

## 1. The Corpus of Hungarian School Metalanguage (CHSM)

1.1. DEFINITIONS OF METALANGUAGE. Van Leeuwen (2004): two groups of definitions: (1) representation-oriented definitions (as a specific register with a scientific nature; representation of cognitive representations; tool of telementing inner beliefs); (2) communication-oriented definitions like Laihonen's (2008: 669): „From an interactional point of view, talk about language is a part of conversational action, such as answering, defending, blaming, accusing and apologizing”. Using this approach, we can define metalanguage as a socially constructed, (self-)reflective and dynamic discourse on language as a system or as a communication practice (cf. Laihonen 2008). Metalanguage is a part of a tradition of texts (this tradition is strongly ideological). Thus, we can analyse metalanguage with Conversation Analysis or Discourse Analysis methods.

1.2. EVALUATIONS, ATTITUDES, IDEOLOGIES, NARRATIVES. Evaluations and attitudes are important parts of metalanguage. Explanations can be identified as ideologies (cf. Laihonen 2008). Ideologies explain language use and legitimize or dispute communication practices. Metalanguage is not stable but dynamic: we can observe a continuous construction and deconstruction of ideologies, attitudes and narratives while conversation emerges. Discursive Social Psychology theory can explain these phenomena (cf. Potter-Edwards 2001, 2003).

1.3. WHY IS THIS CORPUS NEEDED? The primary goal was to carry out a complex sociolinguistic research on mother tongue education for my PhD dissertation. Ideologies on learning, following and disseminating language rules were at the focus of the survey, because Hungarian is a standard language culture (Milroy 2001) and in such a culture, talking about rules has a specific role. We can find (1) a cult of a privileged dialect called standard; (2) continuous standardization (dictionaries, prescriptivist handbooks etc.); (3) prescriptivist elements in curricula of formal education; (4) activities for developing language awareness or consciousness (e. g. movements of language cultivation); (5) other- and self-repair as a common activity, used for language socialization. But there is a heterogeneity in formal training: prescriptivist views are often disputed and descriptivist and prescriptivist ideologies are learnt simultaneously by students. A survey of Hungarian school metalanguage was needed to make a description of current trends in ideology making in a formal educational context.

1.4. HOW WAS IT BUILT? I used three methods during my 2009 fieldwork, that is why CHSM has three subcorpora.

1.4.1. Questionnaires (N = 1195). Groups: students on year 7 and 11. Questionnaires were analysed as mediums by which a discourse emerged between the researcher and the informant in the form of question-response sequences. That is why it can be used as a corpus of metalinguistic interactions.

1.4.2. Notes on classroom observation. Data: 61 school lessons on year 7 and 11. Notes focused on the organization of a lesson and on the patterns of teacher-student communication, with a special regard to interactional routines for regimenting classroom discourse. The corpus consists of cca. 29,000 tokens, stored in XML format.

1.4.3. Interview corpus (IC) is a collection of semi-structured research interviews. (This method is marginal in Hungarian mother tongue education studies). 74 interviews were made with 133 interviewees. Groups: students and their teacher of Hungarian grammar and literature at (a) elementary schools (year 1–4, 7, aged 6–11, 13–14); (b) vocational high schools (year 11, aged 17–19) and (c) grammar schools (year 1–4, 7, 11, aged 6–11, 13–14, 17–19). This corpus contains cca. 47.7 hours of speech. Its transcription consists of 346,500 tokens, stored in XML format:

Groups and classes	Number of tokens	Proportion of researcher's utterances (tokens)
Students 1–4.	46 236	—*
Students 7	91 185	42%
Students 11	82 054	31%
Teachers 7	55 350	15%
Teachers 11	71 675	15%
<b>Total:</b>	<b>346 500</b>	<b>Average**: 28%</b>

\* Not the whole material was transcribed word by word, that is why proportion of utterances was not calculated.

\*\* For Students 7, 11 and Teachers 7, 11

The proportion of researcher's utterances is normal if we compare these data to other Hungarian corpora of semi-structured interviews. In the 2<sup>nd</sup> version of Budapest Sociolinguistic Interview (“Budapesti Szociolingvisztikai Interjú”, abbreviated as BUSZI-2), a corpus made at RIL HAS, 35% of the transcribed corpus is the speech of the interviewees (Borbély-Vargha 2010). In a segment of another corpus built at RIL HAS called Database of Spoken Language (“Beszélt Nyelvi Adatbázis”, abbreviated as BEA), this number is 23% (Bata-Grácsi 2009).

The structure and topic of the interviews: (1) Students: one (rarely), two (regularly), three (sometimes), or more than three (in extreme cases) students were interviewed at the same time. The topics of the interviews were stereotypes, rules, and linguistic evaluations. Tasks differed according to the topic investigated. Students had to read a text initiating the topic, and they had to answer my questions. As for stereotypes, my questions were like these: „If you meet somebody speaking 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? Would you like the way he or she speaks? Would you evaluate or correct explicitly his or her language use?” [Asked separately, for each category.] Or, as for the topic of rules: “What do you think is a rule? Do you know optional rules? Do you know obligatory rules? Give me examples!” In these cases, students had to create narratives on language

use and they had to explain the situation described. (2) Teachers answered my questions on the evaluation of their students' speech, textbooks and other materials used during classroom activities. Another topic was methodology used in practice (with a special regard to repair as a tool of language socialization and evaluation).

1.5. ON PERSONS INVESTIGATED. I collected data in 34 schools (11 in capital city Budapest, 19 in County Bács-Kiskun, Baranya, Békés, Csongrád, Fejér, Győr-Moson-Sopron, Pest, Somogy, Vas and Zala, and 4 in Serbia and Slovakia). As a sum, students and teachers from 80 classes were involved (questionnaires were used in 52 classes).

## 2. A complex research tool: the Interview Corpus (IC)

2.1. ANNOTATION SYSTEM. (1) Thematic annotation (main topics of the interviews), e. g. language use of different groups; practice in other-repair; conversation on Hungarian grammar lessons; learning methods; teaching methods (for teachers); conversation on the present survey etc. (2) Annotation of the characteristics of spoken language, e. g. hesitation; overlap; pause; repair; nonverbal markers; word as a reference (as in "We never use *hát*"); pronunciation (e. g. "gogl [= Google]); noise, unintelligibility etc.

2.2. FOR WHAT PURPOSES CAN IC BE USED? (1) Scientific research on... (a) the construction of language ideologies (cf. Laihonon 2008); (b) argumentation techniques; (c) agency (cf. Aro 2009; Karasavvidis et al. 2000); (d) the grammar of spoken Hungarian. (2) Classroom application: (a) collecting data for classroom debates or for a presentation of different language ideologies; (b) observing spoken Hungarian with a special attention to sociolinguistic variables (status [student, teacher, researcher], age, gender etc.)

2.3. PUBLIC ACCESS. Plans: (1) public access (masking of personal information in voice recordings is needed; transcription is already masked); (2) a regulation of access should be elaborated (a registration and a declaration on ethical issues should be required).

## 3. A case study: discourse marker *hát* 'well' as a clause starter in spoken Hungarian

### 3.1. AN OVERVIEW OF THE LITERATURE.

3.1.1. A prescriptivist tradition. A handbook of language cultivation (NyKk. 1980, I: 323) claims: "We can observe that many people, especially in spontaneous speech, conversations and particularly at meetings, debates or oppositions construct their sentences in a way that they *lard them* with elements *without content and useful linguistic function*, or with unnecessary elements of lost function. These are expletives, phrases and fragments of sentences. Generally, these are used *just* to gain time for the speaker for constructing the message and to maintain the (*apparent*) continuity of the speaker's 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.*" (emphasis added; translation is mine: T. P. Sz.)

3.1.2. A descriptivist approach. (1) First corpus-based investigations on *hát*: (a) Itzész (1981): rich functionality of *hát* (as an expletive); position as a clause starter. (b) Keszler (1983) claims that views on linguistic functionless and the term *expletive* is problematic ↔ *hát* is defined as a "totally functionless [...] unnecessary [...] expletive" (p. 178). (2) Recent studies: (a) *hát* as a discourse marker:

Dér (2010); Dér-Markó (2007); Schirm (2011); (b) as a conversation marker (Kugler 2000). Neutrality.

3.2. A COMPLEX STUDY BASED ON IC. Data: examples from a section of the interviews concerning repair. Questions: "Have you been repaired for your language use?", "How was it done?", "How was it reasoned?", "What was told was the problem with the word/expression repaired?"

3.2.1. A qualitative corpus analysis. Excerpts – my translations (T. P. Sz.) – as illustrations for ideology construction in metadiscourses. ***Bold italics*** are added when *hát* is not used as a reference but as a discourse marker in its primary function. *IR* is for "interviewer".

(a) Quoting and building a "we-group" identity. Dynamics of common ideology construction. Early development: elaborated statements on language use, following prescriptivist tradition (7-year-old girls!).

**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:** Yeah. And what do you think, why we don't start sentences that way? (2 secs) Doesn't they tell you why?

**451:** No.

**452:** Don't used to.

**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.

**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:** Yeah. 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 chocolate ice cream']

**IR:** Yeah, is it an ordinary sentence? And how could it be look 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.* == ['same']

**452:** == *Én elmentem a |jé| fagyizóba ööö vettem egy (2 secs) gombóc (4 secs) fagyilaltot fagyilaltot* == ['same']

**IR:** Yeah, 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* ['same'], 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 – *whispering*) Huh? (7 secs – *whispering*)

**451:** That we can't really == get a conversation with people ==.

**452:** == Get a conversation with people ==.

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

(b) *IR* assimilates to a "we-group" in questioning. Difference between ideology and performance is reacted upon.

[092 does not like clauses started by *hát* or by *és* 'and']

**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, class 7, females)

(c) A teacher creates a narrative on her own repair practice. Positioning herself in a "we-group" of *hát*-users, while punishing the use of *hát* during classes.

**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 ask students, s/he 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 grammar and literature in class 11, female)

(d) Primary ideology (as a general rule): *hát* should not be used → narrative on real-life practice → secondary ideology: *hát* is used by every speaker in certain situations.

**211:** There is no written rule for that.

**IR:** Yeah.

**211:** everybody knows by himself/herself that [laughs it mustn't be done.]

**IR:** Yeah. 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 can gain time by this. ==

**211:** in almost every cases *hát tudod, a nem tudom, mi, hát nyolckor, hát este* ['well, you know I don't know what, well, at eight, well, at the evening'], *hát* almost everybody talks this way, I guess.

(Pest county, city, grammar school, class 11, females)

3.2.2 A quantitative corpus analysis. In the recorded speech of the above-cited interviewees I collected cases where *hát* was used as a clause starter (and not as a reference). These data just show how real language use differs from rules constructed in a prescriptivist manner. Number of turns transcribed word by word show the size of the subcorpus of an interviewee's speech. 671 was a teacher.

Interviewee's ID	Number of utterances transcribed word by word	Frequency of discourse marker <i>hát</i> (count)
091	170	44
671	141	31
211	145	29
092	117	26
701	171	25
212	129	8
451	29	2
452	27	1

We can see that even dozens of occurrences were registered. It means that Labov was right claiming that speakers may produce variants they evaluate negatively (cf. Labov 1976). One should evaluate this situation controversial, but analysing IC excerpts with a Conversation Analysis approach, so-called controversies between ideologies and recorded performance do not seem to be controversies. We can say instead that different traditions of metalanguage are used simultaneously in these texts. Students and teachers are exposed to several metadiscourses and a rich folklore of normative texts, and they import elements from those. Meanwhile, they construct narratives on their language use, and their experience can differ from traditional descriptions. (Constructing narratives has its traditions and routines as well.)

As a control of qualitative ideology study, a statistical analysis of the texts of all 133 interviewees was made. My utterances were filtered out and a stop list of articles and other function words was used. Analysis was made without lemmatizing subsorpora, because in interactional studies, the form of a word is important (cf. agency analysis,

analysis of routinized interactional patterns etc.). Statistical data show that *hát* is the most common word in the interviewees' utterances. Other discourse markers are common as well (e. g. *mondjuk* 'say', *például* 'as an example', *szerintem* 'I guess', *szóval* 'so' etc.). Teachers have a richer repertoire of discourse markers than students.

#### 4. Future plans

(1) Making an interactional description on the grammar of discourse markers, with a special regard to *hát*. (2) data collection outside Hungary (mostly in Slovakia and Romania) for a better understanding of the interactional patterns of classroom interaction and metalinguistic socialization in a bilingual context.

#### 5. References

- Aro, Mari 2009. *Speakers and Doers. Polyphony and Agency in Children's Beliefs about Language Learning*. Jyväskylä, University of Jyväskylä.
- Bata, Sarolta-Grácsi, Tekla Etelka 2009. A beszédpartner életkorának hatása a beszéd szupraszegmentális jellegzetességeire. In: Keszler, Borbála-Tátrai, Szilárd (eds.) *Diskurzus a grammatikában – grammatika a diskurzusban*. Budapest, Tinta. pp. 74–82.
- Borbély, Anna-Vargha, András 2010. Az *l* variabilitása öt foglalkozási csoportban. Kutatások a Budapesti Szociolingvisztikai Interjú beszélő nyelvi korpuszban. *Magyar Nyelv* pp. 455–470.
- Dér, Csilla Ilona 2010. „Töltelékelem” vagy új nyelvi változó? A *hát, úgyhogy, így és ilyen* újabb funkciójáról a spontán beszédben. *Beszédkutató* pp. 159–170.
- Dér, Csilla Ilona-Markó, Alexandra 2007. A magyar diskurzusjelölők szupraszegmentális jelöltsége. In: Gecső, Tamás-Sárdi, Csilla (eds.) *Nyelvelmélet – nyelvhasználat*. Székesfehérvár–Budapest, Kodolányi János Főiskola–Tinta. pp. 61–67.
- Ittész, Nóra 1981. *Szövegszerkesztési kérdések Esztergom regionális köznyelvében*. MA thesis. Budapest.
- Karasavvidis, Ilias et al. 2000. Investigating how secondary school students learn to solve correlational problems: quantitative and qualitative discourse approaches to the development of self-regulation. *Learning and Instruction* 267–292.
- Keszler, Borbála 1983. Kötetlen beszélgetések mondat- és szövegtani vizsgálata. In: Rácz, Endre-Szathmári, István (eds.) *Tanulmányok a mai magyar nyelv szövegtana köréből*. Budapest, Tankönyvkiadó. pp. 164–202.
- Kugler Nóra 2000. A mondatszók. In: Keszler, Borbála (ed.) *Magyar grammatika*. Budapest, Nemzeti Tankönyvkiadó. pp. 292–304.
- Labov, William 1976. *Sociolinguistique*. Paris, Editions de Minuit.
- Laihonen, Petteri 2008. Language ideologies in interviews: A conversation analysis approach. *Journal of Sociolinguistics* 668–693.
- Milroy, James 2001. Language ideologies and the consequences of standardization. *Journal of Sociolinguistics* 530–555.
- NyKk. = *Nyelvművelő kézikönyv*. Eds.: Grétsy, László-Kovalovszky, Miklós. Budapest, Akadémiai, 1980–1985.
- Potter, Jonathan-Edwards, Derek 2001. Discursive Social Psychology. In: Robinson, W. Peter– Giles, Howard (eds.) *The New Handbook of Language and Social Psychology*. 103–118. Chichester–New York, Wiley and Sons.
- Potter, Jonathan-Edwards, Derek 2003. Sociolinguistics, Cognitivism, and Discursive Psychology. *International Journal of English Studies* 1, 93–109.
- Schirm, Anita 2011. *A diskurzusjelölők funkciói: a hát, az -e és a vajon elemek története és jelenkori szinkrón státusza alapján*. [‘The function of discourse markers: The history and present synchronic status of the Hungarian elements *hát*, *-e* and *vajon*’] PhD dissertation. Szeged.
- Van Leeuwen, Theo 2004. Metalanguage in social life. In: Jaworski, Adam– Coupland, Nikolas– Galasiński, Darius (eds.) *Metalanguage. Social and Ideological Perspectives*. Berlin–New York, Mouton de Gruyter. pp. 107–130.

Further information on this research (summary of PhD dissertation, papers etc.): [http://sztp.hu/index\\_eng.htm](http://sztp.hu/index_eng.htm)

Please feel free to contact me: [sztp@nytud.hu](mailto:sztp@nytud.hu)