

Quantitative and Qualitative Research Methods in Sociolinguistics



VYTAUTAS MAGNUS
UNIVERSITY
— M C M X X I I —

VYTAUTAS MAGNUS UNIVERSITY
FACULTY OF HUMANITIES
DEPARTMENT OF LITHUANIAN LANGUAGE

Quantitative and Qualitative Research Methods in Sociolinguistics

STUDY GUIDE
A Resource Book for Students

Vilma Bijeikienė, Aurelija Tamošiūnaitė

Kaunas 2013



VYTAUTAS MAGNUS
UNIVERSITY
— M C M X X I I —

Reviewed by Assoc. Prof. Violeta Kalėdaitė and Prof. Dr. Nemira Mačianskienė

Approved by the Department of Lithuanian Language of the Faculty of Humanities at Vytautas Magnus University on the 5th of November 2013 (Protocol No. 7-1)

Recommended for printing by the Council of the Faculty of Humanities of Vytautas Magnus University on the 7th of November 2013 (Protocol No. 5-2)

Publication of the study guide is supported by the European Social Fund (ESF) and the Government of the Republic of Lithuania. Project title: "Preparation and implementation of the joint degree international master programme in 'Sociolinguistics and Multilingualism'" (project No. VP1-2.2-ŠMM-07-K-02-091)

TABLE OF CONTENTS:

Introduction	7
Chapter 1:	
Research Methods in Sociolinguistics and Multilingualism: Key Concepts and Paradigms	9
Chapter 2:	
Links between Research Design, Method, and Data Collection	20
Chapter 3:	
Study Design: Cross-sectional, Longitudinal, Case Group, Experimental, Quasi-experimental	31
Chapter 4:	
Data: Types of Data and Sampling	44
Chapter 5:	
Ethical Issues	56
Chapter 6:	
Quantitative Research: Concepts and Definitions. Use of Corpora	63
Chapter 7:	
Questionnaires	80
Chapter 8:	
Sociolinguistic Interview	88
Chapter 9:	
Qualitative Techniques: Observations and Recordings (Audio & Video). Transcription	100
Chapter 10:	
Social Network Analysis	117
Chapter 11:	
Conversation Analysis	126
Chapter 12:	
Speech Act Theory	136
Chapter 13:	
Interactional Sociolinguistics	145
Chapter 14:	
Critical Discourse Analysis	155
Literature	167
Answer key	175

INTRODUCTION

These study guidelines are intended for graduate students taking a course in Quantitative and Qualitative Research Methods in Sociolinguistics offered at Vytautas Magnus University. The course is part of the program curriculum of the joint-degree master program “Sociolinguistics and Multilingualism” offered in co-operation with Johannes Gutenberg University in Mainz (Germany), Stockholm University (Sweden) and University of Tartu (Estonia).

The guidelines focus on two main research paradigms related to sociolinguistic research, namely quantitative analysis (related to the variationist tradition) and qualitative analysis (related to the ethnographic tradition). The study guide intends to support in-class activities, as well as to serve as a source for independent study. It does not, however, aim to substitute existing materials (textbooks and handbooks devoted to research methods), rather complement them by adding, where relevant, references to sociolinguistic research in the Baltics. As such, this study guide should only be considered as a work in progress. The authors hope to expand it over time, considering students’ comments and suggestions for improvement.

The guidelines consist of 14 chapters. Each chapter presents and elaborates on a different aspect of sociolinguistic research (i.e., different research methods). The first five chapters introduce key concepts and definitions related to research in general (research methods, research design, data, ethical issues), while the remaining nine chapters focus on different research methods applied in sociolinguistic research (e.g., participant observation, interviews, statistical analysis, conversation analysis, critical discourse analysis, and other). All chapters follow the same structure. First, students are introduced to key concepts related to the topic. Key concepts are followed by a pre-reading activity and then a discussion of the key issues on the topic. Where relevant, the discussion is supported by examples and case studies. Each chapter contains comprehension activities (with answers at the end): some of these activities are intended for independent study, some for in-class activities. Group discussions are strongly encouraged. At the end of the chapter we provide students with a glossary and a list of cited material for further reference. Chapters 1 through 10 were prepared by Aurelija Tamošiūnaitė, and Chapters 11 through 14 were prepared by Vilma Bijeikienė.

CHAPTER 1: RESEARCH METHODS IN SOCIOLINGUISTICS AND MULTILINGUALISM: KEY CONCEPTS AND PARADIGMS

1.1. Key concepts

- Dependent variable*
- Empirical research*
- Experimental research*
- Experimental stimuli*
- Field work*
- Hypothesis*
- Independent variable*
- Non-experimental research*
- Observation*
- Qualitative methods*
- Quantitative methods*
- Triangulation*

1.2. Pre-reading activity

Activity 1. In groups discuss the concepts given above. Which concepts are you aware of? How would you define them in your own words?

.....

.....

.....

.....

.....

1.3. Research methods in sociolinguistics

The **main aim** of this chapter is to equip you with **key concepts** and **paradigms** related to the research methods used when analyzing sociolinguistic phenomena. In the following chapters we will analyze these research paradigms in more detail.

Sociolinguistics investigates how people use language in different environments. It attempts to link patterns of language use to some kind of non-linguistic reality – that is, to things like class, gender, racial or ethnic identification, gang affiliation, and other. In order to investigate these phenomena one needs to have reliable research techniques. However, research techniques used and applied in sociolinguistic research

are quite different in their nature and always depend on the research question under investigation. For instance, sociolinguists interested in the **variation** of a certain phonological feature across different social classes will depend on the **quantitative** methods used in **variationist sociolinguistics**, while sociolinguists interested in the code-switching practices among the bilingual Estonian speech community members might depend on the **qualitative methods** used in **ethnographic** research. Some research questions might require the application of different research methods. This approach is called a **mixed-method approach (or triangulation)**.

1.3.1. Research traditions in sociolinguistics

Coupland & Jaworski (1997: 70–72) differentiate between **three research traditions** in sociolinguistics:

- a. **variationist** – produces statistical information based on large amounts of observed data. This research technique was highly elaborated by its pioneer William Labov, as well as later by Peter Trudgill, Sali Tagliamonte, and many others;
- b. **social constructivist** – highlights the importance of language in understanding the society and social categories and thus promotes the study of language attitudes, beliefs, judgments and reactions about the language (e.g., the research conducted by Howard Giles);
- c. **ethnographic** – stresses the importance of the insider's viewpoint and subjectivity, thus, quite often relies on qualitative research techniques or a combination of both – qualitative and quantitative (e.g., the research conducted by Monica Heller, Barbara Johnstone, Kathryn Woolard and others).

The **variationist tradition** evolved over the past five decades as a discipline that integrates social and linguistic aspects of language (Tagliamonte 2006: Kindle Location 90). Variationist sociolinguistics grew out of the linguists' interest in variation that exists in language and, especially, its correlation to the social life. Thus, as the word "variationist" implies, sociolinguists working in this tradition are interested in different linguistic variants (variables) used in our everyday speech, in other words, they investigate different existing ways of saying the same thing and look for explanations for such use in our social life. Tagliamonte (2006: Kindle Locations 111–114) mentions three main facts about language that constitute the "core" of variationist research: a) the observation that language varies; b) the observation that language changes perpetually; c) language is more than words: it also conveys social meanings.

Variationist sociolinguistics relies on several concepts (Tagliamonte 2006):

- a. the concept of *vernacular*: the main goal of any variationist research is to elicit "vernacular" speech. Vernacular can be defined as "every day speech" or a "style" in which minimum attention is given to the monitoring of speech;
- b. the concept of the *speech community*: a researcher has to immerse him/herself into a community of speakers who share the same verbal repertoire and the same linguistic behavior. The researcher has to be both: a participant and an observer;

- c. the *linguistic variable*: a linguistic (phonological, morphological, semantic, lexical, syntactical) unit. In other words, it is different ways of saying more or less the same thing, e.g., Lithuanian *ei name* vs. *ei nam* (praes. 1 pl. “go”);
- d. the *quantitative approach*: as Tagliamonte (2006: Kindle Locations 184–186) suggests, variationist studies are “based on the observation that speakers make choices when they use language and that these choices are discrete alternatives with the same referential value or grammatical function. Furthermore, these choices vary in a systematic way and as such they can be quantitatively modeled”;
- e. the *principle of accountability*: “every variant that is part of the variable context, whether the variants are realised or unrealised elements in the system, must be taken into account.” (Tagliamonte 2006: Kindle Locations 195–196).

The **social constructivist tradition** presupposes that everything around us has been constructed by the societies in which they appear (Irwin 2010: 100). As Irwin suggests, the social constructivist asks the following questions: “Who has said what, when, to whom, and how to get us to where we are on any one topic?” (2010: 100). Most of the social constructivist tradition in sociolinguistics is related to a theme of identity. Social constructivists wanted to shift the focus from the language system (structure, e.g., the linguistic variable is a key notion in variationist tradition) to the speaker, its performance and construction. Thus, sociolinguists working in the social constructivist light are usually preoccupied with the study of construction of gender, ethnicity, national identity, minority languages; childhood and adolescent identity, and other (Irwin 2010: 106). In terms of identity construction, the work of Erving Goffman is of importance to sociolinguistic research. Goffman views social interaction in terms of a dramatic performance and therefore implies that identity is also an ongoing interactive construction rather than something given (Irwin 2010: 103). On the other hand, many other poststructuralist theorists (Michel Foucault, Pierre Bourdieu, Mikhail Bakhtin, Edward Said among others) influenced the interpretation paradigms in sociolinguistic research.

The **ethnographic research tradition** is based on participant observation and a researcher’s long-term involvement in a community. The ethnographic approach is not specific just to sociolinguistics; it is used widely in other social sciences, especially anthropological research. In the study of language, ethnography was introduced by Dell Hymes in 1960s. He introduced the term “ethnography of speaking” or “ethnography of communication” and by these two he changed the focus of “ethnographic” research: the language does not only help to find out about the culture of those who speaks it or writes it, but is worth being studied as an aspect of culture itself (Johnstone 2000: 84). Linguistic anthropology (some perceive it as a branch of sociolinguistics) mainly deals with the issues of ethnography of communication.

Over the past decades ethnographic description became one of the most important field methods in sociolinguistics (Wolfram 2010: 307). Wolfram (2010: 307) points to

the following benefits of the ethnographic approach in sociolinguistics: quality and amount of data and researcher's familiarity with the community of practice (under the investigation). Monica Heller, one of the key sociolinguists, who had extensively conducted ethnographic studies, explains, that ethnographies aim to discover how people use language, what they believe about language and why do they believe so (Heller 2008: 250). Ethnographies are useful for in-depth descriptions and explanations, and almost always they involve the study of the groups of people (communities of practice). The ethnographic research tradition emphasizes that speakers beliefs, cultural norms and expectations influence their discourse interactions and in order to understand and interpret the social meanings that language use manifests, one needs to study the cultural and social environment with great care.

1.3.2. Empirical research

The investigation of the relationships between language and society as such is an **empirical science**; i.e., sociolinguists are interested in how language functions in its social (real) life. Therefore, for sociolinguists, "data" should be collected through **observation** (field work) as opposed to the data obtained through introspection (analytical analysis) (Milroy & Gordon 2003; Coupland & Jaworski 1997). Johnstone (2000: 1–2) points out, that sociolinguists have their own analytical methods for collecting, describing and interpreting the data in a systematic way, whether the data consist of speech or signing or writing, by one person or many, on one topic or several. On the other hand, there is a need to differentiate between **field methods** (i.e., how one collects the data, for instance, by conducting interviews) and **analytical methods** (i.e., how one analyzes the data, for instance, by statistical analysis). Both of these are important in sociolinguistic research.

Empirical research is widely used throughout the social sciences. Patten (2005: 3) distinguishes **four main questions** one needs to answer before conducting an empirical study:

1. **why** one wants to make an observation: which leads to the **research purpose**, i.e., research questions and hypothesis;
2. **whom** one wants to observe: which leads to the **subjects of the research**;
3. **how**: which points to the **instruments and techniques** one will choose to conduct the observation;
4. **when**: points to the **time framework** (as an important constraint for empirical research).

The empirical observations result in **data**. For sociolinguists the data is usually a "more or less systematic collection of instances of language in use" (Coupland & Jaworski 1997: 70), for instance the data representing the actual speakers' performance (natural speech); subjective responses to certain linguistic behavior (language attitudes, evaluations of speech) or speakers' self-reports on their language use, and other. Johnstone (2000: 22) defines data as "the result of observation, consisting of the 'given things' that researchers analyze".

1.3.3. Experimental vs. non-experimental research

Empirical research can be defined as experimental and non-experimental. Experimental research relies on the cause-and-effect paradigm (it gives treatments to the research subjects), while non-experimental research does not give any treatments.

In sociolinguistics **experimental research** is mainly used to study people's perception. According to Drager (forthcoming), "there are two main directions for perception experiments in sociolinguistics: one examines how language variation influences what social information is attributed to the speaker, and the other investigates whether social information about a speaker can influence how a linguistic variable is perceived". The most common experimental techniques used in sociolinguistic research are the **matched-guise test** and the **identification task** (Drager, forthcoming). The matched-guise test is frequently used in studies investigating language attitudes, mainly linguistic stereotypes. Identification tasks usually are used to investigate the perception of different phonological phenomena (such as vowel merger or chain shifts).

Non-experimental research is much more common in sociolinguistics. It entails both quantitative (surveys, polls) and qualitative (ethnography, discourse analysis) studies. For quantitative studies the researchers gather data that allows statistical analysis, while in qualitative studies the researchers gather data that allows content analysis (the most usual form of such studies is the semi-structured interview). However, according to Johnstone (2000: 36), most sociolinguistic research usually combines both approaches: "sociolinguistic research projects are neither exclusively quantitative nor exclusively qualitative". In other words, a sociolinguist always has to interpret the data – be it numbers or words – and decide what it means. Analyzing sociolinguistic data, thus, involves, some counting in addition to the descriptive explanation that helps to answer the questions of "how" and "why" certain things happen (Johnstone 2000: 37).

Example #1. Experimental study.

Thomas Purnell, William Idsardi & John Baugh, "Perceptual and phonetic experiments on American English dialect identification," *Journal of Language and Social Psychology* 18, 10–30, 1999.

Research question: *whether* dialect discrimination is possible by using *phonetic cues alone*, and if it is possible, *what cues trigger* discrimination.

The study ran *four* different experiments.

Experiment #1. Trying to rent an apartment in San Francisco.

Research question: whether housing discrimination is exhibited in the absence of visual cues.

Null hypothesis: there is no significant difference in appointments made by locale (neighborhood, White, African American or Chicano/Latino) and by dialect.

Test hypothesis: there is a relation between the racial and ethnic constituency of a geographic area and the success of renting an apartment by dialect type.

Method: One of the researchers (Baugh), who is familiar with three distinct American dialects (African American Vernacular English [AAVE], Chicano English [ChE], and Standard American English [SAE]), conducted a series of telephone interviews with prospective landlords in five different locales (that are also ethnically segregated or are predominantly inhabited by one ethnic/racial group). Baugh telephoned the landlords three different times, each time using different dialect. Each call began with the phrase: “Hello, I’m calling about the apartment you have advertised in the paper.”

Results:

Table 1: Confirmed Appointments to View Apartments Advertised for Rent in Different Greater San Francisco Geographic Areas (in percentages) (adapted from Purnell et al. 1999, 15)

Dialect Guise	Geographic Area				
	East Palo Alto	Oakland	San Francisco	Palo Alto	Woodside
AAVE	79.3	72.2	63.5	48.3	27.7
ChE	61.9	58.3	53.2	31.9	21.8
SAE	57.6	68.7	71.9	63.1	70.1
Total number of calls for each locale	118	211	310	263	87

AAVE = African American Vernacular English;
ChE = Chicago English;
SAE = Standard American English.

Table 2: Population in Different Greater San Francisco Geographic Areas by Race and Ethnicity (in percentages) (adapted from Purnell et al. 1999, 15)

Population	Geographic Area				
	East Palo Alto	Oakland	San Francisco	Palo Alto	Woodside
African American	42.9	43.9	10.9	2.9	0.3
Hispanic	36.4	13.9	13.9	5.0	3.8
White	31.7	32.5	53.6	84.9	94.7

Experiment #2. Macro-linguistic cues as indicators of ethnic identity.

Research question: whether dialect identification is possible at the macro-linguistic and sentential level.

Null hypothesis: each guise (certain feature [phonological, morphological] of each dialect [AAVE, ChE, SAE]) should be identified correctly 16.6 percent of the time.

Hypothesis: these guises are identifiable at the same rate as nontridialectal ones.

Method: 20 speakers of the three target dialects (AAVE, ChE, SAE) recorded stimulus tokens. Each token consisted of the sentence: “Hello, I’m calling to see about the apartment you have advertised in the paper.” The tokens then were randomized.

421 (382 native speakers of English, 39 nonnative speakers) undergraduate and graduate students at Stanford University listened to each token once without response. For the second and the third time of listening to the same stimulus the students had to evaluate the race/ethnicity and gender of speakers. They could choose from the following: “African

American”, “Hispanic”, and “European American” combined with the two choices of gender (thus, total of 6 possible responses to each token).

Results: Participants systematically identified Baugh’s guise as AAVE, ChE or SAE. All three guises were judged as being representative of the target dialect. This indicates that macro-linguistic cues are overt indicators of a speaker’s ethnic identity.

Experiment #3. Ability to recognize dialects at micro-linguistic level.

Research question: How do dialects differ in pronunciation? How do listeners identify dialects by pronunciation?

Null hypothesis: there is no difference between the dialects by identification.

Hypothesis: the phonetic features in a short portion of speech are sufficient to trigger identification across dialects.

Method: The word “hello” from the utterance in all three varieties (AAVE, ChE, SAE) was used for this experiment. This word lacks the environment in which the researchers expected dialectal variations. The experiment was conducted with 50 undergraduate students at the University of Delaware. All the participants were Caucasian native speakers of SAE. Total of 60 tokens were randomized (ten instances of “hello” repeated twice for each dialect). Two sets of tokens (120) were presented to each participant. During a 2 second pause participant indicated which dialect they believed they heard.

Results: participants are able to identify the dialects when only hearing one word. However, it remains unclear which features of the speech act as sociolinguistic markers.

Experiment #4. Acoustic measurements of “hello”.

Research question: What acoustic features in “hello” act as sociolinguistic markers?

Acoustic cues under investigation: the frequency of the second formant in the first vowel, the peak of the pitch, the duration of the first syllable, the harmonic-to-noise ratio.

Null hypothesis: there is no difference between the dialects for any of the phonetic measurements the authors perform.

Hypothesis: there is a significant distinction among the dialects accounting for why the tokens are recognized so well.

Method: The researchers measured the segment, syllable, and word durations. The 30 tokens (10 for each dialect) were compared on 28 variables by running a three-way analysis of variance for the different variables.

Results: From four features only the fronting of /e/ is the best cue for differentiating the three varieties under the investigation. The other phonetic features were not significant.

Example #2. Non-experimental study.

Ruth Wodak, Michał Krzyżanowski & Bernhard Forchtner, “The interplay of language ideologies and contextual cues in multilingual interactions: Language choice and code-switching in European Union institutions”, *Language in Society* 41, 157–186, 2012.

Research questions:

- a. How are language repertoires shaped inside EU institutions (also in relation to official, general, and institutionally specific regulations)?
- b. Who/What determines language choice and code-switching in interactions in EU organizational spaces (both front- and backstage)? (158)
- c. Is multilingualism an essential part of the EU’s “diversity”? (159)

Method: The authors view their research as sociolinguistic ethnography and define it as “a close look at language practices in a specific setting” (following Heller 1999: 14–15). The researchers used the technique of ethnographic observation, which is defined as a process of “gathering inside knowledge about institutional practices by focusing on” actions taken by actors in the observed settings (164).

The field work was conducted at European Parliament and European Commission throughout 2009. Each fieldwork session lasted one working week. During the session the researchers conducted ethnographic observations of publicly inaccessible institution spaces at European Commission, as well as meetings at European Parliament. Twenty two semi-structured interviews (of ca. one hour each) were conducted with mid- and high-level officials responsible for shaping the EU language policies. The observations followed the guidelines that were structured according to the research questions. The researchers looked primarily at:

- a. number and variety of languages used throughout the meetings;
- b. frequency of code-switching;
- c. topic related vs addressee related language choices;
- d. language choices and code-switching vs degrees of formality and informality;
- e. meta-communication on language (issues related to multilingualism and language use).

Each meeting was observed by at least two researchers. Each of them focused on different aspects of interaction.

Results:

Different multilingualisms are being simultaneously performed in the EU contexts. A continuum of (more or less) multilingual practices was observed. Those practices are highly context-dependent and serve different manifest and latent functions, e.g., emphasize the power of the speaker, accommodation to the speaker, struggle to win the debate and other. Language choice depends on many factors. The researchers distinguished several factors that influence code-switching and language choice:

- a. co-text related factors (topics, technical jargon, language of the preceding speakers, politeness);

- b. genre-related factors (structure of meetings);
- c. language-ideology related factors (language choice due to the perceived prestige of a language);
- d. power-related factors (struggle to win, control over the debate);
- e. personality- and relationship-oriented factors (preferred language choice, group dynamics).

Table 3: Overall percentage of languages in the observed meetings at the European Commission (adapted from Wodak, Krzyżanowski & Forchtner 2012: 171)

Language	Percentage
English	47
French	29
German	7
Polish	6
Spanish	5
Danish	4
Slovak	1
Latvian	1
Italian	0
Czech	0

Table 4: Overall percentage of languages in the observed meetings at the European Parliament (adapted from Wodak, Krzyżanowski & Forchtner 2012: 172)

Language	Percentage
English	45
Italian	18
French	10
German	8
Polish	7
Greek	5
Spanish	4
Slovak	1
Bulgarian	1
Portuguese	1

1.4. Comprehension activities

Activity 2. Select the appropriate answer:

A. Which one of the following is a *dependent variable* in example #1?

- a) dialect discrimination;
- b) phonetic cues;

B. The results of which type of research are *not* reduced to numbers?

- a) quantitative;
- b) qualitative.

C. Which research *does not* involve treatments (stimuli)?

- a) experimental;
- b) non-experimental.

D. If a researcher asked students for their opinions on which week during the semester they would like to have a midterm test, would the researcher be conducting:

- a) experimental;
- or
- b) non-experimental **research?**

Activity 3. Tagliamonte (2006: Kindle Locations 73–75) claims that: “linguistic theory focuses on the structure of the language. It does not concern itself with the context in which the language is learned and, more importantly, it does not concern itself with the way the language is used”. Thus, what sociolinguistic theory focuses on? How will you respond to the claim that speech (language) cannot be studied apart of the society? Support your argument.

1.5. Glossary

Dependent variable – response or outcome to the independent variable (e.g., establishing the race, ethnicity and gender of the speaker).

Empirical research – type of research that is based on direct observations or experience.

Experimental research – a type of research where researchers give treatments and observe reactions they cause.

Experimental stimuli – the triggers in the experiment that cause the response (sounds, words, sentences, images, video).

Field work – the stage of a research project in which data is collected.

Hypothesis – a statement which needs to be proved or disproved through empirical study.

Independent variable – stimulus or input that causes the change (e.g., the use of the dialect feature); researchers **manipulate** independent variable in the experiment.

Non-experimental research – a type of research where researchers observe participants in their natural settings without giving any treatments.

Observation – a primary research technique of ethnography (Johnstone 2000: 81), i.e., the action/ process of observing something/ someone carefully in order to gain information.

Qualitative methods – a set of methods that aims to gather an in-depth understanding of the phenomenon of the study. Smaller rather than larger samples are used for qualitative research. Qualitative research aims to answer questions *how* and *why*, rather than *what* and *how many*. Qualitative research is usually inductive.

Quantitative methods – a set of methods that is based on quantification or measurement and that employs statistical, mathematical and computational techniques. Quantitative research is usually deductive.

Research question – a problem statement or interrogative question that addresses the problem examined in the study.

Triangulation – combination and application of more than one sampling method for data collection or use of more than one methodology in a research design (Angouri 2010: 34).

1.6. References

- Jo Angouri, 2010. "Quantitative, Qualitative or Both? Combining Methods in Linguistic Research", in *Research Methods in Linguistics*, ed. by Lia Litosseliti, London, New York: Continuum, 29–45.
- Nikolas Coupland & Adam Jaworski, 1997. "Methods for Studying Language in Society", in *Sociolinguistics: a Reader and Coursebook*, ed. by Nikolas Coupland and Adam Jaworski, New York: Palgrave, 69–74.
- Katie Drager (forthcoming). "Experimental Methods in Sociolinguistics: Matched guise and identification tasks", in *Research Methods in Sociolinguistics: A practical guide*, ed. by Janet Holmes and Kirk Hazen, Hoboken, NJ: Wiley-Blackwell.
- Monica Heller, 2008. "Doing Ethnography", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 249–262.
- Anthea Irwin, 2010. "Social Constructionism", in *The Sage Handbook of Sociolinguistics*, ed. by Ruth Wodak, Barbara Johnstone and Peter Kerswill, London: Sage Publications Ltd., 100–112.
- Barbara Johnstone, 2000. *Qualitative Methods in Sociolinguistics*, New York, Oxford: Oxford University Press.
- Lesley Milroy & Matthew Gordon, 2003. *Sociolinguistics: Method and Interpretation*, Malden, Oxford, Melbourne, Berlin: Blackwell Publishing.
- Mildred Patten, 2005. *Understanding research methods: An overview of the essentials*, Glensdale, CA: Pyrczak Publishing.
- Thomas Purnell, William Idsardi & John Baugh, 1999. "Perceptual and phonetic experiments on American English dialect identification", *Journal of Language and Social Psychology* 18, 10–30.
- Sali A. Tagliamonte, 2006. *Analysing Sociolinguistic Variation*, Cambridge, New York: Cambridge University Press [Kindle Edition].
- Ruth Wodak, Michał Krzyżanowski & Bernhard Forchtner, 2012. "The interplay of language ideologies and contextual cues in multilingual interactions: Language choice and code-switching in European Union institutions", *Language in Society* 41, 157–186.
- Walt Wolfram, 2010. "Fieldwork Methods in Language Variation", in *The Sage Handbook of Sociolinguistics*, ed. by Ruth Wodak, Barbara Johnstone and Peter Kerswill, London: Sage Publications Ltd., 296–311.

CHAPTER 2:
LINKS BETWEEN RESEARCH DESIGN, METHOD, AND
DATA COLLECTION

2.1. Key concepts

Data
Research
Theory

2.2. Pre-reading activity

Activity 1. In groups discuss and define the concepts given above. Think of these concepts in a broader sense. What types of data one can use in sociolinguistic research? How is sociolinguistic research different from other types of research?

.....

.....

.....

.....

.....

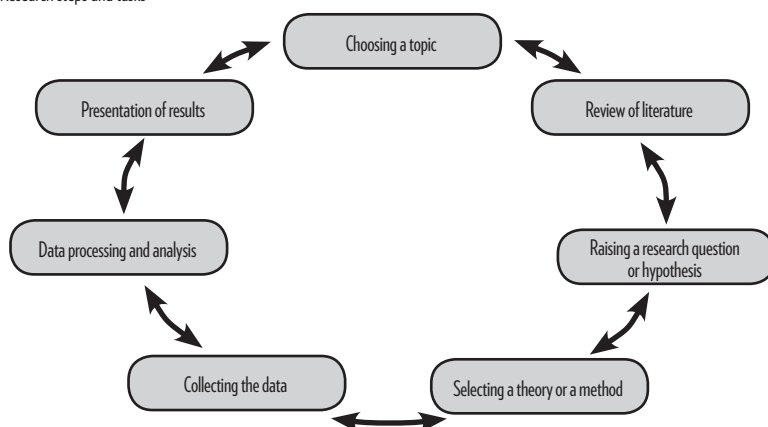
2.3. Linking research design, methods, and data collection

The **main aim** of this chapter is to introduce you to the main steps of research design. It attempts to address the following questions: what is research, how to develop a good research question, how to choose an appropriate research method, how to collect the data and how are method, the research question and data interlinked with each other.

Any type of research is a step-by-step process that involves a series of activities following a certain order (Moyer 2008: 25). Figure 1 visualizes a research process, which involves several constituents:

- a. choosing a topic;
- b. getting familiarized with the literature on the selected topic;
- c. raising a research question (problem statement) or hypothesis;
- d. selecting a theory, a method or a combination of methods for gathering the data;
- e. collecting the data;
- f. data processing and analysis;
- g. presentation of results.

Figure 1: Research steps and tasks



2.3.1. Choosing a topic and reading the literature

There are many interesting topics related to the study of sociolinguistics. Sociolinguists study language in use and its relationship to different layers of a social context, such as situational context of linguistic production, speakers' age, gender, social class, geographical and ethnic origin and other social factors. "Language in use can manifest social phenomena, but language in use can also produce and reproduce social phenomena." (Wodak, Kerswill & Johnstone 2010: 2)

Linguists usually refer to two broader sub-fields of sociolinguistics, namely *macrosociolinguistics* and *microsociolinguistics*. Research in macrosociolinguistics covers such topics as language planning, bilingualism, languages in contact, diglossia, language policies, intercultural and cross-cultural communication, and other, while research in microsociolinguistics addresses such topics as language use in everyday life, face to face communication, linguistic variation (study of the sociolinguistic variable) and other (Wodak, Kerswill & Johnstone 2010: 3). Thus, depending on the research goal, the **topic** can fall either on one or another paradigm of sociolinguistic research.

Sociolinguistic research covers a broad range of topics, however, almost all sociolinguistic research deals with one or another type of linguistic **variation**: variation among different age, gender, social, ethnic groups and so forth. As Moyer (2008: 26) suggests, first of all one needs to decide whether his/her research will focus on **individual or a group**, since this decision will influence the selection of methodology and probably the theoretical framework. One also **needs to select the language or languages** s/he wants to investigate.

Sociolinguistics is a very broad field, thus, it is important to select a specific topic area one wants to investigate. Possible broader topic areas for sociolinguistic research can be the following:

- a. *Geographical variation*: one might want to map a geographical variation of language; attitudes toward different dialects among different layers of society; the use of certain dialect features in a selected geographical area and their social meanings, etc.

- b. *Social variation*: how language varies across different social groups of a given language community; how age, gender and social class affects language use, etc.
- c. *Stylistic variation*: when, how and why an individual modifies his/her speech; what are the means of the modification; which layers of the language are subject to change; how is linguistic variation contextualized; how does it shape the identity of the speaker and index membership to particular community of practice, etc.
- d. *Multilingualism, bilingualism and code-switching*: what is the language choice in a multicultural environment, what determines the use of one language *vis-a-vis* another, what are the patterns of code-switching and how are they contextualized in the spoken interaction, how are they situated in the social context, what is the development of bilingual grammar, what is the role of social networks towards multilingual practices, how are multilingual practices reflected in the linguistic landscape, how are multilingual practices institutionalized, etc.
- e. *Politeness and terms of address*: how do terms of address depend on the situational context, age, gender, ethnic and cultural background; issues of face work, etc.
- f. *Language attitudes*: how is speech perceived by members of different communities of practices (language communities) and what do these beliefs reveal about the social status of different language varieties; how do beliefs towards language shape and motivate a speaker's linguistic behavior; what are the social stereotypes in a selected language community and how do they affect language use; issues in perceptual dialectology, folk linguistics, etc.
- g. *Language and identity*: how does language manifest the identity of a selected community of practice; how does language enregister identity and identities; how do multiple identities correlate with different speech styles; what are the commodification practices of a certain variety in a selected community of practice; and other.
- h. *Language and gender*: do men speak more authoritatively than women, how does gendered language affect/exhibit power relationships, do women use more standard forms than men, how does socioeconomic status correlate with the use of standard and non-standard forms, different stylistic expressions, how do social moves affect the speaker's linguistic behavior, what new social meanings does it entail, and other.
- i. *Language planning* is related to standardization, maintenance and/or language shift: what are the language planning and policy formation strategies in a selected region or language community; how are immigrant and minority language policies and rights predefined; what are the signs of language maintenance and shift among immigrant or minority speakers; how do language ideologies affect language use, and other.
- j. *Language contact*: how do pidgin languages affect the vernaculars of the selected communities; what are the ongoing processes of dialect contact in a

selected language community and what are the factors leading these processes, why do certain language forms focus more quickly in a dialect contact situation than do others; how does age, gender and social class affect linguistic production in a dialect/language contact situation, and other.

These broader topic areas **can serve as a reference point** for selection of literature. For instance, topics related to geographical, social, and stylistic variation will have to be related to the variationist tradition, i.e., one will have to go over the research done by William Labov, Sali Tagliamonte, Penelope Eckert, Peter Trudgill, David Britain, Allan Bell and other sociolinguists (the choice of literature strongly depends on the research question). When you begin reading literature, make sure you review **not only** the “classics” on that particular topic, but also go over **the most recent** studies.

The best way to **start** is to refer to the **introductory anthologies** that exist for sociolinguistic research, for instance,

The New Sociolinguistics Reader by Nikolas Coupland and Adam Jaworski (published by Palgrave Macmillan in 2009),

The Routledge Sociolinguistics Reader by Mirriam Meyerhoff and Erik Schleeff (published by Routledge in 2010).

The book by Coupland and Jaworski is a selected collection of sociolinguistic “classics”. The book contains articles on Language variation (by W. Wolfram, W. Labov, P. Trudgill, J. Milroy and L. Milroy, S. Tagliamonte, P. Eckert), on Language, gender and sexuality (by D. Tannen, S. Kiesling, M. Bucholtz, K. Hall), on Style and identity (by A. Bell, H. Giles, N. Coupland, A. Pennycook), on Language attitudes and ideologies (by E. Ochs, J. T. Irving, S. Gal, A. Jaffe, N. Niedzelski, D. Preston), on Multilingualism, code-switching and diglossia (by J. Fishman, C. Ferguson, S. Gal, C. Myers-Scotton, P. Auer, J. Blommaert), on Language and culture (by D. Hymes, J. Gumperz, J. Holmes, A. Jaworski) and many others.

Meyerhoff’s and Shleeff’s collection of articles combines some of the “classics” in the field of sociolinguistics (foundational research papers by W. Labov, A. Bell, M. Heller, D. Preston, G. Sankoff, P. Trudgill, L. and J. Milroys) with more recent innovative studies in the field (by R. Cameron, Q. Zhang, J. Hay, S. Jannedy and N. Mendoza-Denton, I. Buchstaller, T. Nevalainen and others).

There are many other article collections in the vast market of publishing. Thus, your initial search should not be limited just to the two anthologies mentioned before. The older edition of Coupland’s and Jaworski’s reader is also valuable, as well as the collection entitled *Sociolinguistics: The Essential Readings* edited by Christina Bratt Paulston and G. Richard Tucker and published by Blackwell in 2003.

The **second** step should be related to the overview of articles provided in theoretical handbooks. These types of books (at the moment most of them also have on-line editions) provide a **solid overview of theoretical concepts and methods** related to sociolinguistic research. Several handbooks are worth looking at:

The Routledge Companion to Sociolinguistics, ed. by Carmen Llamas, Louise Mul-lany and Peter Stockwell (published by Routledge in 2006) [available on-line via Taylor and Francis Online];

The Sage Handbook of Sociolinguistics, ed. by Ruth Wodak, Barbara Johnstone and Peter Kerswill (published by Sage in 2010) [available on-line via EBSCO e-books];

The Blackwell Handbook of Historical Sociolinguistics, ed. by Juan-Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre (published by Blackwell in 2012) [available on-line via Wiley Online Library];

The Oxford Handbook of Sociolinguistics, ed. by Robert Bayley, Richard Cameron and Ceil Lucas (published by Oxford University Press in 2013);

The Cambridge Handbook of Sociolinguistics, ed. Rajend Mesthrie (published by Cambridge University Press in 2011);

Sociolinguistics/Soziolinguistik: An International Handbook of the Science of Language and Society / Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft, ed. by Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier and Peter Trudgill (published by Walter de Gruyter in 2008) [available on-line via EBSCO e-books]. This encyclopedia of sociolinguistics has 3 volumes. It is a very good reference source for checking out main definitions, concepts, theories and methods related to sociolinguistic research conducted all over the world.

Most of the articles or excerpts from larger studies that were provided in the handbook or collection of articles **will lead you** to the original or **primary sources** of these studies in **journals** or **books**. The selection of books depends on your research topic. Always ask your advisor for recommendations. There are many journals devoted to sociolinguistic research only. At the same time other types of journals (on general linguistics, pragmatics or semantics), as well as interdisciplinary journals often carry articles touching upon different aspects of sociolinguistic research. Journal articles and conference presentations always provide a researcher with the **latest research** in the field. Everyone who conducts research in sociolinguistics should be familiar with the following scholarly journals: *International Journal of the Sociology of Language*, *Journal of Sociolinguistics*, and *Language in Society*. The following journals (in alphabetical order) also publish sociolinguistic research:

American Speech; *Applied Linguistics*; *Discourse and Society*; *English World-Wide*; *Gender and Language*; *Heritage Language Journal*; *International Journal of Applied Linguistics*; *International Journal of Bilingual Education and Bilingualism*; *International Journal of Multilingualism*; *International Multilingual Research Journal*; *Journal of Language, Identity and Education*; *Journal of Linguistic Anthropology*; *Journal of Multilingual and Multicultural Development*; *Journal of Pidgin and Creole Languages*; *Journal of Pragmatics*; *Journal of the Sociology of Language*; *Language and Communication*; *Language and Education*; *Language Policy*; *Language Problems and Language Planning*; *Language Variation and Change*; *Language*; *Lingua*; *Linguistics*; *Multilingualia*; *Reading and Writing*; *Sociolinguistic Studies*.

Most of these journals are available online; however, they may be referred to in different databases, thus, before the search, check which database carries the journal you require.

Important!

The journals change over time, new journals appear each year, thus, please do not limit your search only to the journals provided in this list.

For a more extensive list of journals carrying articles on sociolinguistics please check the list provided by **LinguistList** (at <http://linguistlist.org/pubs/journals/search-journals2.cfm>).

If you know other languages, it is always useful (and essential) to check similar types of studies **published in** and **for other languages**. Depending on your research focus, you will need to check **the latest studies done** in your target language: if you are conducting research on the use of Lithuanian swear words among the teenage and adult male and female groups, you might need to check more general types of studies on swear words in Lithuanian, as well as any related research done in Lithuanian. In the Baltics currently there is no scholarly journal that specifically aims to discuss just sociolinguistic research, therefore, always check with your advisor who will point you towards the “must-read” literature in the field.

If you want to stay **updated** on the latest works being published in (socio) linguistics, it is useful to join electronic mailing lists. The most popular and known is the worldwide electronic discussion forum *The Linguist List*. You can find further information about the list and how to subscribe to it under the following link: <http://linguistlist.org/>.

2.3.2. Raising a research question or hypothesis and choosing a methodology

Once the initial idea (topic) for research is established, one needs to come up with the research question or hypothesis. You can ask a research question or formulate a hypothesis only after you are done with the review of the literature and have become familiarized with the area of interest. For instance, a student is interested in the use of swear words among male and female speakers. This is **your topic**. In order to design your research, you need to ask questions that will help to narrow down your topic to a few questions and that will help to shape or structure your research. You can start with simple questions, for instance: when males and females use swearing words? Do males use more swear words than females? Why do females use less than males? These simple questions will direct you in the literature and will help you to identify which questions still have not been answered in previous research. As Sunderland (2010: 10) suggests, first of all a researcher needs to ask her/himself “what am I trying to find out in my research project?” and this will provide the basis for a research question. From the very outset of our project we need to specify: **what** we would like to find out and **how** we are going to do this (Rasinger 2008: 16).

One may come across two terms: *research questions* and *hypothesis*. When do we use one and when – another? It depends on the type of research. According to Sunderland (2010: 10), hypotheses are more characteristic of the natural sciences (when we need to prove or disprove something). **Hypotheses** are more common for a quantitative type of

research (when we can count and measure something, hence, prove something correct or incorrect), while **research questions** are more common in qualitative research (they are broader in nature and more exploratory). Thus, the following hypothesis can be formulated for a previously defined topic (the use of swear words among male and female speakers): *teenage male speakers use more swear words than teenage female speakers*. For a hypothesis we need to have **at least two variables**: in this case – one variable is gender, the second one – swear words. The hypothesis can not be tautological, i.e., we **can not** formulate a hypothesis in the following manner – *teenage male speakers can either use more swear words than teenage female speakers or not*.

On the same topic we can also think of a **research question** which would then require a **qualitative type** of study: *What are the swear word using practices among the teenage male and female speakers?*

As one can see from these examples, on the one hand, hypotheses are always more specific and precise than research questions. On the other hand, **hypotheses** are always **statements**, and **research questions** are always **questions**.

Research questions and hypotheses are inseparably linked with theory and methodology. If quantitative research usually begins with a set of research questions or, more likely, a hypothesis, then, research questions in qualitative research may be asked only after some fieldwork experience. Quantitative research often wants to test a certain theory or method. Based on the hypothesis we develop a methodology (instruments), which will allow us to prove or disprove the hypothesis (Rasinger 2008: 11). Such a type of research is *deductive*. The methods/theory we choose must enable us to collect the data that fits our research questions or hypothesis. Thus, if we are testing the hypothesis, most likely we will consider quantitative methods as our methodological/theoretical guidelines and we will base our data collection on the tools used in these methodological frameworks (questionnaires, surveys, recordings, etc.).

If we are working with research questions, most likely we will consider the instruments used in qualitative research (introspection, observation, and other). Qualitative research can often lead to theory building, i.e., from our results we generate new ideas or hypotheses. Such a type of research is *inductive*. It is common to combine quantitative and qualitative methods in sociolinguistic research, as we will see later. Instruments (questionnaires, recordings, interviews, surveys, observations) you choose have to provide you with the data that will enable you to answer the research question / or test the hypothesis. Many different techniques, instruments and methods used in sociolinguistic research will be presented in the following chapters.

2.3.3. Working with the data

Data are the result or outcome of our application of a research method or technique. Data are what sociolinguists collect during their field work. Sociolinguists work with different kinds of data: speech recordings, questionnaires, surveys, images (e.g., of linguistic landscape), written texts (e.g., diaries, letters, autobiographies, and notes), quantification (e.g., data obtained during the experiments), observation notes and recordings, and other.

Work with data entails **3 steps**: *data collection*, *data processing* and *data analysis*. Data collection is a very important research process and should be carefully planned, especially in those cases when one knows that it would be impossible to replicate the data collection. The planning of data collection entails the selection and sampling of the participants (i.e., who will be respondents of your study) or of data sources (i.e., legal documents, recordings, etc.). We have to collect such data than can lead us to *generalizable, reliable, and valid results* (Moyer 2008: 29). There are different ways to collect the data and we will discuss it in more detail in the following chapters.

Once data are collected it needs to be organized and processed. When processing the data the researcher familiarizes her/himself with what s/he has obtained. Processing includes transcription, coding, and quantifying the data. The data also have to be organized according to certain categories, especially if the research deals with different types of data, for instance, recordings, documents, field notes, etc. Speech data often needs to be coded. Thus, the coding system also has to be pre-defined before the data processing begins. Data analysis begins only after the processing is completed. Analysis is always concerned with “searching for explanation” (Moyer 2008: 30). During the analysis we combine our results with the theoretical and conceptual framework in order to explain the results and provide our interpretations. Once analysis is completed, the results of the study have to be presented. The type of presentation depends on the research project: student projects usually grow into term or research papers, which then can be turned into a research article or conference presentation.

Example #1.

Aimee C. Knupsky & Natalie M. Nagy-Bell, “Dear Professor: The Influence of Recipient Sex and Status on Personalization and Politeness in E-Mail”, *Journal of Language and Social Psychology* 30 (1), 103–113, 2011.

The topic: personalization and politeness in e-mail.

Previous research showed that women display greater interpersonal sensitivity, social warmth, emotionality in their e-mails. More first-person singular pronouns were used when writing to opposite sex recipients and women used more intensifiers in e-mails to men. Most of the previous studies suggested that sex of the sender and recipient affect e-mail personalization. However, most of them analyzed e-mails written among friends. Previous research did not look at personalization and politeness in e-mail sent to different audiences and for different purposes.

Therefore, the focus of Knupsky’s & Nagy-Bell’s article is on personalization (formality and expressiveness) and politeness of e-mail requests written by female undergraduates to male and female peers or professors, i.e., the study compared personalization and politeness of e-mails written to peers versus professors.

Narrowing down: the participants were limited to female students only.

Hypothesis: students would write more personalized e-mails to peers; however, when writing to professors, students would adjust their language and construct less personalized and more polite e-mails.

Method and data collection:

- participants – 77 female undergraduates from introductory psychology classes; ages ranged from 18 to 22 years;
- total of 66 e-mails were analyzed: 16 written to male peer, 16 to female peer, 17 to male professor, and 17 to female professor;
- two different scenarios were created and presented to students. One scenario for the letter addressed to male/female professor (asking for help studying and/or study tips for the exam), and one – for male/female peer (asking for help studying and/or study tips for the exam);
- researchers created e-mail accounts using Yahoo!Mail and participants used one of these to send e-mails. Participants received instructions on how to log into Yahoo!Mail and send the message;
- participants were not told about the goals of the study. Participants were tested in a computer lab. Each session was randomly assigned and e-mails to the same recipient were sent in one the same session. Testing took approx. 15 minutes.

Data processing:

- e-mails were coded;
- the judges counted the number of words, personal pronouns, and exclamation points, also positive and negative adjuncts (following Duthler 2006);
- e-mails were scored for overall grammar and for overall personalization using Likert-type scale (0 indicating not at all grammatical/personal and 6 indicating very grammatical/personal). For grammar judges considered punctuation, capitalization, spelling, misused words, and sentence construction. For personalization – formality, disclosure, and word choice. Flesch Reading Ease Readability (FRE) and Flesch-Kincaid Readability (FKGL) scores were calculated to assess writing complexity (the lower FRE scores, the greater complexity, the higher FKGL scores, the greater complexity).

Results:

Table 5: Means and Standard Deviations for Continuous Variables across Sex and Status of the Recipient (adapted from Knipsky & Nagy-Bell 2011: 108)

Variables	Male		Female	
	Peer	Professor	Peer	Professor
Number of words	95.19 (32.92)	82.41 (18.76)	110.13 (45.64)	91.11 (31.47)
Personal pronouns (per word)	6.14 (0.92)	5.49 (0.84)	5.65 (1.12)	4.92 (1.13)
Exclamation points (per word)	1.04 (1.40)	0.40 (0.86)	0.63 (0.45)	0.28 (0.33)
Personalization rating	3.75 (1.28)	2.65 (0.81)	4.00 (0.93)	4.85 (1.07)
Grammatical rating	4.69 (1.24)	5.35 (0.61)	4.03 (1.63)	4.85 (1.07)
Number of positive adjuncts	3.16 (1.17)	2.56 (0.90)	3.38 (1.30)	2.44 (1.00)
Number of negative adjuncts	1.53 (1.02)	0.94 (0.73)	1.88 (0.96)	1.41 (1.52)
FRE	68.38 (5.13)	58.80 (9.60)	67.11 (7.76)	60.09 (8.91)
FKGL	7.26 (0.83)	9.38 (2.11)	7.93 (1.58)	9.35 (2.11)

FRE – Flesch Reading Ease Readability, FKGL – Flesch Kincaid Readability

Discussion: Female participants tended to use greetings more likely when writing to male peers rather than female. This is the only significant affect that sex had on e-mails'

style. However, recipients’ status had a great affect on the linguistic style. E-mails written to professors exhibited less markers of personalization. Formal greetings were used more often in the letters written to professors.

2.4. Comprehension activities

Activity 2. Your research of investigation is the use of English language in higher education in the Baltics. Think of at least two hypotheses and two research questions for this topic. How will you collect the data?

Fill out the table below with your ideas.

Research question/ Hypothesis	Data needed (what data will you need)	Data collection (how will you collect the data)

Activity 3. Read the synopsis of Knupsky’s & Nagy-Bell’s article provided above. Find and read the full article. Answer the following questions:

A) What kind of research the article presents – quantitative or qualitative? Explain your answer.

B) Why researchers chose to raise a hypothesis rather than asking a research question? How did this choice influence the selection of methodology?

C) Try to *reformulate* their hypothesis into the research question. Where will you shift your focus? Will you choose the same methods as Knupsky and Nagy-Bell in order to answer your research question?

D) Imagine that you have to replicate this study in a different context, for instance, you are interested in the personalization and politeness in e-mails among Latvian and Lithuanian speakers. Based on Knupsky’s and Nagy-Bells’ as well as other studies *formulate a focus of your research and a hypothesis*. Indicate what data you will need to test your hypothesis and how will you collect it.

Focus of a research:	
Hypothesis	
Data needed	
Data collection	

2.5. Glossary

Data – the result or outcome of the application of a research method or technique.

Research – a systematic investigation into or study of materials and sources in order to establish facts and reach new conclusions.

Theory – a type of thinking or a result of thinking, a “proven” hypothesis.

2.6. References

- Aimee C. Knuspsky & Natalie M. Nagy-Bell, 2011. "Dear Professor: The Influence of Recipient Sex and Status on Personalization and Politeness in E-Mail", *Journal of Language and Social Psychology* 30 (1), 103–113.
- Melissa G. Moyer, 2008. "Research as Practice: Linking Theory, Method, and Data", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 18–31.
- Sebastian M. Rasinger, 2008. *Quantitative Research in Linguistics. An Introduction*, London, New York: Continuum.
- Jane Sunderland, 2010. "Research Questions in Linguistics", in *Research Methods in Linguistics*, ed. by Lia Litosseliti, London, New York: Continuum, 9–28.
- Ruth Wodak, Paul Kerswill & Barbara Johnstone, 2010. "Introduction", in *The Sage Handbook of Sociolinguistics*, ed. by Ruth Wodak, Barbara Johnstone and Peter Kerswill, London: Sage Publications Ltd., 1–8.

CHAPTER 3:
STUDY DESIGN: CROSS-SECTIONAL, LONGITUDINAL,
CASE GROUP, EXPERIMENTAL, QUASI-EXPERIMENTAL

3.1. Key concepts

- Apparent time*
- Case study*
- Cohort*
- Conditions*
- Control group*
- Cross-sectional*
- Experimental group*
- Fillers*
- Group study*
- Longitudinal*
- Null hypothesis*
- Panel design*
- Quasi-experimental*
- Research design*
- Subject pool*
- Target tokens*

3.2. Pre-reading activity

Activity 1. In groups discuss and define the concepts given above. What could be their role in research? What could be their role in sociolinguistic research?

.....

.....

.....

.....

3.3. Study designs

The **main aim** of this chapter is to focus in more detail on different types of research or study designs used for sociolinguistic research.

Research design, i.e., the actual organization of the research, strongly depends on a number of factors, most importantly on the research questions asked at the beginning. According to Trochim (2006)¹ there are three types of research questions:

1 William M. Trochim. The Research Methods Knowledge Base, 2nd Edition. Internet WWW page, at URL: <<http://www.socialresearchmethods.net/kb/>> (version current as of October 20, 2006).

- a. descriptive (when the research intends to describe what is going on or what exists);
- b. relational (when the research intends to look at the relationship between two or more variables);
- c. causal (when the research intends to determine whether there is a cause and affect relationship between variables).

Thus, the studies can be descriptive, relational or causal. **Temporal order**, **number of subjects**, and **manipulation of variables** determine the types of research designs. In terms of temporal order (or duration of study) there are several types of research designs: *cross-sectional*, *longitudinal* and *panel*. In terms of the number of subjects to be included, research can be designed as *case* or as *group study*. In terms of the manipulation of variables a study can be *experimental* or *quasi-experimental*. Each of these research designs are briefly presented below.

3.3.1. Cross-sectional designs

In cross-sectional studies, a comparatively large amount of data is collected at one particular point of the time. Cross-sectional studies provide “an overview of how a particular variable is distributed across the sample at a particular moment in time” (Rasinger 2008: 36). Cross-sectional studies investigate the relationship between variables. Subjects (respondents, speakers) usually are grouped according to different variables. Their performance and linguistic behavior then is sampled and differences or similarities in their behavior and performance are analyzed.

Example #1.

A typical example of cross-sectional study is Labov’s study on the social stratification of /r/ in New York department stores (Labov [1972]). In one day Labov visited three different department stores in the city of New York. The three stores were selected based on the socio-economic ranking: Saks Fifth Avenue (the highest ranking store, most of the customers in this store are of upper-middle or upper social class), Macy’s (the middle ranking store), and S. Klein (the lowest ranking store). Based on preliminary research by Labov, the advertising and prices of the store were clearly stratified, for instance, Saks did not mention prices in its advertising, while Macy’s advertised prices in large type, and Klein emphasized retail prices.

Once the interviewer entered the store, he located the department that was on the fourth floor, and after he approached a sales person in the store he asked “Excuse me, where are the [women’s shoes]?” followed by the answer “Fourth floor”. Then the interviewer would clarify the answer in order to elicit another utterance said in a careful style under emphatic stress. In 6.5 hours Labov was able to interview 264 respondents. Thus, he was able to collect the utterance with /r/ across different groups of respondents at a given point of time. Labov’s results showed a clear and consistent stratification of /r/ in the three stores: the higher ranking the store, the more presence of /r/ was elicited.

Most of the variationist studies, unless they are preoccupied with process of language change, are cross-linguistic. For language change studies (e.g., language acquisition studies) cross-sectional designs are not suitable. Cross-sectional studies explore the interrelationships between variables (as in Labov's case – postvocalic /r/ and social stratification), in other words, they deal with questions “what” and “when”, but they do not explore cause and effect relationships, in other words they do not attempt to answer questions “why” and “how”. Cross-sectional studies usually deal with a large sample of data in order to have a representative sample that will secure validity of the results. Cross-sectional studies can use a number of different techniques, for instance surveys, questionnaires, observations, tests, interviews, or polls.

3.3.2. Longitudinal designs

Longitudinal studies examine one or more groups of participants at several points of time. If cross-sectional studies explore only the situation at a given point of time, longitudinal studies allow researchers to observe language changes over time. Thus, longitudinal designs are especially valuable in language acquisition studies or studies exploring language changes, changes in linguistic behaviour over time. Due to the demanding nature of the data collection process longitudinal studies usually involve smaller number of subjects compared to cross-sectional designs.

Longitudinal studies can cover such topics as language changes over a period of time, age related language development and change, cause and effect relationships between variables, and others (Hua & David 2008: 93–94).

Longitudinal studies can be classified as *trend*, *panel* and *cohort studies* (Hua & David 2008: 94). This classification is based on the participants involved in the study: whether the same participants are studied at different times and how (Hua & David 2008: 94).

A *trend study* samples **different groups** of people from the same population at different points in time.

A *panel study* measures the **same sample of respondents** at different points in time. For instance, a researcher is interested in the pace of German as a second language acquisition among university students. Thus, s/he will randomly select a small number of students and then will monitor their progress at several different points of time, e.g., in a month, in 3 months, in 6 months, in 9 months and so forth. Data collected at these different points of time will allow making conclusions about the language acquisition of each individual separately, as well as drawing more general patterns of progression for the whole group.

A *cohort study* measures the **same groups** of people over time. Each group shares a particular characteristic or condition, for instance, a year of birth, a year of marriage, a year of enrollment to university; time spent learning a particular language. In the previous example, a panel study would turn into a cohort study if we take not randomly selected students, but students, who begin their language study in the 2012/2013 academic year.

Based on the number of times data are collected longitudinal studies are further described as *repeated measures* (if data are collected twice or a few times) and *time series* (if data are collected many times).

Longitudinal studies are not that common in sociolinguistic research. However, Wagner (2012: 376) suggests, that as the field of sociolinguistics matured, it has become possible for other researchers to replicate previous studies in order to carry out trend studies.

Example #2.

A good example of a longitudinal design is Sankoff's & Blondeau's (2007) study that examined a change in progress in the pronunciation of /r/ in Montreal French. The study employed both trend and panel designs sampled in 1971 and 1984 to document the real-time changes in the community as well as the changes in the /r/ pronunciation of a number of individual speakers who were observed over the period of time. For panel sample same 16 female and same 16 male speakers were interviewed in 1971 and 1984. For trend sample – 32 speakers (16 females and 16 males) were interviewed in 1971 and the same amount in 1984 (12 new respondents were added in 1984). All speakers were stratified by age, sex, and social class. For each speaker the researchers coded approx. one hundred tokens of /r/. Overall 12 400 tokens were analyzed (115 per speaker sample). The panel study revealed that the use of innovative pronunciation [R] increased over time among the same speakers. Early adopters of the innovative form are young, belong to a higher social class and are female. Late adopters are also young, but mainly male. The trend study showed a real-time change in the community: the use of the innovative form increased between younger and older respondent groups over time. Sankoff and Blondeau made a conclusion that the change in the community is a result of individual speakers, especially younger speakers, being added to the pool of majority-users of innovative form (categorical [R]). In order to understand the dynamics of a speech community, one needs to research language across the lifespan. The linguistic structure of individuals in their thirties, forties or fifties is different from their childhood or puberty.

3.3.3. Pseudo-longitudinal designs

According to Rasinger (2008: 41), cross-sectional studies in certain cases can be used to simulate longitudinal studies. We can design a synchronic study that simulates diachronic development. This approach was pioneered by Labov who proposed to use the present in order to explain the past (*the apparent time approach*). Relying on the assumption that speakers' linguistic patterns are largely fixed by early adulthood, the researcher examines the use of certain variables across different generations by treating the age as a proxy for historical time (Wagner 2012: 372). In other words, the researcher assumes that the speech of older respondents represents the language of their youth; i.e., speakers of different ages represent different times.

According to Milroy & Gordon (2003: 35–36), a typical example of an **apparent time study** is Britain's (1992) analysis of New Zealand English where he focused on the intonation contour involving sharp rise at its end. He compared speakers of three genera-

tions (aged 20–29, 40–49, 70–79). The results indicated an increase in the use of the intonation pattern across the generations: the younger the generation, the more occurrences of intonation patterns were observed. Apparent type hypothesis was supported by the real time reports from 1966 that pointed to the use of this rising contour pattern among the Maori children. In order to prove apparent type hypothesis a researcher always has to contrast it with a certain point in a real time (using the dialectological data from previous studies, atlas, and other). Generational differences can also be the cases of age-grading, where the use of a certain form is associated with a particular stage of life and these are stable patterns (Milroy & Gordon 2003: 36). Apparent time studies, however, often (due to the nature of the data collection) observe only the change across the sample, i.e., it does not observe the change at the level of an individual (Rasinger 2008: 41). Apparent time design is widely used in sociolinguistic research, especially in the paradigm of variationist sociolinguistics.

3.3.4. Experimental designs

Experimental designs are very different from cross-sectional or longitudinal designs. Latter designs are based on the data collected in the “natural” environment (“real” data), while experimental designs always rely on the “systematic and deliberate manipulation of one or more variables by the researcher” (Rasinger 2008: 42). In most of the cases, experiments are conducted in laboratories; i.e., the data is collected not in a “natural” setting.

According to Drager (forthcoming), the most important aspect of experimental design is to keep it simple. Experimental designs usually “consist of the comparison between the *experimental group* and a *control group*” (Rasinger 2008: 42).

The assignment of the respondents to experimental or control group is completely random. This is true for all experimental designs. The difference sometimes is made between “within-subject design” and “between-subject design” (Drager (forthcoming) calls it “across-subject design”). In a “within-subject design” the same group of people receive two types of stimuli, while in a “between-subject design” two (or more) groups of people receive different stimuli (Rasinger 2008: 43).

Experimental designs are quantitative in nature (Rasinger 2008: 42), therefore, one needs to have a statistically significant number of respondents in order to make empirically and statistically valid conclusions. Most of the research questions for experimental designs come from the previous research. For example, you have read somewhere, that Lithuanian (or Latvian, Estonian, German) dialect speech is often perceived by young Lithuanian speakers as incorrect (cf. Aliūkaitė & Merkytė 2008) and, therefore, stigmatized. You formulate your **experimental hypothesis**: Lithuanian dialect speech is perceived by young Lithuanian teenagers as incorrect. You can also formulate a **null hypothesis**: Lithuanian dialect speech is not perceived by young Lithuanian teenagers as incorrect.

In order to carry out the experiment you need to think, how will you conduct it: will you be using computer, questionnaires, etc. Since your hypothesis involves evalu-

ation of the speech, most likely you will conduct your experiment in a laboratory, which has computer, speakers and/or headphones. It is always the best to replicate the experiment conducted previously. For instance, Aliūkaitė & Merkytė (2008) used 5 different recordings of 5 different varieties (including standard); they played it to two groups of high-school students (46 and 40) from two different regions in Lithuania. Students had to evaluate the correctness of the speech on a scale from 1 (incorrect, ugly, non-standard) to 7 (very correct, very beautiful, standard) (see Aliūkaitė & Merkytė 2008: 262–264). Different from the classical matched-guise technique, which requires the same text and the same speaker to be used as stimulus for the respondents, in the current study 5 different texts read by 5 different speakers were used. Two respondent groups were formed: one from an urban area school and one from a rural (small town) area (that has a distinctive dialect) school. Thus, the **dependent variable** in this study is the respondent's geographical/dialectal environment (Aliūkaitė & Merkytė 2008: 264). Respondents were selected randomly, however, genders were shared between the two groups.

There are several different ways on how you could conduct the experiment:

- a. one can have two different groups of people reacting to the stimulus. For instance, Group A consists of students who learned about dialects (underwent training), while Group B consist of students, who have not;
- b. one can have the stimuli presented to the respondents in two different conditions, for instance, a dialect speech of an older women telling the stories about her youth and dialect speech in a comedy skit.

It is possible to do experiments with more than two groups, however, it is not recommended. It has been noticed, that respondents tend to get “better” as they get accustomed with the task (the “practice” effect) and worse as they got tired (the “fatigue” effect) (Wray & Bloomer 2006: 149). In order to avoid that, you can train your respondents briefly on your tasks; make your tasks short; divide your stimuli randomly through the test, so that most of them do not appear at the very beginning or end of the test (Wray & Bloomer 2006: 149). Any experiment should be run **as a pilot** first in order to check any discrepancies or mistakes that have been made.

According to Drager (forthcoming), the two most common types of experimental designs used in sociolinguistics are the **matched-guise technique** and **identification task**.

3.3.4.1. The matched-guise technique was developed for sociolinguistic research by Canadian scholar W. Lambert in the 1960s. The technique involves asking respondents to evaluate, for instance, personal qualities, of speakers whose voices are recorded. In a classical way, all of the recordings are done by the same speaker, who uses a different linguistic variety (hence, stimuli) (remember Purnell et al. study of 1999, where the same speaker used different varieties on the phone to rent the apartment). Thus, respondents evaluate personal qualities of the individuals without knowing that it is the same person. Such evaluation then enables researchers to come up with the conclusions on which language ideologies, stereotypes and prejudices are associated with which varieties. Often researchers find someone who is able to speak in

several varieties (e.g., a bilingual or bidialectal speaker). Additionally, a few other respondents, who speak just one variety or language, can be recorded as well. Their linguistic input will serve as fillers. Tokens (linguistic production: sentences, words) should be broken down to two groups and later randomized.

3.3.4.2. The identification task involves asking respondents to listen to the word and to indicate (hence, identification) what they heard (Drager, forthcoming). A researcher usually exposes the respondents to certain social information (regional data, photographs, toys) and this allows him/her to shift respondent's answers (Drager, forthcoming).

For instance Niedzielski in her study of 1999 asked whether social information about the speaker has a significant effect on speech perception. In order to answer this research question she ran an experiment based on an identification task. Forty-one Detroit area residents participated in the experiment. They were asked to “choose from a set of computer synthesized vowels the tokens that they felt best matched the vowels they heard in the speech of a fellow Detroiter” (Niedzielski 1999: 64). Each respondent listened to about 50 sentences, and they were asked to focus on the vowel in a particular word from that sentence and later matched that vowel “to a set of six computer-resynthesized vowels; and were told to choose the one that best matched the one that they heard the speaker produce” (Niedzielski 1999: 64). Half of the speakers were told that the speaker (they listened to) was from Detroit, whereas half were told that the same speaker was from Windsor, Ontario, Canada. However, all of the respondents heard the same speaker from Detroit. Thus, the researcher in this study exposed respondents to different regional information; i.e., manipulated the social data in order to shift respondent's answers.

Table 6 summarizes designs of identification tasks (based on Drager, forthcoming).

Table 6: Designs of identification task (adapted from Drager, forthcoming)

Task	Description
identification task	listeners hear tokens between which there is some distinction (e.g., <i>pin</i> vs. <i>bin</i>) and they identify what word (or other variable) they heard;
discrimination task	listeners indicate whether two words (e.g., <i>pin</i> and <i>bin</i>) sound the same or different;
commutation task	listeners are played back their own (or similar) productions; the tokens in a commutation task are more easily confusable;
coach task	listeners hear a narrative in which the target sound appears in a word that changes the interpretation of the narrative.

Advantages and disadvantages of experimental research are presented in Table 7

Table 7: Advantages and disadvantages of experimental research (adapted from Wray and Bloomer 2006: 150-151).

Advantages	Disadvantages
validity of research	some of the results may not be valid in non-experimental conditions
evidence for or against a pre-specified hypothesis	the danger of hidden variables (concerning the respondents) that a researcher may not be aware of
easy to replicate and to improve	not all respondents make good experimental subjects (i.e., children)
results easy to relate to each other and other experiments done in the same way	the danger that a subject did not understand the task properly
easier to process the data and evaluate the results	task ambiguities or difficult instructions

3.3.5. Quasi-experimental designs

Quasi-experimental designs are similar to experimental; however, Rasinger (2008: 44) lists at least two crucial differences between the two:

1. respondents are not assigned to experimental and control group randomly, but rather “naturally”.
2. in quasi-experimental studies an “external force” changes a particular variable.

Usually, any experimental study, that has “naturally given” respondents is considered quasi-experimental.

Below is the example of a hypothetical quasi-experimental design provided by Rasinger (2008: 44–45).

Imagine, for ease of comprehension that at this moment in time there was no ESOL (English as a Second Language) provision available for migrants. Now, let us also assume the government passes a law which, from next week onwards, significantly improves ESOL provision but also forces migrants to gain an English language proficiency equivalent to the Cambridge Advanced English Certificate (CAE). However, in this first instance, provision will only be provided for migrants of Pakistani descent. Hence, Pakistanis in Britain ‘automatically’ become our experimental group, that is, the group that will be subject to a change in a particular variable (namely ESOL provision), while other migrants will not undergo any changes (i.e., are our control group). We would design our fictional study around two points of data collection: measuring proficiency of both experimental group and control group before the new law comes into effect, and afterwards, let us say in 6 months time. (Rasinger 2008: 44–45)

3.3.6. Case studies

Case studies are in-depth investigations of an individual participant or a small group of participants (Hua & David 2008: 98), usually case studies are used when the circumstances or variables observed are so specific, complex (or rare) that there is no benefit in combining the results of observations of that particular person with those of others (Wray & Bloomer 2006: 170). Case studies are perfect for longitudinal design, for instance, for the investigation of language change in progress one can select one individual or a small group of individuals and observe the same people over time. This type of design would combine case and longitudinal study.

Hua & David (2008: 98) name several characteristics of case studies:

- a. case studies usually consist of a very small number of cases (respondents); there can be single case, double case or multiple case studies;
- b. case studies do not aim to investigate general patterns of a population;
- c. case studies often rely on qualitative methods of data analysis (in-depth analysis);
- d. case studies rely on naturally obtained data (experiments are rarely applied);
- e. case studies are not temporally limited or restricted.

Data for case studies can be obtained using different tools (surveys, tests, interviews, as well as secondary sources, such as reports, novels, etc.) (Hua & David 2008: 99). However, very often case studies rely on extensive observations, which provide researchers with rich and in-depth data, covering variety of different aspects. The data obtained for case studies can be analyzed in two ways: using within-case and cross-case analysis (Hua & David 2008: 100). Within-case analysis investigates the same pattern with the data for the single case, whereas cross-case analysis investigates similar pairs of cases for differences or dissimilar ones for similarities (Hua & David 2008: 101).

Table 8 summarizes the advantages and disadvantages of case study design

Table 8: Advantages and disadvantages of case studies (adapted from Hua & David 2008: 101).

Advantages	Disadvantages
multiple sources of data	limitations on generalizability
in-depth and rich data	time consuming
intensive study and study of rare, extreme cases (e.g., people after certain injuries)	susceptibility to bias
	not suitable for statistical analysis

3.3.7. Group studies

Group studies are a recent development in research design. They grew out from cases studies. Group studies focus on the investigation of a group of subjects “who are examined in parallel with each other” (Hua & David 2008: 102). Different from multiple-case studies, which investigate each respondent as a single case, group-case studies are focused on the group trend (Hua & David 2008: 102). There are several characteristics of group studies:

- they examine only one group (whereas cohort and cross-sectional studies examine several groups);
- they are aimed at comparing individuals within that one group (whereas cohort studies investigate differences between groups);
- they try to focus on the group trend without overlooking individual variation (whereas cross-sectional studies focus only on generalizations);
- as in case studies, they are not limited in time-span (whereas cohort studies have to be carried out over a period) (Hua & David 2008: 102).

An example of a group study can be Heller’s ethnographic study (2006) of one French language minority school in Ontario, Canada. A study focuses on a group of franco-phone high-school students and aims to identify what it means to be francophone and to speak French as seen through the life of the French language minority high school (Heller 2006: 17). We will discuss Heller’s study and her approach in more detail in Chapter 9.

Selection of the appropriate research design depends on a variety of factors. However, usually the **research question and hypothesis determine** what study design one needs to follow (Rasinger 2008: 45; Hua & David 2008: 106).

3.4. Comprehension activities

Activity 2. What are the advantages and disadvantages of cross-sectional design? Think of Labov’s study. What are the limitations of his study? How are these limitations predetermined by research design?

.....

.....

.....

.....

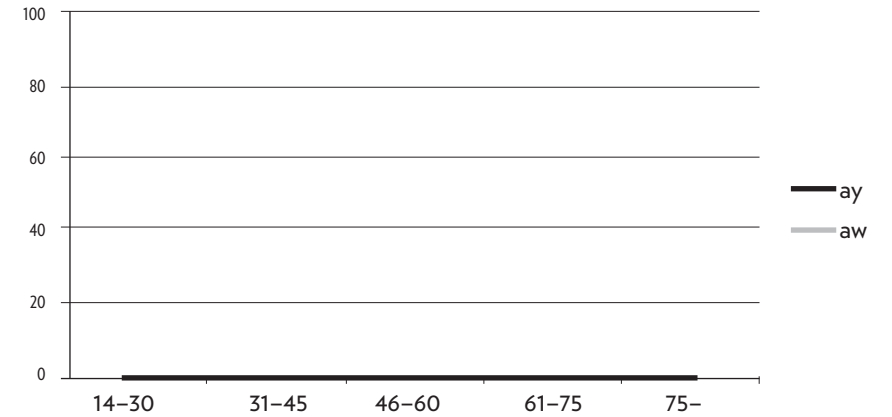
.....

Activity 3. Look at the data adapted from Labov’s Martha’s Vineyard study. Using the data draw a figure that would show the changing pattern across generations. What does the curve look like?

Table 9: Centralization of (ay) and (aw) by age level

Age	(ay)	(aw)
75–	25	22
61–75	35	37
46–60	62	44
31–45	81	88
14–30	37	46

Adapted from William Labov, *The Social Motivation of a Sound Change*, in *The Routledge Sociolinguistics Reader* (eds. M. Meyerhoff, E. Schlee), 2010, New York: Routledge, 305.



Activity 4. Discuss with your friends and create an experimental research design to test the following hypothesis: Younger women tend to have more positive attitudes toward dialect speech than younger men.

Activity 5. Consider these research questions proposed by Drager (forthcoming). Which of the experimental techniques you will choose in order to investigate the problem?

1. What social characteristics are attributed to speakers who use different styles of speaking?

- a) matched-guise;
- b) identification task.

2. Are speakers who use pauses (or quotative *like* or who giggle) rated more negatively/positively as potential friends, workmates, or lovers than speakers who don't?

- a) matched-guise;
- b) identification task.

3. Do people from different regions (or socioeconomic backgrounds) perceive sounds differently from one another?

- a) matched-guise;
- b) identification task.

4. Can listeners' expectations about a speaker's gender (or ethnicity or social class) affect their identification of sounds produced by that speaker?

- a) matched-guise;
- b) identification task.

3.5. Glossary

Apparent time – studies that attempt to investigate language changes as they happen by comparing the speech of older speakers with the speech of younger speakers in a given community.

Case study – in-depth investigation of an individual participant or a small group of participants.

Cohort – a group of subjects with a common defining characteristic or condition.

Conditions – are different groupings of experimental stimuli or different dimensions in which stimuli appear.

Control group – a group of people who do not receive the treatment.

Cross-sectional – a study design that involves observation of respondents at one specific point in time.

Experimental group – a group of people who receive the treatment, e.g., the one that is given the new medicine, etc.

Fillers – are tokens that are used to “fill in” the experiment or distract the participant, so that he does not know what is being researched; responses to fillers are not analyzed.

Group study – an investigation of a group of subjects examined in parallel with each other.

Longitudinal – a type of observation study that involves repeated observation of the same respondent over the long periods.

Null hypothesis – a hypothesis that states that the selected stimuli have no effect or significant difference.

Panel design – a form of a longitudinal study that investigates the same sample of respondents at different points in time.

Quasi-experimental – experimental type of study in which respondents are assigned to experimental and control group “naturally”.

Research design – the actual structure according to which the research (study) is organized.

Subject pool – a group of potential research participants.

Target tokens – are the tokens (sounds, words, sentences, etc.) that are the most important in the research; responses to target tokens will be analyzed to test the research question.

3.6. References

- Daiva Aliūkaitė & Erika Merkytė, 2008. “Sociolingvistic tarminių kodų taisyklingumo tyrimas”, *Kalbos kultūra* 82, 259–274.
- Katie Drager (forthcoming). “Experimental Methods in Sociolinguistics: Matched guise and identification tasks”, in *Research Methods in Sociolinguistics: A practical guide*, ed. by Janet Holmes and Kirk Hazen, Hoboken, NJ: Wiley-Blackwell.
- Monica Heller, 2006. *Linguistic Minorities and Modernity: A Sociolinguistic Ethnography*, London: Continuum.
- Zhu Hua & Annabelle David, 2008. “Study Design: Cross-sectional, Longitudinal, Case, and Group”, in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 88–107.
- William Labov, [1972, reprinted in 1997]. “The Social Stratification of (r) in New York City Department Stores”, in *Sociolinguistics: a Reader and Coursebook*, ed. by Nikolas Coupland and Adam Jaworski, New York: Palgrave, 168–178.
- William Labov, [1972, reprinted in 2010]: *The Social Motivation of a Sound Change*, in *The Routledge Sociolinguistics Reader*, ed. by Miriam Meyerhoff & Erik Schleeff, New York: Routledge, 292–322.
- Lesley Milroy & Matthew Gordon, 2003. *Sociolinguistics: Method and Interpretation*, Malden, Oxford, Melbourne, Berlin: Blackwell Publishing.
- Nancy Niedzielski, 1999. “The Effect of Social Information on the Perception of Sociolinguistic Variables”, *Journal of Language and Social Psychology* 18(1), 62–85.
- Thomas Purnell, William Idsardi & John Baugh, 1999. “Perceptual and phonetic experiments on American English dialect identification”, *Journal of Language and Social Psychology* 18, 10–30.
- Sebastian M. Rasinger, 2008. *Quantitative Research in Linguistics. An Introduction*, London, New York: Continuum.
- Gillian Sankoff & Hélène Blondeau, 2007. “Language Change across the Lifespan: /r/ in Montreal French”, *Language* 83, 560–588.

- William M. Trochim, 2006. *The Research Methods Knowledge Base*, 2nd Edition. Internet WWW page, at URL: <<http://www.socialresearchmethods.net/kb/>>.
- Suzanne Ewans Wagner, 2012. "Age Grading in Sociolinguistic Theory", *Language and Linguistic Compass* 6/6, 371–382.
- Alison Wray & Aileen Bloomer, 2006. *Projects in Linguistics. A Practical Guide to Researching Language*, London: Arnold Hodder.

CHAPTER 4:
DATA: TYPES OF DATA AND SAMPLING

4.1. Key concepts

- Ethnographic approach*
- Interview*
- Population*
- Questionnaire*
- Random or probabilistic sampling*
- Representativeness*
- Sample*
- Sampling*
- Sampling error*
- Sampling techniques*
- Snowball or social network sampling*
- Stratified random sampling*
- Systematic sampling*

4.2. Pre-reading activity

Activity 1. In groups discuss and describe the concepts given above. What do you think could be the role of each in research design and why are they important for sociolinguistic research?

.....

.....

.....

.....

.....

4.3. Data

The **main aim** of this chapter is to explore in more detail different types of data used in sociolinguistic research.

Sociolinguistic research is usually based on **empirical data** obtained through observation (Milroy and Gordon 2003: 2). Different from other linguistic subfields, sociolinguists rarely rely on the data obtained through introspection (for instance, generative

linguistics relies on this type of data). The “nuts and bolts” of sociolinguistic research is the research of “real” or “actual” data, i.e., the data that represents speaker’s actual, real performance (spontaneous speech, vernacular). Speaker’s actual performance is the data most widely used by **variationists** (e.g., Labov, Trudgill, Milroy, Tagliamonte). Depending on the research question, however, sociolinguists might need to obtain other type of data. As Milroy and Gordon argue (2003: 49): “what constitutes ‘good data’ depends on the research objectives, as do the methods for collecting such data.” Further we will discuss types of data employed by sociolinguists and sampling techniques. Different methods of data collection will be discussed in the following chapters.

4.3.1. Types of data

Schleef & Meyerhoff (2010: 3) list **four** kinds of different data used for a sociolinguistic study:

1. Naturally occurring data. This is a speech that occurs naturally, thus, independently whether a researcher would have recorded it or not. Examples of such data could be everyday conversations, radio or TV broadcasts, theater plays, etc. This type of data provides researcher with a great source of unmonitored speech, thus, actual, real performance. Naturally occurring data are widely used in variationist studies. However, if you are interested only in a particular variable (the use of certain lexeme) you might need to collect a lot of naturally occurring data in order to get information you need (i.e., the lexeme you are interested can be used just once in an hour-long conversation, thus, you might need much more data in order to have a representative sample).

2. Interview data. Interviews allow researcher to control the elicitation of data. Interviews can be audio and video recorded. Usually a researcher prepares a set of questions that s/he wants to address in his/her study and during the interview s/he asks these questions. Interviews usually are topic-based. An implication for a classic sociolinguistic interview is that it should sound as natural as possible, thus, questions should be posed as naturally as possible in order to give a feeling of an informal chat. We will discuss the structure of the Labovian interview in more detail in Chapter 8. Interviews enable the researcher to extract required linguistic variable and control the context in which these variables occur (Schleef & Meyerhoff 2010: 4). However, interview data will never be as natural as naturally occurring conversation.

3. Questionnaire data. Questionnaires are usually used not to elicit actual language data but “to collect data on attitudes about language or qualitative sociolinguistic information” (Schleef & Meyerhoff 2010: 4).

4. Experimental data. This type of data is elicited in an experimental setting (on experiments see Chapters 1 & 3). Matched-guise tests are most commonly used in sociolinguistic research. As questionnaire data, experimental data usually do not elicit actual language data but **data about language**.

Thus, if your research design involves investigation of an actual speech, most likely you will need to rely on the naturally occurring or interview data. If your research de-

sign involves investigation about language attitudes, language use and language perception, most likely you will need to rely on the questionnaire or experimental data.

Other sources of data can also be included in sociolinguistic studies:

- a. *written or spoken language corpora*, for instance, historical sociolinguistic studies heavily rely on written corpora. These corpora are designed in such a way that they enable researchers to perform sociolinguistic analysis. One example of such corpora is *Corpus of an Early English Correspondence*², compiled at the University of Helsinki. It is a large corpus that covers English correspondence from the 1400s until 1800s. The linguistic data (written language) is supported with an extensive database that contains background (social) information about letter writers;
- b. *official documents* (remember Wodak's, Krzyżanowski's & Forchtner's study of 2012);
- c. *other written sources*: novels (for code-switching, use of borrowings), newspapers (studies of advertisements, headlines), Internet (forum sites, commentary sections, etc.). Bilingual forum sites, according to Nortier (2008: 50), provide researchers with a handy set of bilingual data. The advantage of forum sites is that they usually do not require a password or registration. The material is also often "more informal than in other written sources" (Nortier 2008: 50). The drawback of forum data is that it usually lacks background information about the participants ("speakers"). Such information is essential for sociolinguistic study.
- d. *census data*. Censuses not only help to reveal linguistic information (which language/s are used at home, how many foreign languages people know, what is the distribution of different native tongues, and other), but also provide researchers with other useful social information. Censuses can include information on the linguistic proficiency of informants (how well do they know a certain language, etc.). Census data are always analyzed quantitatively (using computer software), thus, the questionnaires usually include multiple-choice questions; yes/no questions, or evaluation scales (Nortier 2008: 36). Censuses and other demographic data on a particular region provide researcher with background information on linguistic, ethnic composition of a particular area. As Nortier (2008: 38) concludes, "surveys can give information about the scale on which certain languages are spoken and to a certain degree about proficiency in the languages involved".

As Schleef & Meyerhoff (2003: 5) put it, "many sociolinguistic studies combine data from different sources, a practice we recommend as it allows you to triangulate your data analysis and explore an issue from different angles".

4.3.2. Sampling: key concepts

Any kind of empirical research has to deal with a sampling issue. Sampling involves selection of the participants for our research. In other words, before we start collecting

2 See: <http://www.helsinki.fi/varieng/domains/CEEC.html>.

the data, we have to decide *whom we would like to research*. Rasinger (2008: 45) lists several important concepts related to sampling: *population*, the *sample* and *sampling techniques*. Population, according to him, can be defined as a group of people we are generally interested in (e.g., Latvian speakers living in Lithuania, Lithuanians living in Chicago-area, Germans in Estonia in the 19th century, etc.) and who share certain characteristics (Rasinger 2008: 45–46). Often in sociolinguistic research *population* is related to minority or immigrant communities, dialect speaker communities, urban communities, etc. Since it is impossible to collect the data from “all of the population,” it is usually better to go smaller and minimize the population to manageable numbers (e.g., first generation Latvian speakers living in Vilnius, or teenage Lithuanian speakers in Chicago-area, etc.).

“Populations are usually too large to be studied in their entirety” (Rasinger 2008: 47), in other words, it might be impossible to collect the data from all of the teenage Lithuanian speakers living in Chicago-area. Therefore, one needs to break down the population into smaller groups. These groups are defined as *sample*. According to Rasinger (2008: 47), “a sample is a part of our population, and crucially, the sample must be an adequate reflection of our population”. To put it simple: the *sample* represents the *population*.

Sampling (choosing the right representatives of our population) is a difficult task: the larger the population the more heterogeneous it is, the smaller the population the more homogeneous it is. Thus, when we draw the respondents for our study, we want to keep the balance (proportion): if our population is heterogeneous, we want to have heterogeneous sample and *vice versa*. Let us assume that we have 100 American-Lithuanians, 50 males and 50 females. If we select a sample of 20 American-Lithuanians, it has to maintain the same proportions and have 10 males and 10 females.

According to Rasinger (2008: 48), the most valid sampling technique is *random* or *probabilistic sampling*. Random sampling is based on the idea that “anyone within the sample frame has an equal chance of being selected” (Milroy & Gordon 2003: 25). Random sampling technique is similar to the lottery (or name-in-a-hat technique): if we have 100 different names in a hat and we draw only 25, any name has an equal chance to be drawn. *Systematic sampling* is based on the selection according to certain criteria, e.g., we decide to choose every second or every fifth name in a phone book, without paying attention to whom we select (Rasinger 2008: 49).

When we deal with sampling, it is inevitable to encounter *sampling errors*. “Sampling error is the phenomenon whereby our sample does not have the same characteristics as our population” (Rasinger 2008: 49). Statistically, *sampling error* can be defined as the *difference* between the sample and population values. For instance, if our population includes 100 American-Lithuanians, 50 males and 50 females, we draw a random sample and get 15 females and 5 males. Such sample does not fully represent our population; however, we have a small sampling error and limited representativeness. If we draw a random sample and get 20 females and no males, we will get a large sampling error and no representativeness. In other words, if our research

question involves gender differences, we cannot rely on the last sample. On the other hand, if gender is not crucial to our research, the sample might fit us.

Smaller case studies sometimes employ non-probabilistic sampling. Such sampling is not random and non-probabilistic, since not all members of the population can become a member of the sample; members are deliberately chosen by the researcher (Rasinger 2008: 51). The most popular of the non-random samples are *opportunity* or *convenience samples* (Rasinger 2008: 51). Participants are chosen on the basis of who is available at the given time and place. *Snowball* or *social network sampling*, that is discussed further, is also a type of non-random sampling technique.

4.3.3. Sampling techniques in sociolinguistics

Different sampling techniques are used in sociolinguistic studies. Schleeff & Meyerhoff list at least four most commonly applied in sociolinguistic research (2010: 6):

4.3.3.1. Random sampling. It is used when “the goal of study is to provide a snapshot of variation in a city (*or other linguistic community* – A. T. and V. B.) as a whole” (Schleeff and Meyerhoff 2010: 6). As discussed previously, in random sampling every person in a defined population has an **equal chance** to be interviewed. According to Tagliamonte (2006: Kindle Location 268), “[r]andom sampling means that the fieldworkers do not know the individuals they are talking to. In fact, interviewer and interviewee are usually strangers to one another and, since most data collection endeavours only interview a speaker once, the interaction is limited and represents a ‘one-off’”. Random sampling often involves *sample frame*, namely “any list which enumerates the relevant population, simple examples being electoral registers and telephone directories” (Milroy & Gordon 2003: 25). A researcher then assigns random numbers to the names on a list and draws every fifth, sixth etc. (i.e., *systematic sampling*) individual from the frame (Milroy & Gordon 2003: 25). A typical example of random sampling study in sociolinguistics is Labov’s Philadelphia study. He conducted a telephone survey of “individuals drawn from a random selection of telephone listings” (Labov 2001: 40). Telephone survey results supported the neighborhood survey which was designed in a different way, involving face-to-face interviews.

Here is how Labov explains the procedure:

The sample for the telephone survey was constructed by a random selection from telephone listings in the Philadelphia directory. The number was called, and the first person who responded was asked if he or she would like to help in a brief survey of Philadelphia English. [...] The telephone interview lasted from 15 to 30 minutes, as compared to 45 to 90 minutes for the neighborhood study. (Labov 2001: 69)

A total of 60 respondents were randomly selected and interviewed for the telephone survey.

According to Tagliamonte, “many problems arise in applying strict random sampling methodology to sociolinguistic studies. Perhaps the most important of these is the of-

ten extreme difficulty in finding the subsection of the population that you wish to study, whether this is due to socioeconomic, ethnic or other demographic reasons” (2006: Kindle Locations 305–306). Moreover, post-Labovian research discovered that “sampling methods for studies in the speech community did not actually require traditional random sampling” (Tagliamonte 2006: Kindle Location 306). Even small samples sometimes can be sufficient to indicate linguistic variation or ongoing language change in a community. The diverse nature of the linguistic community also limits the representativeness of the population. Thus, “sampling methods in variation analysis were modified to embrace the most relevant strategies of random sampling alongside more anthropological approaches” (Tagliamonte 2006: Kindle Location 313).

4.3.3.2. Stratified random sampling. The sample is “manipulated according to several secondary variables” (Schleef & Meyerhoff 2010: 7), in other words, sociolinguists stratify their sample based on different social factors (gender, age, social class, etc.). The procedure of the stratified random sampling is as follows (following Schleef & Meyerhoff 2010: 7):

- a. a researcher divides the population into strata, namely individuals that are important to the study;
- b. a researcher collects the data from each of the group and combines the data with the samples from other groups in order to form a full sample.

Tagliamonte (2006: Kindle Locations 317–318) distinguishes two fundamental practices for sampling: “the researcher 1) identifies in advance the **types of speakers** to be studied; and 2) seeks out a quota of speakers who fit the specified categories.” According to her, “a minimum requirement for any sample is that it has a degree of representativeness on the bases of **age, sex**, and (some way of determining) **social class, education level**, or both” (Tagliamonte 2006: Kindle Locations 319–320, *emphasis added*).

A typical example of a sociolinguistic research, using stratified random sampling, is Trudgill’s (1974) study of speakers in Norwich. How did Trudgill form his sample? First, he took the local register of electors (of the city of Norwich). Trudgill selected four wards (electoral neighborhoods) which resembled the city as a whole. The wards were not selected at random, but they were chosen based on social, geographical, economic and housing characteristics that would better represent the city as a whole (Trudgill 1974: 22). Equal number (25) of informants was chosen from each ward (Trudgill also added a fifth suburban ward for the comparison). Out of the 125 people contacted for interviews some were not from Norwich, several had moved or had died, several refused to be interviewed etc. Therefore, the final sample of the informants consisted of 60 people (Trudgill 1974: 26–27). Trudgill selected his informants randomly; on the other hand, he applied social and geographical criteria in order to form his sample groups (he stratified his groups).

It is always useful to create a template for sample design (especially for the study involving stratified random sampling). Let us imagine that we have to conduct a study on the use of formal and informal “you” among 20 and 60 year-old Lithuanians liv-

ing in Kaunas. We want our sample to be balanced in terms of age, sex, and education level. First we have to set up our age groups. Age stratification depends on the research question, some of the studies might require smaller age interval, whereas some might be satisfied with larger interval. Let us agree that we form five different age groups and we interview people of two educational levels: high school education and university education. Table 10 shows our sampling grid.

Table 10: Sampling grid

Age/Gender	High school education		University education	
	Female	Male	Female	Male
20-30	5	5	5	5
31-40	5	5	5	5
41-50	5	5	5	5
51-60	5	5	5	5
Subtotal	20	20	20	20
Grand total				80

It is usually recommended to have at least 5 or 6 respondents in each of the group. Thus, for this type of study we might need at least 80 respondents. If we divide speakers into age groups using 5 year intervals, our grand total will rise to even larger amount of respondents.

4.3.3.3. Ethnographic research. Ethnographic approach is very different from the first two. According to Schleef & Meyerhoff (2010: 7), “in the course of data collection, the researcher often develops personal associations with member of a community and becomes a member of it, i.e., collects data as a participant-observer. According to Tagliamonte, “participant observation, evolving from anthropological linguistic studies, is when the analyst integrates themselves within the community under investigation, either by engagement in local affairs and/or developing personal associations with members” (2006: Kindle Locations 274–275). Usually ethnographic approach, namely participant observation, is combined with other research techniques. There are no strict sampling methods associated just with ethnographic research. Selection of participants in ethnographic research sometimes is a later process. The researcher first enters the community, gets acquainted with it and then selects his/her respondents. Eckert had put it very nicely: “while survey fieldwork focuses on **filling in a sample**, ethnographic fieldwork focuses on **finding out what is worth sampling**.” (Eckert 2000: 69, quoted from Tagliamonte 2006: Kindle Locations 281–282, *emphasis added*).

4.3.3.4. “Friend of a friend” or social network sampling. This sampling method focuses on the study of some pre-existing social group rather than individual speaker (Tagliamonte 2006; Schleef & Meyerhoff 2010: 7). Using “friend of a friend” technique researchers establish a circle (social network) of contacts – respondents, who later are interviewed (or observed, recorded). As the researcher establishes this circle of people, usually s/he begins participant observation. Large amounts of spontaneous speech usually are collected during the observations.

“Friend of a friend” principle is very useful in any kind of research, especially when the researcher enters the community s/he is not familiar with. Tagliamonte (2006: Kindle Locations 295–300) nicely puts that:

In the social network method the investigator must find a means of approaching a group to which the investigator (typically) may have no pre-existing personal ties. Techniques that can be used for a circle of personal friends are too limited. This is when the ‘friend of a friend’ becomes most useful. These are people with a status that is neither that of an insider nor that of outsider, but something of both. With a ‘friend of a friend’ you do not go into a situation cold. You have some ‘in’ into the situation. Naming yourself a ‘friend’ means that you have an entry into the relationships of the network you have attached yourself to.

As in ethnographic approach, a fieldworker, using a “friend of a friend” technique, becomes part of the community – an observer and a participant at the same time. When making contacts in a speech community, it is always important to contact the right people. Tagliamonte (2006: Kindle Location 302) recommends to avoid contacting people with official status, e.g., priests, teachers, community leaders, because the linguistic production of the people in their social networks may be homogeneous and reflect standard speech styles (sociolinguists usually aim for non-standard speech). Thus, your sample then will not be representative of the whole community.

A classical sociolinguistic study, that developed the use social network approach in sociolinguistics, is Milroy & Milroy study on language variation and change in Belfast (see, for instance, Milroy 1992). We will refer to it later in Chapter 10.

Before we proceed to techniques of data collection, here is a checklist recommended by Tagliamonte (2006: Kindle Locations 329–330):

- Identify the target population.
- Determine where they are likely to be found. Engage in extensive background reading, demographic and archival research.
- Circumscribe the location of the speech community.
- Determine its boundaries.
- Devise appropriate means to ‘enter the community’.

4.4. Comprehension activities

Activity 2. Take a look at an excerpt from the Lithuanian Census Questionnaire³ (of 2011). Discuss the questions with your friends. Which data will be valuable for sociolinguistic research? Why? Would you reformulate some of the questions? Which and why?

3 Questionnaire retrieved from: <http://osp.stat.gov.lt/en/surasyms-lapas>.

PART II. DATA ON THE PERSON

<p>Name</p> <div style="border: 1px solid black; height: 15px; width: 100%;"></div> <div style="border: 1px solid black; height: 15px; width: 100%;"></div> <p>Surname</p> <div style="border: 1px solid black; height: 15px; width: 100%;"></div> <div style="border: 1px solid black; height: 15px; width: 100%;"></div> <p>Personal number </p> <p>1. Your sex:</p> <p><input type="checkbox"/> male <input type="checkbox"/> female</p> <p>2. Your date of birth:</p> <div style="display: flex; align-items: center; gap: 10px;"> <div style="border: 1px solid black; width: 30px; height: 20px;"></div> <div style="border: 1px solid black; width: 30px; height: 20px;"></div> <div style="border: 1px solid black; width: 30px; height: 20px;"></div> <div style="border: 1px solid black; width: 30px; height: 20px;"></div> </div> <div style="display: flex; justify-content: space-around; width: 100%;"> year month day age, in years </div> <p>3. Your place of birth:</p> <div style="display: flex; align-items: center; gap: 10px;"> <div style="border: 1px solid black; padding: 2px;">Lithuania</div> <div style="font-size: 2em;">}</div> <div style="display: flex; flex-direction: column; gap: 5px;"> <input type="checkbox"/> urban area <input type="checkbox"/> rural area </div> </div> <div style="display: flex; align-items: center; gap: 10px; margin-top: 5px;"> <input type="checkbox"/> abroad <div style="font-size: 1.5em;">↓</div> </div> <div style="border: 1px solid black; width: 100%; height: 15px;"></div> <div style="text-align: center; font-size: 0.8em;">country</div> <p>4. What is your marital status?</p> <p><i>To be answered by a person aged 15 and older</i></p> <div style="display: flex; align-items: center; gap: 10px;"> <div style="display: flex; flex-direction: column; gap: 5px;"> <input type="checkbox"/> married <input type="checkbox"/> divorced <input type="checkbox"/> widowed </div> <div style="font-size: 2em;">}</div> <div style="border: 1px solid black; width: 50px; height: 20px;"></div> <div style="display: flex; flex-direction: column; gap: 5px;"> year of registration of the first marriage </div> </div> <div style="display: flex; align-items: center; gap: 10px; margin-top: 5px;"> <input type="checkbox"/> never married <div style="font-size: 1.5em;">↓</div> </div> <p>5. How many children have you given birth to?</p> <p><i>To be answered by a WOMAN aged 15 and older</i></p> <div style="display: flex; align-items: center; gap: 10px;"> <div style="border: 1px solid black; width: 30px; height: 20px;"></div> <div style="border: 1px solid black; width: 30px; height: 20px;"></div> <div style="border: 1px solid black; width: 30px; height: 20px;"></div> </div> <div style="display: flex; align-items: center; gap: 10px;"> number of children <input type="checkbox"/> none </div> <div style="display: flex; align-items: center; gap: 10px; margin-top: 5px;"> <div style="font-size: 1.5em;">↓</div> </div> <div style="border: 1px solid black; width: 50px; height: 20px;"></div> <div style="display: flex; align-items: center; gap: 10px;"> year year of birth of your first child </div>	<p>6. What is your citizenship?</p> <p><input type="checkbox"/> Lithuanian</p> <p><input type="checkbox"/> other</p> <div style="text-align: center; font-size: 1.5em;">↓</div> <div style="border: 1px solid black; width: 100%; height: 15px;"></div> <div style="border: 1px solid black; width: 100%; height: 15px;"></div> <p><input type="checkbox"/> stateless</p> <p style="text-align: center; font-weight: bold; margin-top: 10px;">MIGRATION</p> <p>7. Where did you live one year prior to the census (on 1 March 2010)?</p> <p><input type="checkbox"/> in the same locality</p> <p><input type="checkbox"/> in another locality within Lithuania</p> <div style="text-align: center; font-size: 1.5em;">↓</div> <p><input type="checkbox"/> in an urban area</p> <div style="border: 1px solid black; width: 100%; height: 15px;"></div> <div style="text-align: center; font-size: 0.8em;">city/town</div> <p><input type="checkbox"/> in a rural area</p> <div style="border: 1px solid black; width: 100%; height: 15px;"></div> <div style="text-align: center; font-size: 0.8em;">municipality</div> <div style="border: 1px solid black; width: 100%; height: 15px;"></div> <div style="text-align: center; font-size: 0.8em;">ward (seniūnija)</div> <p><input type="checkbox"/> abroad</p> <div style="border: 1px solid black; width: 100%; height: 15px;"></div> <div style="text-align: center; font-size: 0.8em;">country</div> <p>8. Have you ever lived abroad for a year or longer?</p> <p><input type="checkbox"/> yes ⇒ question 9</p> <p><input type="checkbox"/> no ⇒ question 13</p> <p>9. Have you ever lived abroad for a year or longer?</p> <div style="border: 1px solid black; width: 100%; height: 15px;"></div> <div style="text-align: center; font-size: 0.8em;">country</div> <p>10. When did you arrive in or return to Lithuania for the last time?</p> <div style="display: flex; align-items: center; gap: 10px;"> <div style="border: 1px solid black; width: 30px; height: 20px;"></div> <div style="border: 1px solid black; width: 30px; height: 20px;"></div> <div style="border: 1px solid black; width: 30px; height: 20px;"></div> </div> <div style="display: flex; align-items: center; gap: 10px;"> year <div style="display: flex; flex-direction: column; gap: 5px;"> <input type="checkbox"/> arrived ⇒ question 12 <input type="checkbox"/> returned ⇒ question 11 </div> </div> <p>11. What was the main reason for departure?</p> <div style="display: flex; justify-content: space-between;"> <div style="display: flex; flex-direction: column; gap: 5px;"> <input type="checkbox"/> work <input type="checkbox"/> studies </div> <div style="display: flex; flex-direction: column; gap: 5px;"> <input type="checkbox"/> family reasons <input type="checkbox"/> other </div> </div> <p>12. What was the main reason for arrival in or return to Lithuania?</p> <div style="display: flex; justify-content: space-between;"> <div style="display: flex; flex-direction: column; gap: 5px;"> <input type="checkbox"/> work <input type="checkbox"/> studies <input type="checkbox"/> family reasons </div> <div style="display: flex; flex-direction: column; gap: 5px;"> <input type="checkbox"/> wish to live in the native land <input type="checkbox"/> other </div> </div>
---	---

ETHNO-CULTURAL DATA	EDUCATION
<p>13. What is your ethnicity?</p> <p><input type="checkbox"/> Lithuanian <input type="checkbox"/> Polish <input type="checkbox"/> Russian <input type="checkbox"/> another ↓</p> <p><input type="text"/></p> <p><input type="checkbox"/> not indicated</p> <p>14. What is your mother tongue?</p> <p><input type="checkbox"/> Lithuanian <input type="checkbox"/> Polish <input type="checkbox"/> Russian <input type="checkbox"/> another ↓</p> <p><input type="text"/></p> <p><input type="text"/></p> <p><input type="checkbox"/> not indicated</p> <p>15. What other languages do you know, i. e. are able to speak and (or) write?</p> <p><input type="checkbox"/> Russian <input type="checkbox"/> Polish <input type="checkbox"/> Spanish <input type="checkbox"/> English <input type="checkbox"/> German <input type="checkbox"/> other <input type="checkbox"/> Lithuanian <input type="checkbox"/> French ↓</p> <p><input type="text"/></p> <p><input type="text"/></p> <p><input type="text"/></p> <p><input type="checkbox"/> no command of other languages</p>	<p>17. Do you attend a school of higher education, vocational or general school?</p> <p><i>To be answered by a person aged 6–64</i></p> <p><input type="checkbox"/> yes <input type="checkbox"/> no</p> <p>18. What institution have you graduated from?</p> <p><i>To be answered by a person aged 10 and older</i></p> <p><input type="checkbox"/> doctoral studies <input type="checkbox"/> university (academy, institute, seminary, etc.) <input type="checkbox"/> college <input type="checkbox"/> professional college <input type="checkbox"/> special secondary school (technicum) <input type="checkbox"/> vocational school (after graduating from a general upper secondary school) <input type="checkbox"/> general upper secondary school, gymnasium <input type="checkbox"/> vocational school after graduating from a general lower secondary school where, together with a profession, received general upper secondary education <input type="checkbox"/> vocational school after graduating from a general lower secondary school where only obtained a profession <input type="checkbox"/> vocational (trade) school without completion of a general lower secondary school where, together with a profession, received general lower secondary education <input type="checkbox"/> vocational (trade) school without completion of a general lower secondary school where only obtained a profession <input type="checkbox"/> general lower secondary school <input type="checkbox"/> primary school <input type="checkbox"/> unfinished primary school <input type="checkbox"/> literate (no schooling) <input type="checkbox"/> illiterate</p>
<p style="text-align: center;">RELIGION</p> <p>16. Which religious community would you attach yourself to?</p> <p><input type="checkbox"/> Roman Catholic <input type="checkbox"/> Old Believer <input type="checkbox"/> Greek Catholic (Uniate) <input type="checkbox"/> Judaist <input type="checkbox"/> Evangelical Lutheran <input type="checkbox"/> Sunni Muslim <input type="checkbox"/> Evangelical Reformed <input type="checkbox"/> Karaite <input type="checkbox"/> Orthodox Believer <input type="checkbox"/> other ↓</p> <p><input type="text"/></p> <p><input type="text"/></p> <p><input type="text"/></p> <p><input type="checkbox"/> not any <input type="checkbox"/> not indicated</p>	

Activity 3. Think of the random sampling technique in Labov’s Philadelphia study (1970s) (see the example above). List at least three limitations of such approach in sociolinguistic studies:

1.
2.
3.

Activity 4. Create a sampling design (sampling grid) for the following study: you want to investigate attitudes toward minority language rights in Latvia, Estonia and Lithuania among 30 to 60 year-old males and females. Fill in the table with your sample design.

Subtotal				
Grand total				

Activity 5. Tagliamonte (2006) does not recommend contacting people of official status when initiating the first contact with the community. Discuss with your friends and think, why these might be good initial contacts?

4.5. Glossary

Data – factual information organized for analysis or used to reason or make decisions.

Ethnographic approach – a sociolinguistic approach where the researcher integrates him/herself within the community under investigation and conducts participant observation.

Interview – a conversation conducted by the researcher with a respondent in order to elicit linguistic information (or sometimes information about the language).

Population – a group of people who share certain characteristics.

Questionnaire – a form containing a set of questions; in sociolinguistic research it is usually used to elicit information about the language.

Random or probabilistic sampling – sampling technique that chooses individuals for a sample randomly, and, thus, anyone within the sample frame has an equal chance of being selected.

Representativeness – how accurately the sample represents the larger population (Milroy & Gordon 2003: 24).

Sample – part of the population that shares the same characteristics as the population.

Sampling – selection of a group of individuals from the population to form a representative group of it – a sample.

Sampling error – the difference between the sample and population values.

Sampling techniques – techniques for sampling from populations to samples.

Snowball or social network sampling – non-random sampling technique that uses “friend of a friend” approach.

Stratified random sampling – random sampling technique when the sample is manipulated according to several secondary variables.

Systematic sampling – sampling technique involving the selection of elements from an ordered sample frame.

4.6. References

- William Labov, 2001. *Principles of Linguistic Change. Vol. 2: Social Factors*. Malden, Oxford: Blackwell Publishers.
- James Milroy, 1992. *Linguistic Variation and Change: On the Historical Sociolinguistics of English*, Oxford: Blackwell Publishers.
- Lesley Milroy & Matthew Gordon, 2003. *Sociolinguistics: Method and Interpretation*, Malden, Oxford, Melbourne, Berlin: Blackwell Publishing.
- Jacomine Nortier, 2008. “Types and Sources of Bilingual Data”, in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 35–52.
- Sebastian M. Rasinger, 2008. *Quantitative Research in Linguistics. An Introduction*, London, New York: Continuum.
- Erik Schleeff & Miriam Meyerhoff, 2010. “Sociolinguistic methods for data collection and interpretation”, in *The Routledge Sociolinguistics Reader*, ed. by Miriam Meyerhoff & Erik Schleeff, London, New York: Routledge, 1–17.
- Sali A. Tagliamonte, 2006. *Analysing Sociolinguistic Variation*, Cambridge, New York: Cambridge University Press [Kindle Edition].
- Peter Trudgill, 1974. *The Social Differentiation of English in Norwich*. London, New York: Cambridge University Press.
- Ruth Wodak, Michał Krzyżanowski & Bernhard Forchtner, 2012. “The interplay of language ideologies and contextual cues in multilingual interactions: Language choice and code-switching in European Union institutions”, *Language in Society* 41, 157–186.

CHAPTER 5:
ETHICAL ISSUES

5.1. Key concepts

- Anonymity*
- Confidentiality*
- Debriefing*
- Informed consent*
- Observer's paradox*
- Surreptitious data collection*

5.2. Pre-reading activities

Activity 1. In groups discuss and define the concepts given above. How are they related with ethical issues concerning sociolinguistic research?

.....

.....

.....

.....

.....

5.3. Ethical concerns and sociolinguistic research

The **main aim** of this chapter is to introduce you to the main ethical issues concerning sociolinguistic research.

Most of the sociolinguistic research involves human beings: sociolinguists observe, record, and analyze human interaction. Thus, all sociolinguists at some point face ethical concerns. Previously (and this is valid for Baltic countries until the 1990s) sociolinguists did not have to disclose fully their research goals to the participants of the study. This was convenient, since sometimes too much of the information about what is being studied can affect the respondent's behaviour. However, at present, especially in North America, researchers have to follow pretty strict, often institutionalized, ethical guidelines. Ethics policies were established in order to ensure that a researcher is handling the data sensitively and that human subjects (respondents) "have sufficient information to make an informed decision about their participation" (Wray & Broomel 2006: 173).

Tagliamonte (2006: Kindle Locations 438–439) indicates that “the main ethical guidelines for collecting informal interviews remain constant: 1) consent for audio-recording; 2) guaranteed anonymity; 3) voluntary participation; and 4) access to researcher and research findings.” In other words, an honest sociolinguist should never record covertly (*surreptitious recording*). In many places it is illegal and such behaviour might jeopardize your research project (you may not be able to use the data you collected). Participants have to have the right to withdraw from the research project any time.

Most Western universities (VMU, to our knowledge, does not have such practice as of 2013) require researchers to go over the process of **ethics approval** (in North American universities there are usually Internal Review Boards that process researcher’s applications), which involves “informed consent” procedure. “Informed consent” procedure requires researchers to “specify procedures by which they will inform human research subjects of what a study is about and how it will be conducted and to obtain their consent, represented by their signatures” (Johnstone 2000: 42). Consent forms usually state: the purpose of the study, what is expected from the respondents, risks and benefits, compensation, procedures for ensuring confidentiality, and voluntary nature. At least most of the Western universities (namely, their IRB’s) have consent form templates that can be adapted to different types of research.

Picture 1 and Picture 2 provide you with two excerpts of the consent form that was used by Aurelija Tamošiūnaitė and Kimberly Potowski (also Lourdes Torres) in a study on language maintenance and shift among heritage speakers in Chicago.

Informed consent, as well as asking for permission to record poses several problems. First of all, when should we ask our informants to sign the consent form – before or after the recording? This question is crucial in sociolinguistic research, since “respondents who know that they are recorded from the outset may or may not deliberately change their linguistic behaviour – a phenomenon well-known as the ‘observer’s paradox’” (Rasinger 2008: 52). For some time, sociolinguists tried to overcome this problem. The best solution – to get to know your informants well: the closer the relationship, the more relaxed they will be during the recording, the more natural their speech will be, the more reliable your data will be.

Another issue with informed consent relates to the disclosure of the project: how much should we tell our participants about our study? On the one hand, if they know too much, they might deliberately change their linguistic behaviour in order to satisfy the researcher, on the other hand, if information is insufficient, they might not agree to participate in the study. Johnstone (2000: 48) suggests the following: “informing subjects means telling them what you are doing in general terms (“describing Texas women’s speech”, for example) and also what are not doing, namely evaluating them or their speech.” Although, this recommendation sounds simplistic, in practice usually it is hard to present the core idea of your project in such a way that respondents would understand the essence of it, while at the same time it will not affect their linguistic performance or answers.

Picture 1: Excerpt from the consent form



University of Illinois at Chicago
Consent for Participation in Research
"Heritage languages in Chicago"

Why am I being asked?

You are being asked to be a subject in a research study about language use and maintenance conducted by Kim Potowski, Department of Spanish at the University of Illinois at Chicago, and Lourdes Torres, Department of Latino Studies at DePaul University. You have been asked to participate in the research because you self-identify as a heritage speaker of a language other than English and may be eligible to participate. We ask that you read this form and ask any questions you may have before agreeing to be in the research.

Your participation in this research is voluntary. Your decision whether or not to participate will not affect your current or future relations with UIC or DePaul University. If you decide to participate, you are free to withdraw at any time without affecting that relationship.

Why is this research being done?

Researchers are becoming increasingly interested in the how non-English languages are spoken and used in the United States. In this study, we will interview Chicago individuals who speak languages other than English, in their non-English language, about some everyday topics and about their language use. States. The interviews will be analyzed for different linguistic structures and also for narratives of language socialization.

What is the purpose of this research?

The purpose of this research is to explore language structure and maintenance.

What procedures are involved?

If you agree to be in this research, we would ask you to do the following things:

Participate in an informal recorded interview in your non-English language.

Approximately 400 participants may be involved in this research.

Picture 2: Last page of the consent form

Signature of Subject or Legally Authorized Representative

I have read (or someone has read to me) the above information. I have been given an opportunity to ask questions and my questions have been answered to my satisfaction. I agree to participate in this research. I have been given a copy of this form.

Signature

Date

Printed Name

Signature of Researcher

Date (must be same as subject's)

Please choose one:

☐ I **agree** to have my recorded interview used (password protected) online as part of this research project.

☐ I do **not** agree to have my recorded interview used (password protected) online of this research project.

Third issue is related to confidentiality. Sociolinguists often audiotape and videotape people, thus, their identities can be potentially recovered even if researchers use pseudonyms (Johnstone 2000: 43). Wray & Bloomer (2006: 174), however, point to the difference between *confidentiality* and *anonymity*. Confidential data is such data where you, as a researcher, *do know* who are the respondents, but you do not disclose that information to the third parties, whereas anonymous data is such where even you as a researcher *do not know* who has provided what (Wray & Bloomer 2006: 174). Surveys, questionnaires usually are anonymous: if you collected more than 200 questionnaires without respondents giving their names, you will not know who provided what. With interviews and observations it is usually different: a researcher knows whom s/he observed or recorded. Anonymous surveys are not useful if you intend to do a longitudinal type of study (observe, test the same informants for a period of time), because they will not lead you back to an identity for each respondent. In that case you will have to promise confidentiality and ask respondents to provide their names. Anonymity should be favored when you are investigating sensitive issues: attitudes toward racist language, knowledge and use of swearwords etc. (Wray & Bloomer 2006: 174).

Wray and Bloomer (2006: 174) provide several tips for ensuring anonymity of your data:

- a. try to use multiple-choice responses rather than open ended questions;
- b. encourage your respondents to write in capital letters (to make writing less identifiable);
- c. use on-line survey tools.

Keep in mind that most sociolinguistic research usually involves working class people or minority groups that sometimes get tired of always being the object of study (e.g., Lithuanian, Latvian, Estonian emigrants in UK, USA or other countries). Therefore, it is always important to give the feedback concerning your results of the study (it does not necessarily mean, that all respondents will be interested in getting the feedback). This is usually called “briefing” or “debriefing” (Rasinger 2008: 54; Johnstone 2000: 48). “Debriefing” is reporting to your participants, after the study is over, what its results were (Johnstone 2000: 48). Reporting on your study results is, however, not always possible, sometimes for practical reasons (you move to a different country or you performed an on-line anonymous survey and you do not know who your respondents were), sometimes because participants are not interested in it.

Sociolinguistic research poses ethical concerns not only to the participants of the study. A fieldworker can also be faced with some ethical implications. Collecting data sometimes can be simply dangerous, especially if a researcher is trying to collect data in an urban area with a high crime rate or if a researcher has to participate in illegal activities (e.g., gang activities) in order to gain the trust (cf. Mendoza-Denton’s study (2008) on Latina Youth Gangs).

To summarize this chapter here are some of the key points, regarding ethics policies and guidelines, adapted from Wray & Broomel (2006: 173–174):

- the welfare of the respondent is paramount;
- all research procedures should follow the legal framework of a community under the investigation;
- all participants should be informed of the purpose, nature, risks and benefits of the research project;
- all explanations of the research project should be presented in a simple manner, so that the participants can understand; in the case of a child the consent from should be signed by parent or guardian;
- informed consent should be obtained in writing from all participants;
- participants must know that they have the right to withdraw from the project at any stage;
- confidentiality and/or anonymity should be assured for all participants.

5.4. Comprehension activities

Activity 2. Conduct a brief research study and try to find out what is the legislative practice (especially on data protection) in the European Union re-

garding research that involves human subjects. Do these guidelines and ethic regulations concern sociolinguistic research? What is the practice in your country of origin?

Activity 3. Cameron et al. (1992: 23–24) argues that the research should be *on*, *for* and *with* the researched: “a) Persons are not objects and should not be treated as objects. b) Subjects have their own agendas and research should try to address them. c) If knowledge is worth having, it is worth sharing.” Elaborate on these three statements in more detail. What is emphasized here? How can sociolinguistic research “empower” people?

Activity 4. Discuss with your friends the following issues:

- a. Do you think it is appropriate to pay people for participating in your research? When would you pay and when wouldn't you do that? What would you use instead of money?
- b. What are the advantages and disadvantages of using students as research subjects?

5.5. Glossary

Anonymity – the state of an individual's personal identity being unknown to the researcher.

Confidentiality – a promise that limits access on revealing participant's true identity to the third parties.

Debriefing – feedback to research participants after the study is over.

Informed consent – a procedure that requires researchers to “specify procedures by which they will inform human research subjects of what a study is about and how it will be conducted and to obtain their consent, represented by their signatures” (Johnstone 2000: 42).

Observer's paradox – a phenomenon where the observation is influenced by the presence of the observer/investigator.

Surreptitious data collection – secret/clandestine data collection (e.g., recording).

5.6. References

- Deborah Cameron, Elizabeth Frazer, Penelope Harvey, Ben Rampton & Kay Richardson, [1992, reprinted in 1997]. “Ethics, Advocacy, and Empowerment in Researching Language”, in *Sociolinguistics: a Reader and Coursebook*, ed. by Nikolas Coupland and Adam Jaworski, New York: Palgrave, 145–162.
- Barbara Johnstone, 2000. *Qualitative Methods in Sociolinguistics*, New York, Oxford: Oxford University Press.
- Norma Mendoza-Denton, 2008. *Homegirls: Language and Cultural Practice among Latina Youth Gangs*, Malden, Oxford, Carlton: Blackwell Publishing.
- Sebastian M. Rasinger, 2008. *Quantitative Research in Linguistics. An Introduction*, London, New York: Continuum.

- Sali A. Tagliamonte, 2006. *Analysing Sociolinguistic Variation*, Cambridge, New York: Cambridge University Press [Kindle Edition].
- Alison Wray & Aileen Bloomer, 2006. *Projects in Linguistics. A Practical Guide to Researching Language*, London: Arnold Hodder .

CHAPTER 6:
QUANTITATIVE RESEARCH:
CONCEPTS AND DEFINITIONS. USE OF CORPORA

6.1. Key concepts

- Annotation*
- Categorical (discrete) variable*
- Concordance*
- Continuous variable*
- Corpus*
- Descriptive statistics*
- Inferential statistics*
- Linguistic variable*
- Logistic regression*
- Measurement*
- Operationalization*
- Reliability*
- Tagging*
- Validity*
- Variable*

6.2. Pre-reading activity

Activity 1. In groups discuss and define the concepts given above. What is their role in quantitative research?

.....

.....

.....

.....

6.3. Quantitative tools in sociolinguistics

The **main aim** of this chapter is to introduce you to the main concepts and definitions related to the quantitative research paradigm in sociolinguistics.

6.3.1. Variables, reliability and validity

Most sociolinguistic analysis and all variationist analysis is preoccupied with the use of a particular linguistic item, namely a linguistic variable which is the main object of

investigation. The linguistic variable is a variable that “is represented by two or more variants” (Schleef & Meyerhoff 2010: 10) or, simply put, two or more ways of saying the same thing. The variationist approach relies on the idea that variation that we witness at all levels of language is not random, rather “speaker’s choices among variable linguistic forms are systematically constrained by multiple linguistic and social factors that reflect underlying grammatical systems and that both reflect and partially constitute the social organization of the communities to which users of the language belong” (Bayley 2013: 2). We can recall the example we have discussed before: Labov’s 1966 study on /r/ found that pronunciation of /r/ was constrained by socioeconomic status of speakers of New York, as well as speakers’ age.

According to Grieve, earlier variationist studies did not employ statistical methods (Grieve 2012). These studies (such as Labov’s in the 1960s) “focused on simple relationships between the value of a linguistic variable and the value of a social variable” (Grieve 2012). However, over time variationist studies developed more advanced methods for the analysis of linguistic variation. According to Grieve, since the 1970s the **most common statistical method** in sociolinguistic research has been **logistic regression** (2012). In statistics, *logistic regression* is defined as a type of regression analysis used for predicting the outcome of a dependent variable based on one or more predictor variable. In sociolinguistic studies, logistic regression was employed in order to test whether “the alternation between the variants of the linguistic variable is predicted by various aspects of the social background of a speaker, while also taking into consideration the effect of the linguistic environment in which each token of the variable is produced” (Grieve 2012). For example, Ito & Tagliamonte’s study on English intensifiers *very* and *really* illustrated that these intensifiers are preferred to use with predicative adjectives and this use is favoured mainly by middle-aged women (Ito & Tagliamonte 2003). Thus, statistical analysis on the use of these intensifiers found correlation **between alternation** of linguistic variables, **linguistic context**, and the **social background** of a speaker.

Variationist linguistics is preoccupied with the analysis of **variables**, i.e., “changeable factors, features or elements” (OED). Variables, thus, can be linguistic (phonetic, morphological, syntactic, semantic), as well as social (gender, social class, occupation, etc.). When we define the variable we usually attribute certain criteria to it: phonetic variable is such that occurs on phonetic level; gender variable is based on human’s physical appearance, etc. According to Rasinger (2010: 54), “the attribution of a particular variable outcome (e.g., male or female) to a particular case is made by means of *measurement*: we assign a variable value to a particular case using *predefined criteria*.” What poses a problem is the definition of these criteria. Usually, when we assign the gender to a person we rely on certain physical features of a person, but how accurate can they always be?

Rasinger (2010: 55) provides the following linguistic example:

Milroy in her well-known Belfast study (1987), for example, has looked at, among many other things, how the realization of the vowel /e/ in different linguistic environments, particularly the merge of the /e/ as in *peck* with /æ/ as in *pack* into homophones, relates

to the degree of a speaker's inclusion into the social network, with a high frequency of non-standard /æ/ in both contexts indicating a higher degree of inclusion (1987). To cut a long story short, at the very end it comes down to *measuring* (in this case, count) the number of different realizations of the vowels: how often does /e/ occur in words such as *peck*, and how often is /e/ replaced by /æ/ making it homophone to *pack*.

In phonology, sociolinguists usually use acoustic analysis in order to define the values of the variable. The specification of measures is also called *operationalization*.

Important!

Once you have established your measure and have it operationalized, do not change it over the course of your analysis. This might lead to distorted results.

Measurement is also closely related to *reliability* and *validity* (Rasinger 2010: 55). “Reliability refers to our measure repeatedly delivering the same or near the same results” (Rasinger 2010: 55). In other words, if you use the same measure (let us say, we count /e/ and /æ/ in *peck* and *pack*) with the same population under the same circumstances, you should come up with the same results. However, as Rasinger admits (2010: 55), *test-re-test method* is problematic when we do research with human beings. Just imagine, if you run the experiment five times with the same people, on the one hand, they might improve their performance, on the other hand, humans change and their behaviour even under the same circumstances might be different. A solution to this problem can be ‘split-half’ method:

We take a group of people, measure whatever we like to measure, then randomly split the group into two smaller groups and compare the results. If the measure is reliable, we should get very similar results for both subgroups. If we get substantially different results, we should become very cautious and investigate the reliability a bit further. (Rasinger 2010: 56)

Validity, on the other hand, is considered to be the degree to which the measure tool measures what it is supposed to measure (Rasinger 2010: 56). Validity is extremely important when we measure the abstract concepts, e.g., attitudes, ideologies, beliefs (i.e., validity is important if you intend to conduct a survey using questionnaires).

6.3.2. Quantitative analysis: main concepts

Quantitative analysis, as the name presupposes, deals with *counting* something. According to Levon, in order to conduct quantitative analysis it is necessary a) to have data that is ‘countable’ and b) ‘countable’ data must have a potential to be variable (Levon 2010: 68–69). Thus, before any type of research you should always think, whether the phenomenon you want to investigate is quantifiable or not; and if so, whether it is a variable or not.

Example #1.

You want to investigate representation of the sound /č/ in Lithuanian letters at the beginning of the 20th century. The representations of /č/ are quantifiable (you can simply count all instances when it is used) and it is a variable, since /č/ can be represented with graphs <č>, <c> or digraph <cz>. Thus, this topic might require you to involve some of the quantitative analysis.

Example #2.

You want to investigate the ideological meaning of <cz> in Lithuanian writing. Ideological meaning is not something you can quantify, thus, you might need to turn to discourse or content analysis (in order to find out how the use of <cz> was perceived by Lithuanians) in order to investigate this topic.

There are several ways on how to treat quantitative data. If we want simply to **describe** the situation (as in the first example) we might employ what is called *descriptive statistics*. “Descriptive statistics are indices that give information about the general shape or quality of the data, and include such things as the *mean* (i.e., average) and the *median* (i.e., middle) of the data” (Levon 2010: 70). Descriptive statistics can provide us with the potential patterns of our data set, however, it does not enable us to say, whether these patterns are genuine or is just a product of chance (Levon 2010: 70). In order to have more reliable analysis of our data we have to turn to *inferential statistics*. “Inferential statistics are designed to determine whether apparent patterns in a data set really are patterns – whether they are what we call *statistically significant*” (Levon 2010: 70). Inferential statistics test the correlations, as well as validity of the patterns observed. As Levon (2010: 70) puts it: “descriptive statistics allow us to define patterns in the data,” while “inferential statistics then allow us to determine whether those patterns truly exist in some kind of meaningful way.”

Thus, quantitative methods (statistics) are employed when we are interested in the relationship between the dependent variable and independent variable(s). For example, we might hypothesize, that the use or non-use of a graph <cz> (dependent variable) is affected by the age of the writer (independent variable) and our null hypothesis then would state that there is no correlation between the use of <cz> and age of the writer. According to Levon, quantitative analysis does not test experimental hypothesis, it always tests null hypothesis (2010: 71): “this is because inferential statistics provide a *probabilistic* measure – that is, they measure the likelihood that the null hypothesis is true.” Probability is usually denoted with letter <p> and expressed in percentages.

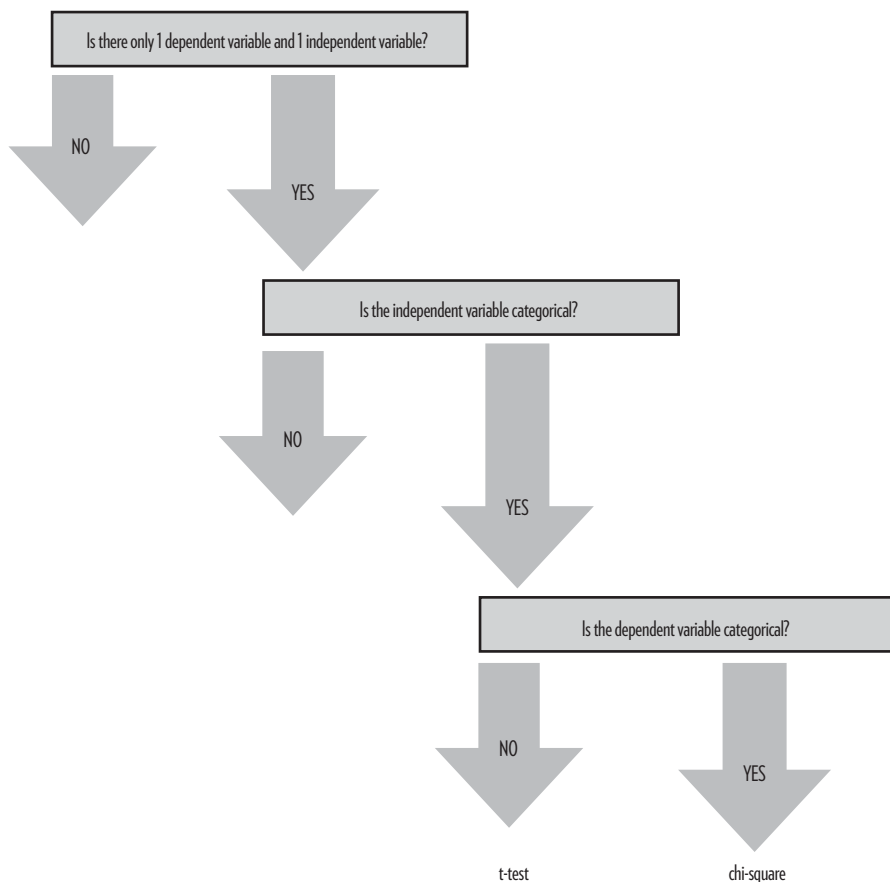
Example #2.

If a statistical test is 80 % sure that the null hypothesis is true, then, $p=0.80$, if a statistical test is 1 % sure that the null hypothesis is true, then, $p=0.01$.

In most of the humanities and social sciences, it is agreed that the threshold percentage, when we cannot reject null hypothesis is 5 % ($p=0.05$), i.e., if p is greater than 0.05 ($p>0.05$) we cannot reject the null hypothesis, if less than or equal ($p\leq 0.05$), we can reject it. Thus, when we reject the null hypothesis, we support our experimental hypothesis, and vice versa.

There are different inferential statistical tests. Thus, which one to use in your analysis?

Figure 2: Decision tree for statistical tests (adapted from Levon 2010: 75)



According to Levon (2010: 72), it depends on the kind of the variable. There are two kinds of variables: *categorical* and *continuous*. “*Categorical variables* are those whose values can be easily separated into discrete categories” (Levon 2010: 72), e.g., <cz>, <c> and <č> in previous examples to represent sound /č/. Sometimes they are also called *discrete variables*. These choices (variables) are distinct and they do not over-

lap. “Continuous variables cannot be easily classified into categories this way; they are variables whose values exist on a mathematical scale” (Levon 2010: 73), e.g., numbers 20, 30, 45, denoting the age, will be continuous variables. In sociolinguistic research, income, distance to the urban area or length of vowels or consonants would also be treated as continuous variables.

Thus, the choice of the statistical test depends on the variables you are examining: whether they are categorical or continuous. When the dependent variable is categorical (in our case <cz>) the statistical test we use is a *chi-square test*. When the dependent variable is continuous (e.g., age), the statistical test we use is a *t-test* (Levon 2010: 73). Figure 2 illustrates the steps one should take in order to select the appropriate statistical test.

6.3.3. Coding the data

All of the linguistic data, obtained in a fieldwork, needs to be **transcribed** and **coded**. There are two ways to code the data: you could just read through your transcripts and mark your linguistic variable where it occurs, or you could compile a searchable corpus and later run a search for that particular linguistic item (Schleef & Meyerhoff 2010: 12).

Coding strategies depend on the data you are working with: it might be easier to code data obtained from questionnaires, while for natural speech data you might need to rely on transcriptions. Variables that occur in numbers (age, year of birth) usually do not require any specific coding; we transfer these numbers to the software we are using for our statistical analysis. However, variables that appear in words (qualitative data) have to be “translated” into numbers. A researcher can think of his/her own conventions on how to “translate” “gender” to numbers, for instance, you can just assign number 1 to female and 2 to male; the same can be applied to answers “yes” and “no”. Some software might require you to put also the number for the missing value (respondents often miss to answer some of the questions in the questionnaires). Multiple choice questions, semantic differentials and Likert (or other) scales usually are encoded using the same manner.

Example #4.

Why do you think, you need to learn Lithuanian (mark all that apply)?

- ☐ because this is the language of your parents = 1
- ☐ because you are a Lithuanian = 2
- ☐ because it is better to know two languages (Lithuanian and English) than one (English) = 3

(adapted from a Lithuanian Saturday School survey, conducted by A. Tamošiūnaitė in 2007)

It is a common practice to code different categories applied in the research. The sociolinguistic analysis often entails different sets of data. For instance, Tagliamonte (2006: Kindle Locations 622–623) lists the following components of her research corpus, used for variationist analysis:

1. recording media, audio-tapes (analogue, digital) or other;
2. interview reports (hard copies) and signed consent forms;
3. transcription files (ASCII, Word, txt);
4. a transcription protocol (hard copy and soft);
5. a database of information (FileMaker, Excel, etc.);
6. analysis files (Goldvarb files, token, cel, cnd and res).

Each file in the corpus needs to be catalogued. Use your own conventions on how to name (label) the files, just make sure that after some time, if you need to go back to your data, you will be able to identify what is what. It is suggested to label audio recordings by **number, date and name of the speaker**. Each of the recordings has to have an “interview report”, a document that includes notes and observations about the speaker (Tagliamonte 2006: Kindle Locations 626–627). You also have to store a consent form from each of the interviewees. Transcription files have to be related to recordings, whereas protocol documents the method how conversations were transcribed. Database of information can be of a general type that list all kinds of information related to data. All sociolinguistic research involves having many analysis files. These are the essentials of our study.

Linguistic data, depending on your variable, will entail another type of coding. First of all, in variationist analysis, the data needs to be extracted. In other words, you might not need to use all of the data you have, but some of the extractions of it. Tagliamonte suggests extracting the data directly to the software (e.g., Goldvarb) (2006: Kindle Location 1299). When you extract the language sample, there are three factor groups you need to encode: “1) the community or a data set; 2) individual speaker; and 3) the dependent variable” (Tagliamonte 2006: Kindle Location 1305). The most common practice is to use letters and numbers to code these factor groups. Tagliamonte recommends to code for each individual factor group through the entire data, e.g., first you code your sample for age, then literacy level, then dependent variable and so forth. Here is a tip provided by Tagliamonte on how to set up a coding system for your data:

The optimal way to set up a coding system is to use mnemonic codes. For example, ‘M’ = main clause, ‘1’ = first person singular, ‘N’ = negation, etc. Further, I typically use capital letters for categories, e.g. ‘S’= subordinate clause; but lower case letters for types within categories, e.g. ‘w’ = subordinate clause with when. If you use the same codes for the same factor groups across studies, these codes will become second nature to you. Of course, your codes cannot always be mnemonic, but making sure the main ones are will make your job a lot easier. (Tagliamonte 2006: Kindle Locations 1319–1322)

Linguistic variables should also be assigned a separate codes, e.g., <cz> can be given the code “P”, <č> – the code “L”, and <c> – the code “B”. Each of the line in your sample that has your variable then has to be coded with these letters. After you have

coded the dependent variable, you should start with independent variables (age, sex, literacy level, occupation, etc.). Coding, according to Tagliamonte (2006: Kindle Location 1348), is the **most time consuming** task of analysis (this is true for all linguistic research). Make sure you keep your coding schemes with you and put them next to the data you coded.

6.3.4. Processing the data

Numerical data usually is processed using statistical tests, while transcriptions of natural speech by variationists usually are processed using concordance programs or a special program, created for variationist research, called Goldvarb. Recent studies of language variation rely on corpus data. Variationists quantify or investigate the frequencies of use of different linguistic variables in a natural speech. Frequencies are *continuous variable* (not categorical), therefore, many corpus studies employ *regression analysis* in order to predict scores on one measure using information from one or more other measures (Tokowicz & Warren 2008: 222, 228).

Processing of data (actual work with data) depends on the statistical tools you will select. For *descriptive statistics* you can simply rely on MS Excel or similar software. For inferential statistics, you might need to rely on a more advanced statistical software, such as IBM SPSS (Statistical Package for Social Sciences), which includes the base software for descriptive statistics, bivariate statistics, including quite widely used statistical model, called ANOVA (Analysis of Variance). IBM SPSS package will allow you to run chi-square test and t-tests, if needed. Variationists often employ a special program, created for the variationist analysis – Goldvarb (or Varbrul). A detailed analysis, on how to use the program for sociolinguistic analysis is provided in Tagliamonte 2006. Which software to use, depends on your research goals and your computer skills: if you are acquainted with SPSS, it might be worth doing statistics with that software package; if your study replicates, e.g., variationist work, it might be worth doing calculations using Goldvarb. Goldvarb is free to use, compatible with both PC and Mac, and it is downloadable from here: <http://individual.utoronto.ca/tagliamonte/goldvarb.htm>.

The most popular statistical tests, applied in (simple) sociolinguistic analysis are chi-square and t-test. Without going into much detail on how these tests run mathematically (refer to Rasinger 2008, Levon 2010 for a more detailed analysis), let us consider just one example, where a statistical test was applied.

Example #5.

Thomas Purnell, William Idsardi & John Baugh, “Perceptual and phonetic experiments on American English dialect identification”, *Journal of Language and Social Psychology* 18, 10–30, 1999.

Purnell et al. conducted a set of experiments (see Chapter 1). Their experiment #3 tested the ability to recognize dialects at micro-linguistic level. Thus, they hypothesized, that phonetic features in a short portion of speech are sufficient to trigger identification

across dialects, and their null hypothesis was that there is no difference between the dialects by identification.

Method: The word “hello” from the utterance in all three varieties (AAVE, ChE, SAE) was used for this experiment. This word lacks the environment in which the researchers expected dialectal variations. The experiment was conducted with 50 undergraduate students at the University of Delaware. All the participants were Caucasian native speakers of SAE. Total of 60 tokens were randomized (ten instances of “hello” repeated twice for each dialect). **Two sets of tokens (120)** were presented to each participant. During a 2 second pause participant indicated which dialect they believed they heard.

The *dependent variable* in this case was the *dialect recognition*, while *independent variable* was the phonetic feature (pronunciation of “hello”). The researchers ran a statistical analysis (both of the variables are categorical, thus, chi-square test is suitable). The results are presented in the table below.

Table 11: *Confusion Matrix and Summary Statistics by Dialect*

	Stimuli				
Response	AAVE	ChE	SAE	Row Total	
AAVE	a 923 (15 %)	b 280 (5 %)	c 196 (3 %)	1,399	(23 %)
ChE	d 235 (4 %)	e 1,607 (27 %)	f 41 (1 %)	1,883	(31 %)
SAE	g 842 (14 %)	h 113 (2 %)	i 1,763 (29 %)	2,718	(45 %)

Note. AAVE = African American Vernacular English; ChE = Chicano English; SAE = Standard American English. $\chi^2 = 4,510$, $df = 4$, $p < .001$; **Accuracy Index (AI)** = .72; percentages = percentage of total for that cell.

χ^2 – refers to chi-square
df – degree of freedom
p – probability

Note that $p < .001$, which strongly suggests that the independent variable influences the dependent variable. It is confirmed by accuracy index, which, according to authors, is quite high in this case. Thus, the null hypothesis is rejected, and hypothesis is supported.

Results: participants were able to identify the dialects when only hearing one word. However, it remains unclear which features of the speech act as sociolinguistic markers.

The software used by sociolinguists (Varbrul or Goldvarb, which is a newer version of the previous one) was created in order to measure the relative contributions of independent linguistic variables to the overall variability of items (Wolfram 2006). Goldvarb, as well as Varbrul, thus, produces weighting values, than can range from 0 to 1. If we take two factors (binomial application) and the weighting value is greater than 0.5, this means that the factor has an effect on the occurrence of the variant, and vice versa. If we have three factors (trinomial application) and the weighting value is more than 0.33, it indicates a favouring effect and vice versa. “The higher the Varbrul weighting is, the stronger the effect of the factor is in accounting for the variability” (Wolfram 2006).

To illustrate this, consider the following data set taken from Tagliamonte 2006: Kindle Location 2818.

Table 12: Variable (ing) in York English

Probability of [n] for occupation	
Occupation	Factor weight (calculated by Goldvarb)
Blue-collar	0.60
Student	0.57
White collar	0.44
Professional managerial	0.44
Range	28

From this data set we can see, that blue-collar people and students seem to favor the [n] variant in (ing) (factor weights higher than 0.5), while white-collar and professional managers do not favor it (factor weights lower than 0.5). Tagliamonte 2006 is a useful step-by-step guide on how to do variationist analysis using Goldvarb. If you intend to do such analysis, we would highly recommend consulting that book.

Most of the tests today are run by software or programs. However, statistical analysis (data processing) is just one step in your analysis. After you come up with your results, you need to interpret them. Several tips on interpreting statistical data:

- a. always check the p-value: if $p \leq 0.05$, your analysis is significant;
- b. always make sure you know what is the dependent and what is the independent variable in your research. This will help you to explain the correlation and possible impact of one on another;
- c. after you have done calculations, you might often need to find explanations for certain phenomena using qualitative evidence. As Levon states (2010: 90), “statistical significance and real-world significance are not always the same thing.” Thus, in order to check the real-world significance you might need to conduct interviews, observations in order to elicit qualitative data.

6.3.5. Corpus tools and use in variationist sociolinguistics

Corpus is another tool that provides sociolinguists (or linguists in general) with a large amount of usually textual data. Corpus analysis enables researchers to test hypothesis about language use, also to raise new questions and theories. Corpora are especially useful if a researcher intends to measure the frequencies or to quantify linguistic patterns. Large corpora highlight the most frequent variables in the language, while at the same time a researcher can find “evidence on rare or unusual cases of language” (Baker 2010: 94).

Corpus can be defined as “any text or collection of texts” (Baker 2010: 95). However, corpus linguists point several main characteristics of corpus: it should be of a finite size, maximally represent variety under examination, be machine readable (Baker 2010: 95). Corpus can be of different sizes: from a transcription of one interview to a 5 mln.-word corpus. Sociolinguistic analysis usually depends on smaller corpora

(due to the nature of the data). On the one hand, sociolinguists might build their own corpus (from the interviews, observations) to use for a specific analysis, on the other hand, they might use already existing corpora. For sociolinguistic research it is important to have balanced corpora (e.g., representing different age groups equally, etc.). Representativeness of the corpora depends on your research questions: for some you might need a smaller size, representing just one variety of language, for other research – several varieties.

Corpora are often annotated (tagged, parsed) for additional information. Bilingual or spoken language data often uses the transcription and data processing system called CHILDES (Child Language Data and Exchange System). It has its own transcription rules (called CHAT). After the data is coded in CHAT, it can be processed using a specialized program called CLAN.

Different corpora annotation strategies can be employed, as different standards exist (considering the morphological and syntactic diversity of different world languages). In sociolinguistic analysis annotation might be optional, especially if you decide to use another coding system, which we have discussed in the previous section (Tagliamonte's example). In any case, there are two ways to go: to use already established system of annotation, or create your own coding system. In both cases, **never forget to keep coding (annotation) conventions next to your data.**

According to Baker, “corpora are normally used in conjunction with analysis software” (2010: 102–103). The most popular ones, used by many linguists are: Word-Smith Tools (developed by Mike Scott, website: <http://www.lexically.net/wordsmith/>, not free of use) and AntConc (developed by Lawrence Anthony, download for free at: <http://www.antlab.sci.waseda.ac.jp/software.html>). Both of these programs enable you to compile index lists, frequency lists, concordance with the variable (a lexeme, orthographic or phonetic feature you are investigating). Corpus analysis is especially valuable for the research on collocations, semantic and lexical analysis, to a certain degree – morphological analysis.

Corpus is very useful for data processing involved in sociolinguistic (variationist) analysis. Here are the main steps what you can do with, e.g., your interview data:

- a. first you have to transcribe (using defined transcription conventions) your data. Any text processor can work for transcription; however, make sure that you transcribe your data into such format, which allows you later (if needed) to convert the transcription to another format. In other words, if you transcribe your file in MS Word document, make sure that later you will be able to convert the file to .txt or ASCII format. Concordance programs **run only** on .txt or ASCII files.
- b. once you have your transcription done, convert (if needed) your file to .txt format.
- c. open your concordance program (e.g., AntConc), feed in one (or several) file you want to analyze and run a concordance analysis. There are several data

- sets you can extract using the program: concordances (with your feature in the middle of the line), word lists, frequency lists, indexes and other.
- d. analyze and count the occurrences of your variable in the concordance. In other words, corpus should be treated only as a quantitative tool, that helps you to count the phenomena, to look for patterns. In order to explain, why there is this and not another pattern, again, you need to turn to the qualitative analysis.

Example #6.

Picture 3 illustrates one simple search of orthographic variable <cz> in Lithuanian letters written between 1901–1910

Picture 3: Sample search using AntConc: use of <cz>.

The screenshot shows the AntConc software interface. The 'Corpus Files' panel on the left lists '1901-1910.txt'. The main 'Concordance' panel displays a list of 27 hits, each with a line number, the text snippet, and the file name. The search term 'cz' is entered in the 'Search Term' field, and the 'Concordance Hits' count is 27. The 'Search Window Size' is set to 50. The 'Files Processed' panel at the bottom shows 'Total No. 1' and 'Files Processed' with a progress bar.

Hit	KWIC	File
1	enksmas daibas wel dirbu tam paczam pabizki kur pirme dirbu boj:	1901-1910.txt
2	is jr siasioje les wisio pirmocziawsz swieikinu asz tamtis si:	1901-1910.txt
3	A. A. Amen jr priaszio taje paczes ziodges del tametiu wisiu :	1901-1910.txt
4	Gerbemasai Pons Dr. J. Basanavicsziul Asz Pranas Zyklus prisiuncas:	1901-1910.txt
5	icseul Asz Pranas Zyklus prisiuncziu per Agenta Obecuna Tamistai	1901-1910.txt
6	s ape savo liki- ma, ir kad ziunczeu 18 d. Balandzio Advakatu Jo	1901-1910.txt
7	ono J. Uilelazio, kad kapp greiczeus padaritumeta man ta malon	1901-1910.txt
8	Amerikos listuvius szirdinga atoziu = Ponu dr. J. Basanavicsziul	1901-1910.txt
9	a atoziu = Ponu dr. J. Basanavicsziul ir Ponu M. Yczu uz jusu gias:	1901-1910.txt
10	z. J. Basanavicsziul ir Ponu M. Yczu uz jusu gias pesiiventima, c	1901-1910.txt
11	ia pesiiventima, del savo Tauteczziul. Teipu- qi esmu nuliudes is:	1901-1910.txt
12	o kad daug musu broliu Amerikieczziul buvo kitokiu palvalgu, kuref	1901-1910.txt
13	j linksmas kad musu Gerbemei szvecziul viska su kantribe pernesze,	1901-1910.txt
14	imdami sunku Junga ant savo peteczziul, jau rodos cze ir kieczzeusiz	1901-1910.txt
15	ga ant savo peteczziul, jau rodos cze ir kieczzeusiz szirdes imogu	1901-1910.txt
16	o peteczziul, jau rodos cze ir kieczzeusiz szirdes imogu sugraudint	1901-1910.txt
17	seneli Gerhema dr. J. Basanavicsziul tunkre amziuje pesiiventima	1901-1910.txt
18	rizdi Gerbemasai dr. J. Basanavicsziul, kad ir mes naftiuretumem lu	1901-1910.txt
19	etaj nog mano pesiiventima, ir cze ne mano kalte bet nesusiprat	1901-1910.txt
20	es in Tievlinia an Paqelbe Vilneczziul stulekti tiek gerover, kad:	1901-1910.txt
21	etuvius ir tt. Sunku isa viska cze isarokoti ir apsakiti ku mu:	1901-1910.txt
22	musu ateitis. Jejgu tiktaj greiczeus pargrizziul taj tikiosiu nur	1901-1910.txt
23	u greito atcekio su tikra viltoze nog Gerbemuju geradeju, kure	1901-1910.txt
24	szirdas vise giaso Visa o ypacz zeikelose Tautisko Namu. Su	1901-1910.txt
25	iko iro Toki mado ant kokis dienzios szventas Kitaz kitam perka	1901-1910.txt

The search on variables <cz> and <cz> to represent /č/ in Lithuanian, in the letters written between 1901–1910 revealed the following results: <cz> was used 27 times, whereas <cz> was used 127 times. These, of course, are just raw numbers, and a further analysis would have to look at the distribution of <cz> and <cz> among different age groups and literacy groups.

Example #7.

Picture 4 illustrates a simple search on the use of the lexeme <know> in the speech of one American-Lithuanian woman.

Picture 4: Sample search using AntConc: use of < know>

The screenshot shows the AntConc software interface. At the top, there are tabs for different views: Concordance, Concordance Plot, File View, Clusters, Collocates, Word List, and Keyword List. The 'Concordance' tab is selected. Below the tabs, the text '10.txt' is visible. The main area displays a list of concordance results. Each result is a line from the text file, with the word 'know' highlighted. The results are numbered 1 through 26. To the right of each line, the file name 'D_Ls' is listed. At the bottom of the window, there is a search bar with the text 'know' entered. To the right of the search bar, it says 'Concordance Hits 179' and 'Search Windows 50'. Below the search bar are buttons for 'Start', 'Stop', and 'Sort'.

Hit	KWIC	File
1	ką daugiau daryt su menu. You know what kind of job I wanted sor	D_Ls
2	ts buvo labai, tėtis buvo you know something something of very :	D_Ls
3	very important ir žinai, you know, man labai patinka su vaikais	D_Ls
4	ka su vaikais dirbti. Ir, you know, pine pear teacher, very dife	D_Ls
5	. Aha. D. L. Jie nešiotų, you know, jeigu yra raudona, viską rai	D_Ls
6	dona, viską raudona arba, you know, raudona ir žalia, arba, you	D_Ls
7	, raudona ir žalia, arba, you know, what ever A. T. Nu taip. D.	D_Ls
8	tu nežinai, it's kind of you know feeling away. A. T. Aha, su	D_Ls
9	. D. L. Vokietijoje. Tada you know, so they taking away. Jo. M	D_Ls
10	inžinierių, kad jie, jie, you know,ėjo į Ameriką ir į tik į Či	D_Ls
11	gyveno? D. L. Jea. Po to, you know, jis su mama take a marriage	D_Ls
12	rs, žinai, ir. Mama buvo, you know, she was oldest than people.	D_Ls
13	sėjo į mokyklą nepatiko, you know, inžinieriai nepatiko ir dakt	D_Ls
14	ur lietuviai, kur norėjo, you know, jų, jų vaikai turėti. Tada	D_Ls
15	lients prieš poor ir per, you know, per. Kai jis dirbo, jis irg	D_Ls
16	s buvom, gal apie ją, I don't know Iam thinking, kai mes buvom :	D_Ls
17	. D. L. Ir ji buvo labai, you know, gera sekretorė. Nes ji žino	D_Ls
18	žinai. A. T. Kodėl? D. L. You know, buvo gerai, buvo smagu when	D_Ls
19	D. L. Maždaug, maždaug. You know, so, mes turime mūsų up and	D_Ls
20	to you. A. T. Aha. D. L. You know, ir viskas buvo worried for fo	D_Ls
21	students. Ir viskas buvo, you know, everything was kirdy. Jeigu t	D_Ls
22	just, everyone was niced. You know, everyone was your age. It wa	D_Ls
23	u child. A. T. Aha. D. L. You know, you have people in same clas	D_Ls
24	ai. A. T. Taip. D. L. Tai you know younion building, tada yra y	D_Ls
25	ounion building, tada yra you know kitos klasės. A. T. Taip. D.	D_Ls
26	i, yra kur yra žolę. Kada you know, so stuf. A. T. Aaa. Taip, s	D_Ls

Search Term ☐ Words ☐ Case ☐ Regex

 Concordance Hits 179 Search Windows 50

The search indicated 179 occurrences of “know” in D. L. speech, in 168 instances it is preceded by “you”. Thus, in most of the cases “know” appears in a collocation “you know”, which is a discourse marker in English. It is retained in Lithuanian conversation. The following step would be to analyze the context in more detail and to compare D. L. results with the speech of other

Important!

Quantitative methods just tell you what people do with language (how do they use it). But they cannot tell you why people do this. In order to answer “why”, you might need to consider qualitative measures.

American-Lithuanians.

6.4. Comprehension activities

Activity 2. Look at the data sets below, adapted from Choi's 2005 [see pages 239–240] study on bilingualism in Paraguay. Indicate in which cases statistical change (over 40 years) is significant, in which not. Which values you will be looking at?

1. What language would you use with your spouse?

Year	Spanish	Guarani	Both	Total
1960–1961	10 (18.9 %)	16 (30.2 %)	27 (50.9 %)	53 (100 %)
2000–2001	20 (32.8 %)	6 (9.8 %)	35 (57.4 %)	61 (100 %)
schicell=15.1292	df=2	p=0.000		

$p < 0.05$.

2. What language would you use with your grandparents?

Year	Spanish	Guarani	Both	Total
1960	6 (11.8 %)	39 (76.4 %)	6 (11.8 %)	51 (100 %)
2000	8 (12.7 %)	33 (52.4 %)	22 (34.9 %)	63 (100 %)
schicell=36.3143	df=2	p=1.302		

$p < 0.05$.

3. What language would you use with your siblings?

Year	Spanish	Guarani	Both	Total
1960–1961	10 (16.7 %)	29 (48.3 %)	21 (35 %)	60 (100 %)
2000–2001	21 (29.6 %)	17 (23.9 %)	33 (46.5 %)	71 (100 %)
schicell=19.0878	df=2	p=0.000		

$p < 0.05$.

4. What language would you use with your friends in the neighbourhood/on the streets of Luque⁴?

Year	Spanish	Guarani	Both	Total
1960–1961	8 (27.6 %)	12 (41.4 %)	9 (31 %)	29 (100 %)
2000–2001	19 (26.8 %)	15 (21.1 %)	37 (52.1 %)	71 (100 %)
schicell=79.0374	df=2	p=0		

$p < 0.05$.

4 Luque – is a town 16 km away from Paraguay's capital Asuncion.

5. What language would you use with your friends in the neighbourhood/on the streets of Asuncion?

Year	Spanish	Guarani	Both	Total
1960–1961	29 (50 %)	6 (10.3 %)	23 (39.7 %)	58 (100 %)
2000–2001	40 (56.3 %)	3 (4.2 %)	28 (39.4 %)	71 (100 %)
schicell=3.61122	df=2	p=0.164		

$p < 0.05$.

Activity 3. Which of these variables are continuous and which discrete (or categorical)?

- “yes” and “no” answers in questionnaire;
- responses from 1 through 5 on a Likert scale (e.g., strongly agree and strongly disagree);
- age;
- social class;
- education;
- gender;
- different allophones of /r/;
- the frequency of vowel /a/;
- a dialect or a variety.

Activity 4. Try to find out what corpora exist for Lithuanian, Latvian, Estonian, and your native languages. How big are they? Would they be suitable for sociolinguistic analysis? How? Are there corpora of minority languages or dialects used in Lithuania, Latvia or Estonia?

6.5. Glossary

Annotation – an extension of the text by addition of various linguistic information.

Categorical (discrete) variable – values that can be easily separated into discrete categories (e.g., gender, social class, etc.).

Concordance – a word or phrase and its immediate context. In corpus linguistics, concordance is used to analyse the use of a single word, word frequency, etc.

Continuous variable – values that cannot be easily classified into categories; values of these variables usually exist on a mathematical (continuous) scale (e.g., age, time).

Corpus – a collection of texts of natural language which is electronically stored and processed.

Descriptive statistics – a type of statistic analysis that aims to provide general information about the sample, e.g., provides mean (average) and median (middle) of the data.

Inferential statistics – a type of statistic analysis that tests correlations, as well as validity of the patterns observed in the data.

Linguistic variable – two or more ways of saying the same thing.

Measurement – an assignment of a value to a particular case (category, variable) using predefined criteria.

Operationalization – specification of measures used in analysis.

Reliability – an estimate used to test whether our measure repeatedly delivers the same or near the same results.

Tagging – adding additional information or classification (tags) to words and other formations in corpora.

Validity – the extent to which the measurement corresponds accurately to the real world.

Variable – a changeable factor, feature or element.

Logistic regression – a type of regression analysis used for predicting the outcome of a dependent variable based on one or more predictor variable.

6.6. References

- Paul Baker, 2010. “Corpus Methods in Linguistics”, in *Research Methods in Linguistics*, ed. by Lia Litosseliti, London, New York: Continuum, 93–113.
- Robert Bayley, 2013. “Variationist Sociolinguistics”, in *The Oxford Handbook of Sociolinguistics*, ed. by Robert Bayley, Richard Cameron and Ceil Lucas, Oxford: Oxford University Press [online version].
- Jinny K. Choi, 2005. “Bilingualism in Paraguay: Forty Years after Rubin’s Study”, *Journal of Multilingual and Multicultural Development* 26(3), 233–248.
- Jack Grieve, 2012. “Sociolinguistics: Quantitative Methods”, in *The Encyclopedia of Applied Linguistics*, ed. by Carol A. Chapelle, Wiley-Blackwell [retrieved from: <https://lirias.kuleuven.be/bitstream/123456789/329022/1/10-0455sociolinguistics2.pdf>]
- Rika Ito & Sali Tagliamonte, 2003. “Well weird, right dodgy, very strange, really cool: layering and recycling in English intensifiers”, *Language in Society* 32, 257–279.
- Erez Levon, 2010. “Organizing and Processing Your Data: The Nuts and Bolts of Quantitative Analyses”, in *Research Methods in Linguistics*, ed. by Lia Litosseliti, London, New York: Continuum, 68–92.
- Thomas Purnell, William Idsardi & John Baugh, 1999. “Perceptual and phonetic experiments on American English dialect identification”, *Journal of Language and Social Psychology* 18, 10–30.
- Sebastian M. Rasinger, 2008. *Quantitative Research in Linguistics. An Introduction*, London, New York: Continuum.
- Sebastian M. Rasinger, 2010. “Quantitative Methods: Concepts, Frameworks, and Issues”, in *Research Methods in Linguistics*, ed. by Lia Litosseliti, London, New York: Continuum, 49–67.
- Erik Schleeff & Miriam Meyerhoff, 2010. “Sociolinguistic methods for data collection and interpretation”, in *The Routledge Sociolinguistics Reader*, ed. by Miriam Meyerhoff and Erik Schleeff, London, New York: Routledge, 1–17.

- Sali A. Tagliamonte, 2006. *Analysing Sociolinguistic Variation*, Cambridge, New York: Cambridge University Press [Kindle Edition].
- Natasha Tokowicz & Tessa Warren, 2008. "Quantification and Statistics", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 214–231.
- Walt Wolfram, 2006. "Variation and Language, an Overview", in *Encyclopedia of Language and Linguistics*, 2nd Edition, ed. Keith Brown, Elsevier Ltd. [online version].

CHAPTER 7: QUESTIONNAIRES

7.1. Key concepts

- Leading questions*
- Likert scales*
- Multiple item questions*
- Piloting*
- Questionnaire*
- Semantic differential*
- Survey*

7.2. Pre-reading activity

Activity 1. In groups discuss and define the concepts given above. How are these concepts related to questionnaire design?

.....

.....

.....

.....

.....

7.3. Questionnaires

The **main aim** of this chapter is to introduce you to key issues related to **questionnaire and survey** design. We will discuss their role in sociolinguistic research, in quantitative and, to a certain extent, in the qualitative tradition.

According to Rasinger (2008: 57), questionnaires today are “the most popular tool in quantitative methodology”. One of the advantages of using questionnaires for data collection is that they help to “accumulate vast amounts of incredibly high quality data” (Rasinger 2008: 57). Indeed, questionnaires are a very efficient means of collecting data: “they allow researchers to gather data from a large number of speakers in a relatively brief amount of time” (Milroy & Gordon 2003: 52), e.g., you can conduct a survey of 80 respondents in just two weeks.

Example #1. Project “Cities and Languages”

A recent Lithuanian sociolinguistic project “Language Use and Ethnic Identity in Lithuanian Cities” (carried out in 2007–2009) heavily relied on the use of questionnaires de-

signed in order to explore the home languages of the Lithuanian cities. This method allowed researchers to collect responses from 23 341 respondents (8–10 year old Lithuanian children) (see Ramonienė 2010: 15 for a more detailed description).

In sociolinguistic research questionnaires are usually used in order to elicit data *about* language, but *not* data on linguistic performance (Codó 2008: 171). Due to this limitation, questionnaires are often combined with other types of data collection, for instance, interviews, participant observation, ethnographic notes, etc. According to Codó (2008: 171), knowledge obtained from surveys can provide you with a general picture of language use in a given community; it can also help you to identify target subjects for participant observation or ethnographic research. Thus, depending on your research questions, your research design might rely just on data obtained from questionnaire survey, or questionnaires can be just one of several data sources you need to use.

Questionnaires generally are of **two types**:

- a. **self-administered**, when the questionnaire is filled in by informants in writing (filled individually or in a group, also returned immediately, on-line or after some time);
- b. **interviewer-administered**, when the questionnaire is administered verbally by the interviewer: the interviewer asks an interviewee questions (face-to-face or over the phone) and fills in the questionnaire by him/herself.

There are several advantages and disadvantages associated with both. Self-administered questionnaires are usually inexpensive, they do not involve as much administrative time, and there are less possibilities of interviewer bias. However, if you mail your questionnaires or use on-line resources, the response rates might be quite low (unless a person is very excited about your topic or feels compationate about your research), you might receive incomplete questionnaires, or respondents might misinterpret or misunderstand some of the questions. Interviewer-administered questionnaires solve the latter problem, since interviewer can always explain the question in more detail, if the respondent has issues with it. Interviewer-administered questionnaires are usually completed (the interviewer makes sure that all the questions were answered), and the response rates are higher. On the other hand, they are expensive to conduct, it takes more time and more administrative effort, and also there are more possibilities of interviewer bias.

7.3.1. Designing a questionnaire: general remarks

Before you start designing a questionnaire, you have to know exactly what kind of information you want to retrieve. In other words, you “must have a clear and precise research question in mind” (Rasinger 2008: 58). The most important aspect of questionnaire design is phrasing our questions. According to Rasinger (2008: 57), our respondents should be able to fully understand the questions and “answer them to the best of their knowledge”. Thus, questions have to be simple and clear. On the other hand, we should also avoid asking questions that do not provide us with useful

information (e.g., asking “Do you think it is a good idea to learn English?” **instead** of “Do you think it is useful to know English for business purposes?” and similar). The most challenging task for a researcher is to elicit people’s comments on their reasoning. In other words, most researchers usually are interested **why** people think as they think. Often, the best option is simply to ask question “Why do you think...?” and provide people with an open space to comment on it. Open questions have their own advantages and disadvantages: they might provide you with very good data or very poor data. Often people try to skip open-ended questions, because they require the respondent to focus in order to formulate his/her thoughts. Just think of yourself, how often do you answer open-ended questions (if they are not obligatory) in on-line surveys?

7.3.2. Multiple choice/item questions

A solution to the problem posed above can be the use of multiple choice or multiple item questions. These are usually questions which provide respondents with a set of prepared answers.

Example #2.

Tamošiūnaitė, in one of her surveys on language use among Chicago Lithuanians, used the following multiple choice question in order to elicit respondents’ approximate language use with their siblings (the questionnaire was bilingual):

Kokia kalba dažniausiai kalbi su savo broliais ar seserimis? / Which language do you use with your **brothers and sisters**?

- ☐ Tik lietuviškai / Lithuanian only
- ☐ Dažniausiai lietuviškai ir truputį angliškai / Mostly Lithuanian and a little English
- ☐ Dažniausiai angliškai ir truputį lietuviškai / Mostly English and a little Lithuanian
- ☐ Tik angliškai / English only

If you intend to use multiple-choice items in your questionnaire, formulate your questions carefully: they should reflect both – “yes” and “no” answer possibilities. In other words, if you ask your respondents “Why do you think English should be taught since the first grade of school?”, then, you should also ask “Why do you think English should NOT be taught since the first grade of school? (Rasinger 2008: 60). It is not so easy to prepare the answers for multiple-choice questions. You cannot just randomly choose and formulate the answers and provide them to your respondents. Your answers should “emanate from previous research findings in a field” (Rasinger 2008: 60–61).

Example #3.

In 2007 Tamošiūnaitė conducted a questionnaire-based study on Lithuanian language use among teenage Lithuanian heritage speakers living in the Chicago area. In her ques-

tionnaire, among many other things, she was interested in students' motivation/needs to learn Lithuanian. Thus, she formulated the following question: *Why do you think you need to learn Lithuanian?* However, in order to provide reasonable answers to this question, she did an overview of literature regarding youth attitudes and motivation to speak Lithuanian. Qualitative research, conducted some time ago with young Lithuanian-Americans, revealed that three factors usually motivate their choice to learn the language: their heritage, parents, and advantages of being bilingual. Considering these outcomes, Tamošiūnaitė came up with the following multiple-choice item:

Kodėl, Tavo manymu, reikia mokytis lietuvių kalbos/ Why do you think you need to learn Lithuanian (*pažymėkite visus atsakymus, kurie tinka/ mark all answers that apply*)?

- ☐ nes tai tavo tėvų kalba/ because this is the language of your parents
- ☐ nes tu lietuvis/ because you are Lithuanian
- ☐ nes geriau mokėti dvi kalbas (lietuvių ir anglų), negu vieną (anglų)/ because it is useful to know two languages (Lithuanian and English) rather than one (English)

Rasinger (2008: 61) suggests, that for smaller-scale projects (such as student projects) it is enough to rely on existing literature, whereas “in studies of larger scale or where potential answers are not as straightforwardly predictable, the construction of multiple choice items is often based on additional studies prior to questionnaire survey”. In those cases, researchers conduct small-scale qualitative studies with interviews in order to establish the guidelines for questionnaire design.

7.3.3. Likert scales or semantic differentials

Likert⁵ scales or *semantic differentials* are often employed in sociolinguistic questionnaires. Actually, these types of questions are especially useful if you are asking people their *attitudes*, *beliefs* and *opinions*. Therefore, if you intend to conduct a study on standard language attitudes among dialect speakers, a questionnaire, comprised of Likert scale answers, might be a good way to collect data. “Semantic differentials ask respondents to indicate their response along a continuum between two opposing terms” (Rasinger 2008: 61–62), for instance, from “very small” to “very big” and etc.

Likert scales are similar to semantic differentials, however, instead of using adjectives on opposing poles, Likert scales ask respondents to *agree* or *disagree* with the particular statement. Likert scales usually have a scale of 5 or 7 choices. The scale depends on “degree of detail required by the researcher” (Rasinger 2008: 62).

Important!

If you intend to use Likert scales in your questionnaire, remember, that **they measure statements**, not **questions**, e.g., you can not **measure** the question “Do you attend Lithuanian Saturday School because you want to improve your Lithuanian?”, instead you should turn it into a statement “I attend Lithuanian Saturday School because I want to improve my Lithuanian”. Below this kind of statement you can provide a scale to choose the answer.

5 The scale is named after psychologist Rensis Likert, who invented it.

Likert scales, as well as semantic differentials, give respondents an odd number of choices. It is common practice to use odd numbers (instead of even, e.g., 6, 8) in order to provide respondents with an opportunity to indicate “a ‘neutral’ or ‘balanced’ opinion” (Rasinger 2008: 62). In order to keep respondents “awake” (so that we avoid them selecting responses at the extreme end of the scales, e.g., *strongly agree* or *strongly disagree*) and to elicit both positive and negative attitudes, a researcher should phrase questions in opposing directions.

Example #4.

a) I am very satisfied with the teaching methods my teacher uses in the class.

Strongly disagree ___1___:___2___:___3___:___4___:___5___ Strongly agree

b) I am very dissatisfied with the knowledge I acquire in Lithuanian Saturday Schools.

Strongly disagree ___5___:___4___:___3___:___2___:___1___ Strongly agree

A typical 5-point Likert item looks like this:

1	2	3	4	5
Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

A typical 7-point Likert item looks like this:

1	2	3	4	5	6	7
Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

7.3.4. Phrasing, layouting, numbering

Phrase your questions (as well as all other textual information) carefully. Try to avoid academic jargon or too much of linguistic (sociolinguistic) terminology. When phrasing your questions, think of the target group you are addressing. If you intend to do a study with adults, you might choose one wording; however, if your respondents will be children, you have to adjust the questions to their level of understanding. Thus, below the statement “I attend Lithuanian Saturday School because I want to improve my Lithuanian” it is better to have a scale like this “*Totally disagree, Somewhat agree, Strongly agree*” rather than “*Strongly disagree* ___:___:___:___:___ *Strongly agree*”.

Always indicate how many responses will be accepted per question. This is very important for multiple-choice answers: in some cases you might need just one answer, in some – several answers can apply (compare examples #2 and #3 above). Avoid asking two questions in one, this will confuse your respondent. So will long and complicated questions. It is preferable to avoid *leading questions*, i.e., questions that *prefer* a particular answer (e.g., Do you think knowing more than one language is an advantage?). And finally, avoid ambiguity: use less abstract nouns and adjectives.

After you are done phrasing your questions, **test** your questionnaire first with several respondents. Piloting is very important, in order to detect any flaws or mistakes

(spelling, numbering, etc.) you might have made in the questionnaire. Some of the phrasing might not work with respondents. You might need to rephrase questions or move questions around. Piloting is essential before carrying out any questionnaire-based study.

Considering the layout of the questionnaire, try adhering to a basic rule: the simpler – the better. Use plain font, large enough to read, do not leave too much space near open-ended questions, so that you don't imply a long answer. It is advisable to use software to create questionnaires, however, a questionnaire, designed in any text processor will work fine as well. On-line survey tools are also very efficient. Tamošiūnaitė in her survey of 2008 used *surveymonkey.com*. The tool helps to design (layout) the questionnaire, publish it, collect responses, and even provides a basic analysis of your results (however, for most of the features you might need to subscribe to a professional plan).

Example #5.

Excerpt from a questionnaire designed in *surveymonkey.com*

***2. Su kuriais iš šių žmonių kalbi kasdien? / To whom of these people you speak every day (pažymėk visus, kurie tinka /check all that apply):**

- ☐ mama/ mother
- ☐ tėtis/ father
- ☐ jaunesnė sesuo/ younger sister
- ☐ vyresnė sesuo/ older sister
- ☐ jaunesnis brolis/ younger brother
- ☐ vyresnis brolis / older brother
- ☐ močiutė/ grandmother
- ☐ senelis/grandfather
- ☐ geriausias draugas(ė)/best friend
- ☐ mokslo draugai/school friends
- ☐ Kita/Other (paaiškinkite/please specify):

***3. Pasirink teiginį, kuris labiausiai Tau tiktų / Select statement which fits the best:**

- ☐ Aš turiu daugiau draugų lietuvių, nei amerikiečių (ar kitų tautybių). / I have more Lithuanian friends than American (or any other nationalities) friends.
- ☐ Aš turiu daugiau draugų amerikiečių (ar kitų tautybių), nei lietuvių. / I have more American (or any other nationalities) friends than Lithuanian friends.
- ☐ Pusė mano draugų yra lietuviai, pusė - amerikiečiai / Half of my friends here are Lithuanians, and another half - Americans (or any other nationalities).

A questionnaire should start with a brief introduction, explaining the goals of the survey. Put simple and easy questions first (such as: age, gender, occupation, etc.), then move on to more complex questions. You can group questions by topic. It is common practice to put some general questions at the end. And finally – do not forget to thank your respondents.

7.4. Comprehension activities

Activity 2. Think and list at least two (or more) cases when you won't be able to use self-administered questionnaire.

Activity 3. Work in groups. Try to come up with a set of semantic differentials indicating character (*good, brave, courageous, etc.*). You can think of semantic differentials in your native language, too. List the words. How strongly positive or negative you would judge these items on the scale from 1 to 5? Compare the results amongst yourselves (task adapted from Rasinger 2008: 67).

Activity 4. Imagine that you have to conduct a sociolinguistic study on an X minority group in the Baltic region. Based on the categories (provided below) design a questionnaire. The following steps will help you to design your questionnaire:

1. identify the subthemes for each category;
2. list these subthemes;
3. formulate a set of questions for each of the subthemes;
4. use different types of questions: multiple-choice, Likert scale, open-ended, etc., however, remember, that the question type depends on the information you want to elicit;
5. find an on-line survey-making tool (e.g., [surveymonkey.com](https://www.surveymonkey.com));
6. create an on-line survey and test it with your friends.

Categories to be addressed in the questionnaire (adapted from Codó 2008: 174).

1. biographical information (on informants, but also parents, children);
2. language use and choice;
3. language proficiency;
4. language dominance;
5. language acquisition;
6. language preferences.

Activity 5. Write down two *main questions* related to your hypothetical research that can be answered with “yes/no”. Now, change them to *open-ended* questions.

7.5. Glossary

Leading questions – questions that prefer a particular answer.

Likert scales – a scale that measures opinion and attitudes along a continuum of items.

Multiple item questions – a form of question in which respondents are asked to select an answer(s) from out of the choices from the list.

Pilot study – a small-scale study or test conducted in order to test validity and reliability of, e.g., your questionnaire.

Questionnaire – a research instrument (tool) consisting of a series of questions.

Semantic differential – a scale used for measuring the *meaning* of things and concepts.

Survey – a method for collecting quantitative data, often involving questionnaires.

7.6. References

- Eva Codó, 2008. "Interviews and Questionnaires", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 158–176.
- Lesley Milroy & Matthew Gordon, 2003. *Sociolinguistics: Method and Interpretation*, Malden, Oxford, Melbourne, Berlin: Blackwell Publishing.
- Meilutė Ramonienė, 2010. "Įvadas", in *Miestai ir kalbos. Kolektyvinė monografija*, mokslinė redaktorė Meilutė Ramonienė. Vilnius: Vilniaus universiteto leidykla, 9–24.
- Sebastian M. Rasinger, 2008. *Quantitative Research in Linguistics. An Introduction*, London, New York: Continuum.

CHAPTER 8: SOCIOLINGUISTIC INTERVIEW

8.1. Key concepts

- Careful speech
- Casual speech
- Conversational modules
- Conversational networks
- Minimal pairs
- Observer's paradox
- Semi-structured interview
- Sociolinguistic interview
- Structured interview

8.2. Pre-reading activity

Activity 1. In groups discuss and define the concepts given above. How are these concepts related to sociolinguistic interviews? Which of these concepts would you not associate with interviews in general?

.....

.....

.....

.....

.....

8.3. The interview

Wolfram calls the sociolinguistic interview “the methodological heart of the sociolinguistic movement” (Wolfram 2010: 302). Sociolinguistic interview in many ways defined the methodology of data collection for sociolinguistic research.

One of the most important goals of sociolinguistic research is **to obtain “natural speech”**, i.e., the language or the vernacular people use in ordinary, everyday interactions with all the variability that this entails (Meyerhoff et al. 2012: 123). Sociolinguists are interested in the “real” language people use. The vernacular (natural speech) has been an object of investigation since the very first sociolinguistic studies. One of the most often used techniques to elicit the vernacular is the sociolinguistic interview. In the following sections we will discuss the key features of this method.

8.3.1. Sociolinguistic interview

The main aim of the **sociolinguistic interview** is to elicit as much natural conversational speech from the interviewee as possible. Sociolinguistic analysis as such (and especially the reasoning behind the use of sociolinguistic interview) is based **on several important premises** coined by Labov (1984: 29–30):

1. *There are no single style speakers: thus, any change in topic, addressee, setting, context might signal a “style shift”;*
2. *Styles can be ranged along a single dimension, measured by the amount of attention paid to style;*
3. *The vernacular (in which the least attention is paid to speech) provides the most systematic data for linguistic analysis: each speaker has a vernacular form (acquired in early (pre-adolescent) years) and the variation in this vernacular, according to Labov, appears to be more regular than variation in later acquired (superposed, i.e., standard) varieties;*
4. *Any systematic observation of a speaker defines a formal context where more than minimum attention is paid to speech: i.e., during the face-to-face interview one does not expect to hear that vernacular which would appear when the researcher (outsider) would not be present;*
5. *Face-to-face interviews are the only means of obtaining the volume and quality of recorded speech that is needed for quantitative (and qualitative) analysis.*

According to Meyerhoff et al. (2012) the “classical” sociolinguistic interview consists of at least **four** parts:

- a. reading a list of minimal pairs (word pairs, e.g., *pin* and *bin*);
- b. reading a list of words in isolation;
- c. reading a short narrative;
- d. talking with the interviewer.

The purpose of the first three tasks is to elicit the range of different styles (see premise #2 on Labov’s list) and they are usually (at least in the earlier studies) combined with the free conversation in order to elicit a **continuum of different styles** for every speaker (Meyerhoff et al. 2012).

Example #1. Trudgill’s Norwich study (1974).

Let us take a look at the method used by Trudgill (1974) in his Norwich study. First Trudgill designed a questionnaire and used it in interviews in order to elicit *formal speech* (1974: 46). Later the respondents were asked to read a passage that included many phonological variables Trudgill was interested in. Reading the passage elicited *reading style* of the respondents (1974: 47). The passage was written in a colloquial style on an informal topic. Next, informants were asked to read aloud, at a normal speed a list of 212 lexical items (*word list*) (1974: 48). It was expected that the informants will pay most of the attention to the pronunciation of the words, thus, they would elicit a formal style (the most artificial and unnatural style). The list included all of the items in the reading passage (1974:48), thus, this later allowed Trudgill to compare stylistically production of different phonologi-

cal variables. And finally Trudgill and his colleague conducted 60 interviews and elicited *casual speech* (speaking about the Norwich). Thus, Trudgill elicited data in four different styles: *word list*, *reading passage*, *formal style* and *casual style*.

8.3.1.1. Goals of sociolinguistic interview

The “classical” sociolinguistic interview should help to achieve a number of goals. Labov (1984) lists at least **ten** of them:

1. to record at least one to two hours of speech of each speaker;
2. to obtain a full range of demographic data (social class, age, gender, school, occupation, income, group membership, and other);
3. to obtain comparable responses to questions that define different experiences and attitudes (such as the famous question of the danger of death);
4. to elicit narratives of personal experience (that will reveal community’s attitudes and norms toward particular phenomena);
5. to stimulate group interaction during the interview (so that the attention is deviated from the interviewer);
6. to allow the interviewee lead in defining the topic of conversation (this might result in longer narratives by the interviewee and less-monitored speech);
7. to trace the patterns of communication among members of the speech community;
8. to obtain overt attitudes toward language, linguistic features and linguistic stereotypes;
9. to obtain linguistic information through formal elicitation: reading text and word pairs;
10. to carry out field experiments on subjective evaluations toward perception of linguistic forms.

According to Labov, in order to achieve these goals a sociolinguist has to follow the use of two technical devices: a) **the module** and b) the **conversational network**.

8.3.1.2. Conversation module

Labov defines the conversational module as a group of questions focusing on a particular topic, i.e., childhood experiences, university experiences, danger of death, aspirations and other (Labov 1984: 33). The conversation modules help to construct the structure of the interview. The module structure of the sociolinguistic interview was developed by Labov during the 1970s, mainly for use on the research conducted in Philadelphia speech community. In his studies Labov relied on the following generalized set of the module structure, called Q-GEN-II:

- Module 1: Demography
- Module 9: Family
- Module 5: Marriage
- Module 15: School
- Module 3: Fights

- Module 6: Danger of Death
- Module 6.1: Fear
- Module 12: Race
- Module 13: City Services
- Module 14: Crime in the Streets
- Module 11: Peer Group
- Module 20: Language (Labov, William. 1973–77. Q-GEN-II: Sociolinguistic Interview Modules).

According to Labov (1984), the questions of each module were shaped over by three main processes:

1. responses to generalized foci of interest (so that questions might be applied for use in different speech communities [different cultures]);
2. colloquial content: most of the questions should be formed using colloquial language so that it would be easier to carry on the conversation with the interviewee;
3. shortening: module questions should be short (take not more than 5 seconds to deliver);
4. feedback: questions should be formulated from an outsider's point of view.

Labov suggests that the interviewer gets very well acquainted with the module questions and later personalizes them and adapts them to his/her colloquial style (1984: 34). The modules are organized in hierarchical structure: they always start with a more general question and then proceed to a more detailed discussion of the topic.

Example #2.

Module “Fear” in Labov’s questionnaire (*Q-GEN-II, Module 6.1, Developed by W. Labov, 20 Mar 1973*)⁶

Fear

1. Have you ever known what it was to be really afraid?
 - 1.1. When was that? What happened?
 - 1.2. How did you feel afterwards?
 2. Did you ever dream about something that happened to you before, like that?
Have it all happen again?
 3. Did you ever know somebody that wasn’t afraid of anything?
What kind of a person was he?
 - 3.1. Or is it just that some people can’t admit it when they are afraid?
- [for 10-12 year olds]*
4. Did you ever have a dream that really scared you?

8.3.1.3. Conversational network

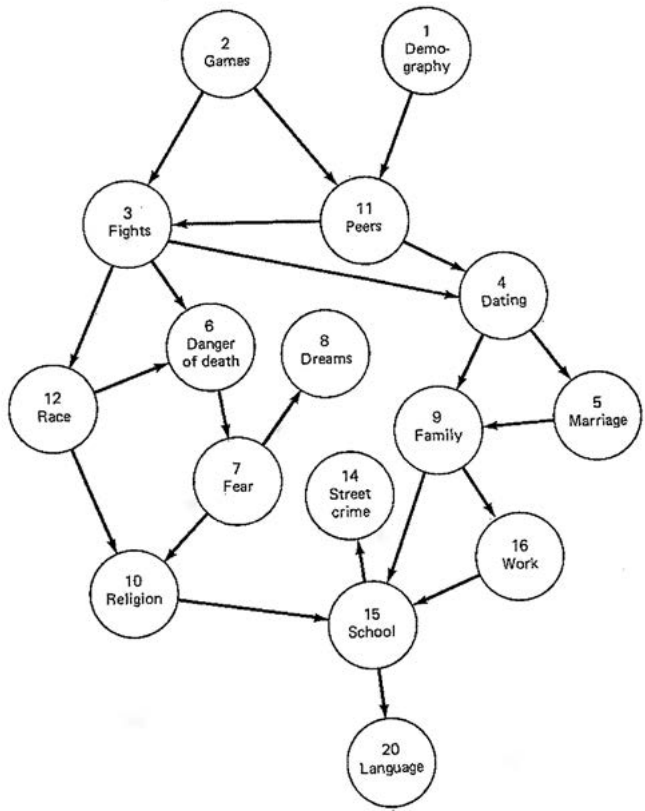
All of the modules are combined into a conversational network. Modules begin and end with the so-called transitional questions that might lead or connect to another

⁶ <http://courses.essex.ac.uk/lglg554/QGEN61%20Fear.htm>.

module in the conversational network. For example, module 2 “Games”, might lead to the questions associated with module 3 “Fights” or module 11 “Peers”; the latter module might lead to module 4 “Dating”, etc. Transitions to other modules might be initiated by the interviewer or might occur naturally (Labov 1984: 34).

According to Milroy & Gordon (2003: 60), “the general idea of this interview schedule is to use interlocking modules to simulate the seamless topic-shift structure of normal conversation.” A researcher always has to design the questionnaire carefully. The questions have to fit community norms, thus, careful attention should be paid in terms of phrasing and ordering within modules (Milroy & Gordon 2003: 60). According to Labov (1984: 37), “the most successful interviews follow the path which is both natural to the speaker and comparable to other paths”. While unsuccessful interviews are those, during which the speakers only answer questions. The interview should simulate a natural conversation. This is a difficult task to accomplish, especially if you are interviewing a complete stranger. The interviewer should pay attention to the topics where the respondent feels the most comfortable and confident. Elaborate on these topics and be flexible. This will help to elicit good data.

Figure 3: An example of conversational network of modules used by Labov in the Philadelphia study (adapted from Labov 1984: 35)



8.3.2. Observer's paradox

No matter how natural the interview might be, the interviewee knows that s/he is being observed. The researcher is then faced with what Labov calls "observer's paradox", that is the "inescapable fact that speakers are more aware of what they are saying and how they are talking as soon as you begin recording them" (Meyerhoff et al. 2012: 123). The observer's paradox is a phenomenon that every sociolinguist, conducting an interview or observation, has to consider. Meyerhoff et al. (2012: 124–125) point out several methods that help to mitigate the effects of the observer's paradox:

1. Modification of the number of people in an interview. Most people feel uneasy when they are interviewed in a one-to-one setting. Thus, inviting several people to participate in a conversation might result in better data and a relaxed interview.
2. Modification of the number of interviewers. This may not seem usual, however, interviewers might frame themselves as friends and their natural conversation might develop into a less-formal interview with the respondent.
3. Removal of the interviewer. This strategy minimizes the effect of outsider presence. However, then there is no control over the recording.

A change in the content of the interview sometimes helps to achieve more casual speech. It has been noticed by many sociolinguists (first of all, by Labov himself), that "when people are emotionally involved (excited, angry, fearful, etc.) in a discussion, they are more concerned with what they say than with how they say it" (Milroy & Gordon 2003: 65). A classical example of such emotional involvement is the "danger of death" question used by Labov. However, this kind of question may not be appropriate to ask in certain situations. If you are conducting an interview with a child and ask whether s/he ever feared for her/his life, you may not receive an expected answer. On the other hand, not all people feel comfortable remembering painful experiences of their lives.

Although interviews may not be the best tools to obtain natural data, if some of the strategies mentioned above are applied, you might collect a pretty good set of data.

8.3.3. Semi-structured vs. structured interviews

The "classical" sociolinguistic interview (used by Labov or Trudgill) can be defined as a semi-structured interview. Researchers prepare specific questions in advance (as, for instance, Labov and Trudgill did) and during the interview they move from one question to another (or from one module to another, if we refer to Labov's conversational network). "A semi-structured interview allows the researcher to offer topics that participants can talk about and lead them to have more dynamic discussions between themselves" (Meyerhoff et al., 2012). Thus, the researcher does not take a full control over the conversation and this helps to create a more relaxed atmosphere during the recording. The drawback of such a technique is that some people might digress to some topics, while others to other topics and the data may not be comparable between respondents. On the other hand, you may end up having a large amount of natural speech data.

In structured interviews (e.g., questionnaire-based) participants do not self-select questions for further discussion. Such interview has certain drawbacks, since usually it is hard to achieve a relaxed atmosphere during the interview. On the other hand, a researcher has the control over the interview and elicits all of the required information. If you interview many people using structured interview, this data becomes comparable between respondents.

8.3.4. Questioning techniques

Interviews are difficult to conduct. Not all respondents are good and compassionate talkers; therefore, sometimes it might seem impossible to elicit answers from someone who is not willing to collaborate. Be patient. Be compassionate. And follow some of these most common tips provided by Tagliamonte 2006:

- tailor your style to suit your interviewee, modify your language if needed, but do not overdo it;
- avoid yes or no questions. In other words do not start questions with “Do you”, but rather, What?, Why? and How?
- use indirect means, e.g., “Is it true that...”, “I’ve heard that...”;
- be careful when asking about person’s age, education, political views, religion;
- word your questions in terms of other people (in general), rather than asking something specific about the interviewee (“Some people say that it was much better to live during Soviet times. What do you think?”);
- do not ask too personal questions (how much money do you make, what are your religious beliefs, etc.);
- organize your questions so that one question naturally leads to the next, prepare your questions in advance;
- **do not read** your questions during the interview;
- do not make a recording next to a source of noise (TV, radio, refrigerator, aquarium, open window to a busy street, etc.);
- let the informant talk, do not interrupt the interviewee;
- use your natural variety (do not sound too formal or too standard); relax and speak informally (it is not easy to do!);
- ask short questions, do not formulate long, complex sentences.

In conclusion, here is Labov’s test that you can apply in order to see whether you conducted a good interview.

Fast-forward an audio-record of an interview.

Listen. Who do you hear? The interviewee? If so, good.

If you hear the interviewer, go forward another five minutes into the interview.

Listen. Who do you hear? The interviewee? Great.

If you hear the interviewer, go forward another five minutes.

Listen. Who do you hear? The interviewee? Wonderful!

If all you hear is the interviewer using this technique, the interview is not so good (adapted from Tagliamonte 2006: Kindle Locations 505–506).

8.4. Comprehension activities

Activity 2. Think of yourself. Are there situations in your life when you do not think of the way you speak? What are those situations? Make a list and discuss them with your friends.

Activity 3. It is not so easy to come up with the reading passage that would involve different representations of the variable you want to study. Challenge yourself. Think of the phonological variable in your native language (or other language you know well). Try to come up with a 100-word narrative that would showcase the variable in as many environments as possible (task adapted from Meyerhoff 2006: 30).

Your variable is: _____

Your narrative:

Activity 4. Take a look once again at the interview module scheme designed by Labov.

- ☐ Module 1: Demography
- ☐ Module 9: Family
- ☐ Module 5: Marriage
- ☐ Module 15: School
- ☐ Module 3: Fights
- ☐ Module 6: Danger of Death
- ☐ Module 6.1: Fear
- ☐ Module 12: Race
- ☐ Module 13: City Services
- ☐ Module 14: Crime in the Streets
- ☐ Module 11: Peer Group
- ☐ Module 20: Language (William Labov. 1973–77. Q-GEN-II: Sociolinguistic Interview Modules).

Based on these modules, create a questionnaire, think of at least three questions you could ask for each module.

Activity 5. Which of these statements are true, which are false.

1. We need to abandon our cultural assumptions in order to see & understand those of another.
 - a) True
 - b) False
2. In order to be a successful interviewer, we need to be as neutral as possible.
 - a) True
 - b) False
3. You need a very very large number of interviews to make your account credible.
 - a) True
 - b) False

Activity 6. Take a look at the protocol of sociolinguistic interview, used by Au-relija Tamošiūnaitė and her colleagues in a study on the use of heritage Lithuanian and Polish in Chicago. Using this interview as a template form, create your own protocol of sociolinguistic interview. Imagine that your respondents are minority speakers in Lithuania, Latvia, Estonia, Germany or any other country. Your goal is two-fold: to elicit their attitudes toward language (on different aspects) as well as natural speech. Think, which language you will use for the protocol: dominant or minority? In which language you will conduct interviews? Why? How might language choice affect the interviewee's responses to some questions?

SOCIOLINGUISTIC INTERVIEW

Questions marked with “” are required. They must be asked to all interviewees. Otherwise, ask questions that you sense will be well received by the interviewee.*

***Do not read down the list mechanically!!
The list is meant to help you generate natural conversation.***

As long as you cover the required questions, you may actually spend the rest of the interview on only one question that you and the interviewee discuss with great detail!! Or you may end up asking many of these questions. Just try to make it feel like a natural conversation as much as possible.

General information

- *How old are you?
- *Where were you born?
- *Where do you live now? What is the zip code?
- *In what other places have you lived?

The family

***Where were your mother and father born?**

In the USA:

Where are your grandparents from?

Outside of the USA:

How old were they when they arrived in the USA?

***Do you know how your parents met?**

***What kinds of jobs have your parents had?**

***How frequently do you travel to Poland/Lithuania? How long do/did you stay each time?**

Do you have siblings? How old are they and what do they do?

Do you have a sibling that you get along with really well? Why?

Do you have children? How old are they? Do they speak Polish/Lithuanian?

The neighborhood

What do you like the most about where you live? And the least?

Has your neighborhood changed a lot in the last 10 years?

Do you feel safe in your neighborhood?

Are there any problems in your neighborhood? For example, are there gangs?

Are the police helpful, or do they create more problems?

Childhood

What did you play when you were a child? How do you play that/what are the rules?

Did you ever fight with siblings or with other children? Can you tell me about one fight in particular that you remember?

What is one of your best memories from when you were a child?

School

***What are the names of the schools you have attended? What do you think about those schools?**

Tell me a story about something that happened in school that you'll never forget.

What do you think about bilingual education?

Work

Are you working now? Where? Have you had other jobs?

Do you ever use Polish/Lithuanian at your job?

What do you like the most about your job? And the least?

What would your ideal job be?

***Language**

All of these questions are required, but should NOT be asked at the beginning of the interview. Wait until some time has passed and the interview feels comfortable.

When you were a child, what languages were spoken in your home? (Who spoke what to whom?)

How did you learn Lithuanian/Polish? How did you learn English?

Did your parents ever say anything about the importance of a particular language?

Nowadays, when/in what situations do you speak Lithuanian/Polish?

What language will you (do you) speak to your children? Do you want them to know Lithuanian/Polish?

Have you ever felt proud because you speak Lithuanian/Polish? Tell me about it.

Describe for me your Lithuanian/Polish abilities. On a scale of 1 to 10, with 10 being very fluent, how well would you say you speak it? Describe why you say that.

Friends

Who are your best friends?

What is necessary in order to form a strong friendship?

What kinds of things cause problems between friends?

Have you ever lost a good friend? What happened?

Romantic relationships

Do you have a partner? Where did you meet?

What are some things that cause fights between partners?

[If the person is married] What was your wedding like?

Other

What is the moment in your life when you felt the most scared?

What would you do if you won 10 million dollars?

8.5. Glossary

Careful speech – formal, self-conscious speech.

Casual speech – informal, relaxed or everyday speech.

Conversational modules – sets of questions organized around certain specific topics.

Conversational networks – sets of conversational modules organized into a network.

Minimal pairs – pairs of words which differ only in terms of their phonological element (e.g., *pin* vs. *bin*).

Observer's paradox – phenomena where the observation is influenced by the presence of the observer/investigator.

Semi-structured interview – a type of interview that does not follow a strict question-answer protocol, but allows researcher to bring up the topics and new ideas on which respondents elaborate. Semi-structured interviews are usually employed in qualitative research.

Sociolinguistic interview – a less structured form of interview, based on conversational modules, conducted by sociolinguists. A classical sociolinguistic interview involves: reading a list of minimal pairs, reading a list of words in isolation; reading a short narrative, and talking with the interviewer.

Structured interview – a type of interview that follows a strict protocol of asking questions, prepared in advance by the researcher. Interviewers read the questions exactly as they appear on the questionnaire. Structured interviews are used in surveys and questionnaires (quantitative research).

8.6. References

- William Labov, 1984. “Field Methods of the Project on Linguistic Change and Variation”, in *Language in Use*, ed. by J. Baugh & J. Sherzer, Englewood Cliffs: Prentice Hall, 28–53.
- Miriam Meyerhoff, 2006. *Introducing Sociolinguistics*, Abingdon, New York: Routledge.
- Miriam Meyerhoff, Chie Adachi, Golnaz Nanbakhsh, Anna Strycharz, 2012. “Sociolinguistic Fieldwork”, in *The Oxford Handbook of Linguistic Fieldwork*, ed. Nicholas Thieberger, Oxford: Oxford University Press [online version].
- Lesley Milroy & Matthew Gordon, 2003. *Sociolinguistics: Method and Interpretation*, Malden, Oxford, Melbourne, Berlin: Blackwell Publishing.
- Sali A. Tagliamonte, 2006. *Analysing Sociolinguistic Variation*, Cambridge, New York: Cambridge University Press [Kindle Edition].
- Peter Trudgill, 1974. *The Social Differentiation of English in Norwich*. London, New York: Cambridge University Press.
- Walt Wolfram, 2006. “Variation and Language, an Overview”, in *Encyclopedia of Language and Linguistics*, 2nd Edition, ed. Keith Brown, Elsevier Ltd. [online version].

CHAPTER 9:
QUALITATIVE TECHNIQUES: OBSERVATIONS AND RE-
CORDINGS (AUDIO & VIDEO). TRANSCRIPTION

9.1. Key concepts

- Audio levels*
- Background noise*
- Community of practice*
- Ethnography of communication*
- Participant observation*
- Resolution*
- SPEAKING*
- Speech community*
- Transcription*
- Transcription protocol*

9.2. Pre-reading activity

Activity 1. In groups discuss and define the concepts given above. How are these concepts related to ethnographic research?

.....

.....

.....

.....

.....

9.3. Doing ethnography

The **main aim** of this chapter is to introduce you to qualitative techniques of socio-linguistic research, namely participant observation. The chapter also discusses tools for data collection (recordings), as well as provides a few hints for data transcription.

9.3.1. Participant observation

Participant observation is one of the primary techniques used for ethnographic research (Johnstone 2000: 81). This technique was developed by cultural anthropologists, who studied (by observing and living in) small, isolated, traditional societies and communities (Johnstone 2000: 81). In sociolinguistics the tradition to use participant observation stems from the work of John J. Gumperz (Labov 1984: 28) and Dell Hymes (Johnstone 2000: 84). Hymes proposed the term “**ethnography of com-**

munication” by pointing out that a community’s speech and writing (language) are themselves an “aspect of culture worth of study” (Johnstone 2000: 84).

Ethnography of communication was favoured by many linguistic anthropologists and folklorists especially during the 1970s. Although most of the early sociolinguists practiced sociolinguistic interview as the main means of eliciting the “best” data, ethnographic techniques have also been involved in their studies. Ethnographic data provide sociolinguists with the necessary background (cultural) knowledge that helps to explain the reasoning behind different phenomena. As Johnstone (2000: 83) puts it, one of the advantages of the ethnography is that it provides “explanations of human behavior that cannot be uncovered through other kinds of research such as experimentation”. In other words, if a researcher wants to find out the reasoning of certain human behavior (e.g., why one dialect is perceived as more beautiful than another or why one language is perceived as having more prestige than another) s/he has to rely or at least support the research with ethnographic methods.

Ethnography studies **groups of people**. These groups can be quite diverse: one can study one speech community (e.g., a Polish community in Lithuania or Russian communities in Estonia or Latvia), one can study a neighborhood community, a city community or just a random group of people (depending on the aims of the research). Usually linguists tend to divide people into groups defined by the names of languages: Lithuanian speakers, German speakers, Estonian speakers, and other. However, such differentiation is rather problematic, since the speakers of those groups might identify themselves differently, with different groups. At the same time people may belong (form) to different kinds of social groups, whether we identify them “through shared practices (*communities of practice*), shared patterns of association (*social networks*) or shared patterns of variation (*speech communities*)” (Meyerhoff et al. 2012).

Participant observation, as a method, has two benefits for sociolinguistic research: a) “the amount and quality of the data collected, and b) the familiarity with community practices gained by the investigator” (Milroy & Gordon 2003: 68). Participant observation works well in small and well-defined communities (school community, minority community). It is crucial for a researcher, involved in participant observation, to develop an in-depth understanding of the community, i.e., to get well acquainted with the norms, traditions, local knowledge, and “sociolinguistic dynamics” (Milroy & Gordon 2003: 71).

Example #1.

Penelope Eckert, *Jocks and Burnouts: Social Categories and Identity in the High School*, New York, London: Teacher’s College Press, 1998.

A classic example of sociolinguistic ethnography is Eckert’s study of “jocks” and “burn-outs” in one of the Detroit-area suburban high schools. Eckert spent two years studying a suburban high school in Detroit area. She remained an outsider in the community, while at the same time she spent a lot of time in the school, outside the classrooms, in the library, cafeteria, i.e., in different settings within the school. She observed student’s

behavior, as well as interacted with them. Part of her ethnographic research was also to interview 200 students (individually or in groups). Eckert earned the students' trust over two years, and this resulted in a very good data in terms of quantity and quality (Milroy & Gordon 2003: 69). The familiarity with students helped to eliminate 'observer's paradox'. Ethnographic research allowed her not only to collect natural speech data, but also to observe student's behavior. Therefore, in her analysis she could support her interpretations relating the use of certain linguistic variables with two different communities of practices that she observed in the school, namely *jocks* and *burnouts*. According to Milroy & Gordon (2003: 69), Eckert's study emphasizes that "the practices associated with being a jock or a burnout can be better understood by a researcher directly observing them than by just hearing about them in interviews". In other words, participant observation not only provides us with our linguistic data, but also provides us with data that explains linguistic behavior.

Participant observation is valuable in studying language used in bilingual or multilingual settings. Code-switching, linguistic choices of bilinguals can be better explained after an in-depth observation of the respondent's daily routines in different settings. According to Milroy and Gordon (2003: 71), participant observation "produces a tremendous supply of high-quality data and crucial insight into community dynamics".

9.3.1.1. Steps

Johnstone (2000: 88) suggests several crucial steps everyone has to consider **before the actual participant observation process**:

- a. selecting an **area** of study (e.g., you are interested in the attitudes toward dialects in urban settlements, thus, you will choose to study urban environment);
- b. selecting the **group** (of people) being studied (e.g., your selection of the group might be determined by age, gender, ethnicity, neighborhood);
- c. **reviewing literature** of what has been done by other researchers in the field;
- d. **finding a way** into the group (i.e., initiating the contact with the group you want to study).

Once you find a way into the community, you select your research participants. Although observations quite often include a group of people, researchers often select a subgroup (several) of people to study more intensively than others (they focus on what might be called "case studies"). Be selective when you choose your "case studies". As Johnstone suggests (2000: 89), the first people we contact in the community are usually those, who are quite different from the rest of the community members. Agar calls them "professional stranger handlers" (quoted from Johnstone 2000: 89). Such people can be: the mayor of the town, a teacher, a priest, a doctor, real estate agent; in other words, all of those people, whom you would contact first in order to establish a connection with the community. But be aware that their linguistic behavior and their attitudes, due to their openness to strangers, might be different than other "core" community members.

9.3.1.2. Strategies

Participant observation usually occurs in two phases: 1) “unplanned and exploratory”; 2) “more systematic” (Johnstone 2000: 90). During the first phase, the researcher observes, conducts interviews, and transcribes some of the data in order to get a sense of what makes sense and what is meaningful in the community. After the initial acquaintance with the data, a researcher should narrow down the focus and employ more systematic techniques in order to collect the data. Thus, you have to build and justify your methodology (why you selected those and not other participants, how did you sample them, what methods did you employ, etc.).

Important!

Always remember that “good ethnography relies on lots of observation, systematic observation, and observation of various sorts” (Johnstone 2000: 95).

In order to get as much valuable data as possible Johnstone (2000: 95, following Erikson 1986), suggests following these tips:

- a. during the observation take a different angle each time: watch a different person, sit in a different place, go at a different time, etc.;
- b. always make notes and always take time to think after each observation. Take notes systematically: note what you see while observing daily routines of your subjects, make charts, schemes, layouts of the settings, put down main key concepts, ideas, i.e., inscribe social practices you see;
- c. try various kinds of participation: sometimes be silent, sometimes participate actively (talk), play with your identities (one time you can be a university student, next time – a local teacher, etc.). Be diverse. And watch what happens;
- d. systematically look for discrepant cases: if everyone says something one way, try to find out who does not do that;
- e. use recording devices (anywhere where possible): audio (and today – video) **recording is an essential part of research design**. It is also useful to make recordings of informal conversations, initially not intended for transcription or analysis. They might become handy later. In sociolinguistic work (different from other fields where participant observation technique is also applied) it is **very important to record exactly what people say**, not paraphrase them (Johnstone 2000: 100).

9.3.1.3. What to observe

The object of observation always depends on your research question. However, there are several key aspects that should be observed. Ethnographic analysis of communicative events often is based on Hymes “SPEAKING paradigm” (Johnstone 2000: 96). Each letter in the SPEAKING acronym represents an important aspect of the communicative event that a researcher should consider:

S – refers to **setting**; thus, it is important to indicate the place where the communicative event is taking a place, e.g., in a classroom, in a court, in a medical institution,

on a street. Some of the communicative events can take place anywhere, some – only in particular places with specific layout (a class usually takes place in a classroom which has a specific layout).

P – refers to **participants**. In each communicative event there are actors who are speaking or acting in certain manner, e.g., in a classroom we usually have students and a professor (or a teacher). Each of these actors might have different roles assigned to them.

E – refers to **ends**, or in other words – the purposes or goals that participants or institutions have. E.g., the purpose of having a class is usually teaching (from the professor's point of view) and learning (from students' point of view). However, the "ends" might be different, this might make the observation even more interesting.

A – refers to **act sequence**. Act sequence indicates the sequence in which certain things during the communicative event happen. For instance, in a classroom setting usually the teacher leads the lecturing and discussion, when s/he enters the class, s/he starts the class, once s/she speaks, students listen, s/he introduces certain topics, there is a homework check or homework assignment at the end, i.e., there is a certain sequence of events that all participants expect to happen.

K – refers to **key**, in other words – the tone, the style or the register of the communicative event. The class is usually formal, although, both the students and the professor might joke. However, one expects the answers to the exam questions to be provided in a formal manner.

I – refers to **instrumentalities** or media of communication, e.g., during class, usually both – oral and written means are used. Lecturing is usually provided orally, while homework assignments, tests are usually done in writing. Lecturing these days can also be done via Skype – on-line, thus, new media technologies might also be used in a classroom.

N – refers to **norms**: both – norms for interaction and norms for interpretation. E.g., in a classroom one expects that the teacher speaks first, that students do not enter and leave whenever they want; agreement is usually reached regarding the assignments (if a student is late, the grade might be lowered, and so on).

G – refers to **genre**, or the way participants categorize the event, e.g., a class can be of different types: lecture, seminar, workshop, and others.

According to Johnstone (2000: 99), this list of aspects "helps to systematize the process of observation and inquiry." On the other hand, ethnographic research does not have such pre-defined tools, techniques or procedures of analysis, as, for instance, variationists have. An ethnographer sometimes has to use his/her own intuitions or imagination mixed with a researcher's "eye" in order to "unpack" studied phenomena and provide the reasoning for it.

Ethnographers tend to think of the culture as a narrative, since culture involves "a set of ways of doing things rather than a set of facts people know" (Johnstone 2000: 100). According to Johnstone (2000: 100), no ethnography can be comprehensive. All ethnographies are researcher's choices, be they conscious or unconscious; it is al-

ways the choice of the researcher on what to observe and what to present (Johnstone 2000: 100). Thus, one of the main tools for the ethnographer is a note-book or a journal (computer or a traditional one) where the researcher puts his/her observations. If you intend to analyze variation, make sure you support your observations with recordings (ideally both, audio and video).

According to Heller (2008: 250), ethnographies often take “an *interpretivist* stance: they aim to discover how people use language, what they believe about language, and why, as aspects of socially constructed reality.” Thus, if you are looking for in-depth explanations for certain phenomena, you are interested in processes (shifting ideologies of language, language practices) rather than objects, or, simply put, you aim to find an answer to a question starting with “why”, you should **rely on the ethnographic approach**.

Important!

Keep in mind that ethnographers do not seek to exactly understand a *community's* (or group's) *perspective*, they seek to understand *how members* of that community understand *their own world* (Kim Potowski's advice, personal communication with A. Tamošiūnaitė).

Example #2.

Monica Heller, *Linguistic Minorities and Modernity: A Sociolinguistic Ethnography*, London: Continuum, 2006.

The scope of ethnography: The book aims to **answer the question:** *what it means to be francophone, and to speak French, as seen through the life of a French-language minority high school in Toronto, Ontario, Canada.*

The study is preoccupied with the following **research questions:** a) How the school, as an institution, constructs and implements its linguistic norms? b) How students are positioned with respect to the school's public discourse on language and identity?

The study was conducted in l'École Champlain high school in Toronto in the early 1990's. Heller together with her research assistants worked in the school for over 3 years (September 1991 – March 1995) (cf. Eckert's study on Detroit high school).

Methodology: the research employs ethnographic techniques in order to describe and explain everyday language practices at school. During the first year researchers started looking at things from the perspective of the school as an institution: they interviewed administrators, teachers, non-teaching staff, read school-related documents (bulletins, callendars, yearbooks, etc.), also 2 days a week they conducted participant observation in the classrooms, in the teachers' lounge, etc. Researchers tape-recorded an average of six each for the six classes observed during the first year. During the second year researchers focused on participant observation and interviews with 15 students. They followed them up around the school, spent time outside the school, and interviewed them. Overall 46

interviews were made with these students. During the 3rd and the 4th year researchers focused on “communities of practices” or “social networks” within the school: a) ‘smokers’ – who were socially and academically marginalized; b) groups of black students (mainly African, also Haitian); c) a group of academically successful but socially marginalized girls, ‘nerds’ or ‘rejects’; d) a group of boys who called themselves ‘multiculturals’; and a group of academically and socially successful students, the ‘populars’.

Several groups of students attend l'École Champlain: a) bilinguals, who grew up in an English-dominated community but go to French-language school. Their L1 is English. Principal link to French is only through school; b) *Lés Québécois*, who are monolingual francophone students coming from Quebec, they have a strong affiliation with Quebec and are usually in Toronto against their will. They are stigmatized by the variety of French they speak; c) students from other francophone countries (mainly Somalia and Haiti). Thus, in terms of student body the school was very diverse, both linguistically and culturally.

Despite many advantages, participant observation (ethnographic approach), similar to quantitative approaches, has its own limitations. Milroy & Gordon (2003: 71–72) point to the following drawbacks:

- a. participant observations (ethnographies in general) are time consuming (just think of Eckert's and Heller's studies, how long did it take to conduct them);
- b. observations are demanding not only in time, but also in “energy, tact, and emotional involvement with community members” (Milroy & Gordon 2003: 71);
- c. in terms of data, participant observation will result in many hours of recorded speech; the thorough analysis of this data will be inefficiently time consuming;
- d. limited availability to generalize and locate the results based on the analysis of one community of practice in a wider sociolinguistic (or global) context (Milroy & Gordon 2003: 72).

In order to minimize these limitations, participant observation is usually combined with other approaches. Thus, ethnographic research as such often involves observations, interviews, recording, document analysis, and even quantification of data (e.g., Mendoza-Denton's study of 2008).

Important!

“No ethnographer is a blank notepad just as no linguist is a tape recorder” (Mendoza-Denton 2008: 48).

9.3.2. Recordings

Recordings are primarily used to record speech in order to analyze the structure of the language (Margetts & Margetts 2012: 14). The workflow with audio and video recording usually goes through different stages. As Margetts & Margetts put it: “the steps typically include making the recording, capturing or copying recordings to a computer, identifying and cutting sessions, transcoding media files to open and com-

mon formats, chunking sessions into units (information units, pause units, intonational units, etc.), transcribing, interlinearizing, and translating the text” (2012: 14, 15).

9.3.2.1. Audio recordings

Audio recordings might be sufficient for different purposes, i.e., different linguistic analysis. However, it is always preferable to have video recording, since it adds more additional (meta) information required for the analysis (especially, if your analysis also involves gestures, intonation, mimics, etc.). If you decide to conduct audio-only recordings, Margetts & Margetts (2012: 17) suggest additionally to record basic meta-data about the recording session, i.e., to indicate the date, the place, the name of the informant, setting, make additional important notes about the session.

Things to consider when making audio recordings:

- a. *resolution*: use high-resolution, non-compressed audio. Compressed recordings diminish the acoustic quality of the data. Each your recording should be valid for different type of research. Thus, even if you are not interested in phonetic production of your recording, in the future your colleagues or you might reuse it for phonetic analysis.
- b. *background noise*: a researcher sometimes has a choice between a noisier (naturalistic) setting and a quieter (less relaxed) setting. Each of these two have their own advantages and disadvantages: the richer setting might delete some of our audio information, while the quieter setting usually implies tension between the informant and the researcher. Be aware that rain, wind, surf, animals and engines of all kinds make problematic background noise (Margetts & Margetts 2012: 18). Thus, try not to conduct your recording sessions around sources of this noise. On the other hand, always remember that the position of the microphone makes a big difference in a noisy recording setting (Margetts & Margetts 2012: 18). Thus, if there is no way for you to escape the background noise; adjust the microphone so that you can get the best data available.
- c. *audio levels*: balancing input levels (setting the recording level as high as possible) is very important and challenging task, however, most of the high quality equipment allows that.
- d. *equipment*: recording quality has to be a top priority for choosing the equipment (Margetts & Margetts 2012: 19). Margetts & Margetts (2012: 19) suggest using the machines which can record non-compressed digital audio (16 bit/44kHz or better resolution) that can be stored as **wav** (Windows) or **aiff** (Mac) files. There are several types of recorders available for use:
 1. **solid-state recorders**: they record directly to flash memory rather than to tape or disk. These are the most popular recorders used in our everyday life (consider your own phone: it might also include one of this kind of recorder). Margetts & Margetts (2012: 20) recommend that you choose the machine that accepts external microphones.
 2. **direct to computer recorders**: generally it is not recommended to have a com-

puter out in the field because your computer will consume more power and it may not have a built-in sound card capable of high quality recordings. One situation where direct to computer recordings might be useful is the group recording.

3. **lower cost MP3 recorders:** the most popular due to their relatively low price. They are also easy to handle, carry, and use. On the other hand, they record data in a compressed format (MP3), thus, for extensive phonological research, recordings done with these types of machines may not be suitable.

Beside recorders, a researcher should always have a microphone (or set of microphones, as external microphones should be the first choice for recordings) and headphones (to listen to the recording on site, in order to evaluate its quality).

9.3.2.2. Video recordings

These days video recordings are a much more advocated method of recording than audio. According to Margetts & Margetts (2012: 32), video contains extralinguistic information, such as gestures, facial expressions, also information on who is speaking, who is present, orientation and seating arrangement, recording location, time of the day, and other;

Things to consider when making video recordings:

- a. *resolution:* different from audio files, it is hard to store and analyze uncompressed video files. Even compressed video files sometimes are too big to work. Those, interested in the recommendations on video resolutions can check the guidelines provided by large linguistic archives such as DoBeS⁷.
- b. *filming techniques:*
 1. make sure that the horizon is flat (and not at an angle);
 2. if possible use a tripod (vs. handheld filming). This way you will avoid the moving and shaking of the camera. As Margetts & Margetts (2012: 36) point out, speakers tend to relax more and forget about the camera if there is no one constantly standing behind it;
 3. you are not required to perform any zooming or panning if your speakers are stationary (they do not move). Thus, try to use less fancy moves as possible, because you might miss the information (gestures, by zooming the face) when needed;
 4. if possible provide plenty of light (it is always easier to deal with a “lighter” picture, than with the dark one). Do not film against the light or a bright background. According to Margetts & Margetts (2012: 39), “the ideal arrangement would be for the speakers to sit at a 45 degree angle to the light, outdoors, and positioned in such a way that bright objects, such as walls, reflect light onto the face”.
- c. *equipment:* use cameras that can record audio in uncompressed linear PCM (pulse-code modulation). Margetts & Margetts recommend using the MiniDV tape cameras (e.g., Sony HVR-A1P). According to them (2012: 42),

⁷ <http://www.mpi.nl/DOBES/>.

it produces the required audio format and quality by default. Solid state/hard-drive cameras are less-preferred choices, although the researcher would spend much less time converting tape to digital format. These cameras might be also used, but one has to show extra care that the recordings are created or converted to archivable formats (Margetts & Margetts 2012: 43).

9.3.3. Transcription

Transcriptions always involve choices. They usually depend on your research goals, namely, what information you will need to extract from the transcripts. For instance, if pausing is important to your research, you will have to transcribe and carefully mark all of the pauses made by respondents; you might even need to indicate the length. On the other hand, if you are interested in the content of the interaction, you might not need to transcribe every pause used by respondent.

A transcript is just a partial representation of a conversation, “and transcribers have to decide what information to include and what to leave out” (Johnstone 2000: 117). These choices can sometimes affect our data, e.g., choices about how to punctuate or how to spell certain words, might affect the way we “inscribe” social status, or the linguistic proficiency of our respondent (Johnstone 2000: 117).

Transcription not only involves the process of representing oral language in written form (Turell & Moyer 2008: 193), but it also involves “transcription” of paralinguistic information, e.g., accents, overlaps, pauses, hesitations, emphasis, mimicry, and gaze. Thus, before the transcription, a researcher has to decide how much of the paralinguistic information s/he needs in her/his transcripts. As Turell & Moyer (2008: 194) suggest, “transcription is already a first step in interpretation and analysis”. Thus, justify your choices before you start transcribing your data.

It is important to remember that spoken language is very different from written, especially in terms of syntax. Be aware of this fact when transcribing. Some of the utterances might not look like the ones you would expect to be “written”, e.g., people often do not finish what they start to say, or they repeat the same phrase several times in a row, or they reformulate the beginning of the sentence (Turell & Moyer 2008: 195).

Transcription is time-consuming work. To transcribe a one hour interview can take up to one day and even longer. According to Tagliamonte (2006: Kindle Locations 660–661), “a standard estimate for a first transcription following the procedures detailed below is that one hour of recorded data equals four hours of transcribing”. It is suggested to aim at the most complete and detailed transcription of recording. Transcription should be “detailed enough to retain enough information to conduct linguistic analyses in an efficient way, and simple enough to be easily readable and relatively easily transcribed” (Tagliamonte 2006: Kindle Locations 671–672).

Before transcription, most of the researchers compile a set of rules (a document) that form the basis of the transcription practice. Such a document is called “transcription protocol”. Tagliamonte (2006: Kindle Locations 676–677) defines the transcription protocol as “a reference document of transcription practice. It is a permanent

record that ensures consistent representation of words, phrases, features of natural discourse, and features particular to the data within and across all the transcriptions in a corpus”. Different aspects or transcription conventions can be touched upon in transcription protocols. However, in most of the cases transcription protocols address the following issues:

- a. orthographic and spelling conventions (capitalization, spelling, contractions, numbers, hyphenated words and compounds, abbreviations, acronyms and spoken letters, punctuation); if you intend to use your transcriptions for a study of phonology, then, most likely you would have to transcribe your data using a standardized phonetic transcription, for instance IPA [International Phonetic Alphabet] notation);
- b. dysfluent speech (filled pauses, hesitation sounds, partial words, restarts, mispronounced or non-standard words, etc.);
- c. additional, paralinguistic information (unclear speech, interjections, etc.).

All transcriptions should begin with a heading section that includes metalinguistic information about the participants, interaction and other important relevant information. For instance, Turell and Moyer (2008: 196) suggest the following information, provided in Table 13, to be included at the beginning of each transcription.

Table 13. Information to be included with transcription (adapted from Turell & Moyer 2008: 196)

Participants	Gender
	Age
	Educational level
	Occupation
Situation	
Languages involved	
Notes	
Dates of recording	
Researchers	
Date of transcription	
Transcribers	

Johnstone (2000: 117) recommends adhering to the following simple tips when transcribing:

- a. to use actual names (i.e., pseudonyms) for the speakers, not initials;
- b. to number the lines of the transcript so that you can refer the reader (especially if you quote your transcript in a paper) to the precise parts you discuss in your paper. This tip is valid for anyone who is involved in conversation analysis, interactional analysis or critical discourse analysis (see Chapters 11, 12, 13, and 14) or ethnographers in general;
- c. when you quote your transcripts, highlight the examples, relevant lines or words you emphasize in your discussion. This will help the reader to follow your reasoning.

Transcription can be done using any of the text processor, e.g., MS Word, Notepad, Pages, etc. Text files later can be fed into software programs (e.g., previously mentioned concordance program AntConc). However, software is also often employed for transcriptions, especially such, that allows the researcher to combine textual files with audio and video recordings. One such popular software, used in social sciences, including sociolinguistics, is ELAN. The software was developed by researchers working at Max Planck Institute for Psycholinguistics. The program is free to use, you can download it here: <http://tla.mpi.nl/tools/tla-tools/elan/>. Before you download it, make sure you have familiarized yourself with the user manual. It addresses key points concerning work with the program.

To illustrate possible transcription formats, let us consider a transcription protocol that was used by Kim Potowski and Lourdes Torres in their study on Spanish in Chicago, as well as one transcription of a 10 minute interview, conducted by Aurelija Tamošiūnaitė in 2009.

Example #3. Transcription protocol

There are two different formats for the voice files: wma and msv. To listen to the wma files, you need to have Windows Media Player installed. To listen to the msv files, you will need to download the program “Sony digital voice zip” freely available at http://www.4shared.com/file/iLari5Vr/Sony_Digital_voice_editor_3.htm.

Both of these players allow you to **rewind the interview by 4 or 5 seconds**. Usually there is a little arrow button that is preprogrammed to do this. It is a very useful feature while you are transcribing.

Some of the following **formatting** issues can be seen in the example that appears at the end of this memo. PLEASE FOLLOW THESE INSTRUCTIONS VERY CAREFULLY, THANK YOU!

Please use single spacing. Place an extra blank space inbetween speakers.

At the beginning of each speaker’s turn of speech, please identify the interviewer’s speech with their initials, and the interviewee’s answer with their initials.

Please use Times New Roman 12 pt. Margins should be **5 inches on top & bottom, 1 inch on left, 1.5 inches on right**. The **page header** should follow the format you see in the example below.

You **do not need to transcribe every single “um” and pause**. But we **do** need every instance of words like “so,” “you know,” “like”, etc.

If you cannot understand something **after listening to it three (3) times**, just put this:
[inaudible].

Do your best to capture the **correct intonation with punctuation and sentence breaks**. For example, the following two transcriptions are **very different** for the purposes of our project:

(a) “Eso lo diferencia de otra persona. So I mean, de mi parte, a mi hijo le voy a enseñar español.

(b) “Eso lo diferencia de otra persona, so... I mean, de mi parte, a mi hijo le voy a enseñar español.

In (a), the pause comes before the word “so.” In (b), the pause comes after the word “so.” This is an important difference in our project. Please just do your best to convey in writing **the way that the interview sounds to you**.

You must use a **Spanish spellchecker**. TOOLS → LANGUAGE → SET LANGUAGE → SPANISH. If a word is underlined in red, right click on it to see whether you have misspelled it.

Example #4. Extract from transcript

S., male, 25, student, ZIP: 60631.

A. So do you remember... or do you know or... does that happened to you... any crazy thing that happened to you in Chicago?

S. Crazy things, that happened to me in Chicago? (laughing) Yeah...

A. What?

S. Yes, a lot of crazy things that happened to me (laughing) that was in Chicago... um... I don't know... things that I would attribute specifically to Chicago... I guess kinda hard to combine, cause I think crazy things happen to me everywhere I know [

A. [so just tell me one...

S. [but one... it was a Thursday morning. And a setting of this story is important cause it was a Thursday morning and like 9 a. m. And I was at Jackson from the Red line waiting for the Blue line to come to UIC, like nine in a morning. And I look over, and there is a woman in a wedding dress, who looks like she had been like beaten up and... (laughing) I immediately think of myself, like well, this is gonna be interesting. So, I kinda like nudge my way and just to hear what's happening and... umm (laughing). She's in this wedding dress, and she is like crying because, I guess, she was going to take the train to her wedding, but end up like falling down outside and missing her train to her wedding. Yet, I don't understand a) why her wedding was on Thursday at 9 a. m. and why she was taking the train to the wedding, and then why she was just sitting there, its like trains continue to pass? She just... She just was like sitting, like sobbing, like crying for herself, it was really bizarre. And anyway, long story short, by the time that I walked up to her and the time that I actually got up on my train to go to school, mm... She had already given out her phone number to some random guy, like... in a subway. She is in

her wedding dress, when he came up, he was like 'hey, baby, how you doin', bla bla bla, just like talking, and... they exchanged numbers, like go date or go party some time. That was pretty [

A. [Oh my god...

S. [weird experience. That was like my third week here in Chicago and I was like, where am I?

9.4. Comprehension activities

Activity 2. Think of Lithuanians living in Lithuania and in the UK. Can we identify them as one group? Would these people identify themselves as one group? When? How? When not? Discuss this with your friends.

Activity 3. Ethnographers always face (ethical and research) issues related to their affiliation to the group or community under study. Discuss with your friends: is it better if the researcher is a member of the group under study? Should the researcher be insider or outsider to the group? Think of at least 3 reasons why is better to be an outsider and why is it better to be an insider.

1.
2.
3.

Activity 4. Watch the following Youtube video about transcription in socio-linguistic research: http://www.youtube.com/watch?v=lroSfpa_XFk. Find and download ELAN or Transcriber. Record 15 minutes of conversation with your family member or your friend. Transcribe it using the selected software.

[illegible]

Activity 6. Take a closer look at transcription protocol and an excerpt from transcription, provided above. What aspects of transcription are addressed in the protocol? How is paralinguistic information addressed? Take a closer look at the transcription. How is the speech transcribed? Name few instances how paralinguistic information is addressed in the transcription.

9.5. Glossary

Audio levels – input levels of the recording.

Background noise – any sound other than the sound being recorded.

Community of practice – a group of people that develop ways of doing things together, activities, common knowledge and beliefs, and ways of talking (Eckert & McConnell-Ginet 2003: 57).

Ethnography of communication – a method of discourse analysis applied in ethnographic (qualitative) research.

Participant observation – a type of data collection (observation) associated with the qualitative research paradigm.

Resolution – a measure of digital audio quality.

SPEAKING – is a model of sociolinguistic study that helps to identify and label components of linguistic interaction. The model was developed by Dell Hymes.

Speech community – a community of speakers who share the same linguistic norms and verbal behaviour (Trudgill 2003: 126).

Transcription – the process of representing oral language in writing.

Transcription protocol – a set of rules or reference document of transcription practice.

9.6. References

- Penelope Eckert, 1998. *Jocks and Burnouts: Social Categories and Identity in the High School*, New York, London: Teacher's College Press.
- Penelope Eckert & Sally McConnell-Ginet, 2003. *Language and Gender*, Cambridge, New York, Melbourne, Madrid, Cape Town: Cambridge University Press.
- Monica Heller, 2006. *Linguistic Minorities and Modernity: A Sociolinguistic Ethnography*, London: Continuum.
- Monica Heller, 2008. "Doing Ethnography", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 249–262.
- Barbara Johnstone, 2000. *Qualitative Methods in Sociolinguistics*, New York, Oxford: Oxford University Press.
- William Labov, 1984. "Field Methods of the Project on Linguistic Change and Variation", in *Language in Use*, ed. by J. Baugh & J. Sherzer, Englewood Cliffs: Prentice Hall, 28–53.
- Anna Margetts & Andrew Margetts, 2012. "Audio and Video Recording Techniques for Linguistic Research", in *The Oxford Handbook of Linguistic Fieldwork*, ed. Nicholas Thieberger, Oxford: Oxford University Press, [online version].
- Norma Mendoza-Denton, 2008. *Homegirls: Language and Cultural Practice among Latina Youth Gangs*, Malden, Oxford, Carlton: Blackwell Publishing.

- Miriam Meyerhoff, Chie Adachi, Golnaz Nanbakhsh, Anna Strylcharz, 2012. "Sociolinguistic Fieldwork", in *The Oxford Handbook of Linguistic Fieldwork*, ed. Nicholas Thieberger, Oxford: Oxford University Press [online version].
- Lesley Milroy & Matthew Gordon, 2003. *Sociolinguistics: Method and Interpretation*, Malden, Oxford, Melbourne, Berlin: Blackwell Publishing.
- Sali A. Tagliamonte, 2006. *Analysing Sociolinguistic Variation*, Cambridge, New York: Cambridge University Press [Kindle Edition].
- Peter Trudgill, 2003. *A Glossary of Sociolinguistics*. Oxford, New York: Oxford University Press.
- Maria Teresa Turell & Melissa G. Moyer, 2008. "Transcription", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 192–213.

CHAPTER 10: SOCIAL NETWORK ANALYSIS

10.1. Key concepts

- Close-knit community*
- Clusters*
- Density*
- First-order network ties*
- Multiplexity*
- Open community*
- Second order network ties*
- Snowball sampling*
- Social network*
- Strong ties*
- Weak ties*

10.2. Pre-reading activity

Activity 1. In groups discuss and define the concepts given above. How are these concepts related to social network analysis?

.....

.....

.....

.....

.....

10.3. Social networks

The **main aim** of this chapter is to outline key concepts and strategies related to the application of social network analysis to sociolinguistic research.

According to Daming et al. (2008: 265), the concept of social networks today is well integrated into sociolinguistics. Social network analysis serves both purposes: as an approach for *data collection*, as well as a theoretical framework for *data analysis* (Daming et al. 2008: 265).

10.3.1. Social network as an approach for data collection

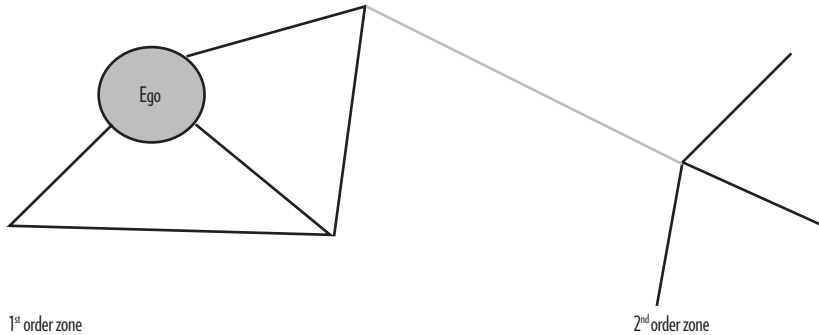
The social network approach for data collection has to be related to James' and Lesley's Milroys' Belfast (Northern Ireland) study, conducted in 1975–1981. This data collection approach is also called “friend-of-a-friend” or to a certain extent, the “snow-bowl approach”. In 1975–1976, Lesley Milroy entered five communities within the city of Belfast: Ballymacarrett, the Clonard, the Hammer (all three of these were low status inner city areas), Braniel (protestant), and Andersonstown (Catholic, both upper-working and lower-middle class neighborhoods) (Milroy & Gordon 2003: 73–74). The choice of the communities was determined by goals of the research: to investigate language change as people moved from one area to another (Milroy & Gordon 2003: 75). This is how Lesley Milroy describes her “entrance” and data collection strategies in the community:

By applying the concept of social network I was able to analyse, and to some extent control, the character of my relationship to the group I was observing. It was possible to equip myself with a status which was neither that of insider, nor that of outsider, but something of both – a friend of a friend, or more technically, a second order network contact. This status enabled me to carry out prolonged observations and record large volumes of varied interaction over a considerable period. (Milroy 1987: 43–44)

Due to political and economical situation of the time, there were certain constraints on who was able to carry out fieldwork in these specific areas of Belfast. According to Milroy (1987: 44), the fieldworker had to be a woman, because women were less likely to be attacked than men; the fieldworker had to enter community alone (in that case a woman was not viewed as a threat); the fieldworker had “to offer some guarantee of good faith of a personal kind” (1987: 44).

The social network approach emphasized the fieldworker's relationship to the community (Milroy 1987: 53). In a social network approach, each person can be viewed as having direct or indirect linkages to other people. “The persons who are linked directly to ego may be characterized as belonging to his *first order network zone*” (Milroy 1987: 46), while distantly, indirectly connected persons form the ego's *second order zone*. According to Milroy, these *first* and *second* order zones are the most important (Milroy 1987: 47). First network zone contacts consist of close friends and family members, while second order contacts consist of acquaintances, or “friends-of-friends”. These “friends-of-friends”, according to Milroy (1987: 53), “perform an important social function by extending the range of goods and services which members of the first order zone are able to provide”. Thus, if you enter a community and you are identified as a friend of a friend, this will enable you to extend your potential network ties in that particular community. Such was Milroy's approach: she entered the community as a friend of a friend.

Picture 5: 1st and 2nd order network zones



When Milroy entered the community, she did not contact community leaders, priests and teachers in order to avoid individuals of an authoritative status (Milroy & Gordon 2003: 75). The reasoning behind this choice is related to the standardized varieties that these people usually are associated with. Milroy explains, that:

The initial 'link' contact knew the purpose of the research, but did not visit the area with me or introduce me personally; he simply provided lists of acquaintances' names and addresses. I subsequently entered each area alone, always made an initial approach to a specific individual whose name had been passed on and introduced myself as 'a friend of X', he thought you might be able to help me'. (Milroy 1987: 53–54)

Later on, when the network of acquaintances began to grow, the researcher would introduce herself as a friend of someone whom she met in the community. According to Milroy, it is very important to name a friend, because it guarantees good faith and a researcher earns the participants' trust (Milroy 1987: 54). Once people agreed to participate, they did not refuse permission. Therefore, Milroy was able to obtain many recordings of different nature and representing different varieties of speech. Researchers used *snowball sampling* in order to obtain recordings of 16 people per community.

The social network approach to data collection is particularly suitable for close-knit communities (e.g., urban or rural communities) (Milroy and Gordon 2003: 78). This approach minimizes the effect of the 'observer's paradox' and enables the researcher to obtain different varieties of speech.

10.3.2. The social network as a theoretical framework for data analysis

Social network analysis stems from the methods developed by social anthropologists in the 1960s and 1970s. The method was adopted by variationists in around the 1970s and 1980s. As Milroy & Gordon (2003: 117) claims, "no canonically correct procedure for analyzing social networks can be identified; scholars from many different disciplines employ the concept for a range of theoretical and practical purposes."

Thus, social network analysis can be defined very generally as “the study of the social structure that individuals and entities construct through interaction” (Velázquez 2013: 190).

Social network analysis relies on the postulate that “people interact meaningfully as individuals, in addition for forming parts of structured, functional institutions such as classes, castes or occupational groups” (Milroy 1987: 45–46). In other words, people create groups, personal communities in order to solve problems of their everyday life (Milroy & Gordon 2003: 117).

Social network analysis deals with “the structural and content properties of ties that constitute egocentric (i.e., that of the ego or individual) personal networks, and seeks to identify ties important to an individual” (Milroy & Gordon 2003: 119). There are several important concepts related to social network theory in general: *weak ties*, *strong ties*, *exchanges*, *density*, *clusters*, and *multiplexity*. Granovetter (quoted from Milroy & Milroy 1997: 201) defines *weak* and *strong ties* as follows: “The strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual condifing) and the reciprocal services which characterise a tie.” Weak ties between people often serve “as bridges through which information and influence are diffused” (Milroy & Milroy 1997: 201). Weak ties are those that connect us with our acquaintances, while strong ties are those that connect us with our friends. Just think of information flows: most of our acquaintances usually provide us with information that we might not *have known*, because we do not share that many interpersonal connections with them (they know more people we do not know), while information that our friends pass to us usually is something that we might already have heard from our common friends.

Wei (1994, following Milardo 1988, quoted from Daming et al. 2008) differentiates between three types of contacts:

1. exchange networks – include those people with whom the “individual could turn to in time of need” (help, advice, support); usually this network consists of kin and close friends;
2. interactive networks – include those people with whom the individual interacts frequently over longer periods of time, but an individual would not contact these people in time of need; this network includes, e.g., co-workers, hairdresser, automechanic, etc.;
3. passive networks are characterized by an absence of regular contact, the individual relies on these relationships for emotional and moral support; this network might include, e.g., immigrants’ relatives at home.

According to Milroy (1987: 47), “a social network acts as a mechanism both for exchanging goods and services, and for imposing obligations and conferring corresponding rights upon its members”. When people interact they send messages to each other, and these messages can be seen as transactions (Milroy 1987: 47). In other words, people exchange greetings, jokes, and information amongst each other in a similar way as they exchange goods, services, and assistance. Thus, every speech act

can be treated as “tokens of exchange” (Milroy 1987: 48). This perception sometimes helps to explain “unsuccessful” or incooperative interaction.

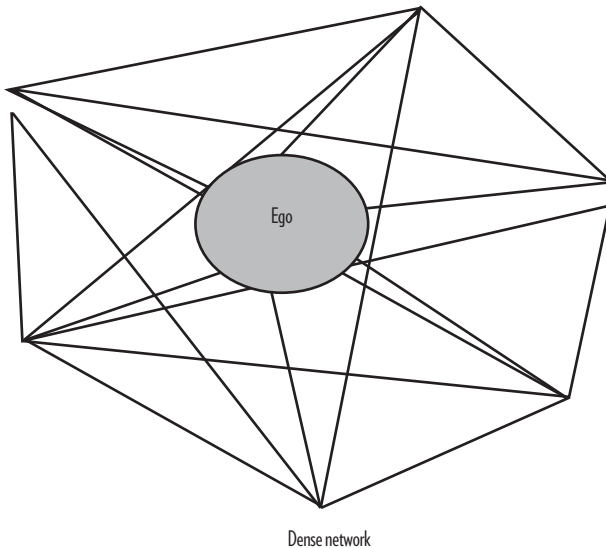
There are several structural and content characteristics of networks that are employed in social network analysis. According to Milroy (1987: 50), the most important characteristic is *density*. Density can be defined as the proportion of direct ties in a network relative to the total number possible. In other words, a network is perceived as dense if a “large number of the persons to whom ego is linked are also linked to each other” (Milroy 1987: 50). Density is calculated using the following formula:

$$D = 100Na/N \%$$

Na – refers to the total **actual** number of links

N – refers to total number of **possible** links

Picture 6: Dense network



Social networks usually consist of many ties. Depending on a density level, some of the relationships in a social network can form a cluster. Clusters “are segments or compartments of networks which have relatively high density: relationships within the cluster are denser than those existing externally and may also be considered as being relationships of like *content*” (Milroy 1987: 50). Examples of clusters in our social networks can be family subnetwork, occupation subnetwork, or specific group membership.

As Milroy points out, some of our direct connections (e.g., family member vs. a janitor in our working place) may have a greater impact on our behaviour (linguistic and non-linguistic) than others. Therefore, according to her, it is important to think of relationships in terms of *content*. e.g., is my sister just a sister, or is she also

my friend, my co-worker and my neighbor? If a relationship to someone has only a single content, this relationship is called *uniplex*. If the connection has more than one content (family member, friend, co-worker, neighbor), this relationship is called *multiplex* (Milroy 1987: 51). The multiplexity score is calculated using the following formula:

$$M = Nm \times 100 / N \%$$

Nm – refers to the number of multiplex links

N – refers to the number of actual links

High multiplexity and density are important in terms of the “effectiveness of the network as a norm-enforcement mechanism” (Milroy 1987: 52). Usually tribal, rural, traditional working class communities are multiplex and dense, because people are linked to each other not only as kin, but also as neighbors, co-workers, and friends, while industrial, urban communities “tend to uniplexity and sparseness” (Milroy 1987: 52).

Density and multiplexity of individual’s social networks are employed in sociolinguistic analysis to explain language change patterns, language ideologies, language choice or language maintenance (e.g., in research concerning minorities). Milroy & Milroy (1997: 202) argue that “weaker intergroup ties are likely to be critical in transmitting innovations from one group to another”. In other words, if you will be investigating language change and spread of a certain linguistic innovation, it is more likely that the innovation will be transmitted from one group (one social network or one cluster) to another by weak ties (Milroy & Milroy 1997: 204). Thus, on the one hand, social network analysis is employed by variationists in order to identify patterns of linguistic innovation. On the other hand, social network analysis is also employed in quantitative research in order to see how certain aspects of language use, language ideologies or language choices depend on a network’s density.

Example #1.

Isabel Velázquez, “Mothers’ social network and family language maintenance”, *Journal of Multilingual and Multicultural Development* 34 (2), 189–202, 2013.

Velázquez used social network analysis to examine a mothers’ social network “as related to Spanish language maintenance in a group of Mexican American families in three communities in the USA” (189). The researcher conducted **two** interviews with **15** Mexican American mothers. The interviews were conducted in respondent’s homes by a “female bilingual who was a native speaker of Spanish” (192). The research was conducted in three Mexican American communities: the city of El Paso (Texas), the neighborhood of La Villita in Chicago (IL), and the city of Lincoln (Nebraska). **Five** families from each of the communities participated in the study.

Velázquez aimed to measure **five** features related to mothers’ networks: *primary language of exchange, density, strength of ties, integration to local, regional and transnational networks, and degree of gender segregation*. (192)

Procedure:

1. respondents were asked to identify:

- five closest friends;
 - five closest relatives not living in their household;
 - five individuals whom they saw more than twice in a normal week;
2. to measure *network density* Velázquez divided the number of concrete exchanges by the number of possible exchanges (193). If every actor knew every actor and exchanged or received support from every other actor, a researcher gave a density score of 1. In order to find out information about exchanges, the respondents were asked whether they borrowed money, provided or received help from other people, gave or asked for advice, called when sad, etc.
 3. to *measure the strength of ties* Velázquez, for each potential exchange, presented respondents with the following options: *I have done this with/for this person, I have not, but would if necessary*, and *I would never do this with/for this person*. (193) This helped to measure respondents' attitudes toward ties. For active exchange ties a researcher gave a score of 2, for virtual or passive ties she gave a score of 1, and for not possible exchanges she gave a score of 0.
 4. for *primary language of exchange* the respondents had to identify the language – Spanish, English, Both or Other.
 5. *integration in local, regional and transnational networks* was determined by asking mothers where exactly the actors lived.
 6. *degree of gender segregation* was measured by determining the percentage of female and male actors in a mother's closest network beyond her household (193).

Results:

1. For 14 out of 15 participants, Spanish was the language of exchange in their closest circle of interaction. Spanish was the most commonly used language in those networks that had most access to community resources (197). Thus, Spanish was “the language” in order to access symbolic and material sources.
2. Only **two** of the 15 respondents belonged to dense personal networks. For most of the respondents the ties to their closest network were virtual or symbolic, i.e., they could use those ties to exchange resources, but did not have to actually do it. (197)
3. The highest percentages of concrete ties were found only in **three** networks, namely of those respondents who had the longest history of residence in the border region. The lowest percentages of concrete ties were found among those who have lived in the locale for the shortest time. (197)
4. All networks involved face-to-face interaction, 78 % of network actors lived in the same community as the respondent, 6 % in the same region and 6 % in another US state, 10 % of network actors lived in Mexico (relatives of the most recent immigrants to the US).
5. 74 % of all network actors in the three communities were female, male actors – 26 %, and only four were not relatives. Thus, individuals on whom mothers “relied upon most outside their household were women” (198).

Results suggest that Spanish transmission is influenced by the mother's perception of benefit, her participation in networks where Spanish is vested with social capital, and her linguistic competence.

10.4. Comprehension activities

Activity 2. Try to draw your social network. Take a piece of paper, draw a circle in the middle (ego or you). Put your closest friends and family members around the circle. Now – put their friends and relatives around their circles. See how deep you can go into your social network.

Activity 3. If you have your “Facebook” account, open it and look under the section “Friends”. How many “Friends” do you have? How many of these “friends” belong to the first order network zone? How many to the second? What do these numbers tell you about your virtual social network? How is it different from the real one? Why?

Activity 4. Calculate the density of a hypothetical social network. Use the formula. Here are the numbers: $N_a = 90$, $N = 150$.

Activity 4. Calculate the density of your social network. Try to think, how many actual links (relationships) do you have and how many possible (potential relationships, think of friends of friends) do you have. Use the formula. What is the number? Compare it to your friends. Whose network is denser?

Activity 5. Calculate the multiplexity score of a hypothetical network. Use the formula. Here are the numbers: $N_m = 30$, $N = 70$.

Activity 6. Calculate the multiplexity score of your social network. Try to think, what is the number of multiplex links that exist in your social network. What is the number of actual links? Use the formula. What is the number? Compare it to your friends? Who has the most multiplex network?

Activity 7. Collect a certain amount of letters (first, ask your family members, whether they have any, then, ask your friends). When you collect the letters, write down their authors and addressees. Count them. Try to establish the social network (who is connected with whom, what kind of ties do they have (uniplex, multiplex)). Once you established the social network of your letters, take a look at the language. Do you see any occurring patterns? If yes, what are they?

10.5. Glossary

Close-knit community – community that is closely united by social and cultural ties.

Clusters – segments of networks which have a relatively high density.

Density – the proportion of direct ties in a network relative to the total number possible.

First-order network ties – a person's direct contacts.

Multiplexity – number of interpersonal ties, relating one person with another.

Open community – community that is not that closely united by social and cultural ties.

Second order network ties – a person's indirect contacts.

Snowball sampling – a type of sampling when existing study respondents recruit future respondents from among their acquaintances.

Social network – a web-like pattern of relationships among individuals (Daming et al. 2008: 263).

Strong ties – network ties that connect friends or kin.

Weak ties – network ties that connect acquaintances.

10.6. References

- Xu Daming, Wang Xiaomei & Li Wei, 2008. "Social Network Analysis", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 263–274.
- James Milroy & Lesley Milroy, 1997. "Network Structure and Linguistic Change", in *Sociolinguistics: a Reader and Coursebook*, ed. by Nikolas Coupland and Adam Jaworski, New York: Palgrave, 199–211.
- Lesley Milroy & Matthew Gordon, 2003. *Sociolinguistics: Method and Interpretation*, Malden, Oxford, Melbourne, Berlin: Blackwell Publishing.
- Lesley Milroy, 1987. *Language and Social Networks*, Oxford, New York: Blackwell.
- Isabel Velázquez, 2013. "Mother's social network and family language maintenance", *Journal of Multilingual and Multicultural Development* 34 (2), 189–202.

CHAPTER 11: CONVERSATION ANALYSIS

11.1. Key concepts

Adjacency pair

Context

First pair part/second pair part

Move

Preferred/dispreferred pairs

Preference organisation

Repair

Sequential analysis

Social order

Talk-in-interaction

Turn/turn taking

11.2. Pre-reading activity

Activity 1. In groups discuss and describe the concepts given above. What do you think could be the role of each in applying conversation analysis?

.....
.....
.....
.....

11.3. Conversation analysis: theory and praxis

11.3.1. Origin of Conversation analysis

As a method of sociolinguistic research, *Conversation analysis*, commonly referred to as CA, owes its derivation to the perspective of *Ethnomethodology* introduced and developed by a famous American sociologist Harold Garfinkel (1967, 1974) and thus relates its origin to the field of sociology. In terms of Garfinkel's ethnomethodological approach, knowledge about society and social order is not given to us as an objective reality but is rather created and displayed by people engaged in everyday activities – linguistic and other:

Ethnomethodological studies analyze everyday activities as members' methods for making those same activities visibly-rational-and-reportable-for-all-practical-purposes, i.e., "accountable," as organisations of commonplace everyday activities. The reflexivity of

that phenomenon is a singular feature of practical actions, of practical circumstances, of common sense knowledge of social structures, and of practical sociological reasoning. (Garfinkel 1967: vii)

From the ethnomethodological perspective, knowledge and understanding of social order is achieved by and could be revealed through the analysis of social action. Focusing mainly on spoken interaction, Harvey Sacks (1984, 1987, 1992), Emanuel Schegloff and Gail Jefferson (1978) aimed at discovering and investigating the methods and principles of how members of a society create social order and coherence through a particular type of social action, namely various types of conversation in different types of interactional environments. From the CA perspective, the use of language is seen as reflecting and producing *social order* as well as the underlying systematic linguistic patterns of societal organization. Discovering and examining these patterns is among the primary academic goals of conversation analysts.

While discussing the origin and the ensuing methodology of CA, Paul Seedhouse (2005) claims that initial relation of CA to sociology and ethnomethodology shapes out its approach to the study of language distinct from general linguistics, which focuses on language as an abstract system:

CA's origins in sociology and specifically ethnomethodology entail a different perspective on the status and interest of language itself from that typical of linguistics. CA's primary interest is in the social act and only marginally in language, whereas a linguist's primary interest is normally in language. In descriptivist linguistics, the interest is in examining how aspects of language are organized in relation to each other. CA, by contrast, studies how social acts are organized in interaction. (Seedhouse 2005: 251)

Despite different perspectives towards language study offered by CA and traditional linguistics, important common touches and synergy can be observed. In terms of Seedhouse, "CA is interested in how social acts are packed and delivered in linguistic terms" (2005: 251). In other words, linguistic structures give an embodiment for a social action carried out in a particular sequencing which becomes the primary focus of conversation analysts.

Another important distinction brought to the study of language as social interaction comes from the general linguistics of the 1970s, namely the distinction between *etic* and *emic*, where the former derives from and signifies *phonetics* and the latter is attributable to *phonemics* in the field of phonology (Pike 1967). In view of a broader scope of application, the *etic* is used to refer to the analysis of a phenomenon outside a particular system, while *emic* symbolises studying a phenomenon from the inside. In regard to research in CA, as claimed by Seedhouse, academic endeavours are aimed at analysing social interaction from an *emic* perspective:

There is no sense that either perspective is inherently superior to the other, and CA does not claim that social actions and *emic* perspectives are inherently more important than language

or *etic* perspectives per se; it is simply the case that CA's unrelenting aim is to portray social action in interaction from an *emic* perspective. What CA means by an *emic* perspective, however, is not merely the participants' perspective, but the perspective from within the sequential environment in which the social actions were performed. (Seedhouse 2005: 252)

Consequently, as analysis of speech in interaction, CA presents the methodology for the study of how participants in interaction try to find solutions to different communicative problems that may arise while completing their communicative tasks such as "opening and closing speech, turn-taking, repair, topic management, information receipt and showing agreement and disagreement" and others (Schiffrin 1994: 239).

11.3.2. Conceptualisation of context

Although having social order as their primary focus, conversation analysts largely disregard the characteristics of *social context* at large. Namely, CA does not embrace the understanding of context in terms of such static social features as social identities of participants, setting and personal attributes, i.e., their social status, age, gender, power, race or ethnicity (cf. Schiffrin 1994, Seedhouse 2005). As generalised by Seedhouse:

CA allegedly refuses to take context into account as it declines to invoke 'obviously relevant' contextual features such as participant's social status, gender, race etc. Since there are an indefinite number of 'external' aspects of cultural, social or personal identity or context that could be potentially relevant to any given instance of talk-in-interaction, an *emic* analysis must show which of these innumerable, potentially relevant characteristics are actually procedurally relevant to those participants at that moment; this can only be accomplished by analysing the details of the interaction. (Seedhouse 2005: 255)

Thus in CA the notion of *context* is relevant as much as it is "grounded in text", or in a conversation (Schiffrin 1994: 236). It is only by looking thoroughly into the text of the conversation that a researcher can determine what contextual aspects are relevant to the interpretation of that particular conversation and could lead to important insights of the analysis. For instance, if a researcher wants to study the role of gender, he or she may code men and women in a particular piece of a conversation that is being analysed and in that way make the variable of gender as "an omnirelevant matter for analysis"; in other words, necessarily having influence on the interpretation. However, in that particular conversation gender *may* or *may not* make relevance, which can only be revealed after a detailed analysis of that particular conversation. Thus any *a priori* generalisations may finally appear analytically misleading (cf. Pomerantz and Fehr 1997).

Given the above conceptualisation of context, in their endeavours, conversation analysts turn to the "naturally occurring conversations" (Sacks, Schegloff and Jefferson 1978: 8), or "the transcriptions of actual occurrences in their actual sequence" (Sacks 1987: 54), as their main source of data. With such choice of data the analysts attempt to avoid any premature generalizations about the use of language. As Sacks

summarizes: “we sit down with a piece of data, make a bunch of observations and see where they will go” (1984: 27).

11.3.3. Sequential analysis and the ensuing categories

In CA language is studied in its interactional environment and not as single utterances removed from its original surrounding discourse (Drew 2005). For conversational analysts, a *turn* is the core unit of a conversation and thereby it is also the main unit of analysis. On a par with taking turns in such social activities as “allocating political office” or “regulating traffic at intersections”, *turn-taking* of a linguistic interaction is seen as a highly significant type of social organization (Sacks, Schegloff and Jefferson 1978: 7). To put it in Hutchby and Drew’s words, it is a task of conversation analysts to examine “how participants understand and respond to one another in their turns at talk”, including interruptions, overlapping and self-repairs, and “how sequences of activities are generated” (1995: 182–3).

A two-unit sequence, called an *adjacency pair*, has been normally seen as the basic sequence, or the central relationship, of a conversation (Sacks 1987, 1992; Sacks, Schegloff and Jefferson 1978). The two units in an adjacency pair are the *first pair part* and the *second pair part*, which are produced by different speakers. The underlying orderliness between the first and the second pair parts is created through their being “type-connected” (Sacks 1987: 55). That is to say, if the speaker produces a particular type of the first pair part, as, for instance, a greeting, an offer or a question, his or her responder has to choose the type of the second pair part that fits the type of the first. It would be a greeting for a greeting, an acceptance or a rejection for an offer as well as different possible answers for a question.

11.3.4. Preference organisation and its strategic potential

As in the case of an offer or a question, there usually exist several alternatives of the second pair part for most types of adjacency pairs. Interested in how a choice of a particular alternative is made, conversation analysts arrived at the notion of *preference organization* (e.g., Atkinson and Heritage 1984; Sacks 1987) or ‘markedness’, as it has been called by Mey (2001: 152). It has been observed that the alternatives for different second pair part slots are not equivalent and can be differentiated into *preferred* and *dispreferred* alternatives. For instance, an acceptance would be a preferred alternative for an offer or an invitation, whereas a refusal would be a dispreferred one (Mey 2001: 152, cf. Levinson 1983: 336):

FIRST PAIR PART	SECOND PAIR PART	
	Preferred	Dispreferred
Request	Acceptance	refusal
Offer or Invitation	Acceptance	refusal
Assessment	Agreement	disagreement
Question	Expected answer	unexpected answer
Blame	Denial	admission

Importantly, as Atkinson and Heritage (1984) maintain, the preferred and dispreferred alternatives differ in the way they are performed. While preferred alternatives are “normally performed directly and with little delay”, the dispreferred ones are carried out “with delay between turns, are commonly delayed within turns, and are variously softened” through the application of hedges or increased indirectness (Atkinson and Heritage 1984: 53). The following example of an adjacency pair is provided by Sacks:

A: Yuh comin down early?

B: Well, I got a lot of things to do before getting cleared up tomorrow. I don't know. I w- probably won't be too early (Sacks 1987: 58).

In his second pair part turn, B gives a disagreeing answer, which can be seen as a dispreferred alternative. Linguistically, his disagreement is softened and made rather indirect with numerous hedges, while, structurally, it is delayed till the very end of the turn.

The use of a dispreferred turn may often result in the need to repair the ‘damaged’ conversation where the damage may take a variety of forms like misunderstanding, broken sequence or other problems (Mey 2001: 148). *Repairs* can be used strategically, for instance, one may opt to correct oneself to gain time to think or to prevent others to enter the conversation. There are two types of repairs: a *self-initiated repair*, namely the speaker's own intention for self-correction, and an *other-initiated repair*, offered by the hearer. Consider the following example between a macho male speaker and a feminist female speaker:

Man: So I was trying to pick up this chick when...

Woman: Excuse ME, did I hear that right?

Man: Awfully sorry, I mean, *woman*...

Woman: PICK UP?

Man: Awfully sorry, I mean, *meet*...

Woman: So you're trying to imply that there actually are women around who would go out with a MALE CHAUVINIST PIG LIKE YOU? (taken from Mey 2001: 149).

The two repairs in the above conversation are *other-initiated*, namely, asked by the female interlocutor. The *other-initiated repairs* can appear to be highly face threatening and the less preferred than the *self-initiated* ones.

11.3.5. Conversational settings

From the discussion above, it can be clearly seen that CA adheres to a structural view of interaction. One of the frequent distinctions made in CA research is the discrimination between mundane and institutional discourse for analysis. Thus CA, on the one hand, aims at a wide range of interactional settings, starting from the analysis of mundane encounters, i.e. informal conversation in casual settings as, for instance, friendly chats or telephone conversations between student colleagues at university

or family members. On the other hand, it spreads into different types of institutional discourse like telephone calls to suicide prevention centres (Sacks 1984, 1987, 1992) TV and radio interviews (Clayman and Heritage, 2002; Hutchby and Drew, 1995), a visit at the doctor's, giving testimony in court, holding meetings at the workplace or consulting students during office hours (Drew and Sorjonen 1997). Regardless of the type of interaction there are not only *verbal* properties of a conversation that a researcher aims to examine, but also the *paralinguistic* ones like pauses, gaps or re-starts (Pomerantz and Fehr 1997: 65, emphasis added).

As has been previously mentioned, interactional structure is necessarily based on a particular preference system so that a question turn would be expected to preferably be followed by an answer turn as well as a greeting would be expected to be answered with a greeting in return. Nevertheless, as indicated by Drew and Sorjonen, even a largely pre-determined turn-taking system is "locally managed" (1997: 102), thus it is the main task of a conversational analyst to identify and examine how this interactional management is achieved. In other words, how the interlocutors orient themselves, on the one hand, towards the discourse pertaining to their institutional identities and roles and, on the other hand, towards the implementation of their specific "task-related roles" (Drew and Sorjonen 1997: 103) in a particular situation or orient themselves to other interactional goals outside the particular constraints of institutional discourse. The conversation in Example 1 can be considered as an illustration of institutional discourse with two colleagues Jim and Kate engaged in a telephone interaction to complete a work-related task.

Example #1.

1. Kate: Hey Jim
2. Jim: How are you Kate Fisher?
3. Kate: How are you doin'
4. Jim: Well I'm doin' all right [thank you very [much
5. Kate: [We- [Well goo:d
6. Jim: And a lo:vely day it is.
7. Kate: Oh:,isn't it gor[geous=
8. Jim: [Yes
9. Kate: =I snuck out at lunch
10. its[really [difficult to come [back
11. Jim: [.hhh [You(h)oo [.hhh that was not-
12. good
13. Kate: See it (was[ese-)
14. Jim: [You're s'pose to stay in your office
15. and work work work [h e h ha:h
16. Kate: [Well-
17. Kate: Jean and I went- she- she works in our office too
18. we went together too: uh-.hhhh u:h do some
19. shopping

20. Jim: [Um hum
 21. Kate: [A:nd we each made each other come ba:ck,
 22. Jim: Atta girl, ye:s I know what you mean
 23. Kate: So maybe that's the ke(h)y of going [like that
 24. Jim: [Huh huh huh
 25. Jim: That's it
 26. Jim: pt.hhhhh [What's up
 27. Kate: [Well-
 28. Kate: Well, I've had a call from Paul toda:y and after
 29. he called, I checked with your- terminal over
 30. there and they said our order's not awarded...
- (Drew and Sorjonen 1997: 93)

Despite the institutional attribution of this particular interaction, the turn with an explicit reference to the institutional roles appears, only at the end of the given interaction in line 28 where Kate informs her colleague Jim about the situation regarding the order, which has not been awarded. The interaction preceding this turn can be regarded as “sociable” (Drew and Sorjonen 1997: 93) rather than institutional entailing an exchange about weather, lunch and shopping experiences. The demarcation for the transition is explicitly fulfilled by the discourse marker *well* in line 28. Regardless of the outwardly explicit changeover from informal topics to the communication of work-related matters, the more in-depth analysis can reveal that the line between institutional and sociable is much more blurred than may appear at first sight. Thus it is the goal of a conversational analyst to go deeper into the contextual, structural and sequential aspects of a particular piece of discourse. With regard to context, as put by Drew and Sorjonen, in Example 1 the institutional identities of conversationalists permeate through “the ways in which the topics of the weather and shopping are set in the context of office routines and employees’ duties” (Drew and Sorjonen 197: 94). The leak of institutional identities is observable through linguistic and sequential features, for instance: *the use of repetition rather than ellipsis* in line 4 and *fronting of the subject complement “a lovely day”* in line 6, which shows a particular degree of formality between the interlocutors pertaining to institutional discourse.

11.3.6. Hints for transcribing

There are many systems for the transcription of recorded data and many conversational analysts add their own corrections to them. As the paradigm of CA itself, the transcription tradition of recorded conversations originates with the names of Sacks, Jefferson and Schegloff (1978). On the basis set by these scholars, other researchers adjust and build their transcription patterns by adding or removing some symbols. Table 1 presents a pattern of transcription adapted from Van Dijk 1997:

Brackets	I don' [know] you don't	-- shows that the indicated parts of utterance were produced simultaneously with the bracket on the left marking the beginning of the simultaneous talk and the bracket on the right signifying the end.
Colons	We::ll now	-- show that the preceding syllable is prolonged and the number of colons shows the length of the prolongation, the larger the number of colons, the longer is the preceding syllable.
Hyphen	But-	-- represents a cutting off short of the preceding syllable.
Capital letters	CAPS	-- shows increased loudness of the marked parts of the utterance.
Italics or underscoring	italics	-- indicates emphasis on the marked parts of the utterance.
Equal signs	'Swat I said= -But you didn't	-- indicates that no time elapsed between the snatches of utterance; the next speaker starts at precisely the end of a current speaker's utterance.
Period encased in parenthesis	(.)	-- denotes a pause of 0.1 of a second.
Numbers encased in parentheses	(1.3)	-- denotes seconds of a pause.
Score sign	(#)	-- denotes a pause of about a second.
Single parentheses	(very slowly)	-- indicate that something was heard but was not clear, could serve as a warning that the transcript may be unreliable.
Double parentheses	((very slowly))	-- not transcribed utterances.
Laughter	Hunh-heh eh-heh hengh	-- indicate laughter particles.
Punctuation marks	.	-- 'falling' intonation.
	?	-- 'rising' intonation.
	.	-- 'falling-rising' intonation.
Degree symbol	°	-- represents softness.
Breathing indicators	.hh	-- breathing.
	hh	-- exhalation.

11.4. Comprehension activities

Activity 2. Indicate whether the given statements are true (T) or false (F) with regard to the application of *Conversation analysis* as a research method. Explain your decisions to your colleagues on the basis of the given theory.

Context is conceptualised as a predetermined social reality for the interlocutors engaged in a conversation.

The choice of naturally occurring conversational sequences as research data allows conversation analysts to avoid and escape advance judgments and premature generalisations in their analysis.

Conversational preference organisation is an inseparable and essential component of sequential order of talk-in-interaction.

The utterance of a preferred alternative of a second pair part is preceded by a longer pause than the delivery of a dispreferred alternative.

The concept of indirectness does not play a significant role in the application of conversation analysis.

Activity 3. Mark the concepts below in the two exchanges and analyse them from a CA perspective.

<i>First pair part</i>	<i>preferred pair part</i>	<i>dispreferred pair part</i>	<i>hedge</i>
------------------------	----------------------------	-------------------------------	--------------

(a)

SALES CLERK: You're over 21, aren't you?

CUSTOMER: Sure.

SALES CLERK: OK, here's your beer.

(b)

SALES CLERK: You're over 21, aren't you?

CUSTOMER: Well, er, yes, my birthday was actually yesterday, and we're having a party tonight...

SALES CLERK: All right, may I see your ID? (taken from Mey 2001: 150)

Activity 4. Make a record of a conversation or find a video record on the Internet. Then transcribe it using the symbols above. Share your record and transcription with a colleague for peer review.

11.5. Glossary

Adjacency pair – a two-unit type-connected sequence in a conversation produced by different speakers and ordered as a first part and a second part sequenced in such a way that a particular first pair part requires a particular second part or parts, e.g., a question and answer.

Context – (in CA) contextual properties grounded in and stemming from the texts rather than pre-determined social attributes.

Dispreferred pair – the second pair part in an adjacency pair consisting of a response to the first pair part that is generally to be avoided by the speaker.

First pair part – the first unit in an adjacency pair which sets a particular requirement for the adjacent turn.

Preferred pair – the second pair part in an adjacency pair consisting of a response to the first pair part that is more or less expected by the hearer.

Repair – the speaker's acknowledgement of a speech error and adequate corrections to what has been said.

Second pair part – the second unit in an adjacency pair which is tied to the first pair part by the principle of relevance.

Sequential analysis – analysis of how the speakers conduct various social actions through their turns in talk and how all aspects of linguistic production such as lexical, syntactic, *etc.* units are organised in terms of, and fitted to, a turn's position in the sequence of a conversation (cf. Drew 2005).

Social order – (in CA) social order, including politics, issues of race, gender, culture or class, is seen as enacted at the level of conversation and interaction, or it would cease to exist.

Turn – the core unit of a conversation and thereby the main unit of analysis.

Turn-taking – systematic exchange of turns between interlocutors.

11.6. References

- Steven Clayman & John Heritage, 2002. *The News Interview: Journalists and Public Figures on the Air*, Cambridge: Cambridge UP.
- Paul Drew & Marja Leena Sorjonen, 1997. "Institutional Dialogue", in *Discourse as Social Interaction*. Vol. 2, ed. by T. A. Van Dijk, London: Sage, 92–118.
- Paul Drew, 2005. "Foreword: applied linguistics and conversation analysis", in *Applying Conversation Analysis*, ed. by Keith Richards and Peter Seedhouse, Houndmills: Palgrave Macmillan, xiv–xx.
- Harold Garfinkel, 1967. *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.
- Harold Garfinkel, 1974. "On the origins of the term 'ethnomethodology'", in *Ethnomethodology*, ed. by Roy Turner, Harmondsworth: Penguin.
- Ian Hutchby & Paul Drew, 1995. "Conversation Analysis", in *Handbook of Pragmatics: Manual*, ed. by Jef Verschueren et al., 182–189. Amsterdam: John Benjamins.
- Stephen C. Levinson, 1983. *Pragmatics*. Cambridge: CUP.
- Kenneth L. Pike, 1967. *Language in Relation to a Unified Theory of the Structure of Human Behaviour*, The Hague: Mouton.
- Anita Pomerantz & Barbara J. Fehr, 1997. "Conversation Analysis: An Approach to the Study of Social Action as Sense Making Practices", in *Discourse as Social Interaction*. Vol. 2, ed. by T. A. Van Dijk, London: Sage Publications, 64–91.
- Harvey Sacks, 1984. "Notes on methodology", in *Structures of Social Action: Studies in Conversation Analysis*, ed. by J. Maxwell Atkinson & John Heritage, Cambridge: Cambridge UP, 21–27.
- Harvey Sacks, 1987. "On the preference for agreement and contiguity in sequences in conversation", in *Talk and Social Organization*, ed. by Graham Button and John R. E. Lee, Clevedon: Multilingual Matters, 54–69.
- Harvey Sacks, 1992. *Lectures on Conversation*, Vol. 2, ed. by Gail Jefferson, with an introduction by Emmanuel A. Schegloff, Oxford: Blackwell.
- Harvey Sacks, Emmanuel A. Schegloff & Gail Jefferson, 1978. "A Simplest Systematics for the Organization of Turn Taking for Conversation", in *Studies in the Organization of Conversational Interaction*, ed. by Jim Schenkein, New York: Academic Press, 7–55.
- Deborah Schiffrin, 1994. *Approaches to Discourse*, Oxford: Blackwell.
- Peter Seedhouse, 2005. "Conversation Analysis as Research Methodology", in *Applying Conversation Analysis*, ed. by Keith Richards & Peter Seedhouse, Houndmills: Palgrave Macmillan, 251–287.

CHAPTER 12: SPEECH ACT THEORY

12.1. Key concepts

- Constative*
- Direct speech act*
- Felicity conditions*
- Illocutionary act*
- Indirect speech act*
- Locutionary act*
- Performative*
- Perlocutionary act*
- Speech act*

12.2. Pre-reading activity

Activity 1. In groups discuss and describe the concepts given above. What do you think could be the role of each in *Speech act theory*?

.....

.....

.....

.....

.....

12.3. Speech act theory

12.3.1. Origin of *Speech act theory* and the conceptualization of a *speech act*

The fifty-year old *Speech act theory* initially evolved out of Austin’s crucial philosophical observation that not all speaking, or language use, is directed towards describing some states of affairs or reporting some facts. Furthermore, not all speaking consists of statements to be judged as true or false. Some speaking corresponds to doing, as, for instance, naming a ship or swearing an oath. Thus Austin made a distinction between a *constative*, i.e. a truth conditional utterance, which describes something and, in that way, can be judged as to being true or false, and a *performative*, which is, or is part of, committing an action as in the following example:

‘I name this ship the *Queen Elizabeth*’ – as uttered when smashing the bottle against the stem (Austin 1962: 5).

Since its introduction, as Green (2007) observes, *Speech act theory* has made impact “not only within philosophy, but also in linguistics, psychology, legal theory, artificial intelligence, literary theory and many other scholarly disciplines”.

Introduced and elaborated by the language philosophers, Austin (1962) and Searle (1969, 1975, 1998), a *speech act* generally refers to what the speaker does when producing an utterance or, in other words, to the *force* this utterance carries; for example, promising, ordering, threatening, criticizing, apologizing, making a statement, etc. Following the principle that “speaking is a rule-governed form of behavior”, Searle describes a *speech act* as “the production or issuance of a sentence token under certain conditions” and also as “the basic or minimal unit of linguistic communication” which is not a word or a sentence, but rather the production of a word or a sentence (1969: 16–17). Green (2007) views a *speech act* as “a staple of everyday communicative life” with its major role in fastening the interlocutors’ utterances into a coherent piece of communication aimed at producing intended activities. For Mey, a *speech act* means performing an activity that changes the world or, as he puts it, “brings about a change in the existing state of affairs” (2001: 95). For instance, uttering *I’ll bet you ten dollars that the buses won’t run on Thanksgiving* is a typical way of performing a *speech act* of making a bet (Mey 2001: 96; italics added). Should the performance of the act of betting be successfully and correctly completed, the successive *speech act* of expressing the ‘uptake’ of the bet, like *You’re on*, is performed (Mey 2001: 97; italics added; cf. Schiffrin 1994).

Thus, instead of being truth conditional as the *constative* utterances are, *performatives* need to satisfy other conditions to be carried out in a *felicitous*, i.e., a correct and appropriate way. Those *felicity conditions* may include, among other things, appropriate circumstances, accepted conventional procedures, particular people as well as the sincerity of the interlocutors. For instance, if somebody produces an utterance like “Happy birthday!”, “Good luck!”, “Congratulations!” or “Well done!”, we can only judge the truth of that person’s feelings and sincerity or the truth of the fact that these utterances were pronounced (Mey 2001: 93). We cannot judge them as being true or false in terms of their propositional content. The reason is that these utterances are wishes and not propositions, namely they are words to carry an action of congratulating, praising or expressing a wish.

12.3.2. Types of *speech acts*

Austin’s (1962) and Searle’s (1969) theories served as the basis to identify five categories of *speech acts* in terms of the functions assigned to them:

Representatives	Directives	Expressives	Comissives	Declaratives
assertions	suggestions	apologies	promises	decrees
claims	requests	complaint	threats	declarations
reports	commands	thanks	offers	

(Jaworowska)

As Schiffrin points out, Austin (1962) ‘systematically dismantles’ his division into *constatives* and *performatives*, because of too much of overlapping between the two types, and concentrates on the performance of *speech acts* (Schiffrin 1994: 51). For Austin (1962), a *speech act* is a tripartite notion so that producing an utterance will inevitably entail the simultaneous production of three acts:

A locutionary act	an act of uttering something with a particular sense and reference, that is making a meaningful utterance, or, in other words, the activity of speaking	<i>E.g.</i> A. It's freezing outside.
An illocutionary act	an act of producing an utterance with a particular function or communicative force like <i>informing, ordering, warning, inviting, greeting, etc.</i>	(depending on the circumstances) = a warning to put on warm clothes, = a request not to go out, etc.
A perlocutionary act	an act or rather an effect possibly intended and achieved by the speaker <i>by</i> making a particular utterance like, like <i>convincing, persuading, deterring, etc.</i> the hearer	B. Thanks, I'll bundle up. = <i>convincing</i> (perlocutionary effect of A's utterance)

Austin 1962; Mey 2001; Yule 1996)

While further elaborating on *Speech act theory*, Searle (1969, 1975) makes some important “analytic connections” functioning within a *speech act* which may guide to particular methodological decisions:

<...> what the speaker means, what the sentence (or other linguistic element) uttered means, what the speaker intends, what the hearer understands, and what the rules governing the linguistic elements are. (Searle 1969: 21)

The identification of what *speech act* has been performed is eased through the use of *Illocutionary Force Indicating Devices* (IFIDs) (Mey 2001; Yule 1996). The most visible devices are performative verbs that would clearly name the *speech act* being performed by filling in the slot in the formula:

I (Vp) you that... (Yule 1996: 49)

The realisation of this formula would result in the following examples: *I predict that..., I promise that..., I warn you that..., etc.* However, in naturally occurring speech performative verbs are very often skipped but the relevant *speech acts* are nonetheless carried out. There are other IFIDs that are in place for the recognition of *speech acts*, namely some prosodic features such as intonation or stress and syntactic features such as word order. For instance, to perform a *speech act* of a warning or threat one might lower his or her voice to indicate the relevant illocutionary force. One way or another, as summarized by Mey, “it is the kind of activity in which people are engaged that makes us count certain utterances as promises, warnings, requests and so on” (2001: 98), or, in other words, the interpretation and understanding of a particular *speech act* being performed highly depends on certain contextual gains (cf. Black-

emore 1995, 2001; Carston 2002; Willson and Sperber 2004 on relevance theory). Consider an utterance provided by Mey:

There is a policeman at the corner.

Without a particular performative verb it can function as: *a warning* if uttered in the context of somebody committing a criminal activity, *a hint* if someone is in need of asking for directions or *a reminder* if someone is about to park his or her car in the place for the handicapped people (Mey 2001: 98).

12.3.3. *Indirect speech acts*

One of the trickiest aspects of *Speech act theory* is the analysis of how *speech acts* are carried out indirectly. Among the first to offer an interpretation of *indirect speech acts*, Searle (1975) proposes the following explanation: when an *indirect speech act* is performed, one *speech act* is understood as performed by the means of another one. In other words, there is a usual link between the main sentence types – declarative, interrogative and imperative – and their typical *illocutionary force*: “an assertive force for declarative sentences, a question force for interrogative sentences and a directive force for imperative sentences” (Verschueren 1999: 25; cf. Ilie 1994: 46; Schiffrin 1994: 70). When this principle is not obeyed as in “Can you pass the salt?”, where an interrogative sentence serves as a request, the language use is usually called *an indirect speech act* (Verschueren 1999).

According to Searle, *an indirect speech act* is performed by uttering “a sentence that contains the illocutionary force indicators for one kind of illocutionary act” but is used to perform “IN ADDITION, another type of illocutionary act” (1975: 59; emphasis in the original). Thus in Searle’s approach, *an indirect speech act* consists of a PRIMARY illocutionary act and a SECONDARY illocutionary act. The mutual operation of the two acts is demonstrated in the following example, described by Searle as “a typical case of the general phenomenon of indirection”:

Student X: Let’s go to the movies tonight.

Student Y: I have to study for an exam.

(Searle 1975: 61)

While the utterance of student X is explicitly a proposal because of its meaning, the answer of student Y can be interpreted as a rejection of the proposal because of the context rather than “in virtue of its meaning” (Searle 1975: 62). Consequently, in this exchange, student Y performs a primary non-literal or indirect illocutionary act of rejecting the proposal by performing a secondary illocutionary act, the direct one, which is literally a statement about him or her having to study for an exam.

For the interlocutors the interpretation of *indirect speech acts* involves intense and complicated inferential processes and requires the adequate inferential effort. Thus

many scholars have been engaged in developing theories to explain not only the use of indirect speech acts, but also the phenomenon of indirect communication in general and proposing methods of its analysis. Dwelling upon conversational logic, Grice (1975) offered the *Cooperative principle* the adherence to which should guide to a successful linguistic exchange. In terms of this principle, a logical and productive conversation can only be achieved if 4 conversational maxims are observed:

- *the maxim of quantity* (i.e., a requirement to be as informative as is required, not less or more),
- *the maxim of quality* (i.e., a requirement not to lie),
- *the maxim of relation* (i.e., a requirement to be relevant),
- *the maxim of manner* (i.e., a requirement to be brief, to avoid ambiguity and obscurity) (Grice 1975: 45–46).

Although Gricean maxims bring important insights into how conversational logic works, they can only be used to account for a “perfectly logical” (Lakoff 1995: 191) and rational conversation which is, however, far from being the only way to communicate one’s message. To account for those cases when the interlocutors prefer to opt out from the framework of the cooperative principle, Grice draws on the phenomenon of *conversational implicature* which he define as various inferential processes necessary to interpret the use of metaphor, irony or other cases of indirectness. Nevertheless, a more productive theory to explain indirect communication and *indirect speech acts* is the *Relevance theory* offered by Sperber and Wilson (1995, cf. Wilson and Sperber 2004). According to the *Relevance theory*, while the three principles or maxims – quantity, quality and manner – can be easily flouted or cancelled, the principle of relevance always holds for a conversation to proceed. Even if the turn in an interaction seems to bring in irrelevant information, its relevance might be explained as, for instance, by the interlocutor’s aim to change the topic, to end the conversation or signal another communicative force. Consequently, a successful interpretation of *indirect speech acts* entails gearing one’s inferential efforts towards the optimization of relevance. For instance, if two workmates were having lunch out and one of them uttered *Damn, I’m out of cash, I forgot to go to the bank today* (Brown and Levinson 1987: 69), under the presumption of relevance the other should not have much difficulty in interpreting this utterance as a request to lend the speaker money.

As suggested by Brown and Levinson (1987), some *indirect speech acts* have lost their indirectness because of their conventional use. Thus on hearing “Can you pass the salt?” one would hardly answer “yes I can” without doing this. *Rhetorical questions* have to be highlighted as one of the most common and important types of *indirect speech act*. For Brown and Levinson, *rhetorical questions* mean asking “a question with no intention of obtaining an answer” (1987: 223). If by putting an interrogative sentence the speaker is not performing a *speech act* of ‘asking a question’ or, in other words, is not requesting information, he or she must be replacing the interrogative force with another one. As Brown and Levinson illustrate, a rhetorical question could serve as an act of excuse like in “How was I to know...?” or as an act of criticism “How many times do I have

to tell you...?” (1987: 223). Ilie (1994) highlights another important feature of *rhetorical questions*. For her, it is not their failure to be used as ‘genuine’, i.e., information eliciting, questions, but their characteristic of being *self-answered* that distinguishes them from other types of questions (Ilie 1994: 2). She gives an example of a rhetorical question from *Time Magazine*: “Why else would Thomson buy a bankrupt LTV in a declining defence market except to get the technology” (Ilie 1994: 2). According to Ilie, the question is self-answered inasmuch as it contains a particular answer within itself. Namely, the only reason for Thomson to buy a bankrupt LTV is to get the technology.

Despite criticism, mostly related to taxonomies and definitions (cf. Mey 2001; Sbisá 1995), Austin and Searle’s *Speech act theory* still remains an important approach to the study of language as “an instrument of action, not just of speaking” (Mey 2001: 124). According to Schiffrin (1994), *speech act theory* has not been initially developed as a framework for a sociolinguistic or discourse analysis because of its focus on utterances, or acts, as the main structural unit. Nevertheless, for the same reason of providing a unit of analysis, as Schiffrin shows in her study, *Speech act theory* has appeared fruitful as a method of analyzing discourse. It is an approach to discourse “in which what is said is chunked (or segmented) into units that have communicative functions that can be identified and labelled” (Schiffrin 1994: 90). With respect to joining those units into sequences, complementary to *Speech act theory* are the methods of *conversation analysis* and *interactional sociolinguistics*. The former generally explores the sequences of utterances, or acts, while the latter focuses on the interpersonal strategies and goals achieved within those sequences.

12.4. Comprehension activities

Activity 2. Match the concepts in box A with the clues to their descriptions in box B:

1..... 2..... 3..... 4..... 5.....

Box A (1) locutionary act (2) felicity conditions (3) illocutionary act
(4) perlocutionary act (5) IFID

Box B (a) doing by saying (b) performative verbs (c) making an effect
(d) making an utterance (e) appropriateness

Activity 3. Consider a piece of *legalese* text given below from a bankruptcy court order (Northern) District of Illinois, Eastern Division, by Judge Jack B. Schmaet-terer, *in re* the petition for relief by the well-known bankrupted bicycle manufacturing company Schwinn, December 28, 1992) and answer the questions:

Enclosed is a form of a Proof of Claim. Each Proof of Claim must be filed ... on or before 4:30 p.m. Chicago, Illinois time, on January 6, 1993...

PLEASE TAKE FURTHER NOTICE THAT... ANY HOLDER OF A CLAIM WHO FAILS TO FILE A PROOF OF CLAIM ON OR BEFORE 4:30 P.M. CHICAGO TIME, ON JANUARY 6, 1993... SHALL BE FOREVER BARRED, ESTOPPED, AND ENJOINED FROM ASSERTING SUCH CLAIM (OR FILING A PROOF OF CLAIM WITH RESPECT THERETO).

Questions:

What kind of discourse is “legalese”?

What speech acts can you identify in the piece of discourse above; how are they expressed?

Why do you think the judiciary uses this kind of language?

Activity 4. Consider the following situation in a restaurant and the exchange between Ann and the waitress.

It is a busy lunch time in a crowded Italian style restaurant in Kaunas. Having received her order of the day menu, Ann realises that her main course of a chicken fillet is barely lukewarm. She manages to get hold of a busy waitress rushing by and opts for tentatively addressing her:

Ann: Is it ok that my chicken fillet is almost cold?

Waitress: No, it is not ok (utters without properly stopping and continues ahead).

Ann continues her lunch...

In pairs or small groups, first of all analyse the given exchange from the perspective of *Speech act theory*. Then add the insights from other sociolinguistic perspectives pondering upon such issues as *gender, politeness, cultural tradition, relevance, etc.*

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

12.5. Glossary

Constative – a truth conditional utterance.

Direct speech act – a speech act showing a direct relationship between the structure and the communicative force of an utterance, e.g. the use of an interrogative structure to place a question *Can you swim?*

Felicity conditions – the appropriate conditions to carry out a speech act.

Illocutionary act – producing an utterance with a particular function or communicative force like *informing, ordering, warning, inviting, greeting*, etc.

Illocutionary force indicating device (IFID) – means indicating communicative force in the speaker's utterance.

Indirect speech act – a speech act showing an indirect relationship between the structure and the communicative force of an utterance, e.g. the use of an interrogative structure to make a request *Can you pass the salt please?*

Locutionary act – the basic act of making an utterance.

Performative – a non-truth conditional utterance which is or is part of doing an action.

Perlocutionary act – an effect possibly intended and achieved by the speaker by making a particular utterance like, like *convincing, persuading, deterring*, etc. the hearer.

Principle of relevance – the principle that human cognition tends to be geared to the maximisation of relevance (Sperber and Wilson 1995).

Speech act – an action performed by the use of an utterance or a minimal functional unit in human communication (Yule 1996; http://instructional.calstatela.edu/lkam-his/tesl565_sp04/troy/spchact.htm).

12.6. References

- John L. Austin, 1962. *How to Do Things with Words: The William James Lectures delivered at Harvard University in 1955*. Oxford: Clarendon.
- Diane Blakemore, 1995. "Relevance Theory", in *Handbook of Pragmatics: Manual*, eds. Jef Verschueren et al., Amsterdam: John Benjamins, 443–452.
- Diane Blakemore, 2001. "Discourse and Relevance Theory", in *The Handbook of Discourse Analysis*, ed. Deborah Schiffrin, Deborah Tannen and Heidi E. Hamilton, Malden: Blackwell, 398–415.
- Penelope Brown & Stephen C. Levinson, 1987. *Politeness: Some Universals in Language Use*, Cambridge: Cambridge UP.
- Robyn Carston, 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*, Oxford: Blackwell.
- Mitchell Green, 2007. <http://plato.stanford.edu/entries/speech-acts/>.
- Paul H. Grice, 1975. "Logic and Conversation", in *Syntax and Semantics. Vol. 3: Speech Acts*, eds. Peter Cole and Jerry L. Morgan, New York: Academic Press, 41–58.
- Cornelia Ilie, 1994. *What Else Can I Tell You?: A Pragmatic Study of English Rhetorical Questions as Discursive and Argumentative Acts*. Stockholm: Almqvist and Wiksell International.

- Joanna Jaworowska. *Speech Act Theory*, retrieved from: http://instructional1.calstatela.edu/lkamhis/tesl565_sp04/troy/spchact.htm.
- Robin Tolmach Lakoff, 1995. "Conversational Logic", in *Handbook of Pragmatics: Manual*, eds. Jef Verschueren et al., Amsterdam: John Benjamins, 190–198.
- Jacob L. Mey, 2001. *Pragmatics: An Introduction*. 2nd ed. Oxford: Blackwell.
- Deborah Schiffrin, 1994. *Approaches to Discourse*, Oxford: Blackwell.
- John R. Searle, 1969. *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge UP.
- John R. Searle, 1975. "Indirect Speech Acts", in *Syntax and Semantics. Vol.3: Speech Acts*, eds. Peter Cole and Jerry L. Morgan, New York: Academic Press, 59–82.
- John R. Searle, 1998. "Indirect Speech Acts", in *Pragmatics: Critical Concepts. Vol. 4: Presupposition, Implicature and Indirect Speech Acts*, ed. Asa Kasher, London: Routledge, 617–638.
- Dan Sperber & Deirdre Wilson, 1995. *Relevance: Communication and Cognition*. 2nd ed. Oxford: Blackwell.
- Jef Verschueren, 1999. *Understanding Pragmatics*. London: Arnold.
- Deirdre Wilson & Dan Sperber, 2004. "Relevance Theory", in *The Handbook of Pragmatics*, eds. Laurence R. Horn and Gregory Ward, Malden: Blackwell, 607–632.
- George Yule, 1996. *Pragmatics*. Oxford: Oxford UP.
- *Text based on the unpublished PhD dissertation: Vilma Bijeikienė, 2007. Gender in Political Discourse: Indirectness in the British Parliamentary Debates*. Bergen University.

CHAPTER 13: INTERACTIONAL SOCIOLINGUISTICS

13.1. Key concepts

Communicative goals/gains

Context

Contextualisation cues

Face

Facework

Move

Negative face

Positive face

Social order

13.2. Pre-reading activity

Activity 1. In groups discuss and describe the concepts given above. What do you think could be the role of each in *Interactional sociolinguistics*?

.....

.....

.....

.....

.....

13.3. Interactional sociolinguistics

13.3.1. The origin and the main principles of *Interactional sociolinguistics*

The origin of *Interactional sociolinguistics* and its fundamental tenets are mostly associated with the names of Erving Goffman (1967) and John. J. Gumperz (1982, 2001) as well as with Dell H. Hymes (1962) as a more distant predecessor through his work within ethnography of speaking and ethnography of communication. As defined by Gumperz:

Interactional Sociolinguistics (IS) is an approach to discourse analysis that has its origin in the search for replicable methods of qualitative analysis that account for our ability to interpret what participants intend to convey in everyday communicative practice. It is well known that conversationalists always rely on knowledge that goes beyond grammar and lexicon to make themselves heard. (Gumperz 2001: 215)

Following Gumperz, *Interactional sociolinguistics* is intended to serve as a methodological framework for qualitative sociolinguistic analysis focusing on strategic aspects of communicative practices as much as on their social properties such as gender, ethnicity, culture, etc. From the perspective of *Interactional sociolinguistics*, speaking is not seen as solely a process of encoding and decoding messages drawing exclusively on grammatical parameters and denotational meaning of lexical items. In *Interactional sociolinguistics* speaking is seen as “an ongoing process of negotiation, both to infer what others intend to convey and to monitor how one’s own contributions are received” (Gumperz 2001: 218). The main focus of analysis is thus the dynamics of the participants’ ‘shared’ and ‘non-shared’ interpretations of messages that may depend on a variety of situational, cultural and social aspects. Consider an example provided by Gumperz:

Following an informal graduate seminar at a major university, a black student approached the instructor, who was about to leave the room accompanied by several other black and white students, and said:

- a. *Could I talk to you for a minute? I’m gonna apply for a fellowship and I was wondering if I could get a recommendation?*

The instructor replied:

- b. *O.K. Come along to the office and tell me what you want to do.*

As the instructor and the rest of the group left the room, the black student said, turning his head ever so slightly to the other students:

- c. *Ahma git me a gig!* (Rough gloss: ‘I’m going to get myself some support.’) (Gumperz 1982:30)

The given exchange can be studied from a variety of linguistic perspectives, if examined from the interactional sociolinguistic perspective, however, it would be first of all analysed with regard to the personal communicative strategies the speaker may be using and the socio-cultural knowledge the interpretation of these strategies could be based on. An important point to start with is to take notice of the fact that the speaker, who is a speaker of Black English variety, resorts to it as much as to the use of Standard English variants. The code-switching between the two varieties must be grounded in certain communicative gains. In order to clarify the basis and effectiveness of such usage, Gumperz asked other students, some of whom participated in this encounter, for their interpretation. The answers could be seen as clustering into the following groups:

- One group of students who have had less encounter with Black English variety did not seem to understand the last utterance or refused to elaborate their interpretations beyond indicating that the speaker “lapsed into dialect”, these students had problems in understanding the words *gig* and *ahma* and considered the sentence in c. as incongruous with the exchange of a. and b.

- The other group regarded the code-switching from the Standard English to Black English variety as “a rejection of the white instructor and of the academic enterprise”.
- The last group, which included black students and one white student with much contact with the black community, presented the following interpretation: “He was trying to justify himself; he was appealing to others in the group, as if to explain his earlier remarks by suggesting: ‘I’m still in control,’ ‘I’m just playing the game as we blacks must do if we are to get along in a white dominated world’”, both the second and the third groups did not see utterance c. as incongruous and explained the coherence of the whole exchange as follows: turn b. is a response to a. while c. serves as a comment on a. and b. interaction (Gumperz 1982: 31–33).

Sharing many of original fundamentals, the methodological approaches of *Conversation analysis* and *Interactional sociolinguistics* are alike in their “concern with the problem of social order” (Schiffrin 1994: 232). Nevertheless, some important aspects of their divergence need to be singled out. While concentrating on the analysis of naturally occurring speech exchanges, however, *Interactional sociolinguistics* does not aim so much at studying structural sequencing of an interaction as it concentrates on its functions and the interpretation of meaning. The general interest of *Interactional sociolinguistics* in the study of language as a social activity is primarily realized by exploring the interrelation between linguistic communication and social context at large. As Gumperz and Cook-Gumperz put it, the focus is on “the role that communicative phenomena play in the exercise of power and control and in the production and reproduction of social identity” (1982: 1). Thus social identity of the interlocutors, or their links with a particular community, including such major social parameters as gender, culture and ethnicity, is of crucial importance in analyzing the use of language.

Interactional sociolinguistics is a method of concentrating on the interpersonal properties of a particular linguistic exchange including the aims, intentions and strategies of the interlocutors. As summarized by Gumperz the major goal of an interactional sociolinguistic study is to investigate “speech exchanges involving two or more actors” and to reveal:

...how individuals participating in such exchanges use talk to achieve their communicative goals in real-live situations, by concentrating on the meaning-making processes and taken-for-granted assumptions that underlie the negotiation of interpretations. (2001: 218; cf. Gumperz 1982)

To achieve this goal, while analysing the interlocutors’ involvement in an interaction, Gumperz encourages researchers to pay major attention to the discovery of *contextualization cues*, i.e., “the means by which speakers signal and listeners interpret what the activity is, how semantic content is to be understood and how each sentence relates to what precedes or follows” (1982: 131). The realization of contextualization cues involves code and style switching as well as various pro-

sodic, lexical and syntactic options. Bayley (2008) describes an example from Gumperz (1982) demonstrating how intonation functions as a cue unshared by two ethnical groups and leads to a misunderstanding. The two groups consist of British and South Asian workers at an airport in the UK:

When an Anglo British cafeteria server in this workplace offered gravy to a person in line, she would say “Gravy?” with a rising intonation contour. Anglo British workers used this prosodic information to interpret the utterance as an offer or question: “Would you like gravy?” In contrast, when recently-immigrated South Asian cafeteria workers asked employees if they wanted gravy, they said “Gravy” with falling intonation. Anglo British workers interpreted the falling intonation as contextualizing a statement (akin to “This is gravy – take it or leave it”), which they found redundant and rude.

Consequently, the sequencing of utterances, or moves, is explored with the basic aim of discovering what communicative gains are achieved or failed. If the understanding and the relevance of the cues are shared by the interlocutors, the cues work unnoticed and the linguistic exchange proceeds smoothly. Otherwise, an exchange can result in various cases of miscommunication or even language-related conflicts.

13.3.2. Face and facework from an interactional perspective

To explore the communication of personal goals by individuals, who are also social and cultural beings, Goffman (1967) introduced the notion of ‘face’ which incorporates two dimensions: the dimension of one’s social identity and the dimension of one’s personal interactional needs. From Goffman’s perspective, a linguistic exchange should be perceived as the interlocutors’ moves towards construction and maintenance of ‘face’. Goffman’s analysis of facework has also served as the basis for a methodological framework in Brown and Levinson’s (1987) politeness theory.

Before its application in contemporary linguistic studies, the notion of ‘face’, as the conceptualization of self, was broadly known in the folklore of various cultures. From there, it was adopted by Goffman (1967) in his theory of social interaction and was further elaborated in the theories of politeness and facework (Brown and Levinson 1987; Locher 2004; MacMartin et al. 2001; Tracy 1990; Watts 2003). Watts considers the ‘face’ metaphor as an embodiment of such abstract human characteristics as honor, respect or esteem and traces it back to the Chinese traditions of politeness as well as to the English idiomatic expressions ‘to put on a good face’, ‘to lose face’, ‘to face up to X’, etc. (2003: 124). Brown and Levinson define ‘face’ as “the public self-image that every member wants to claim for himself” (1987: 61), whereas Watts adds to this definition a feature of being a “positive social image” (2003: 124). For Brown and Levinson, the concept of ‘face’ consists of two interrelated aspects ‘negative face’ and ‘positive face’:

NEGATIVE FACE	Signifies the need to be independent, the freedom of action and the freedom from being imposed on.	E.g. The 'negative faces' of the the hearer is at stake when the speaker is asking the hearer to lend him or her some money.
POSITIVE FACE	Indicates the desire that a person's 'self-image be appreciated and approved of by others; the need to feel connected to a group and to know that his or her wants are shared by others.	E.g. The 'positive face' of the speaker will be harmed when his or her request is denied by the hearer.

(Brown and Levinson 1987; Yule 1996)

Watts further elaborates on the dynamic nature of 'face' suggesting that a self-presentation assumed by interlocutors "may differ from one interaction to the next" or even fluctuate within the same linguistic encounter (2003: 124). Similarly, MacMartin et al. (2001) encourage a more interactional than cognitive approach to the notion of 'face' in linguistic inquiries by suggesting that 'face' should be treated as participants' interactional resource in talk rather than a solely mental construct.

From an interactional perspective, it is generally assumed that the concern about 'face' is a cooperative endeavor of all engaged in communication as long as "normally everyone's face depends on everyone else's being maintained" (Brown and Levinson 1987: 61). Nevertheless, as it also follows from the cooperative treatment of 'face', not only can people cherish each other's 'faces' in the course of a linguistic encounter, but may also threaten them and cause them a certain damage. Moreover, as highlighted by the scholars of facework, such communicative effects as losing 'face', saving 'face', threatening 'face' or enhancing 'face' are intricately related and form patterns of directly proportional tendencies (Brown and Levinson 1987; MacMartin et al. 2001; Tracy 1990; Watts 2003). For instance, Tracy draws a link between minimizing a negative impact on one's personal 'face' and presenting a challenge to another interlocutor: "a basic assumption is that one common way an individual saves face is by attacking an other" (1990: 216). Along the same lines, Brown and Levinson project an intense dynamics between a face threatening and a face enhancing effect claiming that "just as to raise the other is to imply a lowering of the self, so a raising of the self may imply a lowering of the other" (Brown and Levinson 1987: 39).

Brown and Levinson (1987) present a detailed exploration into what kind of language use can be regarded as entailing a threat to the 'faces' of those engaged in a linguistic interaction. In their taxonomy of *Face threatening acts* (FTAs), distinctions are made along two axes: they firstly distinguish between the threats to the hearer's and the speaker's 'faces' and further between a 'positive face' and a 'negative face' of the two. Drawing largely on Brown and Levinson's seminal classification, Partington (2003) observes that the hearer's 'negative face' is threatened by putting "pressure on the hearer to do something" or driving him or her into the feeling of indebtedness (2003: 125). The hearer's 'positive face' can suffer from showing the lack of concern for his or her feelings and goals. In the meantime, the speaker's 'negative face' is put at risk by recognizing his or her debt, as in the case of expressing gratitude and thanks. The speaker's 'positive face' could be damaged by the admission of previous shortcomings and guilt.

TAXONOMY OF FACE THREATENING ACTS (FTAs)

FTA to the hearer's negative face	<i>orders and requests, suggestions, advice, and reminding, threats, warnings and dares;</i>
FTA to the hearer's positive face	<i>disapproval, criticism, contempt or ridicule, reprimands, accusations and insults, challenges, contradictions and disagreement, bringing of bad news about the hearer or good news (boasting) about the speaker, showing non-attention to the hearer's needs;</i>
FTA to the speaker's negative face	<i>expressing thanks, making excuses, acceptance of offers, unwilling promises and offers;</i>
FTA to the speaker's positive face	<i>being disruptively interrupted by the hearer, apologies and self-humiliation, shuffling, cowering, acting stupid and self-contradicting, confessions, admissions of guilt and ignorance, emotional leakage.</i>

(Brown and Levinson 1987: 65–68)

13.3.3. Practical application of *Interactional sociolinguistics* in gender and discourse analysis

Within the perspective of *Interactional sociolinguistics* the different perceptions of *contextualisation cues* and the ensuing communication breakdowns are primarily interpreted as the consequence of cultural and ethnic differences. Nevertheless, in search for a theoretical framework to tackle the problem of gender-based miscommunication in American culture, Maltz and Borker (1998, original article written in 1982) turned to *Interactional sociolinguistics*, more particularly to Gumperz's (1982) model of cross-ethnic communication. They arrived at the following hypothesis:

We argue that American men and women come from different sociolinguistic subcultures, having learned to do different things with words in a conversation, so that when they attempt to carry on conversations with one another, even if both parties are attempting to treat one another as equals, cultural miscommunication results. (Maltz and Borker 1998: 420)

To support their reasoning, Maltz and Borker draw on a number of studies on children's socialization showing that boys and girls develop different rules for engaging in and interpreting a conversation. What boys learn to do with words entails gaining and maintaining an audience as well as expressing power through the use of direct commands like "give it to me", or verbal ridicule like "you're a dolt" (Maltz and Borker 1998: 426). For girls, however, communication is based on the concepts of "best friends", "sharing secrets" and not being "bossy" which result in such co-operative and indirect expressions as "let's", "we gonna" or "we could" (Maltz and Borker 1998: 424–425). These distinct communicative habits, according to Maltz

and Borker, permeate into one's linguistic style in the adulthood and surface out as different use and interpretation of the same linguistic elements and strategies. For women, minimal responses like "yes" and "mm hmm" imply "I'm listening to you", while men perceive them as "I agree with you" (Maltz and Borker 1998: 421). In the same manner, Maltz and Borker name other five tokens of gender-related linguistic variation: questions, linking of utterances, verbal aggressiveness, topic flow and topic shift as well as problem sharing (1998: 430). Consequently, if a conversation is seen as "a negotiated activity" based on "shared assumptions" (Maltz and Borker 1998: 421), the cultural approach seems to reveal the following tendency: some of communicative competence is not equally 'shared' by men and women and thus their interaction might not result in a successfully 'negotiated activity'.

The framework of *Interactional sociolinguistics* was further applied in Tannen's studies (1994a, b, 1998) with her best-selling book *You Just Don't Understand* (1990). As is suggested by the title, Tannen takes over the aims of Maltz and Borker to analyze and explain misunderstandings in cross-gender conversations treating them as cross-cultural differences. Following the cultural model, Tannen considers men and women to speak "different *but equally valid* styles" (1990: 15), as if being members of different cultures: "different words, different worlds" (1990: 23). The dualism, encapsulated in the preceding quote, remains the key feature of Tannen's interpretations of the numerous cross-gender conversations in private (1990) and in public (1994b) domains. For instance, the language women speak and hear is "connection and intimacy", whereas what men speak and hear is "status and independence" (Tannen 1990: 42). Women talk about their problems to take a possibility to just talk them out, to be listened to and to receive understanding, they do not need any solution to be offered. On the contrary, men perceive problem sharing as a call for advice and solution. Men cannot understand how "endless monotonous discussions" (Tannen 1990: 160) could be a sign of intimacy for women. Pursuant to cultural model, Tannen lays much emphasis on societal expectations and restrictions. Thus one of the most salient features of a female linguistic style their conversational cooperativeness and supportiveness is seen as an outcome of teaching girls that better results are reached by presenting ideas in a form of suggestions. By contrast, boys are brought up in a more status-oriented spirit whereby they learn to be assertive and competing as well as anticipating a linguistic challenge.

Within this interactional sociolinguistic approach to gender-related language use, more attention has been paid on the variety of functions that the elements of male and female linguistic styles can perform. As Holmes rightly suggests, "one form may serve many functions, and particular functions are expressed by a variety of forms" (1995: 72, cf. 1998, 2005, Holmes and Stubbe 2003). For women, as she maintains, compliments are the expression of solidarity and politeness, whereas men might find them as an intrusion into their privacy or a threat to their reputation (Holmes 1998). Similarly, linguistic *indirectness*, generally associated with women, could also be used by men but "in different situations and in different

ways” (Tannen 1994b: 79). Women are usually more indirect in giving requests, suggestions and criticism, while men prefer to sound indirect when communicating their weaknesses, problems and errors as well as their emotions other than anger. Although rich in generalisations and discussion of sociolinguistic relevance, Tannen’s books need to be taken with caution, as suggested by some of her critics (Cameron 1998). On the one hand, Tannen tends to overgeneralise by making conclusions about female and male linguistic styles on the basis of one or two exchanges. On the other hand, the explanation that she brings about gender-related variation as cultural variation may be not just inadequate but also misleading.

13.4. Comprehension activities

Activity 2. Indicate whether the given statements are true (T) or false (F) with regard to *Interactional sociolinguistics* as a research method. Explain your decisions to your colleagues on the basis of the given theory.

1. Conversation analysis and Interactional sociolinguistics are similar in the conceptualisation of context.
2. The category of ‘face’ is conceptualised as an interactional as well as a mental construct.
3. There is nothing negative about the concept ‘negative face’, it is simply used as an indicator of the opposite category to ‘positive face’.
4. Minimizing a negative impact on one’s personal ‘face’ inevitably result from minimizing a negative impact on the other’s face.
5. Maltz and Borker accentuate the influence of upbringing on the development of feminine and masculine linguistic subcultures.
6. As result from Interactional sociolinguistic analysis of gender and discourse, linguistic indirectness is an exclusive feature of feminine discourse.

Activity 3. (1) Match the utterances with the function:

Expressing emotions Criticism Compliment Order Expressing thanks Apology

(2) Indicate to which of the four faces (*the speaker’s positive face, the speaker’s negative face, the hearer’s positive face, the hearer’s negative face*) these utterances pose a threat.

Utterances	Functions	FTA to...
1. “I really like you.”		
2. “I think I made a huge mistake.”		
3. “Please give me that book.”		
4. “I think your report was not concise enough.”		
5. “Thank you so much for your help.”		
6. “You’re feeling sad because of your ex-boyfriend, aren’t you?”		

(http://www.glottopedia.org/index.php/Face-threatening_act)

Activity 4.

A. Try to remember a case of either gender- or ethnicity-related miscommunication. Can you also remember why the miscommunication occurred? Can you single out any contextualisation cues that might have been at issue?

B. Find an example of miscommunication in the media. Analyse it from the perspective of *Interactional sociolinguistics* paying attention to the social aspects of the interaction, i.e., gender, race or ethnicity of the interlocutors, the strategies used, the facework, etc.

13.5. Glossary

Context – in IS context is a complicated category; on the one hand, it is understood as situations, occasions, encounters or participation frameworks which have forms and meanings that are partially created and sustained by language; on the other hand, language and context are seen as co-constituting one another so that language does not just function ‘in’ context but also forms and provides context (Schiffrin 1994).

Contextualisation cues – the means by which speakers signal and listeners interpret what the activity is, how semantic content is to be understood and how each sentence relates to what precedes or follows (Gumperz 1982).

Face – a person’s public self-image or an embodiment of such abstract human characteristics as honor, respect or esteem (Watts 2003, Yule 1996).

Facework – the dynamics of face saving, threatening and enhancing strategies in a conversation.

Negative face – a person’s need to have freedom of action and freedom from imposition.

Positive face – a person’s need to be understood, appreciated and accepted by others.

Social order – the normative order indexed by interactional routines and the patterns created within interaction that could be discovered through the analysis of interactional data (Heller 2003).

13.6. References

- Benjamin Bailey, 2008. “Interactional sociolinguistics”, retrieved from: http://works.bepress.com/cgi/viewcontent.cgi?article=1058&context=benjamin_bailey.
- Penelope Brown & Stephen C. Levinson, 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge UP.
- Deborah Cameron, 1998. “‘Is there any ketchup, Vera?’: gender, power and pragmatics”, *Discourse & Society* 9 (4), 437–455.
- Erving Goffman, 1967. *Interaction Ritual: Essays in Face-to-Face Behavior*. New York: Pantheon Books.
- John J. Gumperz, 1982. *Discourse Strategies*. Cambridge: Cambridge UP.
- John J. Gumperz, 2001. “Interactional sociolinguistics: a personal perspective”, in *The Handbook of Discourse Analysis*, eds. Deborah Schiffrin, Deborah Tannen and Heidi E. Hamilton, Malden: Blackwell, 215–228.
- John J. Gumperz & Jenny Cook-Gumperz, 1982. “Introduction: language and

- the communication of social identity”, in *Language and Social Identity*, ed. John J. Gumperz, Cambridge: Cambridge UP, 1–21.
- Monica Heller, 2001. “Discourse and interaction”, in *The Handbook of Discourse Analysis*, eds. Deborah Schiffrin, Deborah Tannen and Heidi E. Hamilton, Malden: Blackwell, 215–228.
 - Janet Holmes, 1995. *Women, Men and Politeness*. London: Longman.
 - Janet Holmes, 1998. “Complimenting- a Positive Politeness Strategy”, in *Language and Gender: A Reader*, ed. Jennifer Coates, Oxford: Blackwell, 100–120.
 - Janet Holmes, 2005. “Power and Discourse at Work: Is Gender Relevant?”, in *Feminist Critical Discourse Analysis: Gender, Power and Ideology in Discourse*, ed. Michelle M. Lazar, Houndmills: Palgrave Macmillan, 31–60.
 - Janet Holmes & Maria Stubbe, 2003. “‘Feminine’ Workplaces: Stereotypes and Reality”, in *The Handbook of Language and Gender*, eds. Janet Holmes and Miriam Meyerhoff, Malden: Blackwell, 573–599.
 - Dell H. Hymes, 1962. “The ethnography of speaking”, in *Anthropology in Human Behaviour*, eds T. Gladwin and W. C. Sturtevant, Washington, D. C.: Anthropological Society, 13–53.
 - Miriam A. Locher, 2004. *Power and Politeness in Action: Disagreements in Oral Communication*. Berlin: Mouton de Gruyter.
 - Clare MacMartin, Linda A. Wood & Rolf O. Kroger, 2001. “Facework”, in *The New Handbook of Language and Social Psychology*, eds. W. Peter Robinson and Howard Giles, Chichester: John Wiley and Sons, 221–237.
 - Daniel N. Maltz & Ruth A. Borker, [1982] 1998. “A Cultural Approach to Male– Female Miscommunication”, in *Language and Gender: A Reader*, ed. Jennifer Coates, Oxford: Blackwell, 417–434.
 - Alan Partington, 2003. *The Linguistics of Political Argument: The Spin-doctor and the Wolf-pack at the White House*. London: Routledge.
 - Deborah Schiffrin, 1994. *Approaches to Discourse*. Oxford: Blackwell.
 - Deborah Tannen, 1990. *You Just Don’t Understand: Women and Men in Conversation*. New York: William Morrow and Company.
 - Deborah Tannen, 1994a. *Gender and Discourse*. New York: Oxford UP.
 - Deborah Tannen, 1994b. *Talking from 9 to 5. Women and Men at Work: Language, Sex and Power*. Lancaster: Virago.
 - Deborah Tannen, 1998. *The Argument Culture: Changing the Way We Argue and Debate*. London: Virago.
 - Karen Tracy, 1990. “The Many Faces of Facework”, in *Handbook of Language and Social Psychology*, eds. Howard Giles and W. Peter Robinson, Chichester: John Wiley and Sons., 209–226.
 - Richard J. Watts, 2003. *Politeness*. Cambridge: Cambridge UP.
 - George Yule, 1996. *Pragmatics*. Oxford: Oxford UP.
 - *Text based on the unpublished PhD dissertation: Vilma Bijeikienė, 2007. Gender in Political Discourse: Indirectness in the British Parliamentary Debates*. Bergen University.

CHAPTER 14:
CRITICAL DISCOURSE ANALYSIS

14.1. Key concepts

- Discourse
- Genre
- Hierarchy
- Hegemony
- Ideology
- Power
- Racism
- Social struggle
- Unequal encounters

14.2. Pre-reading activity

Activity 1. In groups discuss and describe the concepts given above. What do you think could be the role of each in *Critical Discourse Analysis*?

.....

.....

.....

.....

.....

14.3. Critical discourse analysis

14.3.1. The main principles of *Critical discourse analysis*

Critical discourse analysis (CDA hereafter) is another perspective that deals with the study of language in social context. As van Dijk suggests, CDA is neither “a homogeneous method, nor a school or a paradigm, but at most a shared perspective on doing linguistic, semiotic and discourse analysis” (1993a: 131). What this ‘shared perspective’ unites is, firstly, the conceptualization of discourse as “social practice” (Fairclough and Wodak 1997: 258) and, secondly, the *critical* approach to issues under discussion (Wodak 1995). Being neither a method, nor a theory, CDA should be seen as interplay of methods and theories with rich methodological variation and with the choice of a particular method made on the basis of the aims of investigation and the nature of data being investigated. The scope of analytic perspectives may thus include:

- grammatical (phonological, syntactic, lexical, semantic) analysis;
- pragmatic analysis of speech acts and communicative acts;
- rhetorical analysis;
- stylistics;
- the analysis of specific (genre, etc.) structures: stories, news reports, parliamentary debates, lectures, advertisements, etc.;
- conversation analysis of talk in interaction;
- semiotic analysis of sounds, images and other multimodal properties of discourse and interaction (van Dijk 2008: 3).

These different analytic perspectives could be combined in various ways to achieve a specific methodological approach which could serve best in solving particular societal problems, for example, “the semantics of narrative, the rhetoric of political discourse, the pragmatics of conversation, or the semiotics of style” (van Dijk 2008: 3). In the section “Example Research: Critical Discourse Analysis” presented by the University of Sheffield, one is provided with the most usual linguistic aspects attended to by scholars in CDA:

Active or Passive voice	The use of an active verb gives a clear picture of who performed a particular action, and to whom, for example: “Police attack protestors.” The use of a passive verb states what has been done, and to whom, but does not blame anyone in particular for the action, for example: “Protestors attacked.” Alternatively, nominalisation can be used, where the noun form of the verb is used to create even more ambiguity, for example: “Attack on protestors.”
Naming	The ways in which people are named can also perpetuate ideologies. For example, the newspaper headline “five Asian youths involved in armed robbery” creates a very different picture than “five young men involved in armed robbery”. Similarly, the way people are described in texts, or after giving quotes can present two different pictures, for example: “Dr. Sarah Jones” creates a different picture than “Single mother of two, Sarah Jones.”
Pre-modifiers	Pre-modified nouns can present varying views of a topic. For example, “gay marriage” or “same-sex marriage” implies that this is essentially different from hetero-sexual marriage.
Indirect quotes	This is particularly common, when the results of a poll are being used, for example “poll shows 70 % oppose gay marriage”, however there may be no evidence of reported speech saying this.

(Example Research: Critical Discourse Analysis in *All About Linguistics*)

The exploitation of these or other linguistic elements may serve for an analyst as important strategic choices made by the speaker, often showing the speaker’s ideological attitudes.

Scholars in CDA are mostly critical of various societal inadequacies like racial, gender and minority discrimination, legitimating power and ideology and other forms of dominance of one social group over another, as it is implemented through and reflected in linguistic behaviour (Fairclough 2001; Fairclough and Wodak 1997; Lazar 2005; van Dijk 1993b, 1997, 2002, 2008; Wodak 1995, 2001). This critical attitude materializes into problem-oriented studies which, as Wodak observes, “aim less to contribute to a specific discipline, paradigm, school or discourse theory, than to address what they believe to be pressing social issues which they hope to understand better, and possibly change, through their analysis” (1995: 206). Therefore, scholars in CDA,

additionally to linguistic analysis, assume a rather political stance in support of those who they see as a dominated group and crystallise their general goals accordingly:

- relations of domination are studied primarily from the perspective of, and in the interest of the dominated group;
- the experiences of (member of) dominated groups are also used as evidence to evaluate dominant discourse;
- it can be shown that the discursive actions of the dominant group are illegitimate;
- viable alternatives to the dominant discourses can be formulated that are consistent with the interests of the dominated groups (van Dijk 2008: 6).

In line with the goals set by van Dijk, Fairclough (2001) praises CDA for surpassing in its aims and methods other perspectives in the study of language use such as sociolinguistics and pragmatics. As he explains, sociolinguistics, while answering the question of ‘what?’ about sociolinguistic variation, fails to give the answers to ‘why?’ and ‘how?’ it relates to power and dominance (Fairclough 2001: 6). Pragmatics, on the other hand, tends to see “cooperative interaction between equals” (Fairclough 2001: 8) as a prototypical form of social interaction and thus to ignore the constraints of power and social conventions structuring linguistic behaviour. In Wilson’s (2001) opposing view, however, because of such close involvement in political issues, representatives of CDA stop being just the analysts of political texts, but become political actors themselves (cf. Bijeikienė 2008). Moreover, as pointed out by Chilton (2005), the major shortcoming of CDA is the lack of strong theoretical foundation, which makes this perspective of questionable input into the social sciences as well as linguistics.

14.3.2. Interrelation between language and power

Exploring power or a powerful discourse has been on target by the analysts of rhetoric and political language since the epoch of antiquity. In modern times, the concept of power has gained new important meanings and functions in the analysis of language use in various social contexts. Starting with Brown and Gilman’s (1960) ‘power and solidarity semantics’ and then following with the perspective of ‘critical linguistics’ (Fowler and Kress 1979) to the current studies in ‘critical discourse analysis’ (CDA), the concept of *power* has been one of the chief factors in interpreting the motives and reasons for why people speak or choose to speak the way they do.

Brown and Gilman (1960) introduced the power and solidarity dichotomy to analyze the use of second person pronouns in a two-choice pronominal system where at least two pronominal alternatives, a familiar form and a deferential one, are available to address a person. In Brown and Gilman’s theory, the choice of the deferential pronominal address form is motivated by the asymmetrical, i.e., power-governed, relations of the interlocutors. The socially superior person uses the familiar form to address his or her subordinate but receives the deferential form in return. By contrast,

the socially equals exchange the familiar form (in a relatively informal context) or the deferential one (in a relatively formal context) thereby showing the solidarity relation and the absence of power difference. Brown and Gilman's theory of power and solidarity has widely spread into sociolinguistic studies reaching far beyond the issue of 'address forms'. For instance, Tannen, who considers this dichotomy as "fundamental to sociolinguistic theory" (1994: 22), applies it to the investigation of linguistic strategies in the field of gender and discourse. In addition, this dichotomy is referred to in the perspective of 'critical linguistics' (Fowler and Kress 1979), in the studies of politeness strategies (Brown and Levinson 1987; Holmes 1995; Watts 2003) and in the research on political argumentation (Partington 2003).

In CDA the concept of power is filled with a more versatile content. As put by van Dijk:

If we define Critical Discourse Studies (CDS) as a scholarly movement specifically interested in theory formation and critical analysis of the discursive reproduction of power abuse and social inequality, a detailed examination of the concept of power is central task of CDS. Yet, as is the case for many fundamental notions of the social sciences, the notion of power is as complex as it is fuzzy. (van Dijk 2008: 1)

With the relationship between language and power being among their primary goals, critical discourse analysts examine various manifestations of power in society through discursive practices. On the one hand, a relative power, based on one's social status, is an important marker of difference when language is used between the *empowered* and the *powerless* (Partington 2003: 128) in the so-called *unequal encounters* (Fairclough 2001: 36). This could be, for instance, an interaction between a doctor and a medical student (Fairclough 2001: 37) or the police interrogation of a female rape victim (Fairclough 1995: 28–30). There, the unequal distribution of power results in "powerful participants *controlling and constraining the contributions of non-powerful participants*" (Fairclough 2001: 38–9). In that respect, the conceptualisation of *power* is to some extent similar to Brown and Gilman's theory (1960). On the other hand, however, in CDA the importance of the concept of *power* in linguistic analysis exceeds the limits of a face-to-face interaction where a small number of individuals are involved. The interest in the functioning of power reaches out to the widespread and complex issues of the language-related racial, ethnical or gender conflicts and discrimination. In CDA the concept of power is closely intertwined with such broad social phenomena as ideology, control and social struggle. Power is seen as a given social attribute which an interlocutor (or a particular social group) either possesses or not, or possesses to a certain degree, before entering the interaction. It is the type of power that results from and reflects social asymmetries. On the basis of this relative power or powerlessness the interlocutors choose and adapt their linguistic means and strategies.

14.3.3. Focus on ideology

Along with the concept of power, CDA necessarily deals with the concept of ideology by investigating “the ways in which language mediates ideology” which, in its own turn, serves to preserve “unequal power relations” (Wodak 2001: 10; cf. Fairclough 2001). Moreover, examining power and ideology as the dominant and closely intertwined phenomena in the production of political discourse is frequently the initial step in the critical analysis of political texts (Fairclough 2001; van Dijk 1993b, 2002). Fairclough describes ideologies as contrasting meaning systems. For example, the Marxist ideology based on the ideas of “material interests” and “the struggle for power” can be contrasted with the post-war American ideology, which is designed as “a weapon against Marxism” (Fairclough 2001: 78). Most importantly, as Fairclough points out, for ideologies to function effectively, their functioning should be “least visible” (2001: 71). They should be based on common sense and the taken-for-granted assumptions that are not within conscious awareness and are rarely questioned. Consequently, when this commonsensical grounding of unequal power relations surfaces out, the ideology “ceases to be common sense, and may cease to have the capacity of sustaining power inequalities, i.e. to function ideologically” (Fairclough 2001: 71).

Van Dijk echoes Fairclough’s reasoning with his ideas about power and ideology which, in his view, can be exercised through the control of *action* and *cognition* (van Dijk 1993b: 254). Restricting people’s freedom of action illustrates the former, whereas managing their minds, which is usually achieved through text and talk, is an example of the latter. This “mind management” technique is what van Dijk considers the “crucial point” to be studied in CDA (1993b: 254). Another link that, according to van Dijk, has been neglected and can be attended to by CDA is the relation between *macro* issues such as social inequality and political dominance and *micro* issues such as an individual text and talk (1993b, 2002). He draws a link between “the socially shared political representations” (macro issues) and “the individual uniqueness” (micro issues) through political cognition (van Dijk 2002: 203). For instance, “a biased text about immigrants may derive from personal beliefs about immigrants, and these beliefs in turn may be related to the shared racist attitudes or ideologies of a larger group” (van Dijk 2002: 203–204). These observations are further developed in his analysis of racism and immigrant discrimination in parliamentary speeches (van Dijk 1993b, 2002). For example:

In the past 25 years, we have allowed hundreds of thousands of immigrants into this small island so that we now have ethnic minorities of several million people and in some cases, as we all know, their birth rate far exceeds that of the indigenous population. <...> What is the future of our country to be in another 25 years, even if all immigration is stopped tomorrow? What will be the effect on our religion, morals, customs, habits and so on? Already there have been some dangerous eruptions from parts of the Muslim community <...>. (van Dijk 2002: 205; taken from Hansard 1989)

In van Dijk's interpretation, the MP, Sir John Stokes, first of all, expresses his personal mental model, i.e. his personal beliefs, of the current immigration situation in England, spicing it up with his private judgment, as for instance, 'dangerous eruptions'. Simultaneously, however, the speaker weds his attitude with common knowledge through the words 'as we all know'. Therefore, because of the power given to him by his status of an MP and the power of the institution he is talking from, his personal model can be spread via the media and in the long run turn into a 'shared knowledge' for the general public (van Dijk 2002). Moreover, since parliamentary discourse is a means to communicate the highest level of power in a democratic country, it passes down ideologies to the lower level institutions, as, for instance, the police or schools.

14.3.4. Analysis of societal inadequacies: racism and gender discrimination

Discussing the role of CDA in tackling the yet acute societal problem of racism, van Dijk (2008) claims that while discourse being just words cannot break people's bones as the use of sticks and stones, nevertheless, texts and talk play a pivotal role in advancing contemporary racism. Consider an example from Wodak and Reisigl (2001):

Profil: You will not recommend Karl-Heinz Grasser to give in?

Haider: We never thought differently and will continue to do so. The indignation, of course only comes from the side of those like the Carinthian guild masters for construction, a socialist who makes money out of cheap labor from Slovenia and Croatia. And if, today, one goes by one of Hans Peter Haselsteinger's "Illbau" building sites, and there, the foreigners, up to black African, cut and carry bricks, then the Austrian construction worker really thinks something. Then one must understand if there are emotions. (Wodak and Reisigl 2001: 387)

From CDA perspective, Heider's answer is soaked with the speaker's, and possibly those who he represents, negative attitude towards immigrants, especially immigrant workers, in Austrian labour market. Wodak and Reisigl call Heider's speech as "a blatant racist utterance" (2001: 388). The racist ideology first of all permeates through the reference to the workers from Africa as 'black Africans' by explicit mentioning of their skin colour. In the meantime, the Austrian workers are referred to as 'the Austrian construction workers' where the focus is placed on the group's professional skills, namely being skilled workers in the field of construction. As a result, the two reference types put the two groups of construction workers in a sharp juxtaposition to each other – aiming to present the former in the negative light and placing the latter in the positive light. At the end of the extract one comes across the expression of racist attitudes in a more implicit way. The words 'one must understand if there are emotions' are aimed at setting a commonsensical basis for and justification of racial and ethnical discrimination. The urge to 'understand' the 'emotions', which under the presumption of relevance belong to the Austrian workers, relates to such discriminatory discourse and prejudice as foreigners taking away the working places from Austrians and Austrians having privileges in terms of employment (Bijeikienė 2008).

Along the same lines, CDA has been applied to tackle the problems of gender-related inequality and discrimination. Dwelling upon the principles of feminist CDA, Lazar lays out its specific goal, namely the critique of discourses “which sustain a patriarchal social order: that is, men as a social group and disadvantage, exclude and disempower women as a social group” (Lazar 2005a: 5). In comparison to the rise of feminism in the 1970s, gender based discrimination has changed with regard to its scope and its more subtle and implicit forms of manifestation (cf. Lazar 2005b; Wodak 2005). Lazar (2005b) demonstrates this manifestation through her analysis of the construction of parenthood in Singaporean advertising campaign intended to boost Singapore’s birth rate (cf. Bijeikienė 2008). Consider four examples which demonstrate the divergence in the construction of images of a working mother (examples 1 and 2) and a working father (examples 3 and 4):

Example 1.

I’m really excited about parenthood, but I also love my job.

Example 2.

One of my major concerns right now is balancing family and career. But I have friends who have shown me that it can be done.

Example 3.

Family life helps.

Example 4.

Even though my work takes me away, when it comes to joy and dreams, my children are the key (Lazar 2005b: 154–157).

The conjunction ‘but’ in the utterance of example 1 carries a presupposition of contrast showing that for women the two spheres of parenthood and professional career are still seen as co-existing in conflict. In the same way, the word ‘balancing’ in the utterance of example 2 demonstrates a constant urge for a woman to strive for a balance between her two spheres of responsibility – family and career – and in doing so to live in the atmosphere of constant struggle and tension. By contrast, male utterances demonstrate a generally more relaxed and comfortable situation of working fathers with positively connoted words such as ‘joy’ and ‘dreams’ ascribed to their relation to family life and childcare.

14.3.5. CDA in practice: textual analysis

The section “Example Research: Critical Discourse Analysis” published online by the University of Sheffield presents how CDA can be applied in the analysis of a media article taken from *The Guardian*.

Clashes in Beirut after army kills members of anti-Assad group

Fighting ensues after Lebanese soldiers shoot Sunni cleric and another member of March 14 political alliance

Residents of a Beirut suburb fired heavy machine guns and rocket-propelled grenades at each other on Sunday, the latest violence to raise fears that Syria's turmoil is spilling over the border into its neighbour.

Earlier, army soldiers in northern Lebanon shot dead two members of a political party in an alliance against Syrian president Bashar al-Assad.

Many Sunni Muslims in Lebanon's north sympathise with Syria's Sunni-led uprising against Assad and say the Lebanese army is taking orders from Damascus.

Syrian troops were garrisoned in Lebanon, a tiny country still recovering from its own 15-year civil war, until 2005. "Nine people were wounded in the Tariq al-Jadeide district in southern Beirut," a security source said.

A witness said that two Sunni groups were fighting each other, but it was not clear how the violence had started.

Earlier on Sunday, Sheikh Ahmed Abdul Wahid, a Sunni cleric, and Muhammed Hussein Miraib, both members of the Lebanon-based March 14 political alliance, were shot in their car as they sped through an army checkpoint without stopping.

Residents of the northern region of Akkar blocked off roads and burned tyres to protest against the deaths. The main coastal highway and roads in the capital Beirut were also blocked by enraged residents.

The army released a statement confirming the deaths but not giving any information on who was responsible or what led up to the shootings.

"The leadership of the army expresses deep regret for the death of the two victims ... it will immediately form an investigative committee comprised of senior officers and military police under the relevant court," the statement said.

Prime minister Najib Mikati tried to quell the growing tensions. "The government is determined to continue to shoulder its national responsibilities amid this critical period in Lebanon and the region, and it will take all measures necessary to preserve civil peace," he said in a statement. Some troops pulled out of Akkar to prevent tensions from escalating, security sources and residents said.

Khaled Daher, a member of parliament from the Future Movement party, part of the March 14 alliance, said the two men had been assassinated.

"If shots were fired at the tyres, we would say there was a mistake. But we consider this a direct targeting from the army. Frankly, we do not want to see the army here because it works at the service of the Syrian regime," he said.

The application of CDA revealed the following generalizations with regard to the delivery of the message of the given text. At the micro level of analysis:

- Regular use of indirect quotations from vague sources such as ‘a witness’ or ‘a security source’;
- A tendency to use collective nouns, rather than naming the individuals responsible;
- Two instances of active verbs ‘kills’ and ‘shoot’, rather than the sentences being put in the passive. This could be to emphasise who was responsible;
- Evidence of pre and post modification, for example ‘shot **dead**’ and ‘**Lebanese** soldiers’ for emphasis.

(Example Research: Critical Discourse Analysis in *All About Linguistics*)

At the meso level of analysis the important aspect for CDA is the fact that the article is taken from *The Guardian*, which has a centre-left political affiliation. At the macro level of analysis it was concluded that the Conflict in the Middle East is a prominent topic in world news and a report where two nations are uniting against Assad, rather than fighting each other, would be of social relevance at this time.

14.4. Comprehension activities

Activity 2. Discuss the following questions in pairs or small groups. To come up with a more in depth analysis you might have to study more theoretical resources (follow the reference list provided below).

1. *How can CDA be defined as a perspective of sociolinguistic analysis?*
2. *How can Fairclough’s commonsensical functioning of ideology be explained? What is its relevance to sociolinguistic analysis?*
3. *What forms does racial, ethnical, gender or sexuality based discrimination gain nowadays? What other forms of discrimination is the contemporary society ill with? Can they be cured with the help of CDA?*

Activity 3. Read an example of a linguistic exchange taken from Fairclough (1989: 44) and analyse it from the perspective of CDA. Consider such aspects as power relations, formality, genre, linguistic strategies, etc.

(1) Doctor: <...> the first of the infants – now what I want you to do is to make a basic. neo- natal examination just as Dr Mathews has to do as soon as a baby arrives in the ward. all right so you are actually going to get your hands on the infant. and look at the key points and demonstrate them to the group as you’re doing it will you do that for me please. off you go

(2) Student: well first of all I’m going to [

(3) Doctor: [first . before you do that is do you wash your hands isn’t it <...>

(4) Student: just going to remove this.

(5) Doctor: very good . it’s putting it back that’s the problem isn’t it eh.

.....

(12) Student: erm we’ll see if he’ll respond to.

(13) Doctor: now look did we not look at a baby with a head problem yesterday.

Activity 4. Find a media text on topical social issues (possibly related to racism, gender- or sexuality-related discrimination or other) and analyse it or its excerpt from the perspective of CDA using the example presented above. Share your work with a colleague for a peer feedback.

14.5. Glossary

Discourse – the use of language conceived as social practice (Bennett).

Discourse practice – the production, distribution and consumption of a text (Bennett).

Genre – the use of language associated with a particular social activity (Bennett).

Ideology – (1) a form of social cognition, political philosophy or ideas shared by the members of a group, class, or other social formation (van Dijk 2008); (2) a representation of aspects of the world which can be shown to contribute to establishing, sustaining or changing relations of power, domination and exploitation (Fairclough 2003).

Order of discourse – totality of discursive practices of an institution and relations among them (Bennett).

Power – (1) as social power, a property of the relationship between groups, classes or other social formation, or between persons as social members and control of one group over other groups; (2) as interactional power, a dynamic property exercised and expressed through discourse and through different access to various genres, contents and discursive styles (van Dijk 2008).

Racism – a complex societal system of ethnically or racially based domination and the ensuing discrimination and inequality, not reduced to merely racist ideology (van Dijk 2008: 103)

Unequal encounters – cases of interaction between people from less powerful and more powerful groups in terms of social hierarchy, e.g., a police officer and a suspect, a teacher and a student, etc.

14.6. References

- Juliet Bennett. <http://www.julietbennett.com/2011/03/03/critical-discourse-analysis/#.UhloZ9JYPfl>.
- Vilma Bijeikienė, 2008. "Critical Discourse Analysis: an Overview and Appraisal", *Respectus Philologicus* 13(18), 104–113.
- Penelope Brown & Stephen C. Levinson, 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge UP.
- Roger Brown & Albert Gilman, 1960. "The Pronouns of Power and Solidarity", in *Style in Language*, ed. Thomas A. Sebeok, New York: MIT, 253–276.
- Paul Chilton, 2005. "Missing Links in Mainstream CDA: Modules, Blends and the Critical Instinct", in *New Agenda in (Critical) Discourse Analysis: Theory, Methodology and Interdisciplinarity*, Amsterdam: John Benjamins, 19–51.
- "Example Research: Critical Discourse Analysis", in *All about Linguistics*. The University of Sheffield, <https://sites.google.com/a/sheffield.ac.uk/all-about-linguistics/branches/discourse-analysis/example-research>.

- Norman Fairclough, 1989. *Language and Power*. London: Longman.
- Norman Fairclough, 2001. *Language and Power*. 2nd ed. Harlow: Longman.
- Norman Fairclough, 2003. *Analysing Discourse: Textual Analysis for Social Research*. London: Routledge.
- Norman Fairclough & Ruth Wodak. 1997. "Critical discourse analysis", in *Discourse as Social Interaction*. Vol. 2: *Discourse Studies: A Multidisciplinary Introduction*, ed. Teun A. van Dijk, London: Sage, 258–284.
- Roger Fowler and Gunther Kress, 1979. "Critical linguistics", in *Language and Control*, eds. Roger Fowler et al., London: Routledge & Kegan Paul, 185–213.
- Janet Holmes, 1995. *Women, Men and Politeness*. London: Longman.
- Michelle M. Lazar, 2005a. "Politicizing Gender in Discourse: Feminist Critical Discourse Analysis as Political Perspective and Praxis", in *Feminist Critical Discourse Analysis: Gender, Power and Ideology in Discourse*, ed. Michelle M. Lazar, Houndmills: Palgrave Macmillan, 1–28.
- Michelle M. Lazar, 2005b. "Performing State Fatherhood: The Remaking of Hegemony", in: *Feminist Critical Discourse Analysis: Gender, Power and Ideology in Discourse*, ed. Michelle M. Lazar, Houndmills: Palgrave Macmillan, 139–165.
- Alan Partington, 2003. *The Linguistics of Political Argument: The Spin-doctor and the Wolf-pack at the White House*. London: Routledge.
- Deborah Tannen, 1994a. *Gender and Discourse*. New York: Oxford UP.
- Teun A. van Dijk, 1993a. "Editor's foreword to critical discourse analysis", *Discourse & Society* 4(2), 131–132.
- Teun A. van Dijk, 1993b. "Principles of critical discourse analysis", *Discourse & Society* 4(2), 249–283.
- Teun A. van Dijk, 1997. "What is political discourse analysis?", in *Political Linguistics*, eds. Jan Blommaert & Chris Bulcaen, Amsterdam: John Benjamins, 11–52.
- Teun A. van Dijk, 2002. "Political discourse and political cognition", in *Politics as Text and Talk: Analytic Approaches to Political Discourse*, eds. Paul A. Chilton and Christina Schäffner, Amsterdam: John Benjamins, 203–237.
- Teun A. van Dijk, 2008. *Discourse and Power*. Houndmills: Palgrave Macmillan.
- Richard J. Watts, 2003. *Politeness*. Cambridge: Cambridge UP.
- Ruth Wodak, 1995. "Critical Linguistics and Critical Discourse Analysis", in *Handbook of Pragmatics: Manual*, eds. Jef Verschueren et al., Amsterdam: John Benjamins, 204–210.
- Ruth Wodak, 2001. "What CDA is about – a summary of its history, important concepts and its developments", in *Methods of Critical Discourse Analysis*, eds. Ruth Wodak & Michael Meyer, London: Sage, 1–13.
- Ruth Wodak & Martin Reisigl, 2001. "Discourse and Racism", in *The Handbook of Discourse Analysis*, ed. Deborah Schiffrin, Deborah Tannen & Heidi E. Hamilton, Malden: Blackwell, 372–397.
- Ruth Wodak, 2005. "Gender Mainstreaming and the European Union: Interdisciplinary, Gender Studies and CDA", in *Feminist Critical Discourse Analysis: Gender,*

Power and Ideology in Discourse, ed. Michelle M. Lazar, Houndmills: Palgrave Macmillan, 90–113.

- *Text based on the unpublished PhD dissertation*: Vilma Bijekienė, Vilma, 2007. *Gender in Political Discourse: Indirectness in the British Parliamentary Debates*. Bergen University and Vilma Bijekienė, 2008. “Critical Discourse Analysis: an Overview and Appraisal”, *Respectus Philologicus*. 13(18), 104–113.

- Daiva Aliūkaitė & Erika Merkytė, 2008. "Sociolingvistinis tarminių kodų taisyklingumo tyrimas", *Kalbos kultūra* 82, 259–274.
- Jo Angouri, 2010. "Quantitative, Qualitative or Both? Combining Methods in Linguistic Research", in *Research Methods in Linguistics*, ed. by Lia Litosseliti, London, New York: Continuum, 29–45.
- John L. Austin, 1962. *How to Do Things with Words: The William James Lectures delivered at Harvard University in 1955*. Oxford: Clarendon.
- Benjamin Bailey, 2008. "Interactional sociolinguistics", retrieved from: http://works.bepress.com/cgi/viewcontent.cgi?article=1058&context=benjamin_bailey.
- Paul Baker, 2010. "Corpus Methods in Linguistics", in *Research Methods in Linguistics*, ed. by Lia Litosseliti, London, New York: Continuum, 93–113.
- Robert Bayley, 2013. "Variationist Sociolinguistics", in *The Oxford Handbook of Sociolinguistics*, ed. by Robert Bayley, Richard Cameron, and Ceil Lucas, Oxford: Oxford University Press [online version].
- Juliet Bennett. <http://www.julietbennett.com/2011/03/03/critical-discourse-analysis/#.UhloZ9JYPfl>.
- Vilma Bijeikienė, 2007. *Gender in Political Discourse: Indirectness in the British Parliamentary Debates*. Bergen University.
- Vilma Bijeikienė, 2008. "Critical Discourse Analysis: an Overview and Appraisal", *Respectus Philologicus* 13(18), 104–113.
- Diane Blakemore, 1995. "Relevance Theory", in *Handbook of Pragmatics: Manual*, eds. Jef Verschueren et al., Amsterdam: John Benjamins, 443–452.
- Diane Blakemore, 2001. "Discourse and Relevance Theory", in *The Handbook of Discourse Analysis*, ed. Deborah Schiffrin, Deborah Tannen and Heidi E. Hamilton, Malden: Blackwell, 398–415.
- Penelope Brown & Stephen C. Levinson, 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge UP.
- Roger Brown & Albert Gilman, 1960. "The Pronouns of Power and Solidarity", in *Style in Language*, ed. Thomas A. Sebeok, New York: MIT, 253–276.
- Deborah Cameron, 1998. "'Is there any ketchup, Vera?': gender, power and pragmatics", *Discourse & Society* 9 (4), 437–455.
- Deborah Cameron, Elizabeth Frazer, Penelope Harvey, Ben Rampton & Kay Richardson, [1992, reprinted in 1997]. "Ethics, Advocacy, and Empowerment in Researching Language", in *Sociolinguistics: a Reader and Coursebook*, ed. by Nikolas Coupland and Adam Jaworski, New York: Palgrave, 145–162.

- Robyn Carston, 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*, Oxford: Blackwell.
- Paul Chilton, 2005. "Missing Links in Mainstream CDA: Modules, Blends and the Critical Instinct", in *New Agenda in (Critical) Discourse Analysis: Theory, Methodology and Interdisciplinarity*, Amsterdam: John Benjamins, 19–51.
- Jinny K. Choi, 2005. "Bilingualism in Paraguay: Forty Years after Rubin's Study", *Journal of Multilingual and Multicultural Development* 26(3), 233–248.
- Steven Clayman & John Heritage, 2002. *The News Interview: Journalists and Public Figures on the Air*, Cambridge: Cambridge UP.
- Eva Codó, 2008. "Interviews and Questionnaires", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 158–176.
- Nikolas Coupland & Adam Jaworski, 1997. "Methods for Studying Language in Society", in *Sociolinguistics: a Reader and Coursebook*, ed. by Nikolas Coupland and Adam Jaworski, New York: Palgrave, 69–74.
- Xu Daming, Wang Xiaomei & Li Wei, 2008. "Social Network Analysis", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 263–274.
- Katie Drager (forthcoming). "Experimental Methods in Sociolinguistics: Matched guise and identification tasks," in *Research Methods in Sociolinguistics: A practical guide*, ed. by Janet Holmes and Kirk Hazen, Hoboken, NJ: Wiley-Blackwell.
- Paul Drew & Marja Leena Sorjonen, 1997. "Institutional Dialogue", in *Discourse as Social Interaction. Vol. 2*, ed. by T. A. Van Dijk, London: Sage, 92–118.
- Paul Drew, 2005. "Foreword: applied linguistics and conversation analysis", in *Applying Conversation Analysis*, ed. by Keith Richards and Peter Seedhouse, Houndmills: Palgrave Macmillan, xiv–xx.
- Penelope Eckert & Sally McConnell-Ginet, 2003. *Language and Gender*, Cambridge, New York, Melbourne, Madrid, Cape Town: Cambridge University Press.
- Penelope Eckert, 1998. *Jocks and Burnouts: Social Categories and Identity in the High School*, New York, London: Teacher's College Press.
- "Example Research: Critical Discourse Analysis", in *All about Linguistics*. The University of Sheffield. <https://sites.google.com/a/sheffield.ac.uk/all-about-linguistics/branches/discourse-analysis/example-research>.
- Norman Fairclough & Ruth Wodak, 1997. "Critical discourse analysis", in *Discourse as Social Interaction. Vol. 2: Discourse Studies: A Multidisciplinary Introduction*, ed. Teun A. van Dijk, London: Sage, 258–284.
- Norman Fairclough, 1989. *Language and Power*. London: Longman.
- Norman Fairclough, 2001. *Language and Power*. 2nd ed. Harlow: Longman.
- Norman Fairclough, 2003. *Analysing Discourse: Textual Analysis for Social Research*. London: Routledge.
- Roger Fowler and Gunther Kress, 1979. "Critical linguistics", in *Language and Control*, eds. Roger Fowler et al., London: Routledge & Kegan Paul, 185–213.
- Harold Garfinkel, 1967. *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.

- Harold Garfinkel, 1974. "On the origins of the term 'ethnomethodology'", in *Ethnomethodology*, ed. by Roy Turner, Harmondsworth: Penguin.
- Erving Goffman, 1967. *Interaction Ritual: Essays in Face-to-Face Behavior*. New York: Pantheon Books.
- Mitchell Green, 2007. <http://plato.stanford.edu/entries/speech-acts/>.
- Paul H. Grice, 1975. "Logic and Conversation", in *Syntax and Semantics. Vol. 3: Speech Acts*, eds. Peter Cole and Jerry L. Morgan, New York: Academic Press, 41–58.
- Jack Grieve, 2012. "Sociolinguistics: Quantitative Methods", in *The Encyclopedia of Applied Linguistics*, ed. by Carol A. Chapelle, Wiley-Blackwell [retrieved from: <https://lirias.kuleuven.be/bitstream/123456789/329022/1/10-0455sociolinguistics2.pdf>].
- John J. Gumperz & Jenny Cook-Gumperz, 1982. "Introduction: language and the communication of social identity," in *Language and Social Identity*, ed. John J. Gumperz, Cambridge: Cambridge UP, 1–21.
- John J. Gumperz, 1982. *Discourse Strategies*. Cambridge: Cambridge UP.
- John J. Gumperz, 2001. "Interactional sociolinguistics: a personal perspective", in *The Handbook of Discourse Analysis*, eds. Deborah Schiffrin, Deborah Tannen and Heidi E. Hamilton, Malden: Blackwell, 215–228.
- Monica Heller, 2001. "Discourse and interaction", in *The Handbook of Discourse Analysis*, eds. Deborah Tannen, Deborah Schiffrin and Heidi E. Hamilton, Malden: Blackwell, 215–228.
- Monica Heller, 2006. *Linguistic Minorities and Modernity: A Sociolinguistic Ethnography*, London: Continuum.
- Monica Heller, 2008. "Doing Ethnography", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 249–262.
- Janet Holmes & Maria Stubbe, 2003. "'Feminine' Workplaces: Stereotypes and Reality", in *The Handbook of Language and Gender*, eds. Janet Holmes and Miriam Meyerhoff, Malden: Blackwell, 573–599.
- Janet Holmes, 1995. *Women, Men and Politeness*. London: Longman.
- Janet Holmes, 1998. "Complimenting- a Positive Politeness Strategy", in *Language and Gender: A Reader*, ed. Jennifer Coates, Oxford: Blackwell, 100–120.
- Janet Holmes, 2005. "Power and Discourse at Work: Is Gender Relevant?", in *Feminist Critical Discourse Analysis: Gender, Power and Ideology in Discourse*, ed. Michelle M. Lazar, Houndmills: Palgrave Macmillan, 31–60.
- Zhu Hua & Annabelle David, 2008. "Study Design: Cross-sectional, Longitudinal, Case, and Group", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 88–107.
- Ian Hutchby & Paul Drew, 1995. "Conversation Analysis", in *Handbook of Pragmatics: Manual*, ed. by Jef Verschueren et al., 182–189. Amsterdam: John Benjamins.

- Dell H. Hymes, 1962. "The ethnography of speaking," in *Anthropology in Human Behaviour*, eds T. Gladwin and W. C. Sturtevant, D. C. Washington: Anthropological Society, 13–53.
- Cornelia Ilie, 1994. *What Else Can I Tell You?: A Pragmatic Study of English Rhetorical Questions as Discursive and Argumentative Acts*. Stockholm: Almqvist and Wiksell International.
- Anthea Irwin, 2010. "Social Constructionism", in *The Sage Handbook of Sociolinguistics*, ed. by Ruth Wodak, Barbara Johnstone and Peter Kerswill, London: Sage Publications Ltd., 100–112.
- Rika Ito & Sali Tagliamonte. 2003. "Well weird, right dodgy, very strange, really cool: layering and recycling in English intensifiers", *Language in Society* 32, 257–279.
- Joanna Jaworowska. *Speech Act Theory*, retrieved from: http://instructional1.calstatela.edu/lkamhis/tesl565_sp04/troy/spchact.htm.
- Barbara Johnstone, 2000. *Qualitative Methods in Sociolinguistics*, New York, Oxford: Oxford University Press.
- Aimee C. Knupsky & Natalie M. Nagy-Bell, 2011. "Dear Professor: The Influence of Recipient Sex and Status on Personalization and Politeness in E-Mail", *Journal of Language and Social Psychology* 30 (1), 103–113.
- William Labov, [1972, reprinted in 2010]: *The Social Motivation of a Sound Change*, in *The Routledge Sociolinguistics Reader*, ed. by Miriam Meyerhoff & Erik Schleeff, New York: Routledge, 292–322.
- William Labov, [1972, reprinted in 1997]. "The Social Stratification of (r) in New York City Department Stores", in *Sociolinguistics: a Reader and Coursebook*, ed. by Nikolas Coupland and Adam Jaworski, New York: Palgrave, 168–178.
- William Labov, 1984. "Field Methods of the Project on Linguistic Change and Variation", in *Language in Use*, ed. by J. Baugh & J. Sherzer, Englewood Cliffs: Prentice Hall, 28–53.
- William Labov, 2001. *Principles of Linguistic Change. Vol. 2: Social Factors*. Malden, Oxford: Blackwell Publishers.
- Robin Tolmach Lakoff, 1995. "Conversational Logic", in *Handbook of Pragmatics: Manual*, eds Jef Verschueren et al., Amsterdam: John Benjamins, 190–198.
- Michelle M. Lazar, 2005a. "Politicizing Gender in Discourse: Feminist Critical Discourse Analysis as Political Perspective and Praxis", in *Feminist Critical Discourse Analysis: Gender, Power and Ideology in Discourse*, ed. Michelle M. Lazar, Houndmills: Palgrave Macmillan, 1–28.
- Michelle M. Lazar, 2005b. "Performing State Fatherhood: The Remaking of Hegemony", in: *Feminist Critical Discourse Analysis: Gender, Power and Ideology in Discourse*, ed. Michelle M. Lazar, Houndmills: Palgrave Macmillan, 139–165.
- Stephen C. Levinson, 1983. *Pragmatics*. Cambridge: CUP.
- Erez Levon, 2010. "Organizing and Processing Your Data: The Nuts and Bolts of Quantitative Analyses", in *Research Methods in Linguistics*, ed. by Lia Litosseliti, London, New York: Continuum, 68–92.
- Miriam A. Locher, 2004. *Power and Politeness in Action: Disagreements in Oral Communication*. Berlin: Mouton de Gruyter.

- Clare MacMartin, Linda A. Wood & Rolf , O. Kroger, 2001. "Facework", in *The New Handbook of Language and Social Psychology*, eds. W. Peter Robinson and Howard Giles, Chichester: John Wiley and Sons, 221-237.
- Daniel N. Maltz & Ruth A. Borker, [1982] 1998. "A Cultural Approach to Male- Female Miscommunication", in *Language and Gender: A Reader*, ed. Jennifer Coates, Oxford: Blackwell, 417-434.
- Anna Margetts & Andrew Margetts, 2012. "Audio and Video Recording Techniques for Linguistic Research", in *The Oxford Handbook of Linguistic Fieldwork*, ed. Nicholas Thieberger, Oxford: Oxford University Press, [online version].
- Norma Mendoza-Denton, 2008. *Homegirls: Language and Cultural Practice among Latina Youth Gangs*, Malden, Oxford, Carlton: Blackwell Publishing.
- Jacob L. Mey, 2001. *Pragmatics: An Introduction*. 2nd ed. Oxford: Blackwell.
- Miriam Meyerhoff, 2006. *Introducing Sociolinguistics*, Abingdon, New York: Routledge.
- Miriam Meyerhoff, Chie Adachi, Golnaz Nanbakhsh, Anna Strylcharz, 2012. "Sociolinguistic Fieldwork", in *The Oxford Handbook of Linguistic Fieldwork*, ed. Nicholas Thieberger, Oxford: Oxford University Press [online version].
- James Milroy, 1992. *Linguistic Variation and Change: On the Historical Sociolinguistics of English*, Oxford: Blackwell Publishers.
- James Milroy & Lesley Milroy, 1997. "Network Structure and Linguistic Change", in *Sociolinguistics: a Reader and Coursebook*, ed. by Nikolas Coupland and Adam Jaworski, New York: Palgrave, 199-211.
- Lesley Milroy & Matthew Gordon, 2003. *Sociolinguistics: Method and Interpretation*, Malden, Oxford, Melbourne, Berlin: Blackwell Publishing.
- Lesley Milroy, 1987. *Language and Social Networks*, Oxford, New York: Blackwell.
- Melissa G. Moyer, 2008. "Research as Practice: Linking Theory, Method, and Data", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 18-31.
- Nancy Niedzielski, 1999. "The Effect of Social Information on the Perception of Sociolinguistic Variables", *Journal of Language and Social Psychology* 18(1), 62-85.
- Jacomine Nortier, 2008. "Types and Sources of Bilingual Data", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 35-52.
- Alan Partington, 2003. *The Linguistics of Political Argument: The Spin-doctor and the Wolf-pack at the White House*. London: Routledge.
- Mildred Patten, 2005. *Understanding research methods: An overview of the essentials*, Glensdale, CA: Pyrczak Publishing.
- Kenneth L. Pike, 1967. *Language in Relation to a Unified Theory of the Structure of Human Behaviour*, The Hague: Mouton.
- Anita Pomerantz & Barbara J. Fehr, 1997. "Conversation Analysis: An Approach to the Study of Social Action as Sense Making Practices", in *Discourse as Social Interaction*. Vol. 2, ed. by T. A. Van Dijk, London: Sage Publications, 64-91.
- Thomas Purnell, William Idsardi & John Baugh, 1999. "Perceptual and phonetic

experiments on American English dialect identification”, *Journal of Language and Social Psychology* 18, 10–30.

- Meilutė Ramonienė, 2010. “Įvadas”, in *Miestai ir kalbos. Kolektyvinė monografija*, mokslinė redaktorė Meilutė Ramonienė. Vilnius: Vilniaus universiteto leidykla, 9–24.
- Sebastian M. Rasinger, 2008. *Quantitative Research in Linguistics. An Introduction*, London, New York: Continuum.
- Sebastian M. Rasinger, 2010. “Quantitative Methods: Concepts, Frameworks, and Issues”, in *Research Methods in Linguistics*, ed. by Lia Litosseliti, London, New York: Continuum, 49–67.
- Harvey Sacks, 1984. “Notes on methodology”, in *Structures of Social Action: Studies in Conversation Analysis*, ed. by J. Maxwell Atkinson & John Heritage, Cambridge: Cambridge UP, 21–27.
- Harvey Sacks, 1987. “On the preference for agreement and contiguity in sequences in conversation”, in *Talk and Social Organization*, ed. by Graham Button and John R. E. Lee, Clevedon: Multilingual Matters, 54–69.
- Harvey Sacks, 1992. *Lectures on Conversation*, Vol. 2, ed. by Gail Jefferson, with an introduction by Emmanuel A. Schegloff, Oxford: Blackwell.
- Harvey Sacks, Emmanuel A. Schegloff & Gail Jefferson, 1978. “A Simplest Systematics for the Organization of Turn Taking for Conversation”, in *Studies in the Organization of Conversational Interaction*, ed. by Jim Schenkein, New York: Academic Press, 7–55.
- Gillian Sankoff & Hélène Blondeau, 2007. “Language Change across the Lifespan: /r/ in Montreal French”, *Language* 83, 560–588.
- Deborah Schiffrin, 1994. *Approaches to Discourse*, Oxford: Blackwell.
- Erik Schleeff & Miriam Meyerhoff, 2010. “Sociolinguistic methods for data collection and interpretation”, in *The Routledge Sociolinguistics Reader*, ed. by Miriam Meyerhoff & Erik Schleeff, London, New York: Routledge, 1–17.
- John R. Searle, 1969. *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge UP.
- John R. Searle, 1975. “Indirect Speech Acts”, in *Syntax and Semantics. Vol.3: Speech Acts*, eds. Peter Cole and Jerry L. Morgan, New York: Academic Press, 59–82.
- John R. Searle, 1998. “Indirect Speech Acts”, in *Pragmatics: Critical Concepts. Vol. 4: Presupposition, Implicature and Indirect Speech Acts*, ed. Asa Kasher, London: Routledge, 617–638.
- Peter Seedhouse, 2005. “Conversation Analysis as Research Methodology”, in *Applying Conversation Analysis*, ed. by Keith Richards & Peter Seedhouse, Houndmills: Palgrave Macmillan, 251–287.
- Dan Sperber & Deirdre Wilson, 1995. *Relevance: Communication and Cognition*. 2nd ed. Oxford: Blackwell.
- Jane Sunderland, 2010. “Research Questions in Linguistics”, in *Research Methods in Linguistics*, ed. by Lia Litosseliti, London, New York: Continuum, 9–28.
- Sali A. Tagliamonte, 2006. *Analysing Sociolinguistic Variation*, Cambridge, New York: Cambridge University Press [Kindle Edition].
- Deborah Tannen, 1990. *You Just Don’t Understand: Women and Men in Conversation*. New York: William Morrow and Company.

- Deborah Tannen, 1994a. *Gender and Discourse*. New York: Oxford UP.
- Deborah Tannen, 1994b. *Talking from 9 to 5. Women and Men at Work: Language, Sex and Power*. Lancaster: Virago.
- Deborah Tannen, 1998. *The Argument Culture: Changing the Way We Argue and Debate*. London: Virago.
- Natasha Tokowicz & Tessa Warren, 2008. "Quantification and Statistics", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 214–231.
- Karen Tracy, 1990. "The Many Faces of Facework", in *Handbook of Language and Social Psychology*, eds. Howard Giles and W. Peter Robinson, Chichester: John Wiley and Sons., 209–226.
- William M. Trochim, 2006. *The Research Methods Knowledge Base*, 2nd Edition. Internet WWW page, at URL: <<http://www.socialresearchmethods.net/kb/>>.
- Peter Trudgill, 1974. *The Social Differentiation of English in Norwich*. London, New York: Cambridge University Press.
- Peter Trudgill, 2003. *A Glossary of Sociolinguistics*. Oxford, New York: Oxford University Press.
- Maria Teresa Turell & Melissa G. Moyer, 2008. "Transcription", in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, ed. by Li Wei and Melissa G. Moyer, Malden, Oxford, Melbourne: Blackwell Publishing, 192–213.
- Teun A. van Dijk, 1993a. "Editor's foreword to critical discourse analysis", *Discourse & Society* 4(2), 131–132.
- Teun A. van Dijk, 1993b. "Principles of critical discourse analysis", *Discourse & Society* 4(2), 249–283.
- Teun A. van Dijk, 1997. "What is political discourse analysis?", in *Political Linguistics*, eds. Jan Blommaert & Chris Bulcaen, Amsterdam: John Benjamins, 11–52.
- Teun A. van Dijk, 2002. "Political discourse and political cognition", in *Politics as Text and Talk: Analytic Approaches to Political Discourse*, eds. Paul A. Chilton and Christina Schäffner, Amsterdam: John Benjamins, 203–237.
- Teun A. van Dijk, 2008. *Discourse and Power*. Houndmills: Palgrave Macmillan.
- Isabel Velázquez, 2013. "Mother's social network and family language maintenance", *Journal of Multilingual and Multicultural Development* 34 (2), 189–202.
- Jef Verschueren, 1999. *Understanding Pragmatics*. London: Arnold.
- Suzanne Ewans Wagner, 2012. "Age Grading in Sociolinguistic Theory", *Language and Linguistic Compass* 6/6, 371–382.
- Richard J. Watts, 2003. *Politeness*. Cambridge: Cambridge UP.
- Deirdre Wilson & Dan Sperber, 2004. "Relevance Theory", in *The Handbook of Pragmatics*, eds. Laurence R. Horn and Gregory Ward, Malden: Blackwell, 607–632.
- Ruth Wodak & Martin Reisigl, 2001. "Discourse and Racism", in *The Handbook of Discourse Analysis*, ed. Deborah Schiffrin, Deborah Tannen & Heidi E. Hamilton, Malden: Blackwell, 372–397.
- Ruth Wodak, 1995. "Critical Linguistics and Critical Discourse Analysis", in *Handbook of Pragmatics: Manual*, eds. Jef Verschueren et al., Amsterdam: John Benjamins, 204–210.

- Ruth Wodak, 2001. "What CDA is about – a summary of its history, important concepts and its developments", in *Methods of Critical Discourse Analysis*, eds. Ruth Wodak & Michael Meyer, London: Sage, 1–13.
- Ruth Wodak, 2005. "Gender Mainstreaming and the European Union: Interdisciplinary, Gender Studies and CDA", in *Feminist Critical Discourse Analysis: Gender, Power and Ideology in Discourse*, ed. Michelle M. Lazar, Houndmills: Palgrave Macmillan, 90–113.
- Ruth Wodak, Michał Krzyżanowski & Bernhard Forchtner, 2012. "The interplay of language ideologies and contextual cues in multilingual interactions: Language choice and code-switching in European Union institutions", *Language in Society* 41, 157–186.
- Ruth Wodak, Paul Kerswill & Barbara Johnstone, 2010. "Introduction", in *The Sage Handbook of Sociolinguistics*, ed. by Ruth Wodak, Barbara Johnstone and Peter Kerswill, London: Sage Publications Ltd., 1–8.
- Walt Wolfram, 2006. "Variation and Language, an Overview", in *Encyclopedia of Language and Linguistics*, 2nd Edition, ed. Keith Brown, Elsevier Ltd. [online version].
- Walt Wolfram, 2010. "Fieldwork Methods in Language Variation", in *The Sage Handbook of Sociolinguistics*, ed. by Ruth Wodak, Barbara Johnstone and Peter Kerswill, London: Sage Publications Ltd., 296–311.
- Alison Wray & Aileen Bloomer, 2006. *Projects in Linguistics. A Practical Guide to Researching Language*, London: Arnold Hodder.
- George Yule, 1996. *Pragmatics*. Oxford: Oxford UP.

ANSWER KEY

Chapter 1:

Activity 1.

Creative task – see the glossary.

Activity 2.

A. a); **B.** b); **C.** b); **D.** b)

Activity 3.

Creative task, involving group discussion.

Chapter 2:

Activity 1.

Creative task – see the glossary.

Activity 2.

Creative task.

Activity 3.

A) Suggested answer: Quantitative, since the research involves testing a hypothesis. Authors quantified selected words, used scoring system and Likert-type scales in order to evaluate e-mails' grammar and overall personalization. Results are presented in numbers, using means and deviations (i.e. concepts used in employed in statistics).

B) Suggested answer: Researchers wanted to prove (or disprove) the findings of the earlier studies which suggested that the sex of the sender and recipient affect e-mail personalization. The best way to test the interrelationship between different social and linguistic variables is to run a statistical analysis (that indicates probabilities, whether one variable has an affect toward another or no). Therefore, raising the hypothesis rather than a research question guided authors toward the quantitative research paradigm.

C) Creative task.

D) Creative task.

Chapter 3:

Activity 1.

Creative task – see the glossary.

Activity 2.

Suggested answers:

1) Advantages of cross-sectional design:

time and cost: can be conducted in a short period of time for low cost;
statistically valid: usually deal with a representative sample of data;
investigate multiple variables and multiple subjects;
easy to replicate.

2) Disadvantages of cross-sectional design:

they do not explore cause and effect relationships;

not suitable for longitudinal designs;

do not assess individual differences, therefore, more chances of error.

3) Limitations of Labov's 1966 study:

data was not recorded;

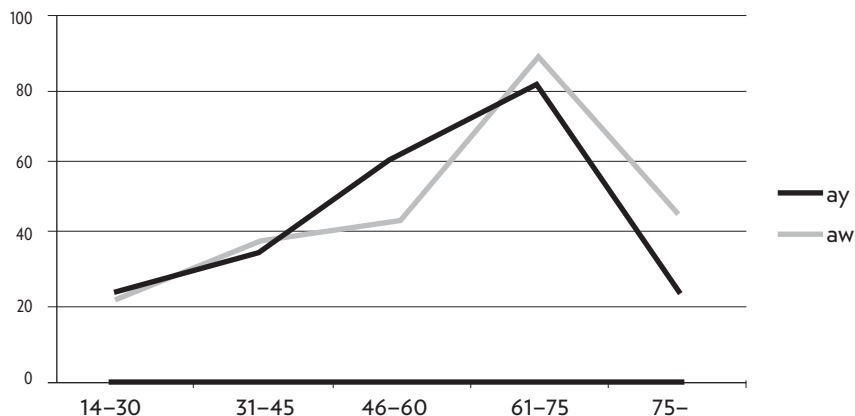
sample was accidental (whoever entered the store);

size of the sample;

problems with social categories, etc.

Activity 3.

Graph:



Activity 4.

Creative task.

Activity 5.

1. a); 2. a) ; 3. b); 4. b).

Chapter 4:

Activity 1.

Creative task – see the glossary.

Activity 2.

Creative task.

Activity 3.

Suggested answers:

- limited quantity of data;
- formality of style;
- not everyone's telephone number is usually listed in a phonebook, thus, some of the members of community are excluded and this affects representativeness of the population;
- unlisted telephones might correlate with person's socioeconomic status, in other words, lower and lower middle class people could not afford to have a telephone back in the 1970s;
- quality of the telephone signal;
- non-typical members of the community might be randomly selected and such data might reverse the full linguistic picture of the community;
- there is no gender and social class balance.

Activity 4.

Creative task.

Activity 5.

Suggested answer: In certain communities people of official status have authority in the eyes of others. It might be good for a researcher to get approval from such an authoritative figure in order to be trusted by the rest of the member of the community. Just don't forget to expand your social network out of the one introduced by these people.

Chapter 5:

Activity 1.

Creative task – see the glossary.

Activity 2.

Creative task.

Activity 3.

Creative task.

Activity 4.

Creative task.

Chapter 6:

Activity 1.

Creative task – see the glossary.

Activity 2.

- 1 – change is statistically significant, p is less than 0.05;
- 2 – change is statistically insignificant, p is more than 0.05;
- 3 – change is statistically significant, p is less than 0.05;
- 4 – change is statistically significant, p is less than 0.05;

5 – change is statistically insignificant, p is more than 0.05.

According to Choi (2005: 243), the most significant linguistic change can be noticed regarding the use of Guarani in the streets of Luque: in 2000 speakers were more in favor using both languages (code-switching), while in 1960 most of them preferred Guarani.

Activity 3.

- a) categorical;
- b) categorical;
- c) continuous;
- d) categorical;
- e) categorical;
- f) categorical;
- g) categorical;
- h) continuous;
- i) categorical.

Activity 4.

Creative task.

Chapter 7:

Activity 1.

Creative task – see the glossary.

Activity 2.

Suggested answers:

1. if a person has literacy (reading and writing) difficulties;
2. if a person is blind;
3. If a person does not speak the language in which the questionnaire is prepared;
4. for on-line questionnaires – computer literacy might also be an issue, especially if your target group is elderly population.

Activity 3.

Creative task.

Activity 4.

Creative task.

Activity 5.

Creative task.

Chapter 8:

Activity 1.

Creative task – see the glossary.

Activity 2.

Creative task.

Activity 3.

Creative task.

Activity 4.

Creative task.

Activity 5.

1. True; 2. True; 3. False.

Activity 6.

Creative task.

Chapter 9:**Activity 1.**

Creative task – see the glossary.

Activity 2.

Creative task.

Activity 3.**Possible answers:**

- 1) danger of becoming too much native, becoming a subject of the study;
- 2) being a woman in men's group and vice versa;
- 3) outsider can ask dumb-sounding questions (insider may not be able to do that) in order to retrieve needed information.

Activity 4.

Creative task.

Activity 5.

Creative task.

Activity 6.

Suggested answer: Protocol addresses formatting issues and emphasizes the marking of pauses, discourse markers, as well as the use of Spanish spellchecker. Transcribers do not have to mark all pauses and hesitations, thus, paralinguistic information is not at the core of the research. In a transcription pauses are transcribed using ..., overlapping speech is marked with [, laughter is indicated in the (brackets).

Chapter 10:**Activity 1.**

Creative task – see the glossary.

Activity 2.

Creative task.

Activity 3.

Creative task.

Activity 4.

Density – 60 %.

Activity 4.

Creative task.

Activity 5.

Multiplexity – 42.85 %.

Activity 6.

Creative task.

Activity 7.

Creative task.

Chapter 11:**Activity 1.**

Creative task – see the glossary.

Activity 2.

1. *F – in CA context is conceptualised not as a conversational achievement, not as a pre-determined fixed reality.*
2. *T*
3. *T*
4. *F – vice versa; interlocutors usually try to delay the utterance of dispreferred alternatives like the rejection of an offer.*
5. *F – while delivering a dispreferred alternative an interlocutor tends to be indirect or equivocal.*

Activity 3. Suggested answer:

First pair part dispreferred second pair part hedge

(a)

SALES CLERK: You're over 21, aren't you? *First pair part*

CUSTOMER: Sure. *preferred pair*

SALES CLERK: OK, here's your beer. *preferred pair*

(b)

SALES CLERK: You're over 21, aren't you? *First pair part*

CUSTOMER: Well, er, yes, my birthday was actually yesterday, and we're having a party tonight..

Aimed by the speaker as preferred, but delivered structurally as dispreferred

SALES CLERK: All right, may I see your ID? *dispreferred pair part*

Hedges -- well, er, yes, all right, may.

In the given exchange the preferred response of the customer is as given in exchange (a), namely, it is an affirmative straightforward answer to the question about age. The customer's turn being a preferred one is also confirmed by the sales clerk's positive response with regard to the customer's (presupposed) request for beer. In exchange (b) the customer's response is meant by the customer to be an affirmative answer but is structurally delivered as a dispreferred pair part as evidenced by:

- numerous hedges used to present his otherwise positive answers,
- too much of loosely relevant information (birthday, party).

As a result, the customer's answer casts doubts as to its credibility and the sales clerk produces a dispreferred pair part by declining the customer's (presupposed) request for beer.

Activity 4.

Creative task.

Chapter 12:

Activity 1.

Creative task – see the glossary.

Activity 2.

1 d; 2 e; 3 a; 4 c; 5 b.

Activity 3. Suggested answer:

1. Technical legal terms and legal jargon used in legal discourse, very difficult to understand for people outside the field.
2. A **request** to file the proof of claim observing the deadline, a **warning** not to be given another possibility to file the claim if the deadline is not observed. These are indirect speech acts, no speech act verbs are used, typical 'legalese' usage of 'shall' structure.
3. To sound sophisticated, to appear useful to the general public as the latter cannot easily read, interpret and produce such languages, etc.

Activity 4. Suggested answer:

Ann views the situation about her lukewarm food as a problem and aims to solve it by addressing the waitress. By uttering 'Is it ok that my chicken fillet is almost cold?', which has an interrogative structure, she produces an indirect speech act, namely an indirect request to solve her problem of cold food. Under presumption of relevance the solution would be to exchange the cold food with the new meal that is of adequate temperature. Nevertheless, the waitress makes use of Ann's indirectness and opts to interpret Ann's utterance as a direct speech act of placing a question, i.e., a request for information. In such a way, the waitress supplies Ann with an answer relevant to a general question, namely an answer on the no/yes continuum. The given example demonstrates miscommunication, which can be either intentional, or unintentional. Under the presumption of relevance, a waitress should take all necessary action to assure the proper supply of meals to the customers rather than just providing them with information. Therefore, the waitress' turn could be interpreted as her intentional misinterpretation of Ann's request.

Chapter 13:

Activity 1.

Creative task – see the glossary.

Activity 2.

1. (F) *The general interest of Interactional sociolinguistics in the study of language as a social activity is primarily realized by exploring the interrelation between linguistic communication and social context at large, while in CA context is treated as an interactional achievement within a particular conversation.*
2. (T)
3. (T)
4. (F) *Tracy draws a link between minimizing a negative impact on one's personal 'face' and presenting a challenge to another interlocutor: "a basic assumption is that one common way an individual saves face is by attacking an other" (1990: 216).*
5. (T)
6. (F) *Tannen observes that linguistic indirectness, generally associated with women, could also be used by men but "in different situations and in different ways" (1994b: 79), so that women are usually more indirect in giving requests, suggestions and criticism, while men prefer indirectness in "the expression of weakness, problems, and errors, and of emotions other than anger".*

Activity 3.

1. "I really like you."	Compliment	<i>The hearer's negative face.</i> The speaker expresses positive emotions towards the hearer which may involve an anticipation of a positive reaction by the hearer.
2. "I think I made a huge mistake."	Apology	<i>The speaker's positive face.</i> The speaker makes a statement about his/her own shortcomings, thereby 'damaging' his/her own positive self-image/face.
3. "Please give me that book."	Order	<i>The hearer's negative face.</i> The speaker expresses an anticipation of some future action of the hearer and thereby restricts his/her personal freedom
4. "I think your report was not concise enough."	Criticism	<i>The hearer's positive face.</i> The hearer's positive face is threatened because s/he is blamed for having done smth. badly.
5. "Thank you so much for your help."	Expressing thanks	<i>The speaker's negative face.</i> The speaker expresses thanks because he/she feels obliged to do so. His freedom of action is thus threatened in the moment of speaking.
6. "You're feeling sad because of your ex-boyfriend, aren't you?"	Expressing emotions	<i>The hearer's positive face.</i> The speaker addresses a topic which involves a state of emotional weakness on the part of the hearer. Could also be treated as a threat to the hearer's <i>negative</i> face as restricting her want to be left alone and not imposed on.

Activity 4.

Creative task.

Chapter 14:

Activity 1.

Creative task – see the glossary.

Activity 2.

Creative task.

Activity 3. Suggested answer

An example of unequal encounter.

In his analysis, Fairclough shows how the doctor exercises his social power by strictly controlling the student's contribution through various linguistic means. Firstly, the student's contribution is clearly defined by the doctor in the latter's opening turn (1); for instance 'you are actually going to get your hands on the infant' or 'look at the key points'. Secondly, the student's contribution is structured by the doctor through the latter's explicit indication when to start talking; for instance, 'off you go' at the end of (1). Thirdly, the doctor regulates the sequence of the student's actions. For instance, in (3) the doctor interrupts the student in order to remind the latter to wash his hands: 'before you do that is do you wash your hands isn't it'. Fourthly, the doctor openly expresses his evaluation of the student's contribution, as, for instance, 'very good' in (5). According to Fairclough, even though such evaluation is positive and encouraging, it is still a technique of control, and thus power, and would be understood as arrogant if placed to someone equal or more powerful. Finally, the most belittling effect is spotted by Fairclough in the doctor's frequent use of negative questions and negative tags; for instance, 'did we not look at a baby with a head problem yesterday' in (13) and 'that's the problem isn't it' in (5). Such grammatical structures carry a presupposition that the student does not know what he is obviously supposed to know.

Activity 4.

Creative task.

ISBN 978-9955-12-989-9 (Print)

ISBN 978-609-467-041-1 (Online)

Vilma Bijekienė, Aurelija Tamošiūnaitė

Quantitative and Qualitative Research Methods in Sociolinguistics

Editor Albina Strunga

Proofreader Agnė Bolytė

Cover designer Saulius Bajorinas

Graphic designer Kornelija Buožytė

Published by Vytautas Magnus University,
S. Daukanto g. 27, LT-44249 Kaunas

Printed by UAB "ARX Reklama",
Savanorių pr. 363a, LT-44249 Kaunas