# Introduction: The Role of Computational Methods within the Humanities and Social Sciences

Ulf Dalvad Berthelsen
Aarhus University
udb@cc.au.dk

Morten Tannert
Aarhus University
mota@edu.au.dk

## 1. The aim of this special issue

The fast-rising interest in computational methods within the humanities and social sciences is largely driven by the fact that texts in the 21st century are digitized. We digitize cultural heritage and literary classics, and new texts within mass communication, education, social media, journalism and literature are by default produced in digital formats. This development means, firstly, that it has become relatively easy to build large text corpora and, secondly, that computational methods, in particular natural language processing (NLP), have become increasingly important in the study of everything from literature and archeology to education and sociolinguistics. Consequently, explicit and detailed analysis of linguistic structure has become the central starting point of much contemporary research within the humanities and social sciences.

This new linguistic turn obviously has implications for the individual disciplines not only in terms of the methods that are applied, but also at a more fundamental level. With new analytical approaches and research instruments follow opportunities for exploring new areas and research questions, but new methodological approaches also carry with them new types of challenges - challenges of both epistemological, ontological, methodological and ethical character. The aim of this special issue is to explore some of the limitations and possibilities of the new linguistic turn from the perspective of different humanistic and social

science disciplines and on the backdrop of different theoretical and methodological ideas.

## 2. A return to surface structure

In 1987 Giddens famously declared that "Structuralism, and post-structuralism also, are dead traditions of thought. Notwithstanding the promise they held in the fresh bloom of youth, they have ultimately failed to generate the revolution in philosophical understanding and social theory that was once their pledge" (Giddens, 1987, p. 195). This is probably true in the sense that structuralism and post-structuralism as paradigms and normative frames for understanding and interpreting language and social reality have proved both insufficient and problematic. But contrary to what one might think, this has not lead to a decreased interest in linguistic structure within the humanities and social sciences – quite the contrary. What Berry (2011) has termed 'the computational turn' within the humanities and social sciences is, in fact, also an empirical turn with a strong focus on linguistic surface structure as an epistemological stepping stone. It is, however, a methodological interest rather than an ideological or hermeneutic interest in surface structure driven by the fact, that the morpho-syntactic structure of language can be transformed into quantitative data. Thus, surface structure has become the interface between computer science and the humanities and social sciences allowing researchers to apply advanced statistical and computational methods in order to answer what has traditionally been considered qualitative questions.

On the one hand, this holds great promise for empirical research within the humanities and the social sciences. On the other hand, it raises a host of serious methodological and theoretical questions. For instance, how do we operationalize social phenomena, what are the proper theoretical frameworks for interpreting the quantified results, are computational methods normative, will technological advances rather than relevant research questions drive future research, what type of research questions can in fact be answered by means of computational methods?

## 3. The articles in this issue

This special issue comprises three articles that in different ways address some of the questions raised above. In the first article, Morten Tannert

discusses the potential of using methods and tools from computational linguistics within the field of educational writing research. The article covers three main sections. In the first section, Tannert provides a brief overview of educational writing research and positions computational linguistics within the field. In the second section, it is discussed what dimensions of educational writing can be studied by means of tools and methods from computational linguistics. The discussion is based on a model of an educational writing situation and on four metagenres for educational writing. The third section centers on the question of how to accommodate methodological critiques of reductionism and technocentrism in computational linguistic writing research. Tannert suggests the notion of affordance and computational hermeneutics as important starting points for more theoretically grounded and in-depth discussions of how to integrate methods from computational linguistics in educational writing research.

In the second article, Anne Agersnap and Kirstine Helboe Johansen adopt a meta-methodological perspective on computational methods. Agersnap & Johansen present a method for inferring knowledge about cultural and religious patterns from measures of linguistic features by combining distant and close reading approaches. They argue that distant readings of texts – i.e. computational analyses of linguistic features in texts – do not exclude close reading, but rather create new types of texts that require close reading. Agersnap and Johansen illustrate this through two cases from a large corpus study of 11,955 Danish sermons. In the first case, the authors show how it is possible to study patterns of gender discourse by studying the distribution of personal pronouns in nominal and oblique cases in the sermon corpus. In the second case, Agersnap and Johansen study how Danish pastors integrate different cultural and Christian traditions (e.g. biblical, literary, historical or popular characters) in their sermons. The authors round off the article with a discussion of how their own approach relates to other ongoing methodological discussions within the field of digital humanities.

In the third article, Rebekah B. Baglini, Anita Kurm, Lasse Hansen, Kenneth Enevoldsen, and Kristoffer L. Nielbo address the challenge of multilingual sentiment analysis using a traditional lexicon and rule-based sentiment instrument that is tailored to capture sentiment patterns in a particular language. The article focusses on a case study of three closely related Scandinavian languages (Danish, Norwegian, and Swedish) and

using three tailored versions of VADER, Baglini et al. measure the relative degree of variation in valence using the OPUS corpus. They find that scores for Swedish are systematically skewed lower than Danish for translational pairs, and that scores for Norwegian are skewed higher for both other languages. They use a neural network to optimize the fit between Norwegian and Swedish respectively and Danish as the reference (target) language.

**References**

Berry, D. M. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine*, *12*, 1–22.

Giddens, A. (1987). Structuralism, Post-structuralism and the Production of Culture. In Giddens Anthony & Turner Jonathan H (Eds.), *Social theory today* (pp. 195–223). Polity.