force learners to express themselves extensively. However, sometimes the teachers go too far, not thinking about the consequences for the natural character of the discourse they want to teach.

### 3) directing

Directing equals getting pupils to do things, which means ordering and controlling, instructing, organizing the teaching/learning process.

## 4 Types of questions

Johnson (1995) points out that teachers control what goes on in classrooms primarily through the ways in which they use language. Typically, they retain this control through a question-answer mode of interaction. Teacher questions can both "assist and assess student learning" (Cazden, 2001, p. 92).

Brown and Wragg (1993) divide questions into three categories: *managerial* if they are to do with the running of the lesson (e.g. *Who finished all the exercises?*), *information/data* if they involve the recall of information (e.g. *How many legs does an insect have?*), and *higher order* if pupils have to do more than just remember facts, for example, if they have to analyze, make generalizations or infer (*Why is a bird not an insect?*). These *higher order* or *thought* questions promote thinking and *lower order (factual)* questions promote recalled facts, the authors claim.

Tsui (1995) points that the type of question that the teacher asks affects the kind of response that the students produce. Hatch and Long (1980) show that what students say is often severely constrained by the kind of questions they are asked:

Teachers' questions are usually not genuine questions at all, but pseudo-questions whose real function is to make students display knowledge (which the teacher already has) or to reveal that they lack a given piece of information. That is, they serve a pedagogic function (or at least are thought to do so). (p.18)

Cazden (2001) reports some criticism of the nature of teacher questions and of the IRE (IRF) sequence in general. Teachers have been criticized for asking only *display questions*, i.e. questions, to which they already know the answers. They are also referred to as *knowledge-checking questions* (Tsui, 1995) or *recall questions* (Brown & Wragg, 1993). The questions are said to be inauthentic. Through such questions either the teacher is simply testing student knowledge or is co-opting students to participate in what could otherwise be a lecture – transforming a monologue into a dialogue by eliciting short items of information at self-chosen points. On the basis of such criticism, Cazden (2001) points out, teachers are frequently admonished to ask “authentic questions” more typical of informal conversation. The opposite of display questions are so called *referential questions* (i.e. such questions, to which the answer is not known in advance). Some researchers believe that referential questions are more likely to produce natural responses than display questions (Nunan, 1989).

Tsui (1995) believes that display questions generate interactions that are typical of didactic discourse, whereas referential questions generate interactions typical of social communication.

Nunan (1989) claims that in the classroom in comparison with ordinary conversation there is much less negotiation for meaning due to the fact that teachers tend to ask display questions. His research shows that the groups of students in which more referential questions were asked gave significantly longer and more syntactically complex responses.

Tsui (1995) is convinced that teachers who often ask display questions and/or disguise them as referential questions are likely to encourage students to repeat facts or pre-formulated language items and discourage students from trying to communicate their own ideas in the target language. Walsh (2006) believes that the extent to which a question produces a communicative response is less important than the extent to which a question serves its purpose at a particular point in a lesson. The use of appropriate questioning strategies requires an understanding of the function of a question in relation to what is being taught.

Another division may be drawn between *open* and *closed* questions, according to Brown and Wragg (1993) also *convergent* or *divergent* or *narrow* and *broad*. Closed questions, according to Brown and Wragg (1993), allow only one possible response, whereas open questions produce more responses. Excessive use of narrow questions, according to the authors, yields short answers and frequently inhibits discussion. According to Tsui (1995), in terms of students’ language output, closed questions are more restrictive than open questions. Even more restrictive are closed questions where the teacher provides the sentence structure as a clue to solicit an appropriate response. Cazden (2001), on the other hand, calls the questions that are open in form but demonstrably closed in function “*pseudo-open questions*”.

Closed questions are seen by many authors in a broader sense as questions which somehow restrict the possible answer. Open-ended questions, in the classroom,

represent a powerful tool available to teachers to stretch children's curiosity, reasoning ability, creativity, and independence (Denton, 2007). She defines open-ended questions as those with no single right or wrong answer. Instead of predictable answers, open-ended questions elicit fresh and sometimes even startling insights and ideas, opening minds and enabling teachers and students to build knowledge together. Long and Sato (1983, in Walsh, 2006) believe that the length and complexity of learner utterances are determined more by whether a question is closed or open than whether it is a referential or display one.

Mehan (1979) talks about a similar type of questions as those named by Sage as *how* and *why*. He calls them *metaprocess questions* (e.g. *How did you know/remember?*) which ask for different kinds of knowledge and prompt longer and more complex responses. Cazden (2001) calls them *process* or *metacognitive questions*. She explains that these questions call the learners' attention to their own thinking and their own knowledge. In non-traditional lessons, teachers are encouraged to ask more of such questions to encourage students to explain their own thinking and reflect on what others have said, she adds.

## **5 Different types of questions for different purposes**

Having thought of the types of questions teachers ask, we can have a look at reasons why they ask questions.

Long and Sato (1983, in Walsh, 2006) list the complex role played by questions in classrooms; they can serve to signal turns, aid comprehensibility, provide opportunities for non-native speakers to participate or even compel involvement. A teacher's use of questions is the single most-frequently used discourse modification to aid and maintain participation among learners. Classroom discourse differs from normal communication in terms of the number of questions used and their function: to encourage involvement rather than elicit new information.

To summarize the reasons for asking questions in the classroom we can quote Wragg and Brown (2001): "Questions are asked to facilitate learning, so they are linked to the aims of lessons and the underlying purpose of the lesson" (p. 11).

The teacher may control both the form and the content of the students' responses. The former case concerns situations whereby the teacher is more concerned with the grammatical accuracy of students' responses than with their ability to express meaning. Lemke (1985; in Johnson, 1995) characterized this sort of reaction from teachers as exerting *interactional control* versus *thematic control* over the pattern of classroom communication. The only value of the answer is grammatical accuracy, the teacher sacrifices understanding of the student. Accuracy, in this case, is valued more than understanding of the content, Johnson (1995) points out.

As far as the number of questions teachers ask is concerned, Walsh (2006) shows that classroom discourse is dominated by question and answer routines with teachers asking most of the questions.

Sage (2000) shows that teachers shoot questions at students “as if they were firing on the enemy” (p. 64). She reports Susskind (1969) who estimated that in American elementary schools, teachers ask two questions per minute, while students ask questions at the rate of two per hour. Tsui (1995) reports that questions constitute 20 to 40 per cent of classroom talk. Wragg (1993) found that in 20 observed lessons there were fewer than 20 questions from students, and these were procedural.

Thornbury (1996) quotes Van Lier (1988): “It is generally believed that learners should, at least some of the time, be asking the questions. A significant source of motivation and attention is lost when turn taking is predetermined rather than interactionally managed by the participants” (p. 133).

Cazden (2001) points out that the teacher should allow several seconds to pass before calling on students so everyone has an opportunity to develop an answer. Students should be given time to evaluate their own ideas, time to think. After the response, they should be given some time to reflect on their response, to have time for self-correction, as it seems to be the most valuable type of correction especially for more advanced students.

Sage (2000) claims that by analysing discourse along levels of difficulty, a teacher can determine the cognitive and linguistic structures that lead to understanding rather than frustration. It is also argued, though, that we must expose children to complex language if we want them to acquire it for their learning. Sage suggests balancing straightforward questions with comments that use complex linguistic structures.

## **6 Research outcomes**

In our research into classroom discourse carried out in three classes of English at the lower secondary school level we also focused on the number and types of questions asked. To be able to refer easily to the three recorded classes, they were attached labels discourse A, B, and C. Class A was taught by a qualified native speaker teacher, class B by a fully qualified non-native speaker teacher and class C by a non-native speaker teacher, unqualified for teaching English.

The study of the types of questions teachers used brought very interesting results. Teacher A used 120 questions per lesson, teacher B 160 questions and teacher C 51 questions per lesson of 45 minutes. As to the division between referential and display questions, teacher A’s questions are in harmony – 50.8 per cent vs. 49.2 per cent. Teacher B used twice as many referential questions – 67.5 per cent vs. 32.5 per cent. Teacher C asked four times more display questions – 21.6 per cent vs. 78.4 per cent.

Table 13.1  
*Referential and Display Questions*

	A	%	B	%	C	%
*R	61	50.8	108	67.5	11	21.6
*D	59	49.2	52	32.5	40	78.4
total	120	100	160	100	51	100

\*R=referential questions; \*D=display questions

It has been pointed out that it is not very productive to speculate whether it is better to use referential or display questions. They are used for different purposes, as can be seen from the particular stages of the recorded lessons A and B. If the main aim was developing speaking, the teachers employed mostly referential questions and the interaction resembled a real life conversation. Such a conversation was interrupted with occasional display questions in which the teachers concentrated on the form of what had been said. In other parts of the lessons the teachers' focus was the language form, either grammar or vocabulary. In such parts, display questions appeared almost exclusively. In discourse C, where the main purpose was to check comprehension of a text for listening, the teacher asked only display questions.

As far as the division between open and closed questions is concerned, teacher A asked four times as many closed questions – 80.8 per cent vs. 19.2 per cent. Teacher B asked twice as many closed questions – 64.4 per cent vs. 35.6 per cent. Teacher C did not ask any genuinely open question.

Table 13.2  
*Open and Closed Questions*

	A	%	B	%	C	%
*O	23	19.2	57	35.6	0	0
*C	97	80.8	103	64.4	51	100
total	120	100	160	100	51	100

\*O=open questions; \*C=closed questions

It is generally believed that open questions are of “*higher quality*” because they give the learners a chance to express themselves, and in addition they promote thinking and elicit more language. It is important to study whether open questions elicit more language.

The teacher A used truly open questions:

- a) *What are you gonna do during the summer?*
- b) *What will you do there?*
- c) *What will you do during your holidays, Ondra?*
- d) *What do you eat in the summer?*

- e) *What is your favourite thing to do?*
- f) *How is the night life in České Budějovice?*

But the answers of the students were only minimal.

- a) *I will go to the Germany first.*
- b) *Opalovat.*
- c) *I go to Slovakia.*
- d) *I eat fruit.*
- e) *I play guitar.*
- f) (No answer, only laughter).

When the teacher felt he did not elicit enough language from the students, he asked another, often more specific question:

- a) *Where?*
- b) *Will you do this?*
- c) *Where?*
- d) *What kind of fruit, what's your favourite?*
- e) *Do you like to play guitar at a fire?*
- f) *Is it good?*

We can draw similar conclusions from the class taught by teacher B. First, she always asked one or more open questions. If she realized that the student responses were minimal and they were not willing to speak, she began to use closed questions. It is interesting to observe that the pattern of each teacher-student exchange remained the same – at least one open question at the beginning followed by one or more closed questions, e.g.: *And Vašek, anything special happened throughout the four years, the four days, sorry? Nothing special happened? Did you enjoy the holiday? Or were you happy to come to school?*

Very often we can observe that the focus of each subsequent question got narrower: *Báro, what about you? What did you do at your grandma's? Did you buy anything special? Clothes?*

We can conclude the study of questions in discourse B by saying that no matter which type of question the teacher uses (open or closed), the students always provide only minimal answers. Open questions do not, in this case, elicit more language than closed questions. We can only speculate about the reasons: they might be attributed to the age of the students (14 or 15) who do not want to talk about themselves, do not want to share their ideas with others and do not want to 'show off' in front of their classmates or to look too cooperative in their eyes. The question is whether a pair work or a group work exercise would be more productive in this case and whether the students would show a greater willingness to speak with their peers instead of the teacher.

Table 13.3  
Types of Questions

	A	%	B	%	C	%
*DC	59	49.2	43	26.9	40	78.4
*RC	38	31.7	60	37.5	11	21.6
*RO	23	19.1	48	30	0	0
*DO	0	0	9	5.6	0	0
total	120	100	160	100	51	100

\*DC=display closed; \*RC=referential closed; \*RO=referential open; \*DO=display open

If we combine the four aspects from Table 13.3, we can see that teacher B asked more referential open questions than teacher A – 30 per cent (B) vs. 19.1 per cent (A). She also asked the highest proportion of referential closed questions – 37.5 per cent (B) vs. 31.7 per cent (A) and 21.6 per cent (C). On the other hand, teacher C asked the highest proportion of display closed questions – 78.4 per cent (C) vs. 49.2 per cent (A) and 26.9 per cent (B). Teacher B was the only one to use the display open questions. The reason can be found in the teachers' pedagogic skills (Betáková, 2010).

One of the interesting points to look at in connection with teacher questions would be the techniques teachers use to reinitiate.

The teacher repeats the question with the same wording:

T: *But what is your favourite thing to do during holidays?*

S: *I will be with friends and I don't know.*

T: *You will be with friends but what is your favourite thing to do?*

The teacher modifies the original question slightly:

T: *And do your friends play instruments?*

T: *Do your friends also play guitar and drums?* Instead of the more general word *instrument* he uses concrete examples of musical instruments.

The teacher uses a clue to help the learner reach the right answer:

T: *What do we call a book that tells you lots of useful holiday expressions in another language?*

S: *Slovníček.*

T: *It's a dictionary, but a special kind of dictionary. It has an adjective. There is another word before it.*

The teacher did not catch the answer, so he asks the student to repeat, using a directive: *Say again?*

## 7 Errors in questioning

Brown and Wragg (1993) present some common errors in questioning:

- Asking too many questions at once.
- Asking a question and answering it yourself.
- Asking questions only of the brightest or most likeable.
- Asking a difficult question too early.
- Asking an irrelevant question.
- Always asking the same types of questions.
- Asking questions in a threatening way.
- Not indicating a change in the type of question.
- Not using probing questions.
- Not giving pupils enough time to think.
- Not correcting wrong answers.
- Ignoring answers.
- Failing to see the implications of answers.
- Failing to build on answers.

The question to raise is whether the three teachers made any of these mistakes in questioning. Teacher A and B probably asked too many questions. The reason might be that they wanted to involve everybody in the class. That is why they asked all the students the same questions: e.g. *What did you do in your holiday?* which did not elicit enough language. They could have asked each student a different question to make it more interesting (e.g. *What did you enjoy most? What was the most boring thing you did?* etc.). On the other hand, in the three lessons, there were only two initiations by the students, both in the mother tongue.

Teacher C always asked the same type of questions: display and closed. If the students were not able to answer, she answered the questions herself, not giving the students enough time to think or find the answer.

From our own experience with class observation we can mention other types of mistakes made by many teachers. Very often they ask too broad, complex questions which cannot be answered in a simple way. Dvořák (2014) reports questions such as *What is culture? What is literature?* Even though they are open questions they produced very limited answers – e.g. *Reading, books*. In this particular case the problem was that the students were not given enough time to think and the teacher only wanted very short answers to be able to answer herself and define what literature and culture was in her own words. In general, teachers seem to show very little interest in the students' answers, they do not accept the content of the answer in relation to learners' age and interests. Little interest is paid to the development of speaking skills of the students in general, teachers tend to be dominant players in the classroom interaction who focus too much on their own performance compared to

the development of the communicative competence of their learners. Many teachers underestimate learners in advance not giving them enough space to develop answers.

Many problems can be overcome by effective questioning strategies – e.g. discussing the questions asked by teachers in pairs or small groups first through peer interaction and later communicating the answer of the group to the teacher. This simple technique has many advantages – it gives students enough thinking time, the learners interact with peers of the same age and similar language level, they try the language out in a friendly environment.

## 8 Conclusion

Both the results of the research we have carried out and our long-term experience with class observation show that it is absolutely necessary to ask both referential and open questions in every lesson. If teachers want to prepare students for real communication outside classroom, they have to ask them referential open questions and they have to teach them to be able to ask such questions themselves as these questions form a crucial part of everyday discourse. Even if we know that display questions do play a very important role in classroom discourse and we know that they promote learning, they cannot be the only questions asked by the teacher in the lesson. Teachers should be aware of the fact that display questions can be very useful for some purposes but their use is limited because they do not promote communicative competence of the learners.

We have seen that there was not any significant difference in student language output emerging from either open or closed questions. It does not mean though, that teachers should withdraw open questions. Students can be trained to answer open questions using a greater quantity of the target language. They should be explained that they themselves are responsible for their own learning and for their success in language learning. Classroom is quite a “*safe*” place for language use in comparison with the real life supposing that a good classroom atmosphere has been created. The atmosphere is co-created by both the teacher and the learners. If the students do not laugh at each other, if they are supported by their teacher and peers to try the language out, if they get used to simulating language use in everyday, natural situations, if they use the target language as much as possible in classroom discourse, they are likely to be able to transfer the skills to the real life and gain communicative competence. On the other hand, if the learners use the language only to carry out language exercises in the classroom and the target language is not used as a means of communication, they will find it much more difficult to use the language outside classroom. Having found that they are not able to communicate successfully outside classroom, they may lose motivation.

To summarize the role of teacher questioning in student language acquisition, we can say that effective questioning depends on teachers' social skills (Dvořák, 2014), i.e. the ability to support and motivate the learners to be actively involved in classroom interaction. The teachers should show empathy with the learners and their process of learning. Listening skills of the teacher are very important. We should listen and show interest in what the students are telling us. As our previous research into classroom discourse showed (Betáková, 2010), supportive feedback of the teacher is also crucial. If the feedback is negative and critical, the students will not try to answer our questions. This would lead to negative classroom atmosphere, limited learner output and consequently to limited language acquisition.



## Chapter 14

# SUPPORTING ENGLISH ACQUISITION THROUGH THE TEACHER'S SOCIAL SKILLS

*Petr Dvořák*

## 1 Introduction

In this chapter we aim to analyse aspects of foreign language teaching (FLT) which may either facilitate or hinder the learning and acquisition processes. We would like to specifically concentrate on the aspects influenced by the teachers and their classroom management, as well as modification of interactional processes that influence foreign language acquisition (FLA).

The generally accepted approach based on Hymes's theory (1979) of communicative competence closely connects language with its use in social context. The fact that communicative language teaching is directly linked to it underscores its primary function, namely the interaction within which teaching takes place. To study the efficiency of English language teaching (ELT), it seems necessary to research interactions within which the language acquisition occurs. This close link between interaction and FLA has been emphasized by numerous authors (e.g. Allwright & Bailey, 1991; Betáková, 2010; Ellis, 1994; Hall & Verplaetse, 2000; Lynch, 1996; Mackey, 2007; Richard-Amato, 2010; Scrivener, 2011; Shumin, 2002). Most of them agree that interaction creates opportunities for learning, provides learners with the target language (language input in Krashen's (1989) view) and facilitates communication of meaning. Interaction also simulates target-language environment, creates opportunities for language use, enhances acquisition, and leads to language production (language output in Krashen's (1989) preliminaries).

## 2 Theoretical preliminaries

### 2.1 Interaction and discourse in foreign language learning

Since interaction is crucial, it should be present in the classroom setting. The classroom is at the centre of our attention as teaching, learning, and language acquisition all occur there. Acquisition and learning are also the result of interactions between the

teacher (T) and the students (S). In FLT it is important for the teacher to manage the interaction in the classroom, so that each learner is actively involved. All learners should be given the opportunity to learn through interaction (Allwright & Bailey, 1991). The classroom represents a social arena (Dörnyei, 2009) in which the language is being acquired. This environment provides the learners with social context that facilitates learning (e.g. Ellis, 1994; Littlewood, 1981; Dörnyei, 2009).

Discourse is understood to be language in context, as agreed by numerous authors (Cook, 1989; Hatch, 1992; McCarthy, 1991), in our case in the context of ELT classroom-based communication. Language plays an important role as it represents both the means and the end-goal of interaction (McCarthy, 1991). This type of discourse is of a specific type as it serves educational aims. Traditionally, the teacher is the dominant player who decides who will speak and for how long. The traditional model of classroom interaction is referred to as IRE (initiation – response – evaluation), alternatively described as IRF (initiation – response – feedback or follow-up (see Cullen, 2002)). Interaction is initiated by the T's questions or instructions, followed by learner response or reaction, and concluded in the final stages by the T's feedback which has an irreplaceable role and can focus on both the form and meaning (e.g. Cazden, 2001; Sinclair & Brazil, 1982; Sinclair & Coulthard, 1975; Mehan, 1985).

## **2.2 Participants of educational interaction – students and teachers**

Foreign language interaction and discourse, and consequently the whole effectiveness of foreign language teaching and learning, are significantly influenced by the main participants – teachers and students. The interaction reflects the functional communicative approach to language and communicative language teaching. From the socio-psychological perspective, the roles of teachers and students are asymmetric but complementary. The teacher is the leader in the educational processes (e.g. Gillernová, 2009; Helus, 2007; Mothejzиковá, 1988; Wright, 1987). Cazden (2001) asserts that the asymmetry is defined by teachers distributing the right and obligation to speak. Lee and VanPatten (2003) emphasise that the teacher is a supplier of communicative opportunities and his/her role is to interact with students. According to constructivist theories, however, the teachers are not mentors but facilitators. They provide the students with the target language enabling communication, thus creating conditions for language acquisition. This is highlighted by numerous authors (Harmer, 2007; Lee & VanPatten, 2003; Richard-Amato, 2010; Williams & Burden, 1997).

The contemporary approach to teaching works with the concept of the teacher-expert (e.g. Gillernová, 2013; Pišová, 2010; Williams & Burden, 1997) who possesses a wide range of professional skills including the subject-specific ones – in this case the target language. The teacher is supposed to create conditions for optimal

acquisition of knowledge and skills and become engaged not only in the content of teaching (the target language and its methodology) but also in the social milieu.

In FLT literature, the period of adolescence is marked as being crucial to learning success (Ellis, 1994; Podrápská, 2011; Richard-Amato, 2010). Although theories posited by Dörnyei (2009), Ellis (1994), and Littlewood (1984) assert that individual FL learners at an older age might evince worse results in pronunciation, there appear to be no significant differences in other language areas. As for grammar, adolescent students reach better results than children due to their higher cognitive maturity such as metalinguistic knowledge, a wider range of learning techniques and problem-solving skills, their higher capacity of abstraction, concentration and goal orientation. Ellis (1994) emphasises that rather than the age of students there are other decisive factors, such as the quantity and quality of language input, social milieu, motivation and others, that may play a key role in the success rate. Moskowitz (1978) sees one of the key characteristics of adolescents in their identity search for self-acceptance. If the adolescent students' needs do not correspond with the school requirements, their motivation, interest and performance fall, and their behaviour changes (Eccles et al., 1993, in Krejčová, 2011).

In this process of self-identity formation, teachers play a significant role – prevalently in communicative FL teaching (Choděra, 2006; Ellis, 1994; Harmer, 2007; Lee & VanPatten, 2003; Lynch, 1996; Richard-Amato, 2010; Scrivener, 2011). Teachers become model adults for their students. According to Vygotsky's (1978) concept, they accompany the students into the zone of proximal development. The students spend most of their time at school, and their teachers are at hand (Krejčová, 2011). Positive T-S relationships are sources of further development within the meaning of adequate socialization (Scales, Benson & Manes, 2006, in Krejčová, 2011; Vágnerová, 2000). FL students in particular are open to their teachers and tend to communicate personal information to them expecting that they will be heard and their communicative acts will be respected.

### **2.3 Teacher's social skills and their educational style**

Teachers' activities place demands on a wide range of professional skills of a complex character (Kyriacou, 2007; Švec, 2002). Various kinds of the teachers' activities correspond with particular skills supporting their realization. We recognize the following division of the teacher's skills: subject-specific skills, teaching skills, diagnostic skills, and social skills (Gillernová, 2013).

Kozulin (1998) stresses that novel approaches are influenced by social requirements typical of modern society. The students' development is influenced by formal and informal education (including the hidden curriculum). Their learning success is not

mere knowledge accumulation but the development of general and specific cognitive strategies.

The cognitive strategies in the target language interactions are closely linked to social functions and social milieu, in this case the classroom environment. The teacher's social skills play a crucial role in this process. Social skills are understood as being "the ability to interact with other people in a way that is both appropriate and effective" (Segrin & Givertz, 2003, p. 136). A wide range of the T's social skills (empathy, listening, acceptance, encouragement, etc.) is reflected in his/her educational style. Wright (1987) defines educational style as a whole complex of the teachers' attitudes and behaviour used to create optimal conditions for learning. He emphasizes that "educational style lies at the heart of the interpersonal relationship between the teacher and the learner" (p. 68).

### **3 Research into English teachers' social skills and their impact on the target language interaction, communication and discourse**

#### **3.1 The aims and methods of the research**

The main research question is: *What is the impact of the teacher's social skills on the processes of FL interaction, communication and discourse at secondary school level?* The ensuing research study is of mixed methods – a qualitative/quantitative character. Its first part investigates (with the help of a questionnaire) how students perceive the educational style of their English teachers along the dimensions of the *educational relationship* and the *educational management*, both of which reflect the range of the teacher's social skills. Its second part explores selected crucial aspects of classroom FL interaction, communication and discourse by means of an analysis of videotaped instructional sequences, i.e. typological interaction analysis (TIA), (Pelikán, 2011). In the final part of the paper the findings of the two research parts are discussed, and their interrelationships considered.

#### **3.2 Research design**

The research sample consisted of five conveniently selected teachers and 64 of their pupils, all attending the most common types of secondary schools, namely grammar school (teachers A, B, and C) and vocational school (teachers D and E). Teacher A is a qualified English teacher, a female, 53 years old, with 22 years of teaching practice, a non-native speaker of English. Teacher B is a qualified English teacher, a male, 29 years old, with four years of teaching practice, a non-native speaker of English. Teacher C is a qualified teacher of English, a male, 38 years old, with seven years

of teaching practice, a native speaker of English. Teacher D is a qualified teacher of English, a female, 40 years old, with sixteen years of teaching practice, a non-native speaker of English. Teacher E is a qualified teacher of English, a female, 35 years old, with four years of teaching practice, a non-native speaker of English.

The teachers' classes included in this research consisted of:

Teacher A: fourteen students, six girls, eight boys, aged 18 – 19 years, English as a first foreign language, B2 level according to the Common European Framework of Reference for Languages (CEFR, 2001).

Teacher B: eight students, seven girls, one boy, aged 17 – 18 years, English as a first foreign language, B1 level according to CEFR.

Teacher C: fourteen students, eleven girls, three boys, aged 18 – 20, English as a first foreign language, B2 level according to CEFR.

Teacher D: fourteen students, ten girls, four boys, aged 17 – 18 years, English as a first foreign language, B1 according to CEFR.

Teacher E: fourteen students, twelve girls, two boys, aged 18 – 19 years, English as a first foreign language, B1 level according to CEFR.

### 3.3 Research into educational styles

The teachers' educational styles reflect selected aspects of T-S interaction (social skills in particular). Gillernová and Krejčová (2012) emphasize that “social skills are crucial variables of the teacher's educational style” (p. 62).

This research study was carried out in cooperation with the Department of Psychology, Faculty of Arts, Charles University, Prague, through a questionnaire titled “Student about Teacher” (for details see Gillernová, in Mertin & Krejčová et al., 2012, pp. 274-283). In the questionnaire the students first defined an ideal teacher of English, and then reflected on their actual teacher's behaviour in two variables – the *educational relationship* and the *educational management*. The questionnaire was construed for students aged twelve and older, assigned by this paper's author, and filled in anonymously.

The goal of the research focused on students' perceptions of their teachers' educational style as measured by the two above-mentioned variables, both of which reflect social skills. The *educational relationship* refers to the teacher's positive and negative displays of behaviour (emotional characteristics such as empathy, listening, acceptance or misunderstanding of the students' behaviour, educational needs, etc.). On the basis of combination of positive and negative components, the questionnaire defined four types of this relationship: negative, inconsistent, medium (these three are marked as non-positive), and positive. On the basis of two further components – managerial requirements and freedom – the questionnaire defined

four types of the *educational management*: weak, inconsistent, medium, and strong. The *educational management* refers to the teacher’s requirements, feedback and control. The combination of the two dimensions – the *educational relationship* and the *educational management* – formed the T’s educational style. The questionnaire distinguishes six types of the T’s educational styles as shown in Table 14.1:

Table 14.1  
*Types of educational styles*

Educational style / Educational characteristics of teacher – pupil interaction	Positive educational relationship	Non-positive educational relationship
strong and medium educational management	1	2
inconsistent educational management	3	4
weak educational management	5	6

### 3.4 Research findings into educational styles of English teachers

In agreement with previous research studies (Gillernová, 2009; Krejčová, 2011), the findings have demonstrated that students prefer strong *educational management*. Secondary school students require that their teachers provide them with enough tasks to meet their educational needs. They also prefer English lessons with clear goals, structure, and control. Students prefer that their teachers use innovative teaching methods and satisfy their desire to learn the target language. The results for all the teachers analysed in terms of the educational management were similar: 92.9 per cent of the pupils perceive the educational management of teacher A as strong, 87.5 per cent for teacher B, 64.3 per cent for teacher C, 100 per cent for teacher D, and 71.4 per cent for teacher E. On the basis of statistical calculations, the educational management of all the teachers analysed was perceived as strong.

The dimension of the *educational relationship* was perceived by the students with distinctive differences. Although the students stated that they require encouragement, support, praise and acceptance, as well as other positive displays of the T’s behaviour, not all of their teachers satisfied their requirements. 50 per cent of the students perceived the *educational relationship* of teacher A as positive and 50 per cent as medium. 50 per cent of the students perceived the relationship of teacher B as medium, 37.5 per cent as positive, and 12.5 per cent as negative. The most positive assessment in this respect went to teacher C: 92.9 per cent of the students perceived it as being positive, 7.1 per cent as medium. Teacher D was perceived similarly: 92.9 per cent of the students graded it as positive, 7.1 per cent of the students as

inconsistent. 71.4 per cent of the students perceived the relationship of teacher E as positive, 14.3 per cent as medium, 7.1 per cent as inconsistent, and 7.1 per cent as negative. On the basis of statistical calculations, the educational relationship of teachers C, D, E was perceived as positive, while the educational relationship of teachers A and B was perceived as medium.

Using the two variables – the *educational management* and the *educational relationship* – the educational styles of the teachers analysed could be defined. The educational style of teachers C, D, and E could be characterized by positive educational relationship and strong educational management (type 1), whereas the educational style of teachers A and B could be characterized by medium educational relationship and strong educational management (type 2).

### 3.5 Analysing discourse in selected ELT classroom interactions

In this part of the study we wanted to ascertain if the quality of interactions, as well as the distribution of the target language and the mother tongue, were influenced by psychological factors, especially the T's educational style and their social skills.

This will be demonstrated via an analysis of the ELT classroom discourse, in particular of several selected aspects of T-S interaction. We attempted to find out if the three move structure, traditionally referred to as IRE or IRF (Cullen, 2002), is still common in secondary schools. This study focuses in detail on the initiation of interaction, asking whether the T's educational style influences it, on the distribution of communication opportunities, and on the types of questions posed by the teachers.

This part of the research was carried out by the method of indirect structured observation on the basis of TIA (Pelikán, 2011). It focuses on ten-minute teaching sequences of the most intensive T-S interaction (sequences with the highest number of communication turns between the teacher and the students). TIA enables researchers to sort the teacher's displays of behaviour into 25 segments classified into three categories – dominant, neutral or interactive – while dividing the students' reactions into four categories – potential students' agreement, open students' agreement, potential students' disagreement, and open students' disagreement. The three types of the teacher's behaviour reflect a wide range of T's social skills, as applied in each particular interaction. On the basis of the three segments we can form an index of interaction which is then calculated according to the following mathematical formula:

$$I_i = \frac{(\sum Os + \sum Ps) - (\sum Pn + \sum On)}{N},$$

where  $I_i$  is the index of interaction,  $\sum Os$  are all open students' agreements,  $\sum Ps$  are all potential students' agreements,  $\sum Pn$  are all students' potential disagreements,

$\Sigma On$  are all students' open disagreements, and N are all students' reactions. The continuum of interaction is from -1 (extremely negative students' responses to the teacher's behaviour) to +1 (unequivocally positive responses to the teacher's behaviour).

Apart from the research into the level of interaction using TIA, we decided to analyse the following aspects of ELT discourse and interaction within the ten-minute sequences defined by TIA:

- communication turns;
- teachers' initiations;
- students' initiations;
- teachers' questions;
- students' questions;
- teachers' feedback;
- students' feedback;
- teacher/student ratio in the target language communication.

### 3.6 The findings related to particular ELT interactions

First, the findings of the level of T-S classroom interaction as recorded by TIA will be presented.

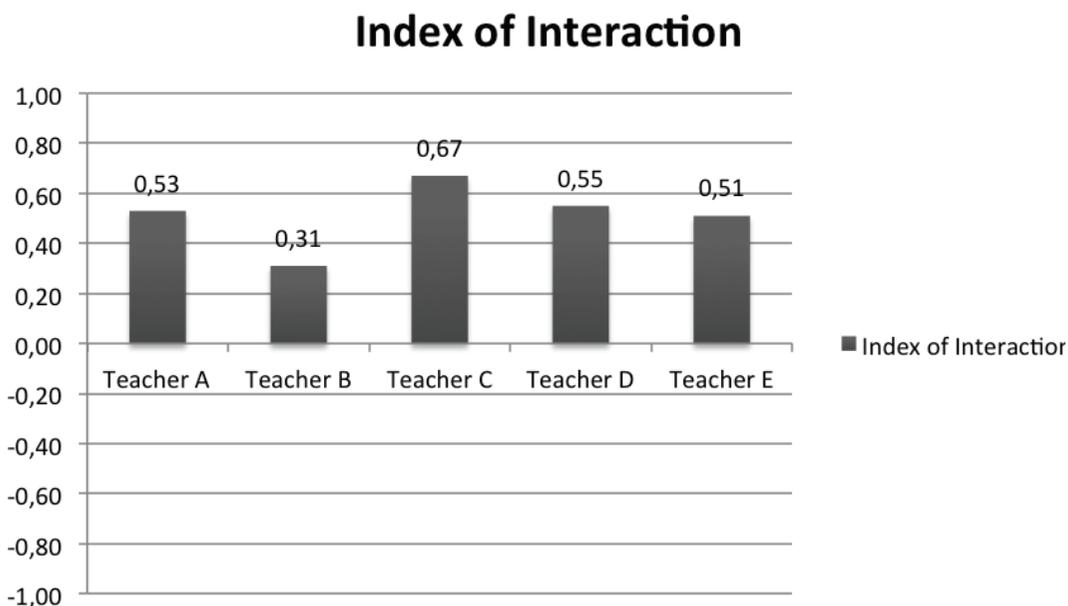


Figure 14.1: Index of interaction analysed

As can be seen from Figure 14.1, all the teachers analysed are in positive figures, which means the students' positive acceptance. The differences between the individual teachers are not extreme. The most positively accepted teacher was teacher C, the least positively accepted one was teacher B. It needs to be emphasized that the higher the index of interaction is, the more interest the students have in the processes of foreign language interaction and communication. The lower the index is, the less willingness there is among the students to participate. The findings into the crucial aspects of ELT discourse are shown in Table 14.2:

Table 14.2

*Discourse analysis*

Aspects of classroom discourse/Teacher	Teacher A	Teacher B	Teacher C	Teacher D	Teacher E
communication turns	100	92	93	52	68
teacher's initiations	36	34	29	17	29
students' initiations	0	0	3	6	2
teacher's questions	58	50	19	13	33
students' questions	0	0	2	6	2
students' answers	47	46	45	21	31
teacher's answers	0	0	1	5	2
teacher's feedback	47	42	36	15	31
students' feedback	0	0	1	2	0

As far as the T-S talk ratio and the types of interaction are concerned, it can be observed that all the five lesson sequences possess similar features. One of the aspects that can be gleaned at first sight is the T's dominant role. In this respect, all five interactional sequences represent the traditional structure of interaction referred to as IRE or IRF where the teachers initiate and allocate turns while the students respond, with the teachers then providing feedback. Whereas in the lessons of teachers A and B there were no occurrences of students' initiations, in the lessons of teachers C, D, and E several such initiations were marked. It also became evident that the communication at the grammar school (teachers A and B) was more dynamic than at the vocational school, mainly due to a much higher number of communication turns.

This research study has produced a number of useful findings that can be applied during teaching practice. Although the number of T's initiations was relatively balanced, the number of their questions posed was significantly different. It is to be noted that teacher A asked 58 questions within 36 initiations during the ten-minute teaching sequence, i.e. almost six questions per minute. Similar figures were derived for teacher B. Furthermore, teachers A and B had to pose more than double the number of questions to get the same number of students' answers as teacher C. Fewer teacher's questions and more students' answers means that the teacher had

given more time to students to think and formulate their answers. It also confirms that more than one student reacted to each T's question – which in turn provides more target language speaking opportunities.

As far as the teacher's feedback is concerned, there was evidence that the students were provided with a substantial quantity of feedback. Its quality, along with other aspects of classroom discourse (types of questions, for example), was analysed as well but cannot be presented in this paper due to the length limitations.

The research findings reveal that the students in the lesson sequences of teachers C, D, and E did not hesitate to pose questions. Moreover, teachers C and D's students were frequently providing the teachers with feedback on their questions.

The study also analysed the teacher/student talk ratio and English/mother tongue ratio. For this part, a VIP computer programme developed by the Department of Psychology of Faculty of Arts, Charles University, Prague, was used. The computer programme enables to precisely record not only the quantity of the selected phenomena, but also their length.

The study of the use of the target language versus the mother tongue showed that teachers B and C did not use Czech at all. In the teacher A's interactional sequence there were three occurrences of Czech (the teacher translated three English words into Czech). Teacher D used Czech five times (three times for word translation and twice for giving instructions). Teacher E used Czech six times (translating two English words and four sets of instructions). To summarize, it is positive to state that the total English/Czech ratio of each individual teacher in the ten-minute interactional sequences analysed did not exceed four per cent of communication in Czech.

Table 14.3

*Time of verbal communication in the target language*

Discourse/ Participants	Discourse A	Discourse B	Discourse C	Discourse D	Discourse E
teacher	0:07:08	0:06:45	0:05:15	0:02:19	0:04:06
students	0:01:06	0:01:46	0:02:37	0:01:15	0:00:57
total	0:08:14	0:08:31	0:07:52	0:03:34	0:05:03

As far as the teacher/student talk ratio is concerned, Table 14.3 shows the T's dominant role in the discourse in all five lesson segments. The figures show the amount of verbal communication in the target language realized by the teachers and the students in the ten-minute interactional sequences. It has been found that a significantly higher volume of communication in English was recorded in the grammar-school lessons. It has also become evident that the total time of students' communication in discourses A and D is comparable, while their teachers' volume of communication considerably differs. In discourse A the T's talk took up more than two thirds of the sequence under study. The highest volume of students' talk

in the target language was marked in the teacher C's lesson – of all the lessons analysed, his teacher/student talk ratio appeared to be the best balanced.

#### **4 Conclusion**

The research process has resulted in a number of useful findings, all relevant to pedagogical practice. The students have agreed that an optimal educational style for the teacher of English can be characterized by positive *educational relationship* and strong *educational management*. Hence the teachers included in the research sample were evaluated according to criteria relevant to those variables. In general the students perceived the teachers' *educational management* as being strong, which tends to correspond positively to their developmental needs and educational requirements. On the other hand, the *educational relationship* was assessed differently: only three of the five teachers were classified as having a positive educational relationship, while the other two were assessed as having a medium educational relationship – implying that the students indicated that these teachers displayed more of negative behaviour, and a lower level of positive reinforcement. Such behaviour indicates less developed teacher's social skills. The teachers whose educational style was perceived as medium received a higher rate of negative reactions in each particular interaction. In the video segments these teachers displayed a limited range and number of interactive traits. They also asked an enormous number of questions – their number and manner of asking prevented the students from responding. The results have proven that any relationship other than what we define as a positive educational relationship tends to dampen the students' efforts to communicate.

On the other hand, those teachers whose pro-socializing manner of expression was evaluated as being primarily positive achieved a higher level of interactions. The students' reactions appeared to be more positive and more frequent. These teachers posed fewer questions, and instead tended to give the students more time to think and formulate their answers. In relation to the number of questions, they tended to receive more answers than the teachers who posed a higher number of questions. All the five interactional sequences represent a typical traditional structure of interaction referred to as IRE or IRF where the teacher initiates communication, students respond, and the teacher provides feedback. The teachers, whose educational relationship was perceived as positive, recorded several students' attempts at initiating communication, including questions and feedback.

As far as the target language/the mother tongue talk ratio is concerned, it could be observed that more than 95 per cent of the communication in each particular interactional sequence, whether by the teachers or the students, was carried out in English. As far as the teacher/student talk ratio is concerned, most of the communication in the target language was performed by the teacher. Nevertheless,

the highest level of student communication was recorded in the instruction of one particular teacher whose educational relationship received the highest evaluation.

To sum up, it can be concluded that social skills play a decisive role among the language teacher's professional competences. Like the target language, these skills are considered to be both the tool and the end goal of teaching. They influence classroom interactions, and therefore also the processes that lead to the acquisition of a foreign language. The findings of this research study have confirmed that "language is socially constructed. Language use, social roles, language learning, and conscious experience are socially situated, negotiated, scaffolded, and guided" (Ellis & Larsen-Freeman, 2006, p. 572).

## RESUMÉ

Odborná kniha s názvem *Learner Corpora and English Acquisition. A collection of studies* (Žákovské korpusy a osvojování si angličtiny. Soubor studií) představuje jeden z výsledků výzkumného projektu *Aspects of English Language Acquisition of Czech Students at the Onset of Teacher Education* (Aspekty osvojování si anglického jazyka u českých studentů vstupujících do přípravného vzdělávání učitelů), který byl v letech 2013-2015 finančně podpořen Grantovou agenturou České republiky (GA ČR 13-25982S).

Čtrnáct studií vybraných do této knihy vychází z příspěvků, které byly předneseny na mezinárodní konferenci pořádané katedrou anglistiky a amerikanistiky Fakulty filozofické Univerzity Pardubice 4. června 2015.

Organizačním vodítkem knihy se stal rámec pro osvojování si cizích jazyků dle Roda Ellise. Ve shodě se strukturou tohoto rámce byly jednotlivé sekce nazvány takto: (1) *Analýza žákovských korpusů*; (2) *Využívání cizojazyčných korpusů k učení se cizímu jazyku*; (3) *Individuální rozdíly mezi jedinci učícími se cizí jazyky*; (4) *Interakce ve třídě a osvojování si cizího jazyka*. Jednotlivé kapitoly skládají dohromady mozaiku reprezentující celou škálu odlišných přístupů ke zkoumání procesů osvojování si angličtiny jako cizího jazyka a různorodé využití jazykových korpusů za účelem lingvistického výzkumu a implementace jeho výsledků do výuky anglického jazyka.

V první sekci nazvané *Analýza žákovských korpusů* se tři autorky zabývají rozbořením mluveného korpusu Czech Students' Spoken English (Mluvená angličtina českých studentů) v rozsahu cca 150 000 slov.

Nejprve Libuše Hornová zkoumá korpus z hlediska lexikální struktury jmenných frází. Dochází k závěru, že v premodifikaci jmenných frází jsou čeští studenti schopni používat jak podstatná jména, tak gerundia v dostatečné míře, že je v mluveném projevu zastoupena většina typů premodifikací a jejich složitost je na poměrně dobré

úrovni. Problémem je naopak malá různorodost využívané slovní zásoby a nízká míra využívání synonym.

V následné studii Jaroslava Ivanová uvádí index obtížnosti chyb ve výslovnosti šesti hlásek s tím, že nejproblematictější je pro české studenty výslovnost velární nazály /ŋ/ na konci slova, realizace znělých souhlásek /d, g/ na konci slova, výslovnost otevřené přední samohlásky /æ/ a redukované středové samohlásky /ə/. Kromě toho zjistila, že většina českých studentů měla méně obtíží při výslovnosti sledovaných hlásek v případě četby izolovaných slov než při četbě souvislého textu a nejvíce obtíží ve spontánním mluvním projevu.

Ve svém výzkumu Šárka Ježková rozebírá funkční aspekty plynulosti promluvy, zejména prostředky váhání. Výsledky ukazují, že v procesu kompenzace nedostatků v jazykové produkci čeští studenti nadužívali extralingvistických zvuků řeči nebo „hluchých“ pauz a daleko méně používali opakování jednotlivých slov, víceslovných pojmenování a ustálených větných spojení, která jsou používána jako vsuvky roditelými mluvčími. Navíc rejstřík diskurzivních markerů používaných studenty je velice omezený a do značné míry ovlivněn mateřským jazykem.

Rozborem psaného projevu se zabývají následující dvě studie. Petra Huschová and Irena Reimannová se soustředí na analýzu praktické části bakalářských prací z hlediska používání modálních sloves jako jazykových prostředků atenuace (tzv. hedges) a na jejich schopnost identifikace vysokoškolskými studenty. Analýzou akademického diskurzu (79 000 slov) autorky zjistily, že ačkoliv dotazovaní studenti měli adekvátní deklarativní znalosti různých jazykových prostředků atenuace, v cizojazyčném psaném projevu převládala tendence přeceňovat úlohu modálních sloves.

V poslední studii první sekce se Silvie Válková a Jana Kořínková zabývají analýzou gramatické a lexikální komplexnosti psaného diskurzu. Z provedeného komparativního rozboru korpusu formálních a semi-formálních textů českých vysokoškolských studentů a kontrolního korpusu představujícího adekvátní písemné projevy roditelých mluvčích vyplynuly následující skutečnosti. Ve srovnání s roditelými mluvčími angličtiny čeští studenti produkovali delší věty, ale s jednodušší stavbou, používali více podřadných a souřadných souvětí, častěji aplikovali spojovací prostředky, jejichž repertoár byl ale značně omezený, preferovali finitní slovesné tvary před nefinitními a méně často tvořili komplexní jmenné fráze.

V druhé sekci nesoucí název *Využívání cizojazyčných korpusů k učení se cizímu jazyku* nalezneme čtyři studie, které kromě popisu používaných korpusů navrhují možnosti integrace výsledků korpusové analýzy či samotné analýzy do výuky angličtiny.

V úvodní studii Christoph Haase představuje, jak specializované odborně zaměřené texty roditelých pisatelů mohou být srovnávány s tematicky paralelními texty populárně-naučnými. Autor navrhuje, jak je možné ve vysokoškolské výuce zjišťovat reálný výskyt konkrétních vyjadřovacích prostředků ve stylu odborném a populárně-naučném pro oblast lexika, vyjádření subjektivity a objektivity prostřednictvím

analýzy psaného korpusu SPACE (the Scientific and Popular Academic Corpus of English), který vytvořil.

V další studii Róberta Boháta, Niny Horákové a Beaty Rödingové je čtenářům představen jiný typ korpusu vybudovaný autory a nazvaný COHAT (Building a Corpus of High School Academic Texts). Jedná se o soubor autentických textů, psaných i mluvených, pocházejících od rodilých i nerodilých mluvčích angličtiny o rozsahu cca 150 000 slov. Autoři zdůrazňují aplikační složku vybudovaného korpusu ve výuce angličtiny, což dokládají několika vybranými aktivitami studentů.

V předposlední studii této sekce John McKenny představuje korpus Buidcorp (British University in Dubai Corpus), který obsahuje autentické znění diplomových prací v angličtině o rozsahu 2,5 miliónu slov napsaných nerodilými pisateli angličtiny. Ukazuje několik možností, jak mohou vysokoškolští studenti pomocí různých typů softwaru zkoumat aspekty mnoha jazykových prostředků v porovnání se psaným referenčním korpusem rodilých pisatelů – BAWE (the British Academic Written English).

V závěrečné studii této sekce Maria Luisa Roca-Varela využívá španělskou část existujícího korpusu ICLE (the International Corpus of Learner English), aby v něm vyhledala a vyhodnotila jazykový materiál týkající se tzv. zrádných slov, což jsou vždy dvě slova ve dvou různých jazycích, která se podobají v písemné či zvukové podobě a která se liší významem a svým užitím. Rozborem korpusu autorka identifikovala nejčastější lexikální chyby španělských studentů v angličtině, což zdůvodnila zdánlivou homografií problémových dvojic slov. Zároveň také navrhla učební aktivitu podporující správné osvojování slovní zásoby.

Třetí sekce s názvem *Individuální rozdíly mezi jedinci učícími se cizí jazyky* obsahuje dvě studie.

V první z nich Gabriela Lojová nastínila přehled vybraných vnitřních faktorů, které mohou ovlivňovat efektivitu učení se cizímu jazyku a které následně přispívají k implementaci principů humanistického pojetí vzdělávání. Celá přehledová studie vychází ze současných poznatků aplikované psycholingvistiky a klade značný důraz nejen na efektivitu učení se cizímu jazyku ve středoevropském kontextu, ale také na roli učitele, který má na zřeteli jak vnitřní faktory ovlivňující žáky, tak individuální rozdíly mezi nimi.

Monika Černá se ve své studii zabývá dvěma aspekty motivace v procesech učení se cizímu jazyku, a to její dynamickou povahou a proměnlivostí v čase. Usiluje o zachycení změn v motivaci učit se angličtinu v případě tří studentů, a to v dlouhodobém časovém horizontu a v různých kontextech učení. Z analýzy jednotlivých příběhů učení vyplynulo, že každý jedinec prochází různým počtem odlišných motivačně transformačních epizod (např. proces vyspívání, postupně se zvyšující zájem o angličtinu, nová životní fáze, zvlnění externích cílů, demotivace a nové nabytí motivace), které se liší strukturou motivačních zdrojů.

Ve čtvrté sekci s titulem *Interakce ve třídě a osvojování si cizího jazyka* jsou prezentovány tři studie. V teoreticko-metodologické studii František Tůma nastiňuje, jak lze pomocí konverzační analýzy video nahrávek zkoumat interakci ve třídě v kurzu angličtiny na vysoké škole. Prezentovaný výzkumný projekt je zaměřen na aspekty organizace střídání partnerů v komunikaci, jejich sled, eventuální selhání a nastolení nové komunikace, dále pak na organizaci střídání jazykových kódů.

Ve studii Lucie Betákové je v centru zájmu angličtina učitele a také otázka, jak její efektivní užívání ve třídě může přispět k osvojování si angličtiny žáky. Její výzkum na nižším stupni střední školy prokázal, že ve třech hodinách angličtiny kladení otevřených či referenčních otázek ne vždy přispělo k produkci delších a/nebo kvalitnějších odpovědí žáka, pokud ve třídě nepanovala pozitivní atmosféra.

Celá kniha končí studií Petra Dvořáka, který prezentuje výsledky výzkumu vlivu sociálních dovedností středoškolského učitele na interakci, komunikaci a celkový diskurz v hodinách angličtiny. Z výzkumu vyplývá, že studenti vysoce cení učitele s těmito dimenzemi sociálních dovedností: se silným edukačním řízením vzdělávacího procesu a pozitivními edukačními vztahy, což se projevilo ve vyšší míře interakce, tj. studenti častěji verbálně reagovali. Učitelé, kteří kladli méně otázek, často dávali studentům více času na odpovědi.

Ačkoliv výsledky jednotlivých studií nelze zobecnit a každá z nich může mít své slabiny, jejich hlavní přínos spočívá v tom, že ukazují některé cesty, kterými se na začátku 21. století ubírá výzkum zaměřený na deskripci a aplikaci žákovských korpusů anglického jazyka ve výuce angličtiny na středních a vysokých školách a na osvojování si angličtiny jako cizího jazyka.

## REFERENCES

- Agustín Llach, P. (2005). The relationship of lexical errors and their types to the quality of ESL compositions: an empirical study. *Porta Linguarum*, 3, 45–57.
- Aijmer, K. (2005). Evaluation and pragmatic markers. In E. Tognini-Bonelli & G. Del Lungo Camiciotti (Eds.). *Strategies in academic discourse* (pp. 83–96). Amsterdam: John Benjamins.
- Aijmer, K. (2009). Corpora and second language acquisition. In K. Aijmer (Ed.). *Corpora and Language Teaching* (pp. 1–10). Amsterdam: John Benjamins.
- Aijmer, K., & Stenström, A. (2005). Approaches to spoken interaction. *Journal of Pragmatics*, 37, 1743–1751.
- Algeo, J. (1995). Having a look at the expanded predicate. In B. Aarts & C. Meyer (Eds.). *The verb in contemporary English: theory and description*. Cambridge: Cambridge University Press.
- Allwright, D., & Bailey, K. (1991). *Focus on the language classroom: an introduction to classroom research for language teachers*. New York: Cambridge University Press.
- Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora*, 4(1), 71–83.
- Alvesson, M., & Karreman, D. (2000). Varieties of discourse. On the study of organizations through discourse analysis. *Human Relations*, 53, 1125–1149.
- Anthony, L. (2014). *AntConc – Version 3.4.3*. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/>
- Atlas.ti7*. (1993 – 2015) Berlin: ATLAS.ti GmbH.
- Auer, P. (1995). The pragmatics of code-switching: A sequential approach. In L. Milroy & P. Muysken (Eds.). *One speaker, two languages: Cross-disciplinary perspectives on code-switching* (pp. 115–135). Cambridge: Cambridge University Press.
- Auer, P. (2014). *Jazyková interakce*. Praha: Nakladatelství Lidové noviny.
- Bachman, L. F. (2004). *Statistical Analysis for Language Assessment*. Cambridge: Cambridge University Press.

- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in Practice*. Oxford: Oxford University Press.
- Baker, P. (2011). Corpus methods in linguistics. In L. Litosseliti (Ed.). *Research Methods in Linguistics*. London: Continuum.
- Bardovi-Harlig, K., & Dörnyei, Z. (1998). Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. *TESOL Quarterly*, 32(2), 233–262. Retrieved from <http://www.jstor.org/stable/3587583>
- Barlow, M. (2002). Corpora, concordancing, and language teaching. In *Proceedings of the 2002 KAMALL International Conference*. Daejeon, Korea.
- Basturkmen, H. (2010). *Developing Courses in English for Specific Purposes*. New York: Palgrave Macmillan.
- Bellack A., Kliebard, H., Hyman, R., & Smith, F. (1966). *The Language of the Classroom*. New York: Teachers' College Press.
- Benson, P. (2004). (Auto)biography and learner diversity. In P. Benson & D. Nunan (Eds.). *Learners' Stories. Difference and Diversity in Language Learning* (pp. 4–21). Cambridge: Cambridge University Press.
- Benson, P., & Nunan, D. (2004). *Learners' Stories. Difference and Diversity in Language Learning*. Cambridge: Cambridge University Press.
- Betáková, L. (2010). *Discourse and Interaction in English Language Teaching*. Praha: Univerzita Karlova v Praze.
- Bialystok, E., & Hakuta, K. (1994). *In Other Words*. New York: Basic Books.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263–286.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Blackmore, S., & Frith, U. (2005). *The learning brain: Lessons for education*. Cambridge, MA: Blackwell Publishing.
- Bohát, R., Horáková, N., & Rödlíngová, B. (2015). *COHAT: Corpus of High School Academic Texts, International School of Prague*. Praha: International School of Prague. Retrieved from <https://sites.google.com/a/isp.cz/isp-corpus/>
- Brown, G., & Wragg, E. C. (1993). *Questioning*. London: Routledge.
- Brown, H. D. (2002). English language teaching in the “post-method” era: Toward better diagnosis, treatment, and assessment. In J. C. Richards

- & W. A. Renandya (Eds.). *Methodology in language teaching: An anthology of current practice* (pp. 9–19). Cambridge: Cambridge University Press.
- Brown, K. (Ed.). (2006). *Encyclopedia of Language and Linguistics* (2nd ed.). Boston, MA: Elsevier.
- Buzzelli, C. A., & Johnston, B. (2002). *The Moral Dimensions of Teaching: Language, power and culture in classroom interaction*. London: Routledge/Falmer.
- Cambridge Dictionaries Online*. (1999). Cambridge: Cambridge University Press. Retrieved from <http://dictionary.cambridge.org/dictionary/english/value>
- Canagarajah, S. A. (2002). *Critical Academic Writing and Multilingual Students*. Ann Arbor, MI: The University of Michigan Press.
- Carroll, J. B. (1965). The prediction of success in intensive foreign language training. In R. Glaser (Ed.). *Training, Research, and Education* (pp. 87–136). USA, Pennsylvania: University of Pittsburgh Press.
- Carroll, J. B., & Sapon, S. M. (1958). *Modern language aptitude test*. New York: The Psychological Corporation.
- Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics*, 16(2), 141–158.
- Carter, R., & McCarthy, M. (1997). *Exploring spoken English*. Cambridge: Cambridge University Press.
- Cazden, C. B. (2001). *Classroom Discourse*. Portsmouth, N.H.: Heinemann.
- Celce-Murcia, M., Brinton, D., & Goodwin, J. (1996). *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Cambridge: Cambridge University Press.
- Čermák, F. (2003). Paremiological minimum of Czech: The corpus evidence. In A. Burger, G. Häcki Buhofer & Greciano, G. (Eds.). *Flut von Texten – Vielvalt der Kulturen. Ascona 2001 zur Methodologie und Kulturspezifik der Phraseologie* (pp. 15–31). Baltmannsweiler: Schneider Hohengehren.
- Černá, M. (2014). Data elicitation through language testing: challenges of test design. In A. Akbarov (Ed.). *Linguistics, Culture and Identity in Foreign Language Education* (pp. 1423–1431). Sarajevo: International Burch University.
- Černá, M., Urbanová, Z., & Vít, M. (2011). Pronunciation awareness of students in English language study programmes. In S. Válková (Ed.). *New Trends in Educating Future Teachers of English VI* (pp. 22–37). Conference proceedings (CD). Olomouc: Pedagogická fakulta Univerzity Palackého.
- Chacón Beltrán, R. (2006). Towards a typological classification of false friends (Spanish-English). *Revista Española de Lingüística Aplicada*, 19, 29–39.
- Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology*, 9(2), 111–125.

- Chamizo, P. (1999). False friends: mousetraps for communication and translation. *International Journal of Communication*, 9, 115–130.
- Chamot, A. U., Barnhardt, S., El-Dinary, P. B., & Robbins, J. (1999). *The Learning Strategies Handbook*. White Plains, N.Y.: Longman.
- Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes*, 25(3), 310–331.
- Chau, M. H. (2012). Learner corpora and second language acquisition. In K. Hyland, M. H. Chau & M. Handford (Eds.). *Corpus applications in applied linguistics* (pp. 191–207). London, Bloomsbury Academic.
- Cheng, W. (2012). *Exploring Corpus Linguistics: Language in Action*. Milton Park, Abingdon: Routledge.
- Choděra, R. (2006). *Didaktika cizích jazyků: úvod do vědního oboru*. Praha: Academia.
- Chráska, M. (1999). *Didaktické testy*. Brno: Paido.
- Chráska, M. (2007). *Metody pedagogického výzkumu*. Praha: Grada.
- Christinson, M. A. (1996). Teaching and learning language through multiple intelligences. *TESOL Journal*, 6(1), 10–14.
- Clemen, G. (1997). The concept of hedging: origins, approaches and definitions. In R. Markkanen & H. Schröder (Eds.). *Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts* (pp. 235–248). Berlin: Walter de Gruyter.
- Cohmetrix version 3.0 idices*. (n.d.). Retrieved from [http://141.225.42.101/CohmetrixHome/documentation\\_indices.html](http://141.225.42.101/CohmetrixHome/documentation_indices.html)
- Comenius, J. A. (1986). *Janua linguarum reserata*. Johannis Amos Comenii Opera Omnia. Volume 15/I. Praha: Academia.
- Comenius, J. A. (1991). *Orbis Sensualium Pictus*. Praha: Trizonia.
- Common European framework of reference for languages: Learning, teaching, assessment*. (2001). Cambridge: Cambridge University Press.
- Cook, G. (1989). *Discourse*. Oxford: Oxford University Press.
- Cook, V. (2010). Prolegomena to second language learning. In P. Seedhouse, S. Walsh & C. Jenks (Eds.). *Conceptualising “learning” in applied linguistics* (pp. 6–22). London: Palgrave Macmillan.
- Cook, V., & Singleton, D. (2014). *Key Topics in Second Language Acquisition*. Bristol: Multilingual Matters.
- Coulthard, M. (1985). *An Introduction to Discourse Analysis*. Harlow: Pearson Education.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research* (2nd ed.). London: SAGE Publications.
- Crompton, P. (1997). Hedging in academic writing: some theoretical problems. *English for Specific Purposes*, 16(4), 271–287.

- Crystal, D. (2010). *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.
- Crystal, D. (2015). Meaning and Style. *Encyclopædia Britannica*. Retrieved from <http://www.britannica.com/EBchecked/topic/329791/language/27170/Meaning-and-style-in-language>
- Cullen, R. (2002). Supportive teacher talk. *ELT Journal*, 52(3), 117–127.
- DeKeyser, R. M. (2012). Age effects in second language learning. In S. M. Gass & A. Mackey (Eds.). *The Routledge handbook of second language acquisition* (pp. 442–460). London: Routledge.
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in second language acquisition*, 22(4), 499–524.
- Denton, P. (2007) *Open-ended questions stretch academic and social learning*. Retrieved from [http://www.educationworld.com/a\\_curr/responsiveclassroom/responsiveclassroom014.shtml](http://www.educationworld.com/a_curr/responsiveclassroom/responsiveclassroom014.shtml)
- Derewianka, B. (2003). Trends and issues in genre-based approaches. *RELC Journal* 34(2), 133–154.
- Dočkal, V., & Schranzová, D. (1992). K prognostickej hodnote prognosticko-diagnostického testu cudzojazyčných schopností. *Psychológia a patopsychológia dieťaťa*, 27(1), 61–75.
- Dörnyei, Z. (2001). *Motivational Strategies in the Language Classroom*. Cambridge: Cambridge University Press.
- Dörnyei, Z. (2005). *The psychology of the language learner*. London: Lawrence Erlbaum Associates.
- Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.
- Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.
- Dörnyei, Z., & Ottó, I. (1998). Motivation in action: A process model of L2 motivation. *Working Papers in Applied Linguistics* (Vol. 4) (pp. 43–69). London: Thames Valley University.
- Doughty, C. J., & Long, M. H. (Eds.) (2003). *The Handbook of Second Language Acquisition*. London: Blackwell Publishing.
- Drew, P., & Heritage, J. (1992). Analyzing talk at work: An introduction. In P. Drew & J. Heritage (Eds.). *Talk at work. Interaction in institutional settings* (pp. 3–65). Cambridge: Cambridge University Press.
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1), 61–74.
- Durik, A. M., Britt, M. A., Reynolds, R., & Storey, J. (2008). The effects of hedges in persuasive arguments: A nuanced analysis of language. *Journal of Language and Social Psychology*, 27(3), 217–234.

- Dvořák, P. (2014). Sociální dovednosti učitelů angličtiny v edukačních interakcích na středních školách (Doctoral dissertation). Available from Filozofická fakulta, Univerzita Karlova Praha.
- Ehrman, M. E. (1996). *Understanding Second Language Learning Difficulties*. Thousand Oaks, CA: SAGE Publications.
- Ehrman, M. E., Leaver B. L., & Oxford R. (2003). A brief overview of individual differences in second language learning. *System*, 31, 313–330.
- Elliott, J. (2005). *Using Narrative in Social Research*. London: SAGE Publications.
- Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27, 305–352. Retrieved from DOI: 10.1017/S027226310505014X
- Ellis, N. (2008). Implicit and explicit knowledge about language. In J. Cenoz & N. H. Hornberger (Eds.). *Encyclopedia of Language and Education*, 2nd Edition, Volume 6: Knowledge about Language, 1–13. Springer Science+Business Media LLC. Retrieved from [http://www-personal.umich.edu/~ncellis/LanguageLearningLab/Publications\\_files/Implicit%20and%20explicit%20knowledge%20about%20language.pdf](http://www-personal.umich.edu/~ncellis/LanguageLearningLab/Publications_files/Implicit%20and%20explicit%20knowledge%20about%20language.pdf)
- Ellis, N., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics – introduction to the special issue. *Applied linguistics*, 27(4), 559–589.
- Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Ellis, R. (1997). *Second Language Acquisition*. Oxford: Oxford University Press.
- Ellis, R. (2008). *The Study of Second Language Acquisition* (2nd ed.). Oxford: Oxford University Press.
- Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis et al. *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching. Second Language Acquisition* (pp. 3–25). Bristol: Multilingual Matters.
- Ellis, R. (2015). *Understanding Second Language Acquisition* (2nd ed.). Oxford: Oxford University Press.
- Ellis, R., & Barkhuizen, G. (2005). *Analysing Learner Language*. Oxford: Oxford University Press.
- Estrada, I. (n.d.). In *Quotes for teachers*. Retrieved from <http://quotes-for-teachers.blogspot.cz/2000/06/learning-education-and-school.html>
- European Commission. (2002). *Barcelona European Council Presidency Conclusions*. Retrieved from [http://europa.eu/rapid/press-release\\_PRES-02-930\\_en.htm](http://europa.eu/rapid/press-release_PRES-02-930_en.htm)
- European Commission. (2011). *A Policy Handbook: Language Learning at Pre-primary School Level: Making it Efficient and Sustainable*. Retrieved

- from [http://ec.europa.eu/languages/policy/language-policy/documents/early-language-learning-handbook\\_en.pdf](http://ec.europa.eu/languages/policy/language-policy/documents/early-language-learning-handbook_en.pdf)
- Farkašová, E. (1998). *Psychologické aspekty pri vyučovaní cudzích jazykov od prvého ročníka základnej školy*. Bratislava: Univerzita Komenského, Filozofická fakulta.
- Farkašová, E. (2002). *Aspekty vyučovania cudzích jazykov u najmladších žiakov*. Trnava: Pedagogická fakulta Trnavskej university.
- Farr, F. (2010). How can corpora be used in teacher education? In A. O’Keeffe & M. McCarthy (Eds.). *The Routledge Handbook of Corpus Linguistics* (pp. 620–632). London: Routledge.
- Firth, A., & Wagner, J. (2007). Second/foreign language learning as a social accomplishment: Elaborations on a reconceptualized SLA. *The Modern Language Journal*, 91, 800–819.
- Fraser, B. (2010). Pragmatic competence: The case of hedging. In G. Kaltenböck, W. Mihatsch & S. Schneider (Eds.). *New Approaches to Hedging* (pp. 15–34). Bingley: Emerald Group Publishing Limited.
- Gardner, H. (1993). *Multiple Intelligences: The Theory in Practice*. New York: Basic Books.
- Gardner, H. (1999). *Intelligence reframed. Multiple intelligences for the 21st century*. New York: Basic Books.
- Gass, S., & Selinker, L. (2008). *Second Language Acquisition: An Introductory Course*. London: Routledge.
- Gee, J. P. (2005). *An Introduction to Discourse Analysis. Theory and method*. London: Routledge.
- Gillernová, I. (2009). Edukační interakce rodičů a dětí a její proměny v reflexi dospívajících. *Československá psychologie*, 53(3), 209–223.
- Gillernová, I. (2013) Edukační styly učitelů. *Řízení školy*, 10, 30–36.
- Gillernová, I., & Krejčová, L. (2012). *Sociální dovednosti ve škole*. Praha: Grada.
- Goodwin, C., & Heritage, J. (1990). Conversation analysis. *Annual Review of Anthropology*, 19, 283–307.
- Graesser, A. C. et al. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behaviour Research Methods, Instruments, & Computers*, 36(2), 193–202.
- Granger, S. (2002). A bird’s eye view of learner corpus research. In S. Granger, J. Hung & E. Petch-Tyson (Eds.). *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3–33). Amsterdam & Philadelphia: John Benjamins.
- Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching. A critical evaluation. In K. Aijmer (Ed.). *Corpora and Language Teaching*. (pp. 13–32). Amsterdam: John Benjamins.
- Green, J., & Harker, J. (Eds.). (1988). *Multiple Perspective – Analysis of Classroom Discourse*. Norwood, N.J.: Ablex Publishing.

- Gumperz, J. J. (1977). Sociocultural knowledge in conversational inference. *Linguistics and anthropology*, 76, 785–798.
- Haase, C. (2008). Clause linking in specialized and popular academic English: An investigation into scientific texts from the SPACE corpus. In B. Lewandowska-Tomaszczyk (Ed.). *Corpus Linguistics, Computer Tools, and Applications – State of the Art* (pp. 481–496). Frankfurt: Peter Lang.
- Haase, C. (2013). Tools for identifying and teaching semantic complexity in academic writing. In Tafazoli, D. (Ed.). *Language & Technology: Computer Assisted Language Teaching* (pp.129–137). Tehran: Khate Sefid.
- Haase, C. (2014). A register approach to analogy in science texts: Popular vs. specialized text types. *Discourse and Interaction*, 7(1), 33–48.
- Haines, S., & Stewart, B. (2002). *Landmark Advanced: Student's Book*. Oxford: Oxford University Press.
- Hall, J. K., & Walsh, M. (2002). Teacher-student interaction and language learning. *Annual Review of Applied Linguistics*, 22, 186–203. Retrieved from <http://doi.org/10.1017/S0267190502000107>
- Hall, J., K., & Verplaetse, L. S. (2000). The development of second and foreign language learning through classroom interaction. In J. K. Hall & L. S. Verplaetse (Eds.). *Second and foreign language learning through classroom interaction* (pp. 1–22). Mahwah, N.J.: Erlbaum.
- Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education*, 5(2), 93–116.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Harleyová, B., & Wangová, W. (2004). Hypotéza kritického obdobia: kde sme teraz? In J. Štefánik (Ed.). *Antológia bilingvizmu*. Bratislava: Academic Electronic Press.
- Harmer, J. (2007). *The practice of English language teaching*. Harlow: Longman.
- Hartl, P., & Hartlová, H. (2000). *Psychologický slovník*. Praha: Portál.
- Hatch, E. M. (1992). *Discourse and language education*. New York: Cambridge University Press.
- Hatch, E. M. (2001). *Discourse and Language Education*. Cambridge: Cambridge University Press.
- Hatch, E. M., & Long, M. H. (1980). Discourse analysis, what's that? In D. Larsen-Freeman (Ed.). *Discourse Analysis in Second Language Research* (pp. 1–40) Rowley, MA: Newbury House Publishers.
- Helus, Z. (2007). *Sociální psychologie pro pedagogy*. Praha: Grada.
- Hendl, J. (2012). *Kvalitativní výzkum*. Praha: Portál.
- Hendrich, J. (Ed.). (1988). *Didaktika cizích jazyků*. Praha: SPN.
- Hinkel, E. (Ed.). (2005). *Handbook of research in second language learning*. Mahwah, N.J.: Lawrence Erlbaum.

- Holub, J. et al. (2008). *Čeština jako cizí jazyk. Úroveň B2*. Praha: Univerzita Karlova, Ústav bohemistických studií. Retrieved from <http://aplikace.msmt.cz/MezinarodniSpoluprace/Dokumenty/cestinaB2.pdf>
- Hornová, L. (2015). Syntactic functions of non-finite verb forms in a learner corpus of Czech students. In M. Adam & R. Vogel (Eds.). *Communication across Genres and Discourses* (Sixth Brno Conference on Linguistics Studies in English, Brno 11–12 September 2014) Conference proceedings. Brno: Masarykova univerzita.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24–44.
- Huddleston, R., & Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Hughes, A., & Lascaratou, C. (1982). Competing criteria for error gravity. *ELT Journal*, 36(3), 175–182.
- Hummel, K. M. (2014). *Introducing Second Language Acquisition. Perspectives and Practices*. Chichester: John Wiley & Sons, Inc.
- Hutchby, I., & Wooffitt, R. (1998). *Conversation analysis*. Cambridge: Polity Press, Blackwell Publishers.
- Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for Specific Purposes*, 13(3), 239–256.
- Hyland, K. (1995). The author in the text: Hedging scientific writing. *Hong Kong Papers in Linguistics and Language Teaching*, 18, 33–42.
- Hyland, K. (1999). *Hedging in Scientific Research Articles*. Amsterdam: John Benjamins.
- Hymes, D. H. (1979). On communicative competence. In Ch. Brumfit & K. Johnson (Eds.). *The communicative approach to language teaching* (pp. 5–26). Oxford: Oxford University Press.
- InterCorp, Parallel Corpus – Release 6*. (2013). Praha: Czech National Corpus. Retrieved from <http://ucnk.ff.cuni.cz/intercorp/?req=doc%3Auvod>
- Isik-Tas, E. E. (2008). A corpus-based analysis of genre-specific discourse of research: the PhD thesis and the research article in ELT (Doctoral dissertation). Available from Middle East Technical University.
- Jensen, E. (1996). *Brain-based learning*. Del Mar, CA: Turning Point Publishing.
- Ježková, Š. (2012). Syntactic deviations and mistakes in learners' spoken English language. In R. Trušník, K. Nemčoková & G. J. Bell (Eds.). *Theories and Practices – Zlín Proceedings in Humanities, Volume 3* (pp. 223–232). Zlín: Tomáš Baťa University in Zlín.
- Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.). *Perspectives on Pedagogical Grammar* (pp. 293–313). Cambridge: Cambridge University Press.

- Johnson, K. E. (1995). *Understanding Communication in Second Language Classrooms*. Cambridge: Cambridge University Press.
- Jones, P. (2007). Talk about literacy in the content areas. In J. Polias & B. Dare (Eds.). *Teaching ESL students in mainstream classrooms: language in learning across the curriculum*. The State of South Australia: DECS Publishing.
- Jucker, A. H., Smith, S. W., & Ludge, T. (2003). Interactive aspects of vagueness in conversation. *Journal of Pragmatics*, 35, 1737–1769.
- Kasper, G., & Blum-Kulka, S. (1993). Interlanguage pragmatics: An introduction. In G. Kasper & S. Blum-Kulka (Eds.). *Interlanguage Pragmatics* (pp. 3–17). Oxford: Oxford University Press.
- Kaweera, C. (2013). Writing error: a review of interlingual and intralingual interference in EFL context. *English Language Teaching*, 6(7), 9–18.
- Keefe, R. (2003). *Theories of vagueness*. Cambridge: Cambridge University Press.
- Keizer, E. (2007). *The English Noun Phrase. The Nature of Linguistic Categorization*. Cambridge: Cambridge University Press.
- Kennedy, G. (2014). *An Introduction to Corpus Linguistics*. London: Routledge.
- Komenský, J.A. (2003). *Dvěře jazyků odevřené*. Praha & Litomyšl: Ladislav Horáček – Paseka.
- Kořínková, J., & Válková, S. (2013). Syntactic complexity in advanced students' writing. *Journal of Interdisciplinary Philology*, 4, 45–52.
- Kotásek, J. (2002). Modely školy budoucnosti. In E. Walterová et al. (Eds.). *Rozvoj národní vzdělanosti a vzdělávání učitelů v evropském kontextu* (pp. 8–24). Praha: Univerzita Karlova, Pedagogická fakulta.
- Kozulin, A. (1998). *Psychological tools: a sociocultural approach to education*. Cambridge, MA: Harvard University Press.
- Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press. Retrieved from [http://www.sdkrashen.com/content/books/sl\\_acquisition\\_and\\_learning.pdf](http://www.sdkrashen.com/content/books/sl_acquisition_and_learning.pdf)
- Krashen, S. (1982). *Principles and Practice in Second language Acquisition*. Oxford: Pergamon Institute of English.
- Krashen, S. D. (1989). *Language acquisition and language education*. London: Prentice Hall International.
- Krashen, S. D., Scarcella, R. C., & Long, M. H. (1982). *Child – adult differences in second language acquisition*. Rowley, MA: Newbury House Publishers, Inc.
- Krejčová, L. (2011). *Psychologické aspekty vzdělávání dospívajících*. Praha: Grada.
- Kwan, B. S. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes*, 25(1), 30–55.
- Kyriacou, Ch. (2007). *Essential teaching skills*. Cheltenham: Nelson Thornes.
- Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2, 458–508.

- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.
- Lee, J. F., & VanPatten, B. (2003). *Making communicative language teaching happen*. Boston, MA: McGraw-Hill.
- Lemke, J. L. (1985). *Using Language in the Classroom*. London: Oxford University Press.
- Lengeling, M. (1995). True friends and false friends. *MEXTESOL Journal*, 19(2), 17–22. Retrieved from [http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/14/b6/d1.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/14/b6/d1.pdf)
- Lenneberg, E. (1967). *Biological Foundation of Language*. New York: Wiley.
- Lewis, M. (1993). *The Lexical Approach: The State of ELT and the Way Forward*. Hove: Language Teaching Publications.
- Linell, P. (1998). *Approaching dialogue. Talk, interaction and contexts in dialogical perspectives*. Amsterdam: John Benjamins.
- Linell, P. (2009). *Rethinking language, mind and world dialogically. Interactional and contextual theories of human sense-making*. Charlotte, N.C.: Information Age Publishing.
- Littlewood, W. (1981). *Communicative language teaching: an introduction*. New York: Cambridge University Press.
- Littlewood, W. (1984). *Foreign and second language learning: language-acquisition research and its implications for the classroom*. New York: Cambridge University Press.
- Lojová, G. (2005/2006a). Vývinové osobitosti a výučba cudzieho jazyka v materskej škole (I). *Cizí jazyky*, 49(4), 134–135.
- Lojová, G. (2005/2006b). Vývinové osobitosti a výučba cudzieho jazyka v materskej škole (II). *Cizí jazyky*, 49(5), 174–175.
- Lojová, G. (2010). Vytváranie pevných základov pre cudzojazyčné zručnosti. In Z. Straková & I. Cimmermanová (Eds.). *Učiteľ cudzieho jazyka v kontexte primárneho vzdelávania* (pp. 6–21). Prešov: Prešovská univerzita.
- Lojová, G. (2011a). *Teória a prax vyučovania gramatiky cudzích jazykov*. Bratislava: Z-F Lingua. Retrieved from [http://www.fedu.uniba.sk/fileadmin/user\\_upload/editors/KAJL/publikacie/Lojova\\_-\\_Teoria\\_a\\_prax\\_vyuc\\_gramat\\_2011.pdf](http://www.fedu.uniba.sk/fileadmin/user_upload/editors/KAJL/publikacie/Lojova_-_Teoria_a_prax_vyuc_gramat_2011.pdf)
- Lojová, G. (2011b). *Individuálne osobitosti pri učení sa cudzích jazykov I*. Bratislava: Univerzita Komenského Bratislava.
- Lojová, G. (2012/2013). Cudzí jazyky v materskej škole. *Cizí jazyky*, 56(4), 9–13.
- Lojová, G. (2015). Focus on internal factors in foreign language learning. *Keynote speech presented at English Acquisition and Corpora Building international conference*, June 2015, Pardubice, Czech Republic.
- Lojová, G., & Kostelníková, M. (2012). Current developments in foreign language teaching in Slovakia. In *Studies in Foreign Language Education* (Vol. 4) (pp.155–173). Köln: Kirsch-Verlag.

- Lojová, G., & Straková, Z. (2012). *Teoretické východiská vyučovania angličtiny v primárnom vzdelávaní*. Bratislava: Univerzita Komenského.
- Lojová, G., & Vlčková, K. (2011). *Styly a strategie učení ve výuce cizích jazyků*. Praha: Portál.
- Lojová, G., Farkašová, E., & Pčolinská, A. (2011). *Využívanie personalizácie vo vyučovaní anglického jazyka v primárnom vzdelávaní*. Bratislava: Z-F Lingua.
- Long, M. H., & Sato, C. J. (1983). Classroom foreigner talk discourse: forms and functions of teachers' questions. In H. W. Seliger & M. H. Long (Eds.). *Classroom Oriented Research in Second Language Acquisition* (pp. 268–285). Rowley, MA: Newbury House Publishers.
- Long, M. H., & Doughty, C. J. (2009). *The handbook of language teaching*. Oxford: Wiley-Blackwell.
- Luoma, S. (2004). *Assessing Speaking*. Cambridge: Cambridge University Press.
- Lynch, T. (1996). *Communication in the language classroom*. Oxford: Oxford University Press.
- Mackey, A. (2007). The role of conversational interaction in second language acquisition. In A. Mackey (Ed.). *Conversational interaction in second language acquisition: a series of empirical studies* (pp. 1–28). New York: Oxford University Press.
- Malamah-Thomas, A. (1987). *Classroom Interaction*. Oxford: Oxford University Press.
- Malíková, M. (1989). *Prognosticko-diagnostický test cudzojazyčných schopností*. Bratislava: Psychodiagnostika.
- Mareš, J. (1998). *Styly učení žáků a studentů*. Praha: Portál.
- Mareš, J. (2009). Posledních dvacet let výzkumu pedagogické interakce a komunikace v České republice (léta 1990–2009). In *Pedagogická komunikace v didaktických, sociálních a filozofických souvislostech. Sborník příspěvků z konference* (pp. 9–29). Hradec Králové: Gaudeamus.
- Marinova-Todd S. H., Bradfor, M. D., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly*, 34(1), 9–34.
- Markee, N. (2000). *Conversation analysis*. Mahwah N. J.: Lawrence Erlbaum Associates.
- Markee, N. (2007). Toward a learning behavior tracking methodology for CA-for-SLA. *Applied Linguistics*, 29(3), 404–427. Retrieved from <http://doi.org/10.1093/applin/amm052>
- Markkanen, R., & Schröder, H. (1997). Hedging: A challenge for pragmatics and discourse analysis. In R. Markkanen & H. Schröder (Eds.). *Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts* (pp. 3–20). Berlin: Walter de Gruyter.
- Marková, I. (1982). *Paradimms, thought, and language*. Chichester: John Wiley & Sons.

- Martin-Jones, M. (1995). Code-switching in the classroom: Two decades of research. In L. Milroy & P. Muysken (Eds.). *One speaker, two languages: Cross-disciplinary perspectives on code-switching* (pp. 90–111). Cambridge: Cambridge University Press.
- Martin-Jones, M. (2000). Bilingual classroom interaction: A review of recent research. *Language Teaching*, 33(01), 1–9. Retrieved from <http://doi.org/10.1017/S0261444800015123>
- McCarthy, M. (1991). *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press.
- McCarthy, M., & Carter, R. (1994). *Language as Discourse: Perspectives for Language Teaching*. London: Longman.
- McCarthy, M., & Carter, R. (2001). Ten criteria for a spoken grammar. In E. Hinkel & S. Fotos (Eds.). *New Perspectives on Grammar Teaching in Second Language Classrooms* (pp. 51–75). Mahwah, N.J.: L. Erlbaum.
- McEnery, T., & Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Mehan, H. (1979). *Learning Lessons*. Cambridge, MA: Harvard University Press.
- Mehan, H. (1985). The structure of classroom discourse. In T. Van Dijk (Ed.). *Handbook of Discourse Analysis, Volume 3: Discourse and Dialogue* (pp. 120–132). London: Academia Press.
- Mercer, N. (2010). The analysis of classroom talk: Methods and methodologies. *British Journal of Educational Psychology*, 80(1), 1–14.
- Mertin, V., & Krejčová, L. et al. (2012) *Metody a postupy poznávání žáka*. Praha: Wolters Kluwer.
- Mishler, E. G. (1996). *Research Interviewing: Context and Narrative*. Cambridge, MA: Harvard University Press.
- Mitchell, R. F. (2009). Current trends in classroom research. In M. H. Long & C. J. Doughty (Eds.). *The handbook of language teaching* (pp. 675–705). Chichester: Blackwell Publishing Ltd.
- Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne, Jr. (Eds.). *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Mahwah, NJ: Lawrence Erlbaum Associates.
- Moon, J., & Nikolov, M. (Eds.). (2000). *Research into teaching English to young learners*. Pécs: University Press.
- Mori, J., & Zuengler, J. (2008). Conversation analysis and talk-in-interaction in classrooms. In M. Martin-Jones, A. M. de Mejía, & N. H. Hornberger (Eds.). *Encyclopedia of language and education. Volume 3: Discourse and education* (2nd ed.) (pp. 15–26). New York: Springer.

- Moskowitz, G. (1978). *Caring and sharing in the foreign language class: a sourcebook on humanistic techniques*. Boston, MA: Heinle.
- Mothežiková, J. (1988). *Methodology for TEFL teachers*. Praha: Státní pedagogické nakladatelství.
- Muñoz, C. (2008). Symetries and asymetries of age effects in naturalistic and instructed L2 learning. *Applied linguistics*, 29(4), 578–596.
- Nádraská, Z. (2013). A pre-diagnostic stage of a research project: selected features of university students' pronunciation of English. *Paper presented at the Fifth Annual International Conference on Anglophone Studies*, September 2013. Zlín, Czech Republic.
- Nikolov, M., & Djigunovic, J. (2006). Recent research on age, second language acquisition, and early foreign language learning. *Annual Review of Applied Linguistics*, 26, 234–260.
- Nunan, D. (1989). *Understanding Language Classrooms*. London: Prentice Hall.
- Nunan, D. (2005). *Handbook of research in second language teaching and learning*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- O'Connell, D. C., & Kowal, S. (2003). Psycholinguistics: A half century of monologism. *The American Journal of Psychology*, 116(2), 191–212. Retrieved from <http://doi.org/10.2307/1423577>
- O'Malley, J. M., & Chamot, A. U. (1990). *Learning Strategies in Second Language Acquisition*. Cambridge: Cambridge University Press.
- Olsen, A. E. (2010). *Academic vocabulary, academic words*. Harlow: Pearson Longman.
- Ortega, L. (2009). *Understanding Second Language Acquisition*. Abingdon: Routledge.
- Oxford, R. L. (1990). *Language Learning Strategies. What Every Teacher Should Know*. Boston, MA: Heinle & Heinle Publishers.
- Paré, A., Starke-Meyerring, D., & McAlpine, L. (2009). The dissertation as multi-genre: many readers, many readings. In C. Bazerman, A. Bonini & D. Figueiredo (Eds.). *Genre in a changing world* (pp. 179–193). Fort Collins, Colorado: The WAC Clearinghouse and Parlor Press.
- Pastorková, J. (2013). Raising English language learners' metacognitive awareness: Focus on learning styles (MA dissertation). Available from Comenius University, Faculty of Education, Bratislava.
- Pelikán, J. (2011). *Základy empirického výzkumu pedagogických jevů*. Praha: Karolinum.
- Phillips, M. (1985). *Aspects of Text Structure: An Investigation of the Lexical Organization of Text*. Amsterdam: North-Holland.
- Pišová, M. (2010). Učitel – expert. Přehled výzkumných trendů a jejich výsledků. *Pedagogika*, 60(3–4), 47–58.
- Podrápská, K. (2011). Celoživotní osvojování cizích jazyků na příkladu výuky u dospělých. In V. Janíková (Ed.). *Výuka cizích jazyků* (pp. 160–177). Praha: Grada.

- Popper, K., & Eccles, J. C. (1984). *The Self and its Brain: An Argument for Interactionism*. London: Taylor & Francis.
- Rampton, B., Roberts, C., Leung, C., & Harris, R. (2002). Methodology in the analysis of classroom discourse. *Applied Linguistics*, 23(3), 373–392. Retrieved from <http://doi.org/10.1093/applin/23.3.373>
- Rayson, P. (2003). Matrix: a statistical method and software tool for linguistic analysis through corpus comparison (Doctoral dissertation). Available from Lancaster University.
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519–549.
- Rayson, P. (2009). *Wmatrix: a web-based corpus processing environment*. Lancaster: Lancaster University. Retrieved from <http://ucrel.lancs.ac.uk/wmatrix/>
- Reid, J. M. (1998). *Understanding Learning Styles in the Second Language Classroom*. Upper Saddle, N.J.: Prentice Hall Regents.
- Richard-Amato, P. A. (2010). *Making it happen: from interactive to participatory language teaching: evolving theory and practice*. White Plains, N.Y.: Pearson Education.
- Richards, J. C., & Smith, R. W. (Eds.). (1974). *Error analysis. Perspectives on Second Language Acquisition*. Harlow: Longman Group UK Limited.
- Richards, J. C., & Smith, R. W. (Eds.). (1985). *Longman Dictionary of Applied Linguistics* (3rd ed.). Harlow: Longman.
- Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Clevedon, Buffalo, Toronto: Multilingual Matters.
- Rinvolveri, M. (2005). *Multiple intelligences in EFL*. London: Helbling Languages.
- Roach, P. (2009). *English Phonetics and Phonology: A Practical Course* (4th ed.). Cambridge: Cambridge University Press.
- Römer, U., & Swales, J. (2010). The Michigan Corpus of Upper-level Student Papers (MICUSP). *Journal of English for Academic Purposes*, 9, 249.
- Rommetveit, R. (1988). On human beings, computers, and representational-computational versus hermeneutic-dialogical approaches to human cognition and communication. In H. Sinding-Larsen (Ed.). *Artificial intelligence and language. Old questions in a new key* (pp. 47–76). Oslo: Tano.
- Ruiz-Garrido, M. F., Palmer-Silveira, J. C. & Fortanet-Gómez, I. (Eds.). (2010). *English for Professional and Academic Purposes*. Amsterdam: Rodopi.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735. Retrieved from <http://doi.org/10.2307/412243>
- Sage, R. (2000). *Class Talk: Effective Classroom Communication*. Stafford: Network Educational Press.
- Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes*, 13(2), 149–171.

- Schegloff, E. A. (1987). Between micro and macro: Contexts and other connections. In J. C. Alexander, B. Giesen, R. Münch & N. J. Smelser (Eds.). *The micro-macro link* (pp. 207–234). Los Angeles: University of California Press.
- Schegloff, E. A. (1991). Conversation analysis and socially shared cognition. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.). *Perspectives on socially shared cognition* (pp. 150–171). Washington, D. C.: American Psychological Association.
- Schegloff, E. A. (1992). In another context. In A. Duranti & C. Goodwin (Eds.). *Rethinking context: Language as an interactive phenomenon* (pp. 191–227). Cambridge: Cambridge University Press.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis I*. Cambridge: Cambridge University Press.
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53(2), 361–382. Retrieved from <http://doi.org/10.2307/413107>
- Schegloff, E. A., Koshik, I., Jacoby, S., & Olsher, D. (2002). Conversation analysis and applied linguistics. *Annual Review of Applied Linguistics*, 22, 3–31. Retrieved from <http://doi.org/10.1017/S0267190502000016>
- Schiffrin, D. (1994). *Approaches to Discourse*. Malden, MA: Blackwell Publishing.
- Schmidt, R. (1993). Consciousness, learning and interlanguage pragmatics. In G. Kasper & S. Blum-Kulka (Eds.). *Interlanguage Pragmatics* (pp. 21–42). Oxford: Oxford University Press.
- Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11, 11–26. Retrieved from <http://nflrc.hawaii.edu/PDFs/SCHMIDT%20Deconstructing%20consciousness%20in%20search%20of%20useful%20definitions.pdf>
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.). *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.
- Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan & I. Walker (Eds.). *Proceedings of CLaSIC 2010* (pp. 721–737). Singapore: National University of Singapore. Retrieved from <http://nflrc.hawaii.edu/PDFs/SCHMIDT%20Attention,%20awareness,%20and%20individual%20differences.pdf>
- Schumann, J. H., Crowell, S. E., Jones, N. E., Lee, N., Schuchert, S. A., & Wood, L. A. (2004). *The neurobiology of learning: Perspectives from second language acquisition*. London: Lawrence Erlbaum Associates.
- Scott, M. (2012). *WordSmith Tools version 6*. Liverpool: Lexical Analysis Software.
- Scott, M., & Thompson, G. (2001). *Patterns of Text*. Amsterdam: John Benjamins.

- Scott, M., & Tribble, C. (2006). *Textual Patterns: keyword and corpus analysis in language education*. Amsterdam: John Benjamins.
- Scovel, T. (2001). *Learning New Languages: A guide to second language acquisition*. Boston, MA: Heinle & Heinle.
- Scrivener, J. (2011). *Learning teaching: The essential guide to English language teaching*. Oxford: Macmillan Education.
- Seedhouse, P. (2004). *The interactional architecture of the language classroom: A conversation analysis perspective*. Malden, MA: Blackwell Publishing.
- Segrin, C., & Givertz, M. (2003). Methods of social skills training and development. In J. O. Greene & R. Brant (Eds.). *Handbook of Communication and Social Interaction Skills* (pp. 135–176). Hoboken: Lawrence Erlbaum Associates.
- Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford: Oxford University Press.
- Selikowitz, M. (1998). *Dyslexia and Other Learning difficulties. The Facts*. Oxford: Oxford University Press.
- Selinker, L. (1994). *Rediscovering Interlanguage*. London: Longman.
- Shoaib, A., & Dörnyei, Z. (2004). Affect in lifelong learning: Exploring L2 motivation as a dynamic process. In P. Benson & D. Nunan (Eds.). *Learners' Stories. Difference and Diversity in Language Learning* (pp. 22–41). Cambridge: Cambridge University Press.
- Shumin, K. (2002). Factors to consider: developing adult EFL students' speaking abilities. In J. C. Richards & W. A. Renandya (Eds.). *Methodology in language teaching: an anthology of current practice* (pp. 204–211). Cambridge: Cambridge University Press.
- Sinclair, J. (2003). *Reading Concordances: An Introduction*. Harlow: Pearson.
- Sinclair, J., & Brazil, D. (1982). *Teacher Talk*. Oxford: Oxford University Press.
- Sinclair, J., & Coulthard, M. (1975). *Towards an analysis of discourse: the English used by teachers and pupils*. Oxford: Oxford University Press.
- Skehan, P. (1989). *Individual differences in second language learning*. London: Edward Arnold.
- Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- Soars, L., Soars, J., & Sayers, M. (2003). *New Headway. Advanced Teacher's Book*. Oxford: Oxford University Press.
- Sparks, R., & Ganschow, L. (2001). Aptitude for learning a foreign language. *Annual Review of Applied Linguistics*, 21, 90–111.
- Sound Forge Pro 10* (2003 – 2015). Sony Creative Software, Inc.
- Stanton, A., & Morris, S. (1999). *Fast track to CAE*. London: Longman.
- Štefánik, J. (Ed.). (2004). *Antológia bilingvizmu*. Bratislava: Academic Electronic Press.

- Sternberg, R. J. (2009). *Kognitivní psychologie*. Praha: Portál.
- Susskind, E. (1969). The role of question asking in the elementary classroom. In F. Kaplan & S. Sarason (Eds.). *The Psycho-Educational Clinic*. New Haven, CT: Yale University Press.
- Švec, V. (2002). *Cesty k učitelské profesi: utváření a rozvíjení pedagogických dovedností*. Brno: Paido.
- Swales, J. (1993). Genre and engagement. *Revue belge de philologie et d'histoire*, 71(3), 687–698.
- Swales, J. M., & Feak, C. B. (1994). *Academic Writing for Graduate Students: Essential Tasks and Skills. A Course for Nonnative Speakers of English*. Ann Arbor, MI: University of Michigan Press.
- Swales, J. M., & Feak, C. B. (2009). *Abstracts and the Writing of Abstracts*. Ann Arbor, MI: University of Michigan Press.
- Tárnyiková, J. (2007). *Sentence Complexes in Text*. Olomouc: Nakladatelství UP.
- ten Have, P. (2007). *Doing conversation analysis* (2nd ed.). London: SAGE Publications.
- textalyser.net (2004). Textalyser.net Retrieved from <http://textalyser.net/>
- Thompson, P. (2005). Points of focus and position: Intertextual reference in PhD theses. *Journal of English for Academic Purposes*, 4(4), 307–323.
- Thoms, J. J. (2012). Classroom discourse in foreign language classrooms: A review of the literature. *Foreign Language Annals*, 45(s1), 8–27. Retrieved from <http://doi.org/10.1111/j.1944-9720.2012.01177.x>
- Thornbury, S. (1996). Teachers Research Teacher Talk. *ELT Journal*, 50(4), 279–289.
- Tsui, A. B. M. (1995). *Introducing Classroom Interaction*. London, Penguin English.
- Tůma, F. (2014a). Dialogism and classroom interaction in English language teaching: A review of Czech research. *Pedagogická orientace*, 24(6), 878–902. Retrieved from <http://doi.org/10.5817/PedOr2014-6-878>
- Tůma, F. (2014b). Dialogismus a výzkum interakce ve třídě: přehledová studie (1990–2012). *Pedagogika*, 64(2), 177–199.
- Urbanová, L. (2003). *On Expressing Meaning in English Conversation: Semantic Indeterminacy*. Brno: Masarykova univerzita.
- Vágnerová, M. (2000). *Vývojová psychologie. Dětství, dospělost, stáří*. Praha: Portál.
- Van Lier, L. (1988). *The Classroom and the Language Learner*. Harlow: Longman.
- Vlčková, K. (2010). *Žákovské strategie učení cizímu jazyku ve všeobecném vzdělávání – průřezový výzkum*. Brno: Masarykova univerzita. Retrieved from <https://is.muni.cz/dok/fmgr.pl?ag=do;furl=%2Fdo%2Frect%2Fhabilitace%2F1441%2FVlckova%2F;info=>.
- Volín, J. (2007). *Statistické metody ve fonetickém výzkumu*. Bratislava: Epoque.
- Volín, J. (2010). Fonetika a fonologie. In V. Cvrček, V. Kodýtek, M. Kopřivová, D. Kovaříková, P. Sgall, M. Šulc, J. Táborský, J. Volín, & M. Waclawičová.

- Mluvnice současné češtiny I. Jak se píše a mluví* (pp. 35–64). Praha: Univerzita Karlova, Nakladatelství Karolinum.
- Volín, J., & Poesová, K. (2008). Temporal and spectral reduction of vowels in English weak syllables. In Grmelová, A. Dušková, L., & Farrell, M. (Eds.). *3rd Prague Conference on Linguistics and literature Studies: proceedings*. Praha: Univerzita Karlova, Pedagogická fakulta.
- Vološinov, V. N. (1973). *Marxism and the philosophy of language*. (L. Matejka & I. R. Titunik, Trans.). New York & London: Seminar Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.
- Vygotsky, L. S. (1981). The genesis of higher mental functions. In J. V. Wertsch (Ed.). *The concept of activity in Soviet psychology* (pp. 144–188). New York: M. E. Sharpe.
- Vygotsky, L. S. (1986). *Thought and Language*. Cambridge, MA: The MIT Press.
- Wales, K. (2001). *A Dictionary of Stylistics*. Harlow: Pearson Education.
- Walsh, S. (2006). *Investigating Classroom Discourse*. London: Routledge.
- Ward, G., & Birner, B. J. (2008). Discourse and information structure. In D. Schiffrin, D. Tannen & H. E. Hamilton (Eds.). *The Handbook of Discourse Analysis* (pp. 119–137). Oxford: Blackwell.
- Wells, G. (2001). *Action, Talk, and Text: Learning and Teaching through Inquiry*. New York: Teachers College Press.
- Wells, J. (2008). *Longman Pronunciation Dictionary* (3rd ed.). Harlow: Pearson Longman.
- Wertsch, J. V. (1991). *Voices of the mind. A sociocultural approach to mediated action*. Herfordshire: Harvester Wheatsheaf.
- Widdowson, H. G. (1998). Communication and pedagogy: The pragmatics of ESP. *English for Specific Purposes*, 17(1), 3–14.
- Widdowson, H. G. (2007). *Discourse Analysis*. Oxford: Oxford University Press.
- Wilkinson, L. C. (Ed.). (1982). *Communicating in the Classroom*. New York: Academic Press.
- Williams, M., & Burden, R. L. (1997). *Psychology for Language Teachers: A Social Constructivist Approach*. Cambridge: Cambridge University Press.
- Wragg, E. C. (1993). *Class Management*. London: Routledge.
- Wragg, E. C., & Brown, G. (2001). *Questioning in the Secondary School*. London: Routledge/Falmer.
- Wright, T. (1987). *Roles of teachers and learners*. New York: Oxford University Press.
- Yin, R. K. (2008). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA: Sage Publications.



## INDEX

- analysis  
  corpus a., *see* corpus  
  conversation a., *see* conversation  
  genre a., 90, 106  
  lexical a., 26
- acquisition, 14, 17, 22, 57–58, 124, 129–131, 137, 187, 189, 198  
  English a., 13, 17, 19, 21, 27, 35, 46, 50, 55, 160, 187  
  foreign language a./FLA, 65, 79, 162, 187  
  L2 a., 14, 17–19, 77, 161  
  language a., 13, 18–19, 25–26, 34–35, 91, 123, 136, 161–162, 173–175, 185, 187–188  
  mother tongue a., 136  
  second language a./SLA, 12–16, 18, 25–26, 47, 49–50, 58, 127, 130–131, 139, 141, 143, 147, 149, 164, 170
- age, 51, 58, 70, 93–94, 127, 133–135, 149–150, 153, 156–157, 181–184, 189, 191
- aptitude, 127, 136–138, 143–144
- assessment, 21, 28, 40, 50–51, 60, 157, 192,
- BAWE, *see* corpus
- BNC, *see* corpus
- Brain, 14, 57, 127, 132–134, 138,
- BUID/Buidcorp, *see* corpus
- classroom interaction, *see* interaction
- COHAT, *see* corpus
- Coh-Metrix, 70–73
- communicative competence, 17, 21–22, 51, 54–55, 140–141, 174, 184
- Common European Framework of Reference for Languages/CEFR, 16, 26, 37–39, 51, 74, 167–168, 191
- Complexity, 17, 21, 23, 25, 34–35, 68–71, 73, 138, 144, 178  
  syntactic c., 17, 23, 68–71

concordance, 100, 102–104, 107, 109, 112, 114–115  
   c. software, 108  
 conversation, 17–18, 22, 28–30, 33, 47–49, 51–53, 55, 89, 99, 117, 170, 176–177, 180  
   c. analysis, 18–19, 51, 161, 163–165, 167, 170  
 corpus, 15–18, 21–23, 25–35, 47, 50–52, 54–55, 63, 68–71, 74, 77–79, 81, 83–87, 89–93, 95–97, 100–114, 119–121, 123, 166  
   BAWE, 78, 93, 110, 115  
   BNC, 92, 102  
   BUiD/Buidcorp, 78, 105–107, 110–111, 114–115  
   COHAT, 18, 77–78, 91–96, 98–104  
   control c., 109–110  
   c. analysis, 15–16, 81, 107, 109–110, 119  
   c. data, 16, 109, 119, 122  
   c. tools, 17, 78  
   CSSE c., 17, 21, 52  
   ICLE, 120–121  
   learner c., 16, 25, 47, 67–68, 77–79, 93, 104–106, 119, 124  
   reference c., 16, 110, 112  
   SPACE, 77, 83–90  
   specialised c., 78, 111  
   sub-corpus, 77, 107, 110–111, 113  
   written c., 17, 67, 97, 101, 103  
 critical thinking, 18, 78, 82, 92, 94  
 CSSE, *see* corpus  
 data-driven learning, 15, 108  
 dialogism, 163–165, 167, 169–170  
 discourse, 16, 18, 21–22, 25, 28–30, 33–34, 47, 49–50, 55, 59, 107, 109, 162, 173–174, 176–181, 184–185, 187–188, 190, 193–196  
   d. analysis, 21, 25, 109, 161, 173, 195  
   d. marker, 48–49, 52, 54–55  
 discovery learning, 18, 78, 91–92  
 dysfluency, 22, 49  
 educational style, 151, 162, 189–193, 197  
 English  
   academic E., 16–18, 78, 81, 91–93, 95, 99, 101, 105, 107  
   E. acquisition, *see* acquisition  
   E. as a foreign language/EFL, 22, 57, 79, 103, 111, 119, 123, 125, 131, 153, 166–168, 170  
   E. as a lingua franca/ELF, 106  
   E. for academic purposes/EAP, 18, 77, 81–82, 90, 105, 109, 112, 114, 167  
   E. language learning, 51, 128

- 
- E. language teaching/ELT, 161, 163, 167–168, 171, 187–188, 193–195  
learner E., *see* learner  
native E., 67–70, 78, 119  
non-native English, 67–68, 79, 93, 117  
error, 14, 16–18, 25, 45, 68, 74, 78, 90, 93, 95, 117–118, 120, 123, 132, 183  
  lexical e., 18, 117–118, 123  
false friends, 18, 77, 79, 117–121, 123–126  
finite verb, 23, 25, 67  
form-focused instruction, 14, 17, 77  
frequency  
  absolute frequency/AF, 97, 99–101  
  f. list, 109–110  
  f. of occurrence, 30, 61, 64, 112, 120  
  raw f./RF, 120  
head noun, 21, 25–30, 32–35, 70  
hedges/hedging, 17, 18, 22–23, 49–50, 57, 59–66, 78, 88, 90, 99–100  
  h. marker, 99–100  
hesitator, 22, 47–50, 53–55  
ICLE, *see* corpus  
initiation, 109, 161, 174–176, 183, 188, 193–195  
  IRE/F, 174, 188, 193, 195, 197  
insert, 22, 49–55, 148  
instruction, 14–17, 22, 50–51, 57–58, 65, 77, 103, 125, 130, 145, 156–157, 162,  
  168, 175, 188, 196, 198  
interaction, 14, 16, 18–19, 28, 48–49, 55, 83, 108, 118, 161–171, 173–180, 183–185,  
  187–198  
  classroom i., 14, 16, 18–19, 161, 163, 166–171, 173–175, 183, 185, 188, 193–194, 198  
interlanguage, 14–18, 39, 77–78, 93, 95  
internal factors, 18, 127, 129–131, 140  
keyword, 18, 78, 82, 97–98, 109, 111–112  
knowledge, 15, 22, 28, 30, 34, 57–60, 62, 65, 82, 96, 109–110, 126–127, 129, 132–140,  
  145, 154, 157, 174, 176–178, 189, 190  
  explicit k. 22, 58, 61  
  implicit k. 23, 58  
language  
  first l., 94, 130  
  foreign l., 13, 18, 21, 23, 37, 55, 57, 59, 78, 117, 159, 161–162, 165–168, 170,  
    176, 187–188, 195  
  l. acquisition, *see* acquisition  
  learner l., 14–18, 78, 118–120, 123–124, 161  
  native l., 13, 16, 47, 55, 118, 123

second l., 13, 16, 78, 93, 122, 125, 130, 134, 137, 143–144, 147, 157, 174  
 spoken l., 16, 22, 48–50, 101, 163  
 written l., 25, 48, 50–51, 119

learner
 

- l. corpus, *see* corpus
- l. English 16, 50, 78–79, 105, 107, 119
- l. language, *see* language

learning strategies, 15, 127, 132–133, 139–140  
 learning styles, 127, 138–140  
 lexical bundle, 81, 113–114  
 lexical item, 21, 26–28, 30–32, 35, 38–40, 83, 118–119, 121, 123–124, 126  
 Longman Grammar of Spoken and Written English/LGSWE, 26, 28–29, 31, 33, 113  
 log likelihood, 110, 112, 115  
 marker, 77, 81–83, 115  
 metacognition, 78, 92  
 mistake, 23, 28, 51, 78, 82, 93, 118, 120, 123–124, 183
 

- lexical m. 120, 123–124

modal verb, 17, 22, 57, 59–65, 99–100  
 modality, 22, 59, 63  
 mode, 19, 176  
 mother tongue, 54, 73, 92–93, 100, 120–124, 126, 130, 132, 134, 176, 183, 193, 196–197  
 motivation, 18, 50, 109, 127–128, 130–131, 133, 136, 143–147, 149–159, 178, 184, 189  
 motivational transformation episodes, 146–147, 155, 158  
 multiword expression, 18, 109, 114  
 native speaker/NS, 17, 22, 51–52, 54, 68, 86, 93, 150, 156, 179, 191  
 Ngram, 109, 113  
 non-finite verb, 17, 23, 25, 30, 68–69, 73  
 non-native speaker/NNS, 50, 54–55, 93, 179, 190–191  
 noun phrase, 17, 21, 23, 25–27, 29–30, 33–35, 68–70, 73, 118
 

- complex n. p., 21, 25, 33–35, 69–70, 73

part of speech/POS, 70, 89, 109, 120  
 passive, 60–62, 64–65, 74, 83, 139  
 popular academic text, 18, 77, 83  
 postmodification, 21, 26, 30–35, 70  
 pragmatics, 16, 21, 57–58, 60, 65  
 premodification, 17, 25–33, 35, 70  
 pronunciation, 16–17, 21–22, 37–46, 50, 117, 126, 189  
 question, 19, 21, 39–40, 45, 48, 50, 54, 61–62, 78, 81, 83, 92, 96, 103–105, 127, 130, 132, 134–137, 143, 148, 155, 161, 174–185, 188, 190, 193–196  
 social skills, 19, 162, 185, 187, 189–191, 193, 197–198

- 
- SPACE, *see* corpus
- Speech, 22, 37–45, 47–51, 54, 70, 89, 109, 120, 132, 138, 164, 166, 173  
    spontaneous s. 22, 37, 43, 47, 49
- spoken performance, 17, 50–51, 54, 155
- taxonomy, 60, 102, 113
- teaching, 17, 23, 30, 37, 45, 55, 58, 67, 74–75, 78, 81, 85, 90–93, 95, 100, 104–105, 108, 110–112, 114, 124–125, 127, 129–130, 132–135, 138–141, 151, 159, 161, 163, 166–168, 171, 174, 179, 187–193, 195
- testing, 38, 45, 109, 130, 136–137, 177
- text, 22–23, 31, 38, 40, 43, 48, 60, 62, 68, 71–72, 82–83, 86–88, 93, 100–101, 105, 107–108, 110, 143, 180  
    t. types, 62, 82, 87, 93, 100–102  
    written t., 22, 34, 47, 77–78, 101, 124
- transfer  
    L1 t., 14, 45, 93  
    negative t., 38–39, 50, 55, 79, 123
- vocabulary, 18, 21, 23, 25–26, 30, 32, 34–35, 67, 78, 91–92, 97–98, 104, 117, 123–126, 132, 151, 180
- Wmatrix3, 15, 17, 78, 109–110, 112, 114
- WordSmith, 15, 17, 27, 52, 78, 109–113
- writing, 15–18, 21–23, 50, 57, 59, 61, 65, 67–70, 72, 74–75, 77–78, 81–85, 87–88, 90, 93–94, 96–97, 99, 101, 103, 105–107, 110–112, 114, 118, 137  
    academic w., 15–17, 22, 57, 59, 61, 65, 77–78, 81, 83–84, 87–88, 90, 93–94, 96–97, 103, 106, 107



## **Learner Corpora and English Acquisition**

A Collection of Studies

Edited by Monika Černá, Jaroslava Ivanová and Šárka Ježková

Published by Univerzita Pardubice, Studentská 95, Pardubice, Czech Republic

2015

First edition

Typeset: Lukáš Vavrečka

Cover design: Marek Chalupník

The publication was approved by the Scientific Board of the Pardubice University Publishing House.

**ISBN 978-80-7395-946-3**

ISBN 978-80-7395-946-3



9 788073 959463