

# A corpus-based analysis of language ideologies in Hungarian school metalanguage

Tamás Péter Szabó<sup>1</sup>

Research Institute for Linguistics, Hungarian Academy of Sciences / Hungary  
University of Jyväskylä / Finland

**Abstract** – The main goal of this paper is to present a recently built interview corpus called *Corpus of Hungarian School Metalanguage – Interview Corpus (CHSM-IC)* and its potential in language ideology studies. This corpus was compiled during a broad survey on Hungarian school metalanguage carried out in 2009 and was recently made available for a wider research community within the CESAR (Central and South-East European Resources) project.

The study investigates interactional routines used in metadiscourses on language use. Printed texts cited from prestigious handbooks and interview data from *CHSM-IC* are compared. Thus, widely used, culturally-inherited text fragments are detected and confronted with the interviewees' narratives on their own communicational experiences. A case study on the discourse marker *hát* ('so', 'well') illustrates that there is a conflict and often a controversy between language ideologies disseminated by the Hungarian school system and the linguistic self-representation in the interviewees' narratives.

Combining Language Ideology, Conversation Analysis, Discourse Analysis and Discursive Social Psychology frameworks, the paper presents a detailed description on the emergence of metadiscourses in a school setting. The paper concludes that metalinguistic utterances (e.g., answers on grammaticality, statements on linguistic accuracy, etc.) and observable, spontaneous (or semi-spontaneous) language use patterns are regularly not in accordance with each other.

**Keywords** – discourse marker, L1 education, language ideologies, standardist language culture

## 1. INTRODUCTION

The present paper summarizes the results of a survey based on a communication-oriented approach of metalanguage. The analysis of interview data taken from a recently built corpus describes the dynamics and interactional structure of school metalanguage, illustrating how language ideologies emerge in metadiscourses. The presented theoretical background and methodology can be widely applied in the analysis of educational communication.

The importance of a corpus-based investigation of Hungarian school metalanguage becomes clear considering the standardist nature of Hungarian language culture (cf. Milroy 2001). In Hungary, curricula used in formal education contain prescriptivist and descriptivist elements. This heterogeneous and often controversial design is broadly criticized

---

<sup>1</sup> I would like to express my gratitude to my anonymous reviewers, to Petteri Laihonon and Alejandro Alcaraz Sintes for their careful reading and useful comments.

by sociolinguists, such as Kontra (2006). As part of this critical discourse, extensive research on school metalanguage has been conducted (for references, see, among others, Csernicskó and Kontra 2008), but interview analysis remained marginal. For such an analysis, interview corpora are needed and the present corpus aims to fill this gap.

Hungarian attitude and ideology research tradition is basically normative. Papers often conclude that attitude A is false and to be avoided, while attitude B is to be disseminated in education. In contrast, the goal of the present survey is not to evaluate ideologies and attitudes, but to investigate them as they emerge in discourses. For the construction and the analysis of the corpus, the methods of Conversation Analysis (CA) and Discourse Analysis (DA) have been followed.

In the second part of this paper, a case study on discourse marker *hát* ('so', 'well') illuminates the benefits of CA and DA approaches. The frequent use of *hát* in a turn-initial position is mentioned as an example of erroneous talk in research interview data, even by teachers and their students. This conception belongs to the prescriptivist tradition. At the same time, *hát* is present in almost all of the interviewee's verbal production in a turn-initial position, regardless of whether a given speaker stigmatizes *hát* or not. That is, the usage of *hát* does not predict its legitimization at an ideological level, nor can the negative evaluation of *hát* foretell its absence in speakers' utterances. The analysis of corpus data suggests that ideologies describing school-related expectations of linguistic accuracy, and observable patterns of spontaneous (or semi-spontaneous) performance are regularly not in accordance with each other.

## 2. ON METALANGUAGE

### 2.1. Approaches to metalanguage

According to Van Leeuwen (2004), at least two groups of definitions of metalanguage can be identified. Definitions from the first group claim that metalanguage is a specific register of language use with a scientific nature. Thus, metalanguage is described as a representation of cognitive representations, a tool for making theories, or telementing<sup>2</sup> inner beliefs or views. In the second group, communication-oriented definitions like Laihonen's can be found. Laihonen (2008: 669) argues that "[f]rom an interactional point of view, talk about language is a part of conversational action, such as answering, defending, blaming, accusing and apologizing". Following this approach, the present paper counts metalanguage as a part of ordinary communication and uses the methods of CA and DA for a dynamic description of metadiscourse patterns. Such analyses are important, because language ideologies emerge during metadiscourses.

Language ideologies can be briefly defined as statements on language use (for a detailed overview of the notion, see Laihonen 2009). These are indispensable elements of a community's language culture and, thus, are part of the linguistic socialization of a speaker acting in the given community.

Metalinguistic utterances in everyday communication routine have their evident antecedents in a huge tradition of metadiscourses. This tradition offers questions and often ready-made answers reflecting everyday language use. Socialization in metalanguage is the exploration of this tradition: speakers are exposed to a high amount of metalanguage (see Berko Gleason 1992) and simultaneously they learn various ways of formulating statements on language. One specific scene of this socialization is formal education.

In schools, several methods are used in order to shape students' metalinguistic performance. Some ideologies are rejected, while others are considered desirable by teachers, who are very influential actors in a formal educational setting. They are positioned as more competent speakers (role models), primary knowers and discourse managers (Lee 2007). Students meet a variety of metalinguistic practice during their training years and they learn how to construct language ideologies by explicit linguistic explanations, evaluations or other- and self-repair (repair often being an implicit way of building ideologies; cf. Laihonen 2008, 2009<sup>3</sup>). Students can assimilate to or differ from practices used in schools in their own communication.

Following a communication-oriented definition of metalanguage, some notes are necessary for the aims of metalanguage studies. Research on attitudes or ideologies is often legitimized by its presupposed usability in changing social structures and maladaptive behavioral patterns (among others, see Bohner 2001). For example, a study on school metalanguage might be considered capable of changing some false practices by teachers. But this argumentation is plausible only if one assumes that the change of attitudes or ideologies can change behavior. However, such a hypothesis involves the oversimplification of the results of cognitive psychology. In this oversimplified approach, reality is the input of cognition, while action is the output: one can recognize something in his or her environment, and then this recognition can be the basis of his or her decisions on behaving in a certain way. Mainstream cognitive

<sup>2</sup> The ideology of telementation is described by Coulter (2005) as a conception which emphasizes the primacy of thoughts. From this point of view, thoughts and feelings readily exist before speaking, and speaking just gives form to them.

<sup>3</sup> Following Schegloff, Jefferson and Sacks (1977: 363), I will prefer the more general term 'repair' rather than 'correction' in data analysis, but when citing interview excerpts, I will use 'correction' because the words *javít*, *javítás* ('to correct', 'correction') were used in the transcribed texts.

psychology rejects this approach, evaluating it as a folk theory based on the Cartesian tradition (see Győri 2008). Győri (2008) argues that awareness is not an antecedent, but an outcome of behavior.

While folk cognitivist tradition explains behavior presupposing the following sequence ‘cerebral activity → behavior → awareness’, Discursive Social Psychology (DSP) does not investigate cerebral activity or mental processes (Potter and Edwards 2001, 2003). According to DSP, one learns to do something somehow, and then learns ideologies explaining, legitimizing or illegitimizing that behavior. These ideologies make behavior meaningful, but they do not govern acting. In a DSP description, ideologies are dynamic in nature and are never stable or finished: they are always maintained by people who construct, reconstruct or deconstruct them.

## 2.2. The role of metalanguage in a standardist community

Hungarian is a standard language culture (see Milroy 2001; Kontra 2006; Sándor 2006). Among others, it means that the cult of a privileged dialect called *standard* is observable. Efforts are continuously made to generate an idealized use of language. Besides, standardist movements aim for a uniform metalanguage as well. To reach this goal, various tools, such as dictionaries, prescriptivist handbooks of good usage and language etiquette (e.g., Grétsy and Kovalovszky 1980,1985) are widely available.

Standardization is a never-ending process, because language is basically heterogeneous. That is why a homogeneous language use, which would form the basis of a standard, cannot exist nor be practised (Milroy 2001). Nevertheless, several communicative habits show that the so-called standard presents an ideal for the majority of Hungarian speakers, and the school system plays a central role in the construction of this position: in formal education curricula, prescriptivist ideologies are disseminated, and language awareness activities are cultivated.

The above-mentioned dominance of standardist ideologies does not mean that there is a total absence of contestants. Descriptivist and prescriptivist ideologies are taught simultaneously, in different contexts. For example, while topicalizing the variety and richness of Hungarian, several vernacular inflections and phonemes are discussed and highlighted. However, these very same features are often evaluated as erroneous and are the object of other-repair, even in Hungarian language and literature lessons. This dynamism shows that language ideologies are embedded in various situations, and that they are used for very different purposes. When the curricular goal is to disseminate information on language varieties in order to develop tolerance, the evaluation of vernacular (dialectal) features are positive, but basically different ideologies are constructed, often implicitly, to legitimize the continuous regulation of classroom discourse, on the grounds that dialectal forms are inadequate in a formal education context.

This duality uncovers a feature of standardist cultures: linguisticism, which can be overt or hidden. The notion of linguisticism has been defined by Skutnabb-Kangas and Phillipson (1989: 455) as “ideologies and structures which are used to legitimate, effectuate, and reproduce unequal division of power and resources (both material and non-material) between groups which are defined on the basis of language”.

Linguistic practices need legitimizing ideologies and these ideologies are formulated through metalanguage. As Milroy (1998: 64-65) summarized, “[i]n an age when discrimination in terms of race, colour, religion or gender is not publicly acceptable, the last bastion of overt social discrimination will continue to be a person’s use of language”.

Linguistic evaluations and repair may thus serve linguisticism, and the Hungarian school system, as an influential institution, routinely disseminates standardist ideologies and repair techniques. Students learn how to reproduce overt defensive statements against another person or against themselves (e.g., other- and self-repair, public negative evaluation of others’ speech or of one’s own speech), and, what is more, they simultaneously learn practices to hide the nature of this language culture. In other words, students might in practice learn, for example, how to repair dialectal forms in interaction, even though learning to appreciate dialects in principle. The positive evaluation of dialects as archaic and regionally/temporally valuable varieties seems to be a kind of tolerance, but restricting the usage of dialects to a really small segment of communication is a type of linguisticism. This practice limits the chances of dialect users to use their vernacular without being stigmatized.

## 3. THE CORPUS OF HUNGARIAN SCHOOL METALANGUAGE (CHSM)

### 3.1. A need for a corpus-based investigation

A systematic description and analysis of implicit and explicit language ideologies in education is needed to achieve an understanding of Hungarian metalinguistic socialization in schools. For this description to be empirically based, the project requires a corpus. The available significant interview corpora of contemporary spoken Hungarian were compiled for other purposes by the researchers at the Research Institute for Linguistics. *Budapest Sociolinguistics Interview* (BSI-2; cf. Kontra and Váradi 1997) was a complex survey on the attitudes and speech performance of the inhabitants of Budapest, carried out between 1987 and 1989. Another larger project at the same institution is *BEA – A*

*Hungarian Spontaneous Speech Database*. Its data collection began in 2007. However, these materials do not include recordings from educational settings, therefore a new corpus was needed.

### 3.2. Data collection

Building the *CHSM* was part of the PhD project of the present author. This project targeted a complex, multiperspectivist investigation of school metalanguage. Thus, data collection followed three different methods in Hungarian elementary schools, vocational high schools and grammar schools, in grades 1-4, 7 and 11. In the Hungarian educational system, children of the same age can learn in different types of schools, so that even in a grammar school very young children can be found in grades 1-4 (aged 6-11), older children in grade 7 (aged 13-14) and in grade 11 (aged 17-19). The sampling does not represent the demographic average of Hungary, but fits the requirements of qualitative representativity (Sántha 2006): it is rather heterogeneous and represents both typical and extreme cases.

#### 3.2.1. Questionnaires

A questionnaire was used to gather background information on the students, concerning (1) the consumption of various cultural goods and language practices in different genres (“How often do you read a book/watch television/write an e-mail/write an official letter/read anything in foreign languages?”, etc.); (2) consumption of standardist culture goods, such as spelling dictionaries, comprehensive dictionaries or television programmes/websites on linguistic accuracy; (3) demography characteristics (age, sex, etc.); and (4) improvement of students at school.

Discussions were initiated on the evaluation of different speech varieties, e.g., dialects, slang and others. Other- and self-repair was recurrently topicalized (e.g., “Do you correct others’ language use? Are you corrected yourself? How? How often?”, etc.).

Some questions generated lively debates in the class community. For example, the students were asked to compare two imaginary young girls who were described by their language use. One of the students was the speaker of an unidentified dialect, and the other was a so-called standard-language speaker. The goal of this task was to initiate a discussion on the possibilities, relevance and legitimacy or illegitimacy of evaluating and classifying others by their use of language. Being an exciting issue, this topic served as an ideal introduction in the interviews.

1,195 students filled in the questionnaires in grades 7 and 11. During the quantitative analysis of the questionnaire data, question–response sequences were taken as turns in a mediated discourse between the researcher and the informants, so questionnaires can also serve the goals of a detailed interactional analysis of Hungarian school metalanguage.

#### 3.2.2. Notes on classroom observations

Notes on classroom observations (grades 7 and 11, N=62 lessons) focused on the organization of a lesson, and on the patterns of teacher–student communication. Emphasis was on interactional routines for regimenting classroom discourse. Scenes from the observed lessons, which were considered as interesting from the interviewer’s angle, were mentioned in the interviews in order to initiate conversation on the given problem (e.g., “Why does your teacher punish slang speakers?”; “What do you think about the student’s response to the teacher in this case?”, etc.).

The corpus contains ca. 29,000 words and it is stored in XML format, ready to be analyzed with corpus linguistics tools, such as CLaRK (Simov et al. 2001).

#### 3.2.3. Interviews: CHSM-IC

Semi-structured research interviews were made with students and their teacher of ‘Hungarian Language and Literature’. Students were interviewed by the present author, generally in the company of one or two of their peers. Regularly, students reacted on the statements of their peers, and their co-constructive routines, having a high importance in the emergence of language ideologies, were observable. Being the largest subcorpus of *CHSM*, the *Interview Corpus (IC)* was published under the name of *CHSM-IC (Corpus of Hungarian School Metalanguage – Interview Corpus)* in 2013, as part of a European Union-funded international project called *CESAR (Central and South-East European Resources)*. The publication process was supervised by Tamás Váradi, chair of the *CESAR* project. An online registration, a research plan and a declaration on ethical issues is required from future users.

Some of the basic topics of the interviews were ‘stereotypes’, ‘language rules’, ‘other-repair’, and the linguistic evaluation of certain language varieties such as dialects (unidentified), slang, obscenity and impediment in speech. Tasks differed according to the topic investigated. Students had to read a text initiating the topic and then answer a few questions. For instance, in the case of stereotypes, the questions were the following: “If you meet somebody speaking a rural dialect / slang / profane words / in a tongue-tied way, what do you think of him or her? Would you like to be his or her friend? Do you like the way he or she speaks? Would you evaluate or correct explicitly his or her language use?”

Or, in the case of rules in general: “What do you think is a rule? Do you know optional rules? Do you know obligatory rules? Give examples!”. Students were invited to deliver narratives on language use and also to explain the situation described.

Teachers answered questions on the evaluation of their students’ linguistic performance, textbooks and other materials used during classroom activities, and on other methodological and pedagogical issues. These interviews could be used as background material for the analysis of classroom discourse or student interviews.

From the viewpoint of language ideology research, the most important subcorpus of *CHSM* is the *IC*. Since it is an annotated transcription of spoken metalanguage, it can be used for various purposes. Data in the *IC* can be searched along two types of annotations:

- (1) Thematic annotations marking the topic of the conversation. These are used primarily when seeking for ideologies emerging on a certain question, e.g., ideologies on other-repair, or, in the case of teachers, ideologies on teaching principles and methods, or the evaluation of children’s knowledge, etc. Annotation is multilevel: within categories, certain subcategories can be found (e.g., other-repair, either initiated by the interviewee or initiated by a communication partner, etc.).
- (2) Annotation of the characteristics of spoken language. This can be useful when analyzing the interactional features of the emergence of an ideology. Comparative studies can also be made, e.g., what interviewees say about repair and how they actually do repair during the recorded conversation. For a detailed presentation of interview structure, see Appendix 1 and Appendix 2.

Table 1 shows the number of interviewees. Interviewee selection targeted a comparative study on metalinguistic performance at different levels of formal schooling. As already mentioned, students were interviewed in small groups of two or three, and for this reason the number of interviewees is much higher than that of the interviews (cf. Table 2). For the selection of interviewees, the so-called *snowball* technique was used (interviewees suggested other potential interviewees), so that the size of the subsamples is not equal.

The interview corpus contains ca. 47.7 hours of recorded speech (346,500 words; see Table 3). Transcription is stored in XML format, ready to be analyzed with corpus linguistics tools such as CLaRK (Simov et al. 2001). Transcript annotation fits the standards published in TEI guidelines (Burnard and Bauman 2012). Personal data such as names, location of the interview, address or residential city of the mentioned persons, etc. are masked. At the present state, the audio files recorded by a digital device have not yet been annotated.

Interview collection of children aged 6 to 11 (grades 1-4) was not transcribed word by word in full length, because discourse topic often deviated from the interview outline to small talk. These secondary topics were summarized in brief notes.

Groups and classes	Capital	City	Village	Σ
Students, grades 1–4.	28	13	6	47
Students, grade 7	26	11	2	39
Students, grade 11	13	10	–	23
Teachers, grade 7	7	3	1	11
Teachers, grade 11	8	5	–	13
<b>Σ</b>	<b>82</b>	<b>42</b>	<b>9</b>	<b>133</b>

Table 1. Number of interviewees

Groups and classes	Capital	City	Village	Σ
Students, grades 1–4.	15	5	3	23
Students, grade 7	11	4	1	16
Students, grade 11	7	5	–	12
Teachers, grade 7	7	3	1	11
Teachers, grade 11	8	5	–	13
<b>Σ</b>	<b>48</b>	<b>22</b>	<b>5</b>	<b>75</b>

Table 2. Number of interviews<sup>4</sup>

<sup>4</sup> A teacher working in the capital city speaks about his experience with grade 7 and 11 students in the interview. This implies that this interview is counted twice.

Groups and classes	Capital	City	Village	Σ
Students, grades 1–4.	492	149	87	728
Students, grade 7	450	173	31	654
Students, grade 11	348	221	–	569
Teachers, grade 7	265	106	21	392
Teachers, grade 11	316	204	–	520
<b>Σ</b>	<b>1871</b>	<b>853</b>	<b>139</b>	<b>2863</b>

Table 3. Length of interviews (minutes)

Interviews were not based on elicitation but had a conversational design. Both the interviewer (the present author) and the interviewees participated intensively, as Table 4 shows. In grade 7, the interviewer often ended up explaining tasks in a more detailed manner than in grade 11. That is the main reason why the proportion of the interviewer's speech is higher in grade 7. Speaking with teachers, the interviewer was not so dominant, at least according to quantitative indicators. As trained and experienced professionals, teachers delivered detailed narratives and explicit evaluative comments, and they often kept the floor for several minutes.

Groups and classes	Number of words in subcorpora	Proportion of words uttered by the researcher
Students, grades 1–4.	46 236	— <sup>3</sup>
Students, grade 7	91 185	42%
Students, grade 11	82 054	31%
Teachers, grade 7	55 350	15%
Teachers, grade 11	71 675	15%
<b>Total:</b>	<b>346 500</b>	<b>Average<sup>4</sup>: 28%</b>

Table 4. Number of tokens in the interview collection<sup>5,6</sup>

The proportion of the researcher's words is not higher than in other Hungarian corpora of semi-structured interviews. In the second version of the *Budapest Sociolinguistics Interview*, 35% of the transcribed corpus is the speech of the interviewers (cf. Borbély and Vargha 2010), while in a segment of *BEA*, this proportion is 23% (cf. Bata and Grácz 2009).

#### 4. A CASE STUDY: DISCOURSE MARKER *HÁT* ('SO', 'WELL') IN A TURN-INITIAL POSITION IN SPOKEN HUNGARIAN

##### 4.1. Traditions in grammar studies: from prescriptivism through controversies to descriptivism

Any kind of linguistic description is a language ideology, but grammars have special importance. Metatexts, published in an academic context, are prestigious sources of language ideology construction: they can serve as a basis of argumentation in metadiscourses. Therefore, it is important to investigate characteristic approaches to *hát* through the history of Hungarian grammars before analyzing the data recorded in *CHSM-IC*.

*Hát* ('so' or 'well' in English) has important interactional functions in turn-taking. In the mainstream Hungarian literature of *hát*, a duality of prescriptivism and descriptivism can be found. Authors from both streams claim that *hát* is used mainly for marking the intention of speaking (the speaker's wish to participate in a conversation), but they differ in the evaluation of the item.

<sup>5</sup> Since not the whole material was transcribed word by word, the last column does not show the proportion of utterances for students, grades 1-4.

<sup>6</sup> The total average is for students (grades 7 and 11) and teachers (grades 7 and 11).

A very typical example of the prescriptivist approach comes from *Nyelvművelő kézikönyv* ('Handbook of Language Cultivation'; Grétsy and Kovalovszky 1980, 1985). The two thick volumes of this prestigious handbook can be found in almost every school library in Hungary, and its content was widely popularized through grammar textbooks or various journal and newspaper articles. The viewpoint of this handbook is characteristic to the normative approach which was dominant in Hungary for the second part of the 20th century. (Most of the manuscript was written in the 1960s, but its publication and editing lasted until the 1980s.) The excerpt cited below can be found in the manual's entry for *beszédtöltelékek* or 'filler words':<sup>7</sup>

Megfigyelhetjük, hogy sokan, főképp az élőbeszédben, társalgásban s különösen értekezleteken, vitában, felszólaláskor, úgy szerkesztik mondataikat, hogy **teletűzdelik** őket **tartalom és hasznos nyelvi funkció nélküli**, ill. funkciójukat vesztett fölösleges elemekkel: töltelékszavakkal, szókapcsolatokkal, mondattörésekkel. Ezek többnyire **csupán** arra valók, hogy a beszélő időt nyerjen mondanivalójának megfogalmazására, megtartsa beszédének (**látszólagos**) folyamatosságát, ill. megakadályozza, hogy a beszélgető társ elvegye tőle a szót. **Nemegyszer azonban a gondolatok kialakulatlanságából, zavarosságából, esetleg hiányából ered a használatuk.**

We can observe that many people, especially in spontaneous speech, conversations and particularly at meetings, debates or speeches, lard their sentences with elements **without content and useful linguistic function**, or with unnecessary elements of lost function. These are filler words, phrases and fragments of sentences. Generally, these are used **just** to gain time for the speaker for constructing the message and to maintain the (**illusionary**) continuity of the speech and to arrest his or her communication partner in continuing. **But, frequently, the motivation of their use is the primitivity, confusion or lack of thoughts.**

(Grétsy and Kovalovszky 1980: 323; emphasis added)

There is no evidence that this text directly shaped metadiscourses, but it appeared in a prestigious handbook, published by the Hungarian Academy of Sciences and edited by reputed linguists, so that it could indeed have an impact. Appearing in a synthesizing handbook, this text is not without antecedents: it summarizes a folk tradition. The text describes *hát* as an element which is insignificant and lacks a function, but which can disturb communication processes (e.g., the flow of discourse). However, the passage is contradictory: some functions are mentioned ("gain time for the speaker for constructing the message and to maintain the (illusionary) continuity of the speaker's speech"), but they are not evaluated as "useful".

As predecessors of present-day corpus-based linguistic analyses, studies on dialectal texts have had a long tradition in Hungarian linguistics. Based on tape recordings, grammars of regional varieties were elaborated. The author of such a regional grammar, Itzész (1981) examined the case of *hát*. She writes about the rich functionality of *hát* and notes that most often it occurs in a turn-initial position, but evaluates it as a filler word.

Using another corpus of recorded and transcribed speech gathered at Eötvös Loránd University, Keszler (1983) notes that every linguistic element has a function, so a zero function (or functionless) can hardly be thought to exist in language use. She adds that the term *töltelékszó* 'filler word' had not been defined in detail in previous Hungarian studies. Keszler's notes foreshadow the basic notions of purely descriptivist sociolinguistics, but surprisingly, as a contradiction, the last note in her paper—and the conclusion of her later account (Keszler 1985) – is that *hát* is often a "totally functionless (...) unnecessary (...) filler word" (Keszler 1983: 178).

In recent papers, a zero function is mentioned as nonsense in linguistic description. Using a heterogeneous terminology, several authors – mainly from the framework of speech act pragmatics – conclude that *hát* has an important and multiple function in the organization of spontaneous speech as discourse marker, mainly in the signaling of turn-taking. Although this approach can now be considered mainstream in the community of Hungarian linguists (cf. Dér 2010; Dér and Markó 2007; Schirm 2011), the case is not the same in public formal education, as interview data will show in the next section.

#### 4.2. A corpus-based investigation of explicit language ideologies on *hát*

In order to investigate language ideologies emerging in research interviews, we can observe explicit evaluations and narratives on communication practice and then compare the findings to the performance of the interviewees. Thus, explicit and implicit ideologies can be compared, because the usage of *hát* can be counted as an implicit ideology claiming that *hát* performs different functions.

The topic of *hát* arose in various contexts during the research interviews, but most often it was framed by narratives on other-repair. As a part of the interview protocol, the following questions concerned this topic: "Have you been corrected for your language use? How was it done? How was it reasoned? What was the problem with the corrected

<sup>7</sup> In the indented examples, Hungarian originals come first, and then my own translations follow.

word/expression?” These questions were targeted to all of the participants in the interview, which is why it was possible to observe the reflections between participants. These cases show the dynamics and interactive characteristics of ideology construction very well. The following excerpt illustrates a widely occurring ideology.<sup>8</sup>

701: Osztályfőnök azt mondta, hogy »háttal nem kezdünk mondatot.«

701: Our teacher said “we don’t start a sentence by *hát*”.

(Csongrád county, city, grammar school, grade 7, female)

In this case, a prescriptive statement on language practice was quoted by the interviewee, and the source (the teacher) was identified. Quoting is a very common form of making ideologies on language use: by mentioning an authority as a source, the relevance and value of the given statement can be increased. Quoting plays an important role in metalinguistic socialization: it marks the first step in the acquisition and internalization of a given metalinguistic procedure, e.g., the construction of an ideology (cf. Aro 2009, 2012).

In the previous excerpt, a direct quotation is formulated as a statement on a “we-group”. This group avoids using *hát*. This common identity can be characterized through other narratives on “how we use language” as well. “We-groups” are often in an opposition to “they-groups”: “their” language use differs from “ours”. This dynamics is crucial in a standardist community as a basis for establishing distinctions between groups.

The following excerpts illustrate that even in the first stage of formal schooling, developed practices of language ideology construction are learnt. The interviewer had a discussion with two girls in grade 2 in a Baranya county elementary school. The evaluations used by teachers during repair were discussed:

IV: [Mit mondtak a tanárok?]

451: == **Hát** azt mondja, ne kezdd hogy »ne kezdd úgy mindig a mondatunkat [!], hogy *hát*« ==

452: == mondja, hogy (2 mp) »ne kezdd úgy (5 mp) *hát* meg *ööö*« ==

451: hogy meg azt is mondják, hogy »nem így kezdjük a mondatokat« .

IV: Ühm. És mit gondoltok, hogy miért nem úgy kezdjük a mondatokat? (2 mp) Azt azt nem mondják el, hogy miért?

451: Nem.

452: Nem szokták.

IV: Aha. És ő ti mit gondoltok? Van ezzel kapcsolatba valami ötletetek, hogy mi lehet ennek az oka? (8 mp – suttognak egymás között)

452: Van.

IR: [What did teachers say?]

451: == **Hát** she says, don’t start that »don’t start our sentences in a way that “*hát*...”« ==

452: == says that (2 secs) »don’t start in a way (5 secs) *hát* and *ööö*« ==

451: and they say as well that »we don’t start sentences this way«.

IR: Uh huh. And what do you think, why we don’t start sentences that way? (2 secs) Don’t they tell you why?

451: No.

452: They don’t.

IR: Yup. And er you, what do you think? Do you have an idea on what could be the reason of that? (8 secs – whispering between each other)

452: We have.

(Baranya county, village, elementary school, grade 2, females)

This excerpt illustrates the process of common ideology construction. The girls in the excerpt cooperate in answering the interviewer’s questions. They quote their teachers in a similar way and the girl numbered 452 often follows her classmate, reformulating her statements. It may also be observed that the interviewer adopts his speech in the use of personal pronouns to the teachers quoted by the girls, e.g., “Why *we* don’t start sentences that way?” and then switches to the students’ position, e.g., “And er *you*, what do you think?”.

After a discussion between each other, the girls create an ideology on why *hát* should not be used. The interviewer continues ideology construction, initiating another perspective on erroneous talk, and asks the girls to make an “ordinary” utterance (in their terminology: *sentence*) “not ordinary”. Girls solve the problem by inserting *hát* and hesitation marker *ööö* to positions where they are normally used in spoken Hungarian:

<sup>8</sup> In the excerpts, bold fonts signal that *hát* is used as a discourse marker in its primary function (many occurrences of *hát* display a secondary, metalinguistic function, as subjects of evaluations). *Ööö* is used for transcribing a common hesitation marker; it is not translated when used as a subject of evaluation. *IR* (in the Hungarian text: *IV*) stands for “interviewer”, a three-digit number identifies interviewees. == marks overlaps. Between » and «, statements counted as quotations can be read. || marks reiteration and correction.

- IV: No, mi az oka? Mire jöttetek rá? (4 mp)
- 451: Hát az, hogy a *hát*tal meg az *ööö*vel meg az ilyenekkel nagyon nem lehet szép, rendes mondatot == alkotni. ==
- 452: == Alkotni. ==
- IV: Ühm. És milyen az a rendes mondat? (1 mp)
- 451: Mondjuk ha *én elmentem a fagyizóba, vettem egy csokis csokis fagyit.*
- IV: Ühm. ez egy rendes mondat? És milyen lenne ez úgy, hogyha ez nem rendes mondat lenne? (4 mp)  
Hogyan mondanátok ezt nem rendesen?
- 451: Hogy == *hát én elmentem a fagyizóba ööö vettem egy (4 mp) egy gombóc ö fagyit.* ==
- 452: == *Én elmentem a |jé | fagyizóba ööö vettem egy (2 mp) gombóc (4 mp) fagylaltot fagylaltot* ==
- IR: Well, what's the reason? What have you arrived at? (4 secs)
- 451: **Hát** that by *hát* and *ööö* nice, ordinary clauses are not really possible to == be created ==.
- 452: == Be created. ==
- IR: Uh huh. And what does an ordinary clause look like? (1 sec)
- 451: Let's say, *én elmentem a fagyizóba, vettem egy csokis csokis fagyit* ['I went to the ice cream shop and I bought a scoop of chocolate ice cream']
- IR: Yeah, is it an ordinary sentence? And what would it be like if it weren't an ordinary sentence? (4 secs) How would you say it in a not ordinary way?
- 451: That == *hát én elmentem a fagyizóba ööö vettem egy (4 secs) egy gombóc ö fagyit.* == ['I *hát* went to the ice cream shop *ööö* I bought an ice cream']
- 452: == *Én elmentem a |jé | fagyizóba ööö vettem egy (2 secs) gombóc (4 secs) fagylaltot fagylaltot* ==  
['I went to the ice cream shop *ööö* I bought a scoop of ice cream']  
(Baranya county, village, elementary school, grade 2, females)

At this point, the interviewer changes to linguistic stereotypes concerning the usage of *hát* and *ööö*. The girls prepare the answer together for some 15 minutes and then construct an ideology: as *hát* users, they can be evaluated as bad communicators in a conversation:

- IV: Ühm, tehát akkor ilyen lenne a nem rendes. És mit gondolsz, mondjuk hogyha így beszélgetnék velem és és öö úgy beszélgetnék velem, hogy mondanátok, hogy *hát elmentem és ööö vettem egy fagyit*, akkor én mit gondolnék rólatok vagy gondolnék-e valamit egyáltalán, hogy ti most ilyen nem rendes mondatban válaszoltatok?
- 451: Igen.
- 452: Igen.
- IV: Mit gondolnék rólatok? (8 mp – suttognak) Hm? (7 mp – suttognak)
- 451: Hogy mi nem tudunk nagyon == beszélgetni az emberekkel == .
- 452: == Beszélgetni az emberekkel ==
- IR: Uh huh, so this would be the not ordinary. And what do you think, let's say, if you would speak that way to me, and and er you would speak to me that way that you would say *hát elmentem és ööö vettem egy fagyit* ['*hát* I went and *ööö* I bought an ice cream'], then what should I think about you or should I think anything about you because you have answered in a not ordinary sentence?
- 451: Yes.
- 452: Yes.
- IR: What should I think about you? (8 secs – girls are whispering) Huh? (7 secs – girls are whispering)
- 451: That we can't really == get a conversation with people ==.
- 452: == Get a conversation with people ==.
- (Baranya county, village, elementary school, grade 2, females)

In the above cited language ideologies, a rather negative evaluation of *hát* can be found in explicit statements. Further investigation proves that similar ideologies were created at different levels of education as well: these can be seen as popular. For example, in another interview done to two girls in grade 7, the students claimed that *hát* marks uncertainty and lack of communication skills.

In another conversation, one of the interviewees claims that she does not like utterances starting by *hát* or *és* 'and'. Subsequently, assimilating to the ideology constructed by this student, the interviewer asks for a legitimization of the negative evaluation of these lexemes:

IV: [...] miből gondoljátok, hogy ő hogy rossz dolog *éssel* vagy *háttal* kezdeni mondatot? (3 mp)

092: *Hát* ő igazából ezzel nem kezdünk mondatot.

091: [*nevet*]

IV: [*nevet*]

092: Már úgy kezdődik.

IR: [...] how do you know that it is a bad thing to start a sentence by *hát* or *és*? (3 secs)

092: **Hát** er actually we don't start a sentence by this.

091: [*laughs*]

IR: [*laughs*]

092: It is started that way, then.

(Budapest, elementary school, grade 7, females)

The student, by creating a “we-group”, answers that utterances (in her terminology: *sentences*) are not started by *hát* (or *és*), but betrays her own tenets by starting her utterance precisely by *hát*. To this self-contradiction the other interviewee and the interviewer respond with laughter and then 092 herself reacts by laughing, too. This can signal that she behaved differently from the members of the previously constructed, ideal “we-group”.

It is not by accident that students make negative evaluations on turn-initial *hát*. As a source of high formality, the teacher's statements contain a rejection of *hát* as well. In a narrative, constructed upon her own practice as a teacher, the interviewee cited below claims that students should not use *hát* in a classroom context. Her opinion is that students should know answers by heart and that is why *hát* is not acceptable:

671: A *hát*, minden mondatot háttal kezdünk, én is sokszor, (1 mp) ő ha szabadon beszélgetünk, persze nem olyan nagy baj. De amikor a diákokat kérdezem, elvileg már őneki meg kellett tanulnia azt a választ, tehát nem kezdi [nyújtva *hát*] ilyen időnyerő válasszal, de mindig *háttal* kezdik, és akkor azt is [...] fölírom, hogy *hát*, és akkor áthúzom. [*nevet*]

671: *Hát*, we start every sentence by *hát*, including me, many times, (1 sec) er when we talk spontaneously, of course, it is not a big problem. But when I examine students, they have had to learn the answer – in principle – so s/he wouldn't start by “*hát...*”, which is a tool for gaining time. But they always start by *hát* and then [...] I write *hát* up on the blackboard and then I score it out [*laughs*]

(Csongrád county, city, grammar school,  
teacher of Hungarian language and literature in grade 11, female)

The ideology presented is educational for at least two reasons:

- (1) The quoted teacher notes that students should know answers “by heart”. This legitimization of school practices supports a traditional assimilation method (cf. Aro 2009, 2012) that gives only one main task for students: the reproduction of normative texts disseminated by the school.
- (2) A narrative on a repair method can be observed as well. The teacher writes examples of erroneous talk on the blackboard and then she scores them out. By using this method, she enforces her verbal instructions with visual ones.

The following excerpt is a typical example of the dynamics of common ideology construction in a research interview context. Talking about phenomena evaluated as errors (turn-initial *hát* and *és* ‘and’, and definite article before proper names), the interviewer starts to investigate the reason of students' statements:

211: Erre nincs írott szabály,

IV: Ühm.

211: ezt mindenki tudja magától, hogy [*nevet* nem szabad.]

IV: Ühm. És akkor hogyha mindenki tudja magától, akkor valóban így is beszélnek? Tehát akkor nem is ő kezd senki *háttal* mondatot?

211: De.

212: De.

IV: [*nevet* És mi] lehet ennek az oka, hogy hogy tudják, s akkor mégis használnak *háttal* mondatot, vagy kezdenek *háttal* mondatot vagy *éssel*, kirakják a név elő a névelőt?

211: Valaki kérdez valamit == ő a választ azt ==

212: == Igen, és ezzel időt nyerünk. ==

211: majdnem mindig *hát* tudod, a nem tudom, mi, *hát* nyolckor, *hát* este, *hát* majdnem mindenki így beszél szerintem.

- 211: There is no written rule for that,  
 IR: Yeah.  
 211: everybody knows by himself/herself that [*laughs* it mustn't be done.]  
 IR: Uh huh. And if everone knows it by himself/herself, do they speak this way in practice? So then nobody starts a sentence by *hát*, huh?  
 211: They do.  
 212: They do.  
 IR: [*laughs* And what] can be the reason of that? They know it and they still use sentences by *hát* or start sentences by *hát* or *és* and they use names with an article...  
 211: Somebody asks something == er the answer is ==  
 212: == Yes, and we gain time by this. ==  
 211: in almost every cases *hát tudod, a nem tudom, mi, hát nyolckor, hát este* ['*hát*, you know, I don't know what, *hát*, at eight, *hát*, at the evening'], **hát** almost everybody talks this way, I guess.  
 (Pest county, city, grammar school, grade 11, females)

Students claim that every speaker knows the rules governing the use of these words and that these words are incorrect. Following this line, the interviewer – in a naive manner – supposes that speakers who know these rules never use the mentioned words. Students make narratives on everyday communication to prove that the opposite is true. By doing so, they construct a linguistic description that is at odds with real life usage. This is a kind of implicit language ideology as well: prescriptivist metalinguistic tradition does not conform to everyday practice.

#### 4.3. An analysis of *hát* in the speech of *CHSM-IC* interviewees

The data stored in *CHSM-IC* can be analyzed from a quantitative approach as well, which, like the qualitative one, also uncovers differences between (1) a school tradition of metatexts, (2) narratives on one's own communication practice and (3) the performance recorded in the interviews. From the interviews cited in section 4.2, cases were collected where *hát* was used in a turn-initial position in its primary function as a discourse marker, and not as the object of a linguistic evaluation. These data show how recorded performance differs from what is presented as an ideal. Table 5 shows the number of turns transcribed word by word from the selected interviewees' speech. This number indicates the size of the subcorpus of the given speaker. Another number shows the occurrence of *hát* in turn-initial position as a discourse marker. Speakers presented in Table 5 were all students, only participant 671 was a teacher. The data in Table 5 confirm Labov's claim that speakers, even teachers, may produce variants they evaluate negatively (Labov 1972; Nardy and Barbu 2006).

Interviewee's ID	Number of turns transcribed word by word, uttered by the interviewee	Frequency of turn-initial <i>hát</i> (count)
091	170	44
671	141	31
211	156	29
092	117	26
701	171	25
212	129	8
451	29	2
452	27	1

Table 5. They said "we don't use *hát*..."

A general percentual description of the use of *hát* in interviews is summarized in Table 6. The goal of this analysis is to illustrate the proportion of *hát* in the speech of the interviewees. The interviewer's utterances were filtered out. Definite and indefinite articles (*a, az* 'the' and *egy* 'a/an'), connectives such as *és* 'and', *hogy* 'that' and *is* 'too', and the negator *nem* 'no, not' were not considered. The analysis was made without lemmatizing subcorpora, since, in interactional studies, the form of a word is important from the point of view of agency analysis and routinized interactional patterns. Table 6 shows the proportion of the most common words occurring in *CHSM-IC* subcorpora.

Students				Teachers			
Rank	Grades 1–4	Grade 7	Grade 11	Rank	Grade 7	Rank	Grade 11
1	<i>hát</i> (2,71%)	<i>hát</i> (3,06%)	<i>hát</i> (2,01%)	1	<i>tehát</i> (‘therefore’, ‘so’, 1,78%)	1	<i>tehát</i> (‘therefore’, ‘so’, 1,65%)
2	<i>akkor</i> (‘then’, 2,46%)	<i>akkor</i> (‘then’, 1,92%)	<i>akkor</i> (‘then’, 1,59%)	2	<i>ez</i> (‘this’, 1,45%)	2	<i>akkor</i> (‘then’, 1,36%)
				4	<i>akkor</i> (‘then’, 1,33%)	4	<i>ez</i> (‘this’, 1,30%)
				10	<i>hát</i> (0,91%)	9	<i>hát</i> (0,99%)

Table 6. Interviewees’ most frequently used tokens (percentage of the total number of tokens)

As may be seen in Table 6, *hát* is the most common token in each student subcorpus and is quite common in teachers’ language use as well. It is noteworthy that *hát*, *akkor* ‘then’, *ez* ‘this’ and *tehát* ‘therefore’ all play a significant role in discourse organization and that *tehát* displays functions similar to those of *hát*.

## 5. SUMMARY AND FUTURE PLANS

As a part of a broad survey on metalanguage, I compiled the *CHSM*. This is a complex research tool which can be used for research on language ideology, for studies on spoken Hungarian or for educational purposes in general. The *CHSM-IC* lends itself to the analysis of the emergence of language ideologies in spoken discourses, with a CA approach (Laihonen 2008).

In the present paper, a case study on *hát* concluded that there exists no consistency between ideologies on language use and language use itself (cf. Krashen 1982). That is, language ideologies have no bearing on performance and *vice versa*: even teachers, who are perhaps the most prestigious authorities in prescriptive discourses, have used *hát* as a discourse marker regularly.

A collection of additional data from regions outside Hungary would be needed for a better understanding of Hungarian metalinguistic socialization in different cultural settings. To reach this goal, fieldwork in a bilingual context should be carried out. Another avenue for further research would be to study institutional multilingual policies in dual language schools. A third study, placing the present discussion into a wider cultural context, should deal with the linguistic landscape of educational spaces (e.g., pictures, cultural symbols, summaries of grammar instructions on the school walls) and its impact on the assimilation of linguistic evaluations (Brown 2012). The *CHSM-IC*, combined with additional material from the *BEA* and *BSI-2* corpora, can be the basis of a detailed CA description of discourse markers such as *hát*.

The applicability of the *CHSM-IC* is versatile, especially for educational purposes:

- (1) A corpus-based analysis of interview discourse could be conducted while dealing with language-planning or sociolinguistics in the classroom. Excerpts from the *CHSM-IC* can serve as models for discussion on the topic and as experiments for teachers while planning their classroom activities.
- (2) Alternatively, the *CHSM-IC* can help both teachers and students to observe spoken Hungarian. Tasks should be given to students, e.g., an analysis of an excerpt with special attention to sociolinguistic variables, such as status (student, teacher and researcher), age, gender, etc. In this case, the *CHSM-IC* can be used as a corpus of spoken language and its metalinguistic character would have a secondary importance.
- (3) A systematic analysis of language ideologies emerging during interview discourses in the *CHSM-IC* can be used for decision-making in language policy and educational policy.

## REFERENCES

### Corpora

*BEA* = *Beszélt nyelvi adatbázis / A Hungarian Spontaneous Speech Database*. Budapest: Research Institute for Linguistics of the Hungarian Academy of Sciences. Project chair: Mária Gósy. Available at <<http://metashare.nytud.hu/repository/search/>>

- BSI-2 = Budapesti Szociolingvisztikai Interjú / Budapest Sociolinguistics Interview. Version 2.* Budapest: Research Institute for Linguistics of the Hungarian Academy of Sciences. Project chairs: Miklós Kontra and Tamás Váradi. Available at <<http://buszi.nytud.hu/>>
- CESAR = Central and South-East European Resources.* Project chair: T. Váradi. Available at <<http://cesar.nytud.hu/>>
- CHSM-IC = Magyar Iskolai Metanyelvi Korusz – Interjúkorpusz / Corpus of Hungarian School Metalanguage – Interview Corpus.* Budapest: Research Institute for Linguistics of the Hungarian Academy of Sciences. Project chair: Tamás Péter Szabó. Available at <<http://metashare.nytud.hu/repository/search/>>

#### Secondary sources

- Aro, Mari. 2009. *Speakers and doers. Polyphony and agency in children's beliefs about language learning.* Jyväskylä: University of Jyväskylä.
- Aro, Mari. 2012. Effects of authority: voicescapes in children's beliefs about the learning of English. *International Journal of Applied Linguistics* 22/3: 331–346.
- Bata, Sarolta and Tekla Etelka Grácsi. 2009. A beszédpartner életkorának hatása a beszéd szupraszegmentális jellegzetességeire [‘Impact of the communication partner's age on the suprasegmental features of speech’]. In Borbála Keszler and Szilárd Tátrai (eds.), *Diskurzus a grammatikában – grammatika a diskurzusban.* Budapest: Tinta, 74–82.
- Berko Gleason, Jean. 1992. *Language acquisition and socialization.* Boston: Boston University.
- Bohner, Gerd. 2001. Attitudes and attitude change. In Miles Hewstone and Wolfgang Stroebe (eds.), *Introduction to social psychology.* Third edition. Oxford: Blackwell, 239–282.
- Borbély, Anna and Andras Vargha. 2010. Az *l* variabilitása öt foglalkozási csoportban. Kutatások a Budapesti Szociolingvisztikai Interjú beszélt nyelvi korpuszban. [‘Variability of *l* in five professional groups. Studies on the corpus of Budapest Sociolinguistics Interview’]. *Magyar Nyelv* 106: 455–470.
- Brown, Kara D. 2012. Minority languages in the linguistic landscape. In Heiko F. Marten, Durk Gorter and Luk van Mensel (eds.), *Linguistic landscapes and minority languages.* New York: Palgrave, 281–298.
- Burnard, Lou and Syd Bauman (eds.). 2012. *TEI P5: guidelines for electronic text encoding and interchange by the TEI Consortium.* Charlottesville, VA: TEI Consortium.
- Coulter, Jeff. 2005. Language without mind. In Hedwig te Molder and Jonathan Potter (eds.), *Conversation and cognition.* Cambridge: Cambridge University Press, 79–92.
- Csernicskó, István and Miklós Kontra (eds.). 2008. *Az Üveghegyen innen. Anyanyelvátváltatok, identitás és magyar anyanyelvi nevelés.* [‘Varieties of Hungarian, identity and Hungarian mother tongue education’]. Ungvár–Beregszász: PoliPrint–II. Rákóczi Ferenc Kárpátaljai Magyar Főiskola.
- Dér, Csilla Ilona. 2010. “Töltelékelem” vagy új nyelvi változó? A *hát, úgyhogy, így* és *ilyen* újabb funkciójáról a spontán beszédben. [‘“Filler words” or new linguistic variables? Newer functions of *hát, úgyhogy, így* and *ilyen* in spontaneous speech’]. *Beszédkutatás* 2010: 159–170.
- Dér, Csilla Ilona and Alexandra Markó. 2007. A magyar diskurzusjelölők szupraszegmentális jelöltsége. [‘Suprasegmental markedness of Hungarian discourse markers’]. In Tamás Gecső and Csilla Sárdi (eds.), *Nyelvelmélet – nyelvhasználat.* Székesfehérvár–Budapest: Kodolányi János Főiskola–Tinta, 61–67.
- Grétsy, László and Miklós Kovalovszky (eds.). 1980. *Nyelvművelő kézikönyv.* [‘Handbook of language cultivation’]. Vol. 1. Budapest: Akadémiai.
- Grétsy, László and Miklós Kovalovszky (eds.). 1985. *Nyelvművelő kézikönyv.* [‘Handbook of language cultivation’]. Vol. 2. Budapest: Akadémiai.
- Györi, Miklós. 2008. Tudatosság és megismerés. [‘Consciousness and cognition’]. In Valéria Csépe, Miklós Györi and Anett Ragó (eds.), *Általános pszichológia.* Vol. 3. Budapest: Osiris, 267–297.
- Ittész, Nóra. 1981. *Szövegszerkesztési kérdések Esztergom regionális köznyelvében.* [‘Features of text construction in the regional standard of Esztergom’]. MA thesis. Budapest: Eötvös Loránd University.
- Keszler, Borbála. 1983. Kötetlen beszélgetések mondat- és szövegtani vizsgálata. [‘Syntactic and discourse features in spontaneous speech’]. In Endre Rácz and István Szathmári (eds.), *Tanulmányok a mai magyar nyelv szövegtana köréből.* Budapest: Tankönyvkiadó, 164–202.
- Keszler, Borbála. 1985. Über die Verwendung der Füllwörten. *Annales Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae. Sectio Linguistica* 15: 11–26.
- Kontra, Miklós. 2006. Sustainable linguisticism. In Frans Hinskens (ed.), *Language variation – European perspectives.* Amsterdam: Benjamins, 97–126.
- Kontra, Miklós and Tamás Váradi. 1997. *The Budapest sociolinguistics interview: version 3.* Budapest: Research Institute for Linguistics of the Hungarian Academy of Sciences. <<http://www.nytud.hu/buszi/wp2/index.html>> (12 June 2013).
- Krashen, Stephen D. 1982 [2009]. *Principles and practice in second language acquisition.* Oxford: Pergamon Press. The 2009 online edition is the author's own: <[http://www.sdkrashen.com/Principles\\_and\\_Practice/Principles\\_and\\_Practice.pdf](http://www.sdkrashen.com/Principles_and_Practice/Principles_and_Practice.pdf)> (12 June 2013).

- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.
- Laihonen, Petteri. 2008. Language ideologies in interviews: a conversation analysis approach. *Journal of Sociolinguistics* 12/5: 668–693.
- Laihonen, Petteri. 2009. *Language ideologies in the Romanian Banat. Analysis of interviews and academic writings among the Hungarians and Germans*. Jyväskylä: University of Jyväskylä.
- Lee, Yo-An. 2007. Third turn position in teacher talk: contingency and the work of teaching. *Journal of Pragmatics* 39: 180–206.
- Milroy, James. 1998. Children can't speak or write properly anymore. In Laurie Bauer and Peter Trudgill (eds.), *Language myths*. London: Penguin, 58–65.
- Milroy, James. 2001. Language ideologies and the consequences of standardization. *Journal of Sociolinguistics* 5/4: 530–555.
- Nardy, Aurélie and Stéphanie Barbu. 2006. Production and judgment is childhood. The case of liaison in French. In Frans Hinskens (ed.), *Language variation – European perspectives*. Amsterdam: Benjamins, 143–152.
- Potter, Jonathan and Derek Edwards. 2001. Discursive social psychology. In W. Peter Robinson and Howard Giles (eds.), *The new handbook of language and social psychology*. Chichester: Wiley and Sons, 103–118.
- Potter, Jonathan and Derek Edwards. 2003. Sociolinguistics, cognitivism, and discursive psychology. *International Journal of English Studies* 1: 93–109.
- Sándor, Klára. 2006. Nyelvtervezés, nyelvpolitika, nyelvművelés ['Language planning, language policy and language planning']. In Ferenc Kiefer (ed.), *Magyar nyelv*. Budapest: Akadémiai, 958–995.
- Sántha, Klára. 2006. *Mintavétel a kvalitatív pedagógiai kutatásban* ['Sampling in qualitative surveys on pedagogy']. Budapest: Gondolat.
- Schegloff, Emanuel A., Gail Jefferson and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language* 53/2: 361–382.
- Schirm, Anita. 2011. *A diskurzusjelölők funkciói: a hát, az -e és a vajon elemek története és jelenkori szinkrón státusa alapján* ['The function of discourse markers: the history and present synchronic status of the Hungarian elements hát, -e and vajon']. PhD dissertation. Szeged: University of Szeged.
- Simov, Kiril, Zdravko Peev, Milen Kouylekov, Alexander Simov, Marin Dimitrov and Atanas Kiryakov. 2001. CLaRK – An XML-based system for corpora development. In Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie and Shereen Khoja (eds.), *Proceedings of the Corpus Linguistics 2001 Conference*. Lancaster: Lancaster University, 553–560.
- Skutnabb-Kangas, Tove and Robert Phillipson. 1989. 'Mother tongue': the theoretical and sociopolitical construction of a concept. In Ulrich Ammon (ed.), *Status and function of languages and language varieties*. Berlin: Walter de Gruyter, 450–477.
- Van Leeuwen, Theo. 2004. Metalanguage in social life. In Adam Jaworski, Nikolas Coupland and Dariusz Galasiński (eds.), *Metalanguage. Social and ideological perspectives*. Berlin: Mouton de Gruyter, 107–130.

#### APPENDIX 1. MAIN INTERVIEW TOPICS IN GRADES 1-4

- Do you like talking (with friends, family, classmates)? Do you talk a lot (with friends, family, classmates)?
- Is there any difference between talking at school/kindergarten and talking at home? Is there any difference between talking to children/adults?
- Is there anything you can do/say at home/with friends but not at school/kindergarten (and vice versa)? Have you ever been told you should not talk like that or should not say something?

## APPENDIX 2. MAIN INTERVIEW TOPICS IN GRADES 7 AND 11

The main interview topics in grades 7 and 11 are presented in the following table, where (+/-) signals that the given topic emerged (+) or did not emerge (-) regularly in interviews made with students and teachers. Abbreviations were used during the XML annotation of the transcription.

<b>Abbreviation</b>	<b>Focus</b>	<b>Students</b>	<b>Teachers</b>
Metadisc	Metadiscourse: speech on the current interview discourse	+	+
CorrInstr	Narratives and ideologies on repair strategies (own/other) Self- and other-repair activity during the interview	+	+
Attitudes	Evaluation of other speakers' practice (dialects, slang, curse, impediment)	+	-
Grammar	Narratives on grammar courses	+	-
Rules	Explicit description of different rules (not linguistic)	+	-
Attitudes	Evaluation of students' speech	-	+
TeachingExp	Teaching experience	-	+
TextBook	Evaluation of textbooks and other materials	-	+
Grammar	Evaluation of students' activity during grammar classes. Ideologies on the meaning and goals of grammar lessons. Own motivations.	-	+