

## Summary

# Corpus research into English interlanguage of Macedonian learners: acquisition and crosslinguistic influence

Liljana Mitkovska  
Marija Kusevska  
Eleni Bužarovska

This book presents the results of a number of investigations into the interlanguage of Macedonian learners of English. The research has been conducted within the project *Analysis of English interlanguage of Macedonian learners at CEFR levels A1, A2, and B1 with a special emphasis on cross-linguistic influence*. The project was funded by the Macedonian Ministry of Education and Science and included researchers from several universities in the country:

- Head researcher: Professor Liljana Mitkovska, PhD (FON University, Skopje)
- Research collaborators: Professor Eleni Buzarovska, PhD (University of Cyril & Methodius, Skopje), Assistant professor Marija Kusevska, PhD (University Goce Delčev, Štip) and Anastazija Kirkova-Naskova, MA (University of Cyril & Methodius, Skopje)
- Junior researchers: graduate students who are working on topics related to the project as part of their degree papers or MA theses at FON University, Skopje and University of Cyril & Methodius, Skopje
- Technical support: Assistant professor Slavko Čungurski, PhD (FON University)

The main goal of this project is to determine, classify and explain the characteristics of the English interlanguage used by Macedonian learners of English. The initial goal was to describe the interlanguage at the first three proficiency levels (A1, A2 and B1) from a morpho-syntactic, lexical, phonological and pragmatic aspect. However, during the project the scope of our investigation expanded to include B2 level as well. Special emphasis is put on the role of Macedonian as a native language or a language of the community (vkluceni se i pripadnici od malcinstvata ili celta e da se potencira deka makedonskiot e voedno i jazikot na sredinata? Izgleda toa 'or' me buni malku).

The initial research was focused on particular structures at morpho-syntactic and pragmatic level in order to:

(1) establish the linguistic competence at each level, i.e. to what extent the structure under examination has been acquired at levels A1, A2, B1, B2;

(2) investigate the development of the given structure across the language levels: how learners make progress in their use of the structure under examination;

(3) identify the problems at each level: those that result from intralingual factors (developmental errors), and those occurring as a result of L1 interference (cross-linguistic influence).

In the analysis of these aspects more emphasis has been placed on explaining errors due to L1 influence. Our research focuses on typical problems that Macedonian speakers encounter when learning English. As the aim of the project is not only to detect, but also to analyze and find ways how to overcome these problems, the current research becomes instrumental in ESL pedagogy. The results of the research should draw teachers' attention to specific problems Macedonian learners of English are faced with, to which the general literature on language acquisition does not offer answers. This will provide them with a more precise picture of what they can expect from their students at each level and give them ideas how to cope with certain problems that are constantly repeated due to the influence of the Macedonian language as L1. It should help them plan how to use available teaching materials currently in use, which are intended for all learners of English irrespective of their L1 and therefore are indiscriminate with respect to the needs of a specific L1 background. Apart from teachers, this research should provide help to both undergraduate and graduate students of English in training programs for future teachers or translators and interpreters.

The investigation into the learner language is based on contemporary theories of second language acquisition with the application of current empirical methods of interlanguage analysis that make use of electronic corpora. For this end, Macedonian English Learner Corpus (MELC)<sup>1</sup> was created with data gathered from about 2000 participants. It contains around 500,000 words but the number of words and participants is constantly on the rise. In the next section, we give a brief description of the principles that

---

<sup>1</sup> This corpus is available at <<http://mkam.fon.edu.mk/>>

underlie the foundation of the corpus and how it is structured. In section 0.2 we give a short overview of the chapters included in this book.

### **The electronic corpus of learner language**

The Macedonian English Learner Corpus (MELC) is an electronic collection of texts written by learners of English in the Republic of Macedonia. MELC was created for the purposes of the project to serve as a research database consisting of authentic materials, a tool which should enable more effective examination of learner language. Beside its key function in the project, the creation of MELC was motivated by the following long term goals:

(1) To create an electronic corpus of the English interlanguage that can be used in future research on second language acquisition, both by domestic and foreign researchers.

(2) To develop a methodology for corpus creation that can serve as a model for establishment of similar corpora for other languages.

(3) To develop a methodology of empirical research on learner language for practical goals.

(4) To promote the use of research methods based on advanced technology in language studies and educate young researchers to implement new techniques.

(5) To introduce the latest corpus-based methods in language research and to define topics for future research using the created corpus.

(6) To connect research and practice.

What does an “electronic learner corpus” represent? Any random collection of learners’ texts cannot be considered as such. According to Granger (2002: 7):

“Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardised and homogeneous way and documented as to their origin and provenance.”

This means that an electronic corpus should be well planned in terms of type of participants and texts, methods of data collection, etc. Having this in mind, we designed MELC following certain principles and defined parameters that make its use effective for scientific and other purposes.

The work on MELC started in October 2010 by the team of researchers from three universities in the country: FON University (Skopje), University of Cyril & Methodius and Goce Delčev University, (Štip). This team designed the

electronic data base in the preparatory stage of the project and in the next stage administered the processes of data gathering and building the corpus. A large number of participants, mainly consisting of graduate students from these universities, were involved in obtaining written data from learners of English from various institutions: elementary and high school students, university students, as well as children and adults attending extra-curricular English courses,

Building up a corpus is a complex and time-consuming endeavor which requires the attention of experts in different disciplines. MELC was created three years ago and more work needs to be done for its improvement and wider use. The following immediate goals should be pursued in the future:

- upgrading of the corpus interface in order to facilitate its use and make the corpus more accessible;
- introduction of search tools for automatic data selection and statistics;
- grammatical annotation of the corpus marking parts of speech (POS) and errors;
- addition of spoken language.

### **Structure and organization of the book**

The project goals set out by the team are long term and will be accomplished gradually. One of the target outcomes of the project is to encourage research activities in the area of second language acquisition using empirical methods and new technology. This book specifically aims at promoting such research and, accordingly, it consists of two parts: the first part gives the theoretical framework and the method used in our research of MELC data, while the second part includes several papers that present the first results of our investigation of this corpus. The papers share several common goals: (a) to provide a description of a particular structure at all levels and show how it develops from one level to another, i.e. its acquisition sequence; (b) to show and explain the role of the Macedonian language as L1 in the acquisition process. All papers use the same theoretical framework and make use of the same or similar research methodology described in the first part of the book. In what follows we present a brief outline of the book's contents.

**The first and the second** chapter include an overview of the theoretical framework used in the research studies based on MELC data. The aim is to present the main principles of second language acquisition and explain the

relevant terms, such as interlanguage, cross-linguistic influence, contrastive analysis, error analysis, developmental and transfer errors. The second chapter presents some basic methodological principles applied in corpus-based interlanguage research (Granger, 2002, Gilquin, 2008), contrastive analysis and error analysis.

**The third** chapter looks into the problem of acquisition of pragmatic competence. The term pragmatic competence is defined and some basic requirements for such competence are discussed. The authors clarify the relation of CEFR to pragmatic competence and describe the main instruments which were used for collection of materials relevant for the research of pragmatic competence.

**The fourth** chapter presents an overview of CEFR which serves as basis for validation of English language ability according to several levels. The authors outline the criteria by which level distinction is made, provide descriptors through which each level is determined, discuss approaches to development of language skills, etc. The authors also provide various activities aimed at improvement of CEFR and its more effective application in pedagogy.

The second part of the book is dedicated to research conducted on MELC data. The **fifth** chapter examines the use of two present tenses (Present Simple and Present Progressive) by Macedonian learners aged from 8 to 15. Given that in Macedonian there is only one present tense that expresses the semantic category of present time as opposed to two present tenses in English, it could be predicted that Macedonian learners would encounter difficulties in delimitation of the functions of the two English tenses. The starting hypothesis is that aspect differences in the two languages lead to generalization: Macedonian learners generalize the form of present progressive tense to express certain functions which native speakers express by simple present. The quantitative data of tense occurrences, which was gathered from the attested examples in MELC, is statistically analyzed and represented in tables, graphs, and histograms at each level. This is followed by a qualitative analysis of form and function. The cross-level comparison of the results confirms the hypothesis that the overuse of present progressive occurs at the expense of present simple tense forms.

The sixth and the seventh chapter are devoted to the analysis of modal verbs use by Macedonian learners of English. The **sixth** chapter deals with the acquisition process of English verbs of obligation: *must, have to, should, need to* at A1, A2, B1 and B2 levels. The distribution of modals at each level is given and explanation is provided for most common formal and functional errors. To find

out whether the use of these modals in learner language differs from their use in native speaker texts the frequency distribution of each modal is compared to its distribution in native corpora.

The **seventh** chapter explores interlanguage influences on the acquisition of English modal verbs of possibility: *can*, *could*, *may* and *might*. By determining the functional distribution of each modal verb across levels the authors provide the acquisition sequence of these modals pointing out the late development of epistemic uses. The prevalent tendency of learners at lower levels to overuse *can* at the expense of *may* is statistically documented and explained.

The **eighth** chapter examines the use of the possessive genitive construction (*the student's exams*) in comparison to the possessive analytic construction (*the exams of the student*). Although there is no significant difference between these two constructions they are not always mutually replaceable. The choice of the construction depends on the interplay of a number of structural, semantic and discourse-pragmatic factors. The goal of this investigation is to explore how Macedonian speakers make a distinction between these constructions in view of the fact that the basic possessive construction in Macedonian is analytic, corresponding structurally to the English *of*-construction.

The last two chapters deal with the acquisition of pragmatic competence by Macedonian learners of English. The **ninth** chapter explores the speech act of request, while the **tenth** analyzes the speech act of apology. Both chapters have a similar structure in that the presentation of the theoretical framework of the given analysis is followed by the description of the strategies used by native speakers in realization of these speech acts. Then the authors examine the strategies of request and apology respectively, as used by the learners of A1, A2, B1 and B2 levels and compare them to the descriptions provided in CEFR in order to see whether the learners' responses match the expected standards of CEFR levels. The analysis of these speech acts reveals insights into learners' verbal behavior and the difficulties in acquisition of pragmatic competence. Among such difficulties are the following: Macedonian learners tend to use the most common markers and constructions characteristic of a given speech act instead of the full inventory of such markers used by native speakers; under the influence of L1 learners do not make a functional distinction between synonymous markers (e.g. *sorry* vs. *excuse*) and exhibit uncertainty in the choice of an appropriate intensifier in a given speech act.