

# HerCoRe

## **Development of language specific lists with lexemes / expressions indicating vagueness / uncertainty**

Cristina Vertan, University of Hamburg

1. The lists were developed for the following three languages: German, Latin, Romanian
2. For German we considered an additional parameter the style and vocabulary of the 18<sup>th</sup> century, reflecting the language in HeRcoRe-Corpus
3. For Latin, we considered the classic Latin, language which Cantemir used mostly in his works
4. For Romanian we considered the modern language, used for the translations included in our corpus
5. The vagueness lists were developed for each language, in a first step independent one from another. They were done based on linguistic resources (dictionaries) and not by analysing the corpus.
6. The lists were then automatically checked against the GermaNet, Romanian WordNet and the Latin WordNet available via BabelNet and improved with Synonyms.
7. In order to ensure calculability for the reasoned and readability for the user, we filtered out the words which may introduce less important vague aspects.
8. Finally, the lists were compared one with the other and improved as necessary with translations from the other 2 languages