**Chroma’s Corpus of Early Czech**

Mgr. Anna Chromá

Department of General Linguistics

Charles University, Prague

[anicka.chroma@gmail.com](mailto:anicka.chroma@gmail.com)

Doc. PhDr. Filip Smolík, Ph.D.

Institute of Psychology

Academy of Sciences of the Czech Republic, p.r.i.

[smolik@praha.psu.cas.cz](mailto:smolik@praha.psu.cas.cz)

Mgr. Kateřina Šormová, Ph.D.

Department od Czech Language and Theory of Communication

Charles University, Prague

[katerina.sormova@ff.cuni.cz](mailto:katerina.sormova@ff.cuni.cz)

Participants: 6

Type of Study: longitudinal

Location: Czech republic

Media type: transcripts (audio)

**Citation Information (not yet available)**

**Corpus**

**Warnings.**

1. During transcription, we did not pay particular attention to the ‘pho’ layer. The ‘pho’ transcription is only approximate. Researchers interested in the phonetical/phonological aspects of the data should ask Anna Chromá (anicka.chroma@gmail.com, the submitter) or Kateřina Šormová (katerina.sormova@ff.cuni.cz, member of the AKCES project) for the access to the audiorecordings.

2. So far, there was no morphological coding.

3. There is no translation to English or any other languages. All the transcribed material as well as all the Comments are in Czech.

**Identity protection.** For each participating child, both his/her parents (as well as other participating caregivers) gave an informed consent for the use of the data. For the target children, we consistently use pseudonyms which respect at least some of the morphological and phonological aspects of their original names. The corresponding nicknames were created from the pseudonyms as well. With few exceptions, the first names of other participants are not replaced. Surnames and addresses are replaced by a code ‘Zzz’ and commented.

**Restrictions.** There are no restrictions on the use of the transcripts.

**Specific Codes.** There are three codes of our own for the usage of interjection with the function of a predicate *@z:ip*, of a nominal phrase *@z:in*, and of an adjective *@z:ia*.

**History and Aknowledgements.** The recording for this corpus began in 2014 without any project funding. In 2016, the project ‘Longitudinal Corpus of Early Language Development’ by Anna Chromá obtained financial support for two years from Faculty of Arts, Charles University within the Specific Higher Education Research. We owe many thanks to prof. Karel Šebesta for his support before the project funding was obtained. Thanks to him, the corpus is integrated into the larger Czech project [AKCES](http://akces.ff.cuni.cz/) oriented to building a set of various Czech acquisitional corpora. The AKCES project is represented here by Kateřina Šormová. We also thank much to Filip Smolík who was the main methodological consultant and who is the coauthor of the Czech Manual for transcription (based on the CHAT manual). The following students have cooperated on the transcriptions and/or revisions: Anna Marklová, Barbora Blahnová, Martina Vokáčová, Jan Henyš, Kateřina Bělehrádková, Denisa Šebestová, Alžběta Macháčková, Jana Segi Lukavská, and Tereza Binderová.

The participants were recruited by word of mouth under the friends and colleagues of the main investigator. The criteria for recruitment were (1) an adequate age of the child, (2) just one predominant (Czech) language in his/her environment and (3) his/her typical development so far.

The recording was done completely by the child’s caregivers without any observer. The caregivers were instructed to record various home situations as eating, home work, reading and playing for 20–40 minutes once per two weeks. This was the ideal frequency to which did the caregivers approximate more or less. The caregivers alone decided which recordings would they hand over for subsequent processing, but usually they sent all of the recorded material to the corpus database.

Typically, one trained transcriber processed continuously all the recordings from one family. Each transcription was once revised by a second trained person and the revision was finally incorporated by the original transcriber. At the end, all the transcriptions were revised once more by means of CLAN checking.

Especially the following phenomena were coded during the transcription: dysfluencies (repetitions, false starts, self-corections), interjections *@i* and specific use of interjections (see Specific Codes), morphological innovations *@n*, errors *[\*]* and child forms *@c*.

**Subcorpora.**

1. **Aneta**

**Caregivers and environment.** The girl with pseudonym Aneta lives the whole life in Prague or the metropolitan area. During the recorded period, she lived in a flat in the broader city center. Until 3 years of age, she spent most of the time with her father, mother and a babysitter. From 3 to 4 years, she went to a daycare center once a week for the whole day and four times a week for half a day. Occasionally, she might spend one or two weeks at her grandparents in Silesia (east of Czechia). Both of her parents come from the eastern part of Czech Republic (Ostrava and Přerov), so in their as well as in her speech, some regionalisms occur. Her father was 41 years old when Aneta was born. He has high school education but he studied university for a long time and he is employed at a research institution. Her mother was 33 years old when Aneta was born. She has university education and is a scientific worker at a university. The first babysitter and close family friend Ivan (pseudonym) was a graduate student during the recording. At that time, he was abroad for almost a year and Aneta spent much time with another babysitter Tereza and her daughter Františka (the same age as Aneta). Aneta has an older brother Libor (pseudonym). He was six years old when she was born.

**Recordings.** Aneta began with recording at 2;2.8 and continued until 3;3.18 with the median interval of 11 days between recordings (range from 2 to 50 days). The median lenght of a recording is 0:18:49 (range from 0:01:25 – in an additional recording when there was more then one in a day – to 0:33:24). Together, there are 38 recordings of total lenght of 11:31:33. The recording was usually provided by Ivan or by the mother. Occasionally, the father, the brother and the grandfather also appear on the recordings. The recorded situations mostly take place at home while drawing, reading, playing or eating. The recorded period was between 2014 and 2016. Aneta’s original audiorecordings are available for other researchers to an analysis.

1. **Anna**

**Caregivers and environment.** The girl with pseudonym Anna lives the whole life in Prague. During the recorded period, she lived in a flat in the city center (shortly before, they moved from a smaller flat in the suburbs). In the first two years of age, she spent most of the time with her father and mother. From 24 months, she went to a daycare center twice a week for the whole day, from 36 months, it was 5 days a week (in another institution). Both of her grandmothers and her aunt looked after her once a week each. Both of her parents come from the southern part of Czech Republic (České Budějovice). Her father was 32 years old when Anna was born. He has university education and is a scientific worker. Her mother was 27 years old when Anna was born. She has university education and is a cultural manager. Anna was the first child. When she was 25 months old, her younger sister Iva (pseudonym) was born.

**Recordings.** Anna began with recording at 1;9.30 and and continued until 2;7.27 with the median interval of 21 days between recordings (range from 14 to 42 days). The median lenght of a recording is 0:27:33 (range from 0:07:00 to 0:55:13). So far, there are 11 recordings of total lenght of 5:09:42. Together, there are 14 recordings of total lenght of 6:29:54. The recording was usually provided by the father, sometimes by a grandmother. Occasionally, the mother and the sister also appear on the recordings. The recorded situations take always place at home while reading, putting to sleep, playing or eating. The recorded period was between 2017 and 2018. Anna’s original audiorecordings are available for other researchers to an analysis.

1. **Jan**

**Caregivers and environment.** The boy with pseudonym Jan lives from birth on in a flat in the broader center of Prague. From 14 months, he went to a daycare center two- or three times per week. Otherwise, he spent most of his time with his father and mother. Approximately once a week, he spent several hours with his grandparents. From 29 to 37 months, he went to another daycare center for 4 days in a week. His father was 32 years old when Jan was born. He lived in Prague for his whole life. Jan‘s mother was 27 years old when Jan was born. She comes from Brno, but no regionalisms usually occur in her speech. Both the parents are lawyers. Jan was a single child during the whole recording period.

**Recordings.** Jan began with recording at 1;7.5 and continued until 2;9.27 with the median interval of 17 days between recordings (range from 1 to 88 days). The median lenght of a recording is 0:19:26 (range from 0:06:20 to 0:51:27). Together, there are 21 recordings of total lenght of 7:11:13. The recording was usually provided by the mother. Occassionally, the father also appears on the recordings. The recorded situations take always place at home while drawing, reading, playing, eating and home work. The recorded period was between 2016 and 2017. Jan’s original audiorecordings are not available.

1. **Julie**

**Caregivers and environment.** The girl with pseudonym Julie lives from birth on in the metropolitan area of Prague. She spend most of the time with her mother. Each day, she spend some time with her father and approximately once a week, she spend several hours with her grandparents. From 36 months, Julie went two times in a week for half a day to a daycare center. From 40 months, she change the institution and went there once for half a day and once for the whole day. Her father was 30 years old when Julie was born. He comes from the place they live till now. He has university education and works as an IT-specialist. Julie‘s mother was 29 when Julie was born. She comes from Prague. She has university education and is a photographer. Julie was the first child. When she was 19 months old, her younger sister Žofie was born.

**Recordings.** Julie‘s recording began at 1;7.5 and continued until 3;9.11 with the median interval of 22 days between recordings (range from 14 to 89 days). The median lenght of a recording is 0:29:23 (range from 0:07:57 to 0:50:18). Together, there are 32 recordings of total lenght of 14:00:53. The recording was always provided by the mother. From her birth, the sister is often present while recording. Occassionally, the father also appears on the recordings. The recorded situations take always place at home while drawing, reading, playing or eating. The recorded period was between 2016 and 2018. Julie’s original audiorecordings are available for other researchers to an analysis.

1. **Klára**

**Caregivers and environment.** The girl with pseudonym Klára lives the whole life in Hradec Králové, the 8th largest Czech city with 94 thousand of inhabitants. During the recorded period, she lived in a flat at the edge of the city. From 32 months, she went regularly from Monday to Friday to a nursery. Before, she had no babysitters and spent her time with her mother and father. Her father comes from the eastern part of Czech Republic. He was 36 years old when Klára was born. He has university education and is an office worker. Klára’s mother comes from a small town in the eastern Bohemia and she studied in Brno. She was 33 years old when Klára was born. She has university education and works as an elementary school teacher. Klára has an older brother Vojtěch. He was 34 months old when Klára was born.

**Recordings.** Klára began with recording at 2;4.22 and continued until 3;4.24 with the median interval of 7 days between recordings (range from 1 to 65 days). The median lenght of a recording is 0:07:00 (range from 0:01:11 – in an additional recording when there was more then one in a day – to 0:17:34). Together, there are 38 recordings of total lenght of 4:41:09. The recording was always provided by the mother. The father and especially the brother often appear on the recordings. The recorded situations mostly take place at home while playing, eating or in a bad. The recorded period was between 2014 and 2015. Klára’s original audiorecordings are available for other researchers to an analysis.

1. **Viktor**

**Caregivers and environment.** The boy with pseudonym Viktor lives from birth on in the city center of Prague. He spend most of the time with his father and mother. During the recorded period, he went two or three days in a week to a daycare center and had three babysitters (family friends). First of them looked after him two or three days in a week for several hours, the two others once a week for several hours. The first babysitter stopped meeting the family shortly before the end of the recording. Viktor‘s father was 45 years old when Viktor was born. He comes from a small town near Prague. He has high school education and produces musical instruments. Viktor‘s mother was 29 when Viktor was born. She was born in northern Moravia, but most of her life she lives in Prague. She has university education and was a postgraduate student during the recording. Viktor was the first child. When he was 34 months old, his younger brother Melichar (pseudonym) was born.

**Recordings.** Viktor‘s recording began at 2;6.23 and and continued until 3;8.06 with the median interval of 14 days between recordings (range from 2 to 36 days). The median lenght of a recording is 0:23:19 (range from 0:07:34 to 0:56:38). Together, there are 32 recordings of total lenght of 14:22:44. The recording was usually provided by the mother, sometimes also by the father. From his birth, the brother was usually present while recording. The recorded situations take usually place at home while reading or playing. The recorded period was between 2014 and 2016. Viktor’s original audiorecordings are available for other researchers to an analysis.