

PoInt Corpus Information and Copyright

© PoInt Team 2002

PoInt Corpus (Polish Intonation) is a collection of digital recordings of Polish speech and relevant meta-data created in the Polish Government-funded research project “Intonacyjna baza danych dla języka polskiego” (Intonation Database for Polish). PoInt Corpus consists of the following sections:

Monologues

- **Music** (moderated monologues on the “Torreador Song” from “Carmen” by G. Bizet)
- **Painting** (moderated monologues on the third section of H. Bosch's triptique “The Garden of Earthly Pleasures”.)
- **Comic strip** (narration of a comic strip)
- **Things** (moderated monologues on a couple of items of organic origin – a shell, some pieces of wood, etc.)

Dialogues

- **Map task** (a map-task dialogue based on specially designed maps)
- **Quiz** (a ten questions game)
- **Discussion** (a discussion on a given topic)

Read Texts

- A piece of **poetry**
- A piece of **prose**

The speakers are native Poles from the region of Greater Poland. However, most of them speak standard Polish with only few traces of local pronunciation. All of them are young and educated people (students, employees of the university, young professionals). Although many steps and precautions were taken in order to create relaxed atmosphere during the recordings, we refer to our material as “quasi-spontaneous” or “semi-spontaneous” because the presence of the microphone and the recording studio condition has certainly influenced the behaviour of the speakers.

The equipment used for recordings:

- two Shure SM58 Beta microphones for dialogues
- single AKG C3000 condenser microphone for monologues and read texts
- Spirit Folio F1 mixing console
- Tascam RW-CD700 digital recorder (working at 44.1kHz with 16-bit resolution)

The team of the PoInt project keeps all the rights to the recordings. The team is willing to share the data for non-commercial applications if the following conditions are met:

- the PoInt Corpus is obtained directly from the PoInt Team and not shared or conveyed further;
- the intended use of the PoInt Corpus is clearly declared;
- the source of the data is clearly declared in publications and presentations (you can refer to the publication mentioned below);
- no responsibility for the quality of the data nor for the consequences of its usage is attributed to the PoInt Team (however, we take full responsibility for the scientific quality of our work).

The speakers agreed that the recordings voices can be used only for scientific research preserving their anonymity. Accordingly, the recordings cannot be publicly presented in any media (web, radio, television) without further negotiation. Any commercial use of the PoInt Corpus should will be negotiated individually.

If you base your research on our data, please cite:

M. Karpiński, J. Kleśta 2001. A project of an intonational database for the Polish language [in:] St. Puppel, G. Demenko (eds.) *Prosody 2000: Speech Recognition and Synthesis*. The Faculty of Modern Languages and Literature AMU, pp. 113-118.

(or any other publication by Maciej Karpiński et al. which refers to the corpus)

© 2002 PoInt Team

Maciej Karpiński (head of the project)

Wiktor Jassem

Janusz Kleśta

Emilia Kim (Szalkowska) (collaborator)

Katarzyna Klessa (Francuzik) (collaborator)

In case of any questions related to the Corpus, please do not hesitate to contact us.

Maciej Karpiński
Institute of Linguistics
Adam Mickiewicz University
maciej.karpinski@amu.edu.pl
<http://www.staff.amu.edu.pl/~maciejk>