Language Sciences xxx (2018) 1–11

Contents lists available at ScienceDirect

Language Sciences

journal homepage: www.elsevier.com/locate/langsci

SenseDisclosure: A new procedure for dealing with problematically ambiguous terms in cross-disciplinary communication

Julie Mennes

Centre for Logic and Philosophy of Science, Department of Philosophy and Moral Sciences, Ghent University, Blandijnberg 2, B-9000 Ghent, Belgium

ARTICLE INFO

Article history:
Received 29 March 2017
Received in revised form 12 January 2018
Accepted 18 June 2018
Available online xxx

Keywords:
Cross-disciplinarity
Interdisciplinarity
Interdisciplinary communication
Terminological ambiguity
Natural language processing
Word sense induction

ABSTRACT

Communication in *cross-disciplinary* (and thus in *inter-*, *multi-* and *transdisciplinary*) projects is frequently challenged by *problematically ambiguous terms* (henceforth 'PATs'), i.e. terms that have multiple meanings and for which it is not always clear what meaning is used, thereby generating communication problems. The reason why communication in cross-disciplinary projects is so sensitive to PATs, is that they often involve disciplines that share one or more terms, yet attribute different meanings to them in an implicit and/or unsystematic manner. The teams of such projects are in need of a *PAT resolution procedure*, i.e. a procedure that helps them to identify and resolve PATs, as they are generally not trained to do this themselves. A first attempt to provide such a procedure consists in the identification of existing candidate procedures and an evaluation of their capacity to resolve PATs in cross-disciplinary communication contexts using a new set of task and performance criteria. It is shown that none of them sufficiently meets all criteria. It also becomes clear that the realization of an efficient PAT resolution procedure requires the ability to automatically process large quantities of linguistic data. Hence, input from the field of applied (computational) linguistics seems necessary. With this need for automation in mind and against the background of the new set of task and performance criteria, a theoretical characterisation of a new PAT resolution procedure called 'SenseDisclosure' is presented. SenseDisclosure is meant to be applicable to all kinds of cross-disciplinary projects (by an external facilitator). Its characterisation incorporates multiple techniques from Natural Language Processing to realize several critical automations. As the techniques were not specifically developed for PAT resolution, some of them require further research and development before they can be reliably integrated. Finally, it is argued that, if this extra research and development yields positive results, SenseDisclosure can be a truly effective PAT resolution procedure.

E-mail address: Julie.Mennes@UGent.be.

https://doi.org/10.1016/j.langsci.2018.06.003
0388-0001/

1. Introduction

Contemporary research projects are often characterised as *cross-disciplinary* (Bordons et al., 2004). A project is considered to be cross-disciplinary when it combines knowledge and know-how from different disciplines to answer a question or solve a problem. Besides 'cross-disciplinary', words like 'inter-', 'multi-' and 'transdisciplinarity' are commonly used. They refer to specific types of cross-disciplinarity that are distinguished on the basis of (i) the way in which knowledge is combined and (ii) the kind(s) of knowledge that is/are combined: 'interdisciplinarity' refers to the integration of knowledge from different academic disciplines, 'multi-disciplinarity' stands for the juxtaposition of knowledge from different academic disciplines, and 'transdisciplinarity' refers to the integration of knowledge from academic disciplines and practitioners (Klein, 2010). Irrespective of the type of cross-disciplinarity, a cross-disciplinary project (henceforth 'CD project') is generally carried out by a cross-disciplinary team (henceforth 'CD team') of researchers and/or practitioners who have a background in one or more of the *source disciplines* of the project, i.e. the disciplines from which knowledge is drawn. Because of their different backgrounds, communication and collaboration among CD team members is not always easy. This paper focusses on a set of communication problems that frequently arise in CD teams. The nature and common cause of the problems are best introduced via some concrete examples.

A first illustrative example is the case of a CD team of knowledge engineers and medical specialists who were working on the evaluation of three medical expert systems for the diagnosis of thyroid disorders (File and Dugard, 1997). Their first results indicated that only 44%–62% of the diagnoses generated by the expert systems matched the diagnoses of the specialists. As this effectiveness ratio was far lower than expected, the team revisited its work and found out that there had been a misunderstanding between the knowledge engineers and the medical specialists regarding the meaning of the terms 'diagnosis' and 'outcome'. The medical specialists used 'diagnosis' in a pre-treatment context, i.e. to describe the state of a patient before any treatment, and 'outcome' in a post-treatment context, i.e. to describe the effect of a given treatment on the health of a patient. However, the knowledge engineers had erroneously equated both terms. Because a 'diagnosis' is the kind of knowledge the expert systems produce, they considered it to be the 'outcome'. As a consequence, they used the wrong dataset for the evaluation of the systems. A re-evaluation based on the right dataset indicated that the efficiency of the systems was between 70% and 83%.

Another interesting example is discussed by Bracken and Oughton (2006). They describe communication problems in a CD team of physical and social scientists. While discussing the aims of their project on sustainable land use, the scientists discovered that they were attributing different meanings to the term 'catchment'. In order to enhance mutual understanding, they decided to share their definitions of the term with the whole team. The physical scientists considered 'catchment' to refer to 'the area of land defined by the watershed (drainage boundaries) of a particular river', as they focus on physical topography (Bracken and Oughton, 2006, p. 378). Yet, for the social scientists, the meaning of 'catchment' also required a link to the (spread of) human activities in the area, as they are interested in economic processes related to the physical landscape.

Finally, some other good examples are given by Ranade et al. (2011). They point out the recurring confusion between collaborating chemical and electrical engineers with respect to the meaning of terms like 'phase' and 'transducer'. Terms posing similar problems in cross-disciplinary communication (henceforth 'CD communication') have been identified in the fields of law (e.g., 'formalism' (Kemp, 1996)), disaster risk reduction (e.g., 'natural disaster' (Halldin et al., 2015)), cartography (e.g., 'representation' (Rossetto, 2016)) and physical science (e.g., 'artificial' (O'Rourke and Crowley, 2013)).

The common cause of this type of communication problem is *terminological ambiguity*, i.e. a term having multiple meanings (or, being *polysemous*) while it is not always clear what meaning is addressed by the term. In cross-disciplinary communication, an important source of terminological ambiguity are the (differences between the) jargons of the source disciplines. Each source discipline has its own jargon, and although some terms are shared by several of these jargons, they do not necessarily have the same meaning, or concept, across those jargons. Needless to say, terminological ambiguity negatively affects the overall efficiency of CD communication. However, it is not the case that every ambiguous term has the potential to cause communication problems like the ones described above.[1] In order to be a *problematically ambiguous term* (henceforth 'PAT'), besides being ambiguous, a term should also play an important role, i.e. its interpretation should affect the course of action. It is clear that CD teams should address PATs, preferably before they cause any misunderstanding, confusion or disagreement. Unfortunately, CD teams generally do not have the time nor training to systematically address the PATs in their communication. Hence, they are in need of a *PAT resolution procedure*, i.e. a procedure that helps them to efficiently identify and resolve PATs.

This paper responds to this need by first evaluating whether certain existing procedures, designed to facilitate CD communication, could also be used for the (more specific) task of PAT resolution. The evaluation is based on a new set of criteria and shows that there is no good PAT resolution tool readily available. It also reveals that an efficient PAT resolution procedure should be (partially) automated. Next, a theoretical characterisation of a new PAT resolution procedure called

[1] Note, however, that some authors argue that (certain levels of) terminological ambiguity can also be advantageous for CD communication. For example, Francl argues that her discipline (viz. chemistry) benefits from terminological ambiguity because it "offers us language that is elastic, stretching to encompass our expanding sense of a field or serving to bridge boundaries between fields' (Francl, 2015: p.534). Hodges (2008) makes a similar argument with respect to research in ecology.

Please cite this article in press as: Mennes, J., SenseDisclosure: A new procedure for dealing with problematically ambiguous terms in cross-disciplinary communication, *Language Sciences* (2018), https://doi.org/10.1016/j.langsci.2018.06.003