

Hungarian Linguistics PhD Program  
Graduate School in Linguistics  
University of Szeged

**Style in sociolinguistics:  
Strategies of child–adult discourse**

*PhD dissertation summary*  
Kinga Jelencsik-Mátyus

Supervisors:  
Dr. Anna Babarczy  
Dr. habil. Anna Fenyesi

Szeged  
2015

## 1. Introduction

The heterogeneity and variability of language is a basic principle of sociolinguistic research. Using heterogeneous structures is not a result of knowing more dialects and not a part of linguistic performance either but should be seen as an inherent part of the linguistic competence of the monolingual speaker (Weinreich et al. 1968: 101). This inherent variability has been demonstrated in numerous Hungarian and international sociolinguistic investigations.

“Style refers to a way of doing something” (Coupland 2007: 1). In this dissertation linguistic style is understood as operating in all linguistic subsystems. In speech – the object of the present study – style is constructed and continually reconstructed in the interaction, with interlocutors mutually influencing each other in styling, all throughout the conversation.

This dissertation studies how sociolinguistic style is represented in the speech of 4;5–5;6-year-old children attending preschools<sup>1</sup> in Budapest, Hungary. The study focuses on how different task types influence their language use, and what linguistic features this variability occurs in. I also investigate how different text level and discourse level features behave in the interviews of different children (inter-speaker variability), or within the individual interviews (intra-speaker variability), and how they affect the discourse dynamics. The dissertation aims to describe what typical co-occurrence patterns these features form in the recorded discourses, and what language strategies can be detected in them.

The dissertation has three main parts. In the first part, quantitative methods are used to detect the variability in the different task types, concerning utterances, turns and vocabulary. Next, interaction structure is studied: turn structure, adjacency pairs and turn-taking. The relationship of style structure and the sociocultural factors is investigated in the framework of cognitive stylistics along the lines of Tolcsvai Nagy (1996, 2004, and 2005). Finally, style is also studied from a metalinguistic point of view: whether children of the studied age are already aware that in different communicative situations different linguistic features and structures are used.

The research is based on the HUKILC (Hungarian Kindergarten Language Corpus), recorded in preschools in Budapest, in the spring of 2012. The corpus consists of 62 interviews with 4;6–5;6-year-old children. The interviews are 20–30 minutes long and follow a strict methodology; 55 of the 62 interviews have been analyzed for this dissertation.

---

<sup>1</sup> In Hungary, preschool is attended by 3-6-year-old children, and is currently compulsory from the age of 5 (from September 2015, it will be compulsory from the age of 3), before starting primary school.

Labov (1966, 1972) was the first to use style as a central concept in sociolinguistics. According to his attention-to-speech model, style depends on the amount of attention paid to speech. Thus, a more formal communicative situation requires more attention, this way style is more formal, whereas in casual talk the vernacular variety is used.

The audience design model (Bell 1984, 2001) was a critical response to Labov's theory. According to Bell's approach, style depends on the interlocutor and the audience. He does not only consider the addressee, but also the auditor, the overhearer and the eavesdropper. The theories of Labov and Bell are mainly reflexive, regarding style primarily as a response to the language use of the interlocutor – and, as such, variability in style, as occurring in the language use of the speaker, comes from the variability in the social domain, and stylistic patterns reflect social patterns (Bell 1984: 151).

According to the social constructivist approach, in contrast to the former models, linguistic variability is not a mere reflection of social patterns and social identity but rather a linguistic practice constructing social landscape (Eckert 2012). Variability is considered to be a principal feature of the language – and so it goes back to Weinreich, Labov and Herzog (Weinreich et al. 1968). The social constructivist approach sees variability as the basis of the social semiotic system, providing an opportunity to express the social relationships of the community. And as these relationships are constantly changing, variables do not index permanent meanings, rather their most important feature is variability, as expressed in style, in language use, as speakers interpret and reinterpret linguistic variables and incorporate them in their communicative practices (Eckert 2012: 94, Silverstein 2003). According to the social constructivist approach, style is a device to construct social meaning (Bartha and Hámori 2010: 302), and the initiative part of style is emphasized (Schilling-Estes 2004). Important studies following this approach (e.g. Coupland 1980, 1985, Eckert 2000 and Schilling-Estes 1999) focus on speaker agency, and the initiative, rather than reflexive nature of style. The dissertation follows the social constructivist approach.

In cognitive stylistics the basis and main requirement of style is linguistic potential, the stylistic opportunities originating in linguistic variability. Language use is influenced by numerous intra-linguistic and extra-linguistic factors, creating variability that can be detected at every level of language, from phonetics to discourse.

Acquisition of variability is not a side effect of language development (as opposed to the view expressed in Labov's 1964 model), it is rather a part of language acquisition (Roberts 2005: 153–154). The theory according to which the acquisition of variability is part of language acquisition is important both from a psycholinguistic and sociolinguistic point of view, as it

explains how the linguistic system of a child is constructed in an extremely diverse and socially structured linguistic environment (Nardy et al. 2013).

## **2. Aims and hypotheses**

In this dissertation I investigate how the task types influence the language use of 4;6–5;6-year-old children in conversation with an adult the children are not familiar with, and what patterns can be found in the dynamics of such conversation. The morphological analysis of the HUKILC corpus (Orosz and Mátyus 2014) is the basis of this part of the study. Using this analysis, I study the number and length of the utterances and turns in the children's speech as well as their vocabulary, and also investigate how these results correlate with the independent social variables such as age of the children, level of education of their parents, the number of books at their homes, or the task types.

In addition to providing a quantitative analysis my dissertation also aims at studying other features of the conversation, namely, how the sociocultural factors of style (Tolcsvai Nagy 1996, 2004, 2005) influence the language use in these conversations. I investigate how the differences of the sociocultural factors occur at the text and discourse levels, and what patterns or strategies they form in the speech of the children.

Finally, I also investigate whether at the meta level a certain awareness of different language usage connected to different speech situations appears in the conversations. This part of the research aims at studying what kind of meta-knowledge these children have about different varieties, to what extent they are aware that different communicative situations imply different elements of our linguistic repertoire. I attempt to identify which part(s) of the communicative situation they think influence(s) our language use, and how this knowledge can be detected in the recorded discourses.

The hypotheses of the dissertation are as follows:

1. I expect the mean length of utterance (MLUm) and mean length of turn<sup>2</sup> (MLTm) results to be influenced by the socio-economic background and the gender of the speaker on the one hand, in a way that children from wealthier backgrounds and girls have higher MLUm and MLTm measures. On the other hand, I also expect the different task types to influence MLUm and MLTm measures with higher results in the picture description tasks and lower results in the free talk.
2. Both gender and socio-economic background influence the size of the vocabulary.
3. There is a great difference between the picture description tasks and free talk concerning the structure of the conversations (sequential organization, turn structure, turn-taking).
4. The sociocultural factors of style highly influence the use of various linguistic elements at the text as well as discourse levels.
5. In the interviews, the style used at the beginning of the interview is different from that used later, progressing towards the informal domain (as style is usually more formal at the beginning, and more informal towards the end of the interviews).
6. Children use several strategies during the interviews, but the most dominant is the “transitional” strategy, where the child who is reserved at the beginning is getting more and more confident as the interview goes on.
7. In the metalanguage part of the dissertation, mainly the topics raised by the interviewer are discussed, but most of the children do not engage in these discussions. There are very few topics occurring during the interviews that cannot be found in the metalanguage question list of the interviewer.
8. Concerning the “voices” appearing in the metalanguage part of the study, general rules and public opinions are the most dominant. Polyphony is not usual in the interviews.
9. Children mostly position themselves as patients in the metalanguage part, as ones who do not have any influence on the rules of language use.

---

<sup>2</sup> Both mean length of utterance and mean length of turn are measured in morphemes.

### **3. Overview of the dissertation**

The Introduction (Chapter 1) is followed by the summary of research questions and the hypotheses. Chapter 2 presents the different sociolinguistic approaches of style, briefly discussing the literature on Hungarian as well. Then, in Chapter 3, the developmental theories of variability are discussed. The summary of Hungarian child language studies, connected to the topic of the dissertation, can be read in Chapter 4.

Chapter 5 discusses the theoretical backgrounds used in the quantitative, qualitative and metalanguage parts of the study, concentrating mainly on interactional stylistics and cognitive stylistics.

The dissertation is based on the analysis of selected interviews of the HUKILC corpus, thus, Chapter 6 aims to introduce this corpus in detail. The chapter also discusses the steps of automatic text processing which, as one of the main aims of the research, makes the research of the HUKILC corpus possible in other topics, and these processing methods can also be used in other corpora. Then, the dependent variables of the dissertation are discussed.

Chapter 8 begins with a general analysis of the interviews, using the methods of conversation analysis. Then, results are analyzed along demographic factors and task types as well. This part is followed by the analysis of the interaction structure and relationships of style structure and sociocultural factors. And finally, the language strategies used by the children are investigated mainly using the methodology of interactional stylistics and cognitive stylistics. In Chapter 9 the quantitative and qualitative results of the study are compared.

Chapter 10 focuses on what metalinguistic knowledge the children have about different language modes, and what features of the communicative situation they believe to affect language usage.

Chapter 11 provides a summary of the results of the dissertation: an overview of to what extent the data of the interviews prove the hypotheses, and also an insight into the further possibilities of the applications of the HUKILC corpus. The dissertation concludes with a list of references and an appendix containing the picture sets of the tasks and the demography questionnaire.

#### **4. The HUKILC corpus**

The HUKILC corpus was compiled primarily to study variability in child language. It consists of 62 20–30-minute-long interviews recorded with 4;6–5;6-year-old children attending preschools in Budapest. The recordings were made in the spring of 2012. The aim of the HUKILC is to collect a large amount of data to help study the language use of children of preschool age, and to develop and provide new automatic methods to analyze and research larger child language corpora by adapting the methodology used with written adult language corpora. Several linguistic levels can be examined in the interviews, and they are also a good starting point to study non-typical language development.

The corpus contains 178,377 tokens in 37,714 utterances. Due to the characteristics of the study and the age of the children, about half of the data comes from the interviewer. In recordings made with young children the role of the interviewer is important, he/she talks much more compared to recordings conducted with adults.

55 of the 62 HUKILC interviews were chosen to be analyzed for this dissertation. Some interviews had to be left out as during the checking process and while processing the demographic questionnaires it turned out that due to different reasons some interviews could not be analyzed. These reasons are, for example, hearing problems, severe articulation deficits that could hinder understanding, and, sometimes, reserved personality or a complete lack of motivation (to an extent that hardly any data could be collected). However, no children were excluded for leaving some tasks uncompleted. The analyzed interviews contain 156,177 tokens in 33,495 utterances.

Before recordings started, all the parents filled in a consent form. They were also asked to answer a demographic questionnaire, but unfortunately not all of these were returned. All the interviews were recorded in a separate room – away from the usual classroom of the children.

As for the methodology of the interviews, two types of tasks were used: a picture description task (with four sets of pictures), and free conversation about topics that were not strictly determined (telling the rules of a game the child knew very well, talking about “nice ways of speaking”, and one topic of choice suggested by the child).

The automatic processing of the texts also played an important part in the research. My aim was, on the one hand, to help other researchers to use the HUKILC corpus for other research purposes, and, on the other, through the methods of analysis developed here, to provide technical support for other corpora. The CHILDES transcription was the basis of the analysis,

then a Hungarian morphological analyzer (HuMor) and a disambiguator (PurePos) were attached and adapted to spoken child language.

## **5. Methodology**

Different methods were used in the three parts of the dissertation. In the first, quantitative section the data from the automatic morphological analysis was used, applying the SPSS 19 statistical software. In the second part mainly qualitative analysis was used, to investigate the discourse structure, the style structure, and the linguistic strategies of the children. According to empirical studies, the effects of style occur not only at the text level but also at the discourse level (Bartha and Hámori 2010, Tannen 2005), thus, discourse level features like turn length, turn-taking or pauses also imply social norms, and, as such, they also carry social meaning. To study this, I mainly applied the methods of interactional stylistics and conversation analysis (for a similar study, see, for example, Boronkai 2009, 2011, 2013). To research the relationship of style structure and the sociocultural factors, I relied on Tolcsvai Nagy's (1996, 2004, 2005) cognitive stylistic methodology.

In the third, metalinguistic part, methodology played a crucial role, as here it was especially important and also very demanding to maintain an objective point of view. Based on the methodology used by Aro (2009), first the topics of the last module were discussed, then the investigation of the voices and polyphony followed, and finally, I investigated what examples of agency could be found in this part of the child–adult discourse.

## 6. Results

In this section the main results of the research are presented, with a special focus on the hypotheses. As expected, the MLUm and MLTm measures are found to correlate with the gender of the speaker so that girls produced longer utterances and turns than boys. Differences have turned out to be bigger in the picture description tasks and smaller in the free topics. However, socio-economic background has hardly any effect on the results, primarily in longer tasks – so this hypothesis was only partially proved. The weak effect of the socio-economic background is one of the most surprising results of the research, as similar studies conducted with older children all show the strong influence of the socio-economic status of the family. It seems that in a spoken language study, with the variables investigated here (compared to the National Competence Survey, or to morphological or morphosyntactic studies) the gender of the speaker shows stronger correlations. No effect of the education of the parents was found in the interviews.

The different tasks (task types) were aimed to elicit different text types (concerning the extent of “planning”) from the children. While in picture description tasks almost monologic texts were expected, in the free topics language use approaches spontaneous speech. In the former there are only very few, but very long turns, while in the latter there are more but shorter turns. In accordance with the initial hypothesis, task types have a strong effect on both MLUm and MLTm measures, with higher results in picture description tasks and lower ones in the free topics. Statistical analysis also supports these differences.

Although vocabulary is an important measure in child language research, in the present study, using language technology applications, there were hardly any differences concerning the lexicon – neither gender, nor socio-economic background have an effect.

In accordance with the hypothesis, in the conversation structure (sequential organization, turn structure, turn-taking) there are considerable differences between the picture description tasks and the free topics.

The quantitative and qualitative differences between the task types together point to the fact that the language usage detected in picture description tasks presents the highest level of the language knowledge of the child, mainly appearing in conversations with adults, while a less structured variety is characteristic of the free topics. Consequently, the two types of language usage, as seen in the two task types, have to be separated. These results may be important in language development and language deficit studies.

The results are also in accordance with the hypotheses concerning the sociocultural factors of style. These factors have a strong effect on the use of various language elements both at the text and discourse levels. The neutral subdomain is dominant as far as attitudes and the other factors as well. At the discourse level the greatest variability is found in turn-taking: in interviews, where the children often apply self-appointed turn-taking, attitude approaches the familiar subdomain. At the text level the addition of adjuncts and the use of more elaborate words point towards the sophisticated subdomain, but sometimes the familiar subdomain appears at the text level as well. The results only partially prove my hypothesis: according to the data there are some interviews where style variance can be found (compared to the starting style), but neutral domains are dominant in most recordings.

Even in the case of the same task types and identical socio-cultural factors, there are great differences in the language use of children in the interviews I analyzed. These differences originate from different language strategies. The children used several kinds of strategies during the interviews, like determined, unsure or the boisterous. But, unlike I expected, transition in strategies (for example, unsure at the beginning, turning more and more confident as the interview progresses) is quite rare in the corpus. At the same time, it is very interesting that the transitional strategy is characteristic of one specific group of the speakers, namely, the group of lower socio-economic background male children.

The hypothesis concerning the metalinguistic study has also been proved: in the metalanguage module mainly the topics raised by the interviewer were discussed, but, unlike in the hypothesis, most of the children were happy to get involved in this kind of conversations – in most interviews I got long answers. Besides, there were several topics typically raised by the children, like the differences of girls' and boys' talk. As for the voices in the metalanguage module, the general rules and public opinions were the most frequent, with the children presenting themselves as patients. Children presenting themselves as agents were rather rare, occurring in very few interviews only, similarly to the occurrence of polyphony.

## 7. Conclusions

The dissertation processes and analyzes 55 of the 62 interviews of the HUKILC corpus in detail. The aim of the study has been to investigate what differences can be found in the language use of children in a child-adult interview as communicative situation, and what features (for example, discourse features) these differences affect, and what demographic factors these differences correlate with. The language strategies the children used during the interviews are also discussed. These features are studied at quantitative, qualitative and metalinguistic levels.

Mainly the frame of interactional stylistics is used in the dissertation, and, closely connected to this, in the qualitative part of the study the current approaches of cognitive stylistics are applied.

The results of the research show the variability occurring in the different task types of the discourse concerning the variables studied here (MLUm, MLTm, vocabulary), and also show what factors may cause the differences. The quantitative analysis demonstrates correlations between the number and length of utterances and turns, and the task types as well as the demographic background. The qualitative analysis focuses predominantly on the structure of interaction and style.

Language technology applications are inevitable to process and analyze a corpus of this size in detail. However, in Hungarian there have been effective language tools only for written adult corpora (lemmatization, morphological analysis and disambiguation). To support the analysis presented in the dissertation, a lot of technical innovations had to be made, where, working together with György Orosz, the above mentioned applications were adapted to spoken child language – and that made the automatic measurement of MLUm, MLTm and vocabulary possible. These new language technology tools help other researchers to study the HUKILC corpus, as well as to process and analyze other spoken child language corpora.

HUKILC was predominantly compiled in order to provide a sufficient amount of data for the present dissertation. However, as there were only very few spoken child language corpora available at the time the dissertation was written, the recordings were anonymized and made available for other researchers as well. The corpus has already been used in other projects, for example, Tamás Pál Szabó and Zoltán Jakab at the ELTE Bárczi Gusztáv Faculty study the understanding of sentences created with the help of sine wave speech synthesis. The list of the sentences was made using the word frequency results of the HUKILC. Mario Brdar, Rita Brdar-Szabó and Nóra Kugler carry out a corpus-based study to investigate the appearance and

construction of the plural marker *-ék* (Brdar et al. 2014). HUKILC was also used in this research.

The data in the corpus may serve as a point of reference for studies investigating when and how children start to plan and structure texts, and when the opposition of planned/unplanned texts occurs in their linguistic knowledge. The corpus also makes studying child discourse (e.g. discourse markers, phatic expressions, turns or turn-taking) possible – as there have been only very few similar studies in Hungarian so far (Tolcsvai Nagy 2001: 50). Besides, studying the sociolinguistic variables not included in the dissertation – like the inessive case ending (bVn) – is definitely important as well.

## References

- Aro, Mari (2009) *Speakers and doers: Polyphony and agency in children's beliefs about language learning*. Jyväskylä: University of Jyväskylä.
- Bartha, Csilla and Hámori, Ágnes (2010) Stílus a szociolingvisztikában, stílus a diskurzusban. *Magyar Nyelvőr*, 134(3): 298–321.
- Bell, Allan (1984) Language style as audience design. *Language in Society*, 13(2): 145–204.
- Bell, Allan (2001) Back in style: Reworking audience design. In: Eckert, Penelope and John Rickford (eds.) 139–169.
- Boronkai, Dóra (2009) *Bevezetés a társalgáselemzésbe*. Budapest: Ad Librum.
- Boronkai, Dóra (2011) Az interakciós szerkezet és a szociokulturális tényezők összefüggése a spontán társalgásokban. *Alkalmazott Nyelvtudomány*, 11: 151–168.
- Boronkai, Dóra (2013) A beszélőváltások és a szociokulturális tényezők összefüggései társalgási szövegekben. *Magyar Nyelvőr*, 137 (1): 68–86.
- Brdar, Mario, Brdar-Szabó, Rita and Kugler, Nóra (2014) A többség *-ék* jeles kifejezésének nyelvvaltozási folyamata. Presented at the 5th Conference “*Generációk nyelve – Új nézőpontok a magyar nyelv leírásában*” November 27, 2014.. ELTE.
- Coupland, Nikolas (1980) Style-shifting in a Cardiff work-setting. *Language in Society*, 9(1): 1–12.
- Coupland, Nikolas (1985) “Hark, hark the lark”: Social motivations for phonological style-shifting. *Language and Communication*, 5(3): 153–72.
- Coupland, Nikolas (2007) *Style: Language variation and identity*. Cambridge: Cambridge University Press.
- Eckert, Penelope (2000) *Linguistic variation as social practice*. Massachusetts: Blackwell.
- Eckert, Penelope (2012) Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, 41: 87–100.
- Labov, William (1966) *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

- Labov, William (1972) *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Mátyus, Kinga and Orosz, György (2014) MONYEK: Morfológiai egyértelműsített óvodai nyelvi korpusz. *Beszédkutatás* 2014. 237–245.
- Nardy, Aurélie, Chevrot, Jean-Pierre and Barbu, Stéphanie (2013) The acquisition of sociolinguistic variation: Looking back and thinking ahead. *Linguistics*, 51(2): 255–285.
- Orosz, György and Mátyus, Kinga (2014) An MLU estimation method for Hungarian transcripts. *Text, Speech, and Dialogue, of Lecture Notes in Computer Science*, 865: 173–180.
- Roberts, Julie (2005) Acquisition of sociolinguistic variation. In: Ball, Martin J. (ed.) *Clinical sociolinguistics*. Oxford: Blackwell. 153–164.
- Silverstein, Michael (2003) Indexical order and the dialectics of sociolinguistic life. *Language and Communication*, 23: 193–229.
- Tannen, Deborah (1987) Conversational style. In: Dechert, Hans W. and Raupach, Manfred (eds.) *Psycholinguistic models of production*. Norwood, NJ: Ablex. 251–267.
- Tannen, Deborah (1999) New York Jewish conversational style. In: Jaworski, Adam –and Coupland, Nikolas (eds.) *The discourse reader*. London and New York: Routledge. 454–466.
- Tannen, Deborah (2005) *Conversational style: Analyzing talk among friends*. New York: Oxford University Press.
- Tannen, Deborah (2007) *Taking voices: Repetition, dialogue, and imaginery in conversational discourse*. Cambridge: Cambridge University Press.
- Tolcsvai Nagy, Gábor (1996) *A magyar nyelv stilisztikája*. Budapest: Nemzeti Tankönyvkiadó.
- Tolcsvai Nagy, Gábor (2004) A nyelvi variancia kognitív leírása és a stílus (Egy kognitív stíluselmélet vázlatja). In: Büky, László (ed.) *A mai magyar nyelv leírásának újabb módszerei VI*. Szeged: SZTE Általános Nyelvészeti Tanszék, Magyar Nyelvészeti Tanszék. 143–160.
- Tolcsvai Nagy, Gábor (2005) *A cognitive theory of style*. Frankfurt am Main: Peter Lang.
- Weinreich, Uriel, Labov, William and Herzog, Marvin I. (1968) Empirical foundations for a theory of language. In: Lehman, W. and Malkiel, Y. (eds.) *Directions for historical linguistics*. Austin: University of Texas Press. 95–188.

## Publications related to the research area of the dissertation

- Mátyus Kinga (2009) Öt- és tízéves gyerekek és a (bVn) változó. Borbély Anna, Vanconé Kremmer Ildikó, Hattyár Helga (szerk.) *Nyelvideológiák, attitűdök és sztereotípiák*. 15. Élőnyelvi Konferencia. MTA Nyelvtudományi Intézet – Gramma Nyelvi Iroda – Konstantin Filozófus Egyetem Közép-európai Tanulmányok Kar: Budapest – Dunaszerdahely – Nyitra. Budapest: Tinta Könyvkiadó, 121–129.
- Mátyus Kinga (2012) Anonimizálási gyakorlat? In: Tanács Attila – Vincze Veronika (szerk.) *Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem, Informatikai Tanszékcsoport. 338–342.
- Mátyus Kinga (2013) The (bVn) variable in the speech of five- and ten-year-old Hungarian children. In: Polczynska, M. et al. (szerk.) *Young Linguists' Insights – Taking interdisciplinary approaches to the fore*. Poznan: Uniwersitet im. Adama Mickiewicza w Poznaniu. 55–62.
- Mátyus Kinga (2014) Stílus a gyermeknyelvben: Az átlagos megnyilatkozáshossz (MLU) mint a stílus jelzője? In: Ladányi, Mária – Vladár, Zsuzsa – Hrenek, Éva (szerk.) *Nyelv – társadalom – kultúra. Interkulturális és multikulturális perspektívák*. Budapest: MANYE–Tinta, Vol. 1: 203–208.
- Mátyus Kinga – Orosz György (2014) MONYEEK: Morfológiailag egyértelműsített óvodai nyelvi korpusz. *Beszéd kutatás* 2014. 237–245.

Budapest, March 2015.

Kinga Jelencsik-Mátyus