The ENRICHMENT OF LEXICAL INFORMATION AND CORPUS RESOURCES
BY USING THE RESULTS OF MORPHOLOGICAL ANALYSIS OF HISTORICAL TEXTS

STEP ONE collecting materials (from a scan to a ready-to-work TEI-tagged text)

A Bout funding:
- main investigator: Bień and his team for the purposes of the IMPACT project
- corpus size: around 12 million tokens
- the aim: the creation of a corpus of 17th and 18th century Polish texts (another stage of development of the Polish National Corpus)
- the creation of an Electronic corpus of 17th and 18th c. Texts, based on Korbeusz – an analyser for contemporary Polish morphological analysis
- the creation of a corpus of 17th and 18th century Polish texts (another stage of development of the Polish National Corpus)
- the creation of a corpus of 17th and 18th century Polish texts (another stage of development of the Polish National Corpus)

STEP TWO from a transliterated text to a transcribed and morphosyntactically tagged text

AMEBA SUPER TOOL (from Polish ameba, Gr. ἀμβέα; change): a tool converting texts from their transliterated to transcribed version

STEP THREE, FOUR, FIVE ... enrichment of lexical information and corpus resources

ANOTORNIA manual disambiguation and creating the corpus

...AND ALMOST OUR GOAL

ABOUT OUR PROJECT:
- title: Electronic corpus of 17th and 18th century Polish texts (up to 1772), acronym KorBa (KORpus BArokowy)
- the aim: the creation of a corpus of 17th and 18th century Polish texts (another stage of development of the Polish National Corpus)
- corpus size: around 12 million tokens
- duration: 2013 - 2018
- cooperation: Institute of Polish Language, Polish Academy of Sciences
- cooperation: Linguistic Engineering Group, Institute of Computer Science, Polish Academy of Sciences
- main investigator: Witoldmierz Gniewczynsky
- funding: A Ministry of Science and Higher Education National Programme for the Development of Humanities grant 0036/NWR2/HS1.91/2012

CONTACT US:
Renata Bronikowska
r.bronikowska@wp.pl
Emanuel Modrzejewski
emodrzejewski@gmail.com

Institute of Polish Language, Polish Academy of Sciences

The Electronic Corpus of the 17th and 18th c. Polish Texts (up to 1772)
http://sgjp.pl/