Promotor Prof. Dr. Erik Weber
Vakgroep wijsbegeerte en moraalwetenschap
Copromotor Prof. Dr. Els Lefever
Vakgroep vertalen, tolken en communicatie

Decaan Prof. Dr. Gita Deneckere
Rector Prof. Dr. Rik Van de Walle

FACULTEIT LETTEREN
EN WIJSBEGEERTE

Facilitating cross-disciplinarity

Joining forces of philosophy of science and computational linguistics

Julie Mennes

Proefschrift voorgelegd tot het behalen van de graad van Doctor in de wijsbegeerte en in de taalkunde
2019

UNIVERSITEIT
GENT

Voor Margaretha

Acknowledgements

Writing a dissertation is one of the most challenging things I ever did, but I did not do it alone. Many people helped and supported me in even more different ways. I would like to (try and) thank them all.

First of all, there are Erik and Els. This dissertation would never have been written if it was not for them. Erik introduced me to philosophy of science, supervised my master dissertation, hired me on a research project and supervised my PhD dissertation. I learned a lot from him in the countless hours we spent talking, discussing, negotiating, questioning, strategizing and evaluating. Els introduced me to computational linguistics and helped me acquire the necessary knowledge and skills. Her enthusiasm has been a great source of motivation and with her lively and cheerful ways, she helped me put many academic worries in perspective. Erik and Els, thanks for having believed in my potential as a researcher and for having supported me in pursuing my cross-disciplinary research interests.

During office hours, I commuted between three working environments. Of those three, the Centre for Logic and Philosophy of Science was my headquarters. Despite the fact that each member of the CLPS has his/her own research project, there never was a lack of collegiality. I am especially thankful to Dingmar for having taken me in tow, to Inge for our Glassboro adventure, to Dietlinde for her wit and friendly ear, to Federico for the delightful coffee breaks, to Leen, Dirk and Joke for regularly checking up on me, and to the logicians for having involved me in their more and less serious discussions. Also colleagues from outside the CLPS, such as Wilfried, Jonathan x2, Bohang, Gitte, Carine, Lut and Emilie made my time at the Blandijn more enjoyable and productive.

The offices of the Language, Translation and Technology Team were my second academic home. I am greatly indebted to LT3, not only because they hired me as a member of the anti-cyberbullying team without any programming experience. I would like to thank Cynthia and Els for the Python time and energy they invested in me, Bart and Nils for their invaluable advice and assistance, Ayla, Arda, Orphee, Peter and Véronique for having been such kind and helpful colleagues, and the whole of LT3 for the Melo-cakes, the sunny lunches and the professional silly photoshoots.

While I did not have any official business at my third working station, I was always welcome at DataWoods. Thank you Faan and Pieter-Jan for your work ethics, which has been contagious, your music selection and the improv.

I also got indispensable professional support from outside (U)Ghent.
Olivia and Martha took me on board of the "Intrepid" COST action and gave me the opportunity to gain firsthand experience in cross-disciplinary collaboration. As a member of Intrepid, I got to know more inspiring researchers from whom I learned a lot. I am especially thankful to Josefin, who showed me how to facilitate cross-disciplinary conversations, and Steve/Stephen, who initiated me in the art of interviewing (and let me practice on his network of cross-disciplinary researchers).

Liesbet was the person who "set me up" with Intrepid. Besides the matchmaking, I'm very grateful for the time she freed up to tell me about the workings of her interdisciplinary research group, Prometheus.

When one's interests are cross-disciplinary, it is not easy to find a community where one feels in place. Yet, I was lucky to meet researchers whose work I could build on, who were open to exchange ideas and who provided advice. I'm especially thankful to Dan(iel), Endla, Hanne, Henk, Michael, Michiru, Miles, Nick/Nicholas, Pyhillis, Sabina and Uskali.

Also on a more personal level, I got a lot of help and support.
First of all, I want to thank Faan for having stood by my side (cheering) from the very start till the finish. He encouraged me to explore and grow. He believed in me when I didn't. I couldn't have wished for a more generous, talented and hilarious companion. I am also thankful to Astrid for the cozy down time and the healthy food (experiments), to Otto for having been such a cute and quirky source of distraction, to I n l a n d for the days we kickstarted with a coffee and ended with freshly created tunes, to Marie, Pieter, Fieke and Roselien for the Ghent quality time, to T2 (and friends) for having been such a warm, enthusiast, creative and welcoming refuge, to my flat mates for having tolerated my limited cleaning activity, to Ralph and Cricri for their hospitality and impressive meringues, to Cameron for regularly having dropped an "egh", to Lotte and Griet for the energizing sleepovers, to Dora for always having been close (notwithstanding the 1400 km that separate us), to Marika for the power and the cross-disciplinary cover art, to Dwight for the beautiful stories and the reintroduction of the pomegranate, and to Devin for the madness. I would also like to thank Fede for having been such a strong and inspiring partner in crime. He calmed me down when I was overwhelmed, but also took a proactive approach and introduced an invaluable alternative perspective on (working) life. Thanks to him, I managed to enjoy the ride. Finally, I am very grateful to my parents, brother and grandparents for not only having understood and forgiven my absences, but for always having answered them with great love and support.

List of Abbreviations

BTE	bone tissue engineering
CD	cross-disciplinarity, cross-disciplinary
FN	false negative
FP	false positive
FWO	Research Foundation Flanders
ID	interdisciplinarity, interdisciplinary
KU Leuven	Catholic University of Louvain
MD	multidisciplinarity, multidisciplinary
NLP	natural language processing
PAT	problematically ambiguous term
POS	part-of-speech
TD	transdisciplinarity, transdisciplinary
tf-idf	term frequency – inverse document frequency
TN	true negative
TP	true positive
UAntwerpen	University of Antwerp
UGent	Ghent University
UHasselt	University of Hasselt
VSM	vector space model
VUB	Free University of Brussels
WSD	word sense disambiguation
WSD-T	word sense disambiguation as a computational task
WSI	word sense induction
WSI-T	word sense induction as a computational task
IWSD	induction-based word sense disambiguation
IWSD-T	induction-based word sense disambiguation as a computational task

List of Tables

List of Figures

Table of Contents

Overview of the included papers

1. Mennes, J. & Weber, E. (2019). How increased conceptual clarity allows for more efficient supporting of cross-disciplinarity. Examples from Flanders. Submitted to *Minerva: a review of science, learning and policy*.

2. Mennes, J. (2019). Putting multidisciplinarity (back) on the map. Submitted to the *European Journal for Philosophy of Science.*

3. Mennes, J. (2018). SenseDisclosure: a new procedure for dealing with terminological ambiguity in cross-disciplinary communication. *Language Sciences*. 69, pp. 57-67.

4. Mennes, J., Pedersen, T., & Lefever, E. (2019) Approaching terminological ambiguity in cross-disciplinary communication as a word sense induction task: a pilot study. *Language Resources and Evaluation*, pp 1-29.

5. Mennes, J., & van der Waart van Gulik, S. (2019) A critical analysis and explication of word sense disambiguation as approached by natural language processing. Submitted to *Linguistics and Philosophy.*

Introduction

Cross-disciplinarity

This dissertation is about cross-disciplinarity or, interchangeably, cross-disciplinary research. Loosely speaking, 'cross-disciplinarity' stands for the combining of knowledge or techniques from different disciplines. In this characterization, 'combining', 'knowledge' and 'discipline' are understood broadly: combining can, for example, amount to integrating, juxtaposing or contrasting, knowledge can, for example, be theoretical or practical; descriptive, methodological, esthetic, or ethical; implicit or explicit, and disciplines can be academic or non-academic.

The phenomenon of cross-disciplinarity has been widely researched. However, in this research it is generally called 'interdisciplinarity'. Thus, there is a divergence between a large part of the research literature and this dissertation with respect to how 'discipline-combining research' is called. Because this divergence is not merely a verbal dispute, it is first explained why the title of this dissertation is 'facilitating cross-disciplinarity', and not 'facilitating interdisciplinarity'.

The contemporary literature on what is called 'cross-disciplinary research' in this dissertation, dates back to the 1970s, when the OECD published the report 'Interdisciplinarity: problems of teaching and research in universities' (Apostel, 1972; Weingart, 2010). One important goal of the report was to define 'interdisciplinarity' as well as all the different forms it can take (Duget, 1972). Erich Jantsch greatly contributed to this goal by introducing a comprehensive "taxonomy of interdisciplinarity" (Jantsch, 1972; Klein, 2010). Later, following Jantsch's example, more taxonomies were proposed, such as the ones by Mayville (1978), Kockelmans (1979), Stember (1991), Klein (1990) and Huutoniemi et al. (2010). While these taxonomies contain varying numbers of types, they all contain a type that is called 'interdisciplinarity', one that is called 'multidisciplinarity' and one called 'transdisciplinarity'. The common denominator of these taxonomies, i.e. the triad of the concepts inter-, multi- and transdisciplinarity, came to be generally accepted as capturing the main types of ... interdisciplinarity (Klein, 2010; Holbrook, 2013). This means that, in the literature, 'interdisciplinarity' is standardly used both as an umbrella term (i.e. to refer to discipline-combining research) and as the name of a specific type of discipline-combining research, thus rendering the term ambiguous. To avoid this ambiguity, Sanford D. Eigenbrode et al. (2007) suggested to use 'cross-disciplinarity' as the umbrella term and to reserve 'interdisciplinarity' as the name of the specific type. In this dissertation, this suggestion is followed (and promoted). Yet, note that the ambiguous use of 'interdisciplinarity' still is the standard in the research literature and also has come to be adopted in policy notes.

The framework underlying this dissertation thus consists of the concepts of inter-, multi- and transdisciplinarity and that of cross-disciplinarity. Following the literature, the first three concepts are characterized on the basis of 'integration' (Lattuca, 2001; Mobjörk, 2010).

- *Interdisciplinarity* is a type of cross-disciplinarity that involves integrating knowledge or techniques from multiple academic disciplines.
- *Multidisciplinarity* is a type of cross-disciplinarity that involves juxtaposing (without integrating) knowledge or techniques from multiple academic disciplines.
- *Transdisciplinarity* is a type of cross-disciplinarity that involves integrating knowledge or techniques from one or more academic disciplines and one or more non-academic disciplines.

This conceptual framework is limited in (at least) two important respects. First, it does not capture the actual diversity of the types of cross-disciplinarity that can be found in practice. For example, with respect to ways of combining (knowledge or techniques from different) disciplines, it only takes integration and juxtaposition into account. Moreover, the framework defines discrete types of cross-disciplinarity while in reality, these types constitute a spectrum. Secondly, its concept definitions are vague because they are based on two notions that are themselves hard to define, i.e. discipline and integration (see Lélé and Norgaard, 2005 on discipline and O'Rourke et al., 2016 on integration).

The reasons for adopting this framework nevertheless, are pragmatic. As is explained in the next section, the goal of this dissertation is to facilitate cross-disciplinarity. Despite its limitations, the conceptual framework provides a good starting point for working towards that goal.

From the perspective of the framework, the scope of this dissertation is restricted in two ways. First, the dissertation focuses exclusively on research projects, which are only one application context of the framework. A 'research project' is an enterprise that aims to answer a question or to solve a problem. Usually, funding is obtained specifically for a given project, and the project is carried out by an individual or by a team. As is explained in Epilogue to Part I, despite a focus on cross-disciplinary research projects, some of the findings that are presented also bear relevance to other contexts in which cross-disciplinary research is performed, such as institutes, ateliers, networks and programmes. Secondly, the scope of the dissertation is restricted with respect to the content of the framework as in a large part, transdisciplinarity is not covered.

Facilitation

This dissertation aims to *facilitate* cross-disciplinarity, which is not the same as aiming to *promote* cross-disciplinarity. In the last decades, cross-disciplinarity has become increasingly popular, up to the point of having become a hype. Because of its popularity, cross-disciplinarity is already heavily promoted. The goal of this dissertation is to provide theoretical analyses and practical tools that can be used by research policy makers, funding bodies and researchers to overcome the challenges that come with cross-disciplinarity, irrespective of a hype status. Most of these challenges result from (the interplay of) two factors.

The first factor is the organization of research infrastructure: universities, funding channels and journals are (still mostly) structured around disciplines (Lowe et al., 2013). This greatly complicates the funding and execution of projects that do not "fit" in a disciplinary structure because they combine knowledge or techniques from multiple disciplines (Bromham et al., 2007) as well as the dissemination of results that transcend the scope of (mono-) disciplinary journals (Kostoff, 2002).[1]

The second factor concerns differences between disciplines. These differences make it difficult to take knowledge or techniques from different disciplines and combine them. Think, for example of:

- different levels of description and hence, different scales and units (Morse et al. 2007);
- different values and hence, different opinions on the formulation of research questions and on the type of answer that should be sought (Lélé & Norgaard, 2005; Barlow et al., 2011);
- different methodologies and hence, different types of argumentation and notions of evidence (Benda et al. 2002; Bruun et al., 2005);

[1] For more comprehensive overviews of challenges for cross-disciplinarity, see (Benda et al., 2002; Bruun et al., 2005; NAS, 2005; Lélé and Norgaard, 2005; Choi & Pak, 2007; MacLeod, 2018).

- different disciplinary languages and hence, different jargon terms and metaphors (Wear, 1999; Bracken and Oughton, 2006).

Both factors cause challenges for a key aspect of cross-disciplinarity, viz. communication.

The first factor mostly affects communication *about* cross-disciplinarity, the second communication *within* cross-disciplinarity teams. Communication *about* cross-disciplinarity goes in two directions. In one direction, i.e. from research policy makers and funding bodies to researchers, it consists in the former proclaiming their interest in, or commitment to cross-disciplinarity, for example by means of vision and mission statements (e.g. NSF, n.d.), foresight reports (e.g. MIT, n.d.) and calls for cross-disciplinary projects. In the other direction, i.e. from researchers to research policy makers and funding bodies, it consists in researchers informing on their cross-disciplinary experiences and requesting means and a suitable environment for carrying out cross-disciplinary projects, for example, by means of research proposals, opinion pieces (e.g. Irani, 2018), special issues (e.g. Lawrence & Després, 2004), essays (e.g. Wilthagen et al., 2018) and editorials (e.g. Nature editors, 2015). Note that the two directions of communication are closely connected, as they are both about questions or demands regarding cross-disciplinarity and answering (to) them. Moreover, policy makers sometimes commission researchers to write reports on the status of cross-disciplinarity (e.g. Bruun et al., 2005). Communication *within* cross-disciplinary teams takes place between the members of a team that is responsible for a given cross-disciplinary project. Usually, each 'source discipline' of a project, i.e. a discipline from which knowledge is drawn, is represented by one or more team members who have a background in that discipline. For this reason, communication within cross-disciplinary teams is also called 'cross-disciplinary communication'. Cross-disciplinary communication is about the content and execution of research projects, i.e. developing strategies, setting up experiments, discussing literature, interpreting results, writing papers, etc.

This dissertation focuses on one specific challenge for each of the two levels of communication.

Regarding communication about cross-disciplinary projects, the focus is on the current lack of an unambiguous, accurate and neutral conceptual framework for cross-disciplinarity that is shared by research policy makers, funding bodies and researchers. While there is a received framework that is shared by all three parties, as is argued in this dissertation, it is ambiguous, inaccurate and biased. Hence, its use does not contribute to effective communication about cross-disciplinarity. The first way in which this dissertation aims to facilitate cross-disciplinarity is by analyzing the problems and limitations of the received framework, and by presenting solutions for them.

Regarding communication within cross-disciplinary teams, this dissertation focuses on terminological ambiguity. Each discipline has its own jargon and while some terms are shared by different jargons, they do not necessarily have the same meaning across those jargons. Misunderstandings may arise when, in the context of a cross-disciplinary project, a term is used that has different meanings across the jargons of its source disciplines and thus, is (unknowingly) understood differently by the team members. The second way in which this dissertation aims to facilitate cross-disciplinarity is by providing cross-disciplinary teams with a tool that allows to resolve terminological ambiguity.

As is explained in the next section, this dissertation consists of two parts, which are dedicated to the first and the second challenge respectively. It is also further clarified how the challenges are approached, and what contribution is made to the facilitation of cross-disciplinarity.

A dissertation consisting of two parts

This dissertation consists of five papers that have been published, or have been submitted to academic journals (see Overview of the included papers). The papers are grouped into two parts as follows:

Part I

- Paper 1: How increased conceptual clarity allows for more effective supporting of cross-disciplinarity. Examples from Flanders, Belgium
- Paper 2: Putting multidisciplinarity (back) on the map

Part II

- Paper 3: SenseDisclosure: A new procedure for dealing with problematically ambiguous terms in cross-disciplinary communication
- Paper 4: Approaching terminological ambiguity in cross-disciplinary communication as a word sense induction task: a pilot study
- Paper 5: A Critical Analysis and Explication of Word Sense Disambiguation as Approached by Natural Language Processing

Part I revolves around facilitating communication about cross-disciplinarity between research policy makers, funding bodies and researchers. For this communication to be effective, all three parties should use the same conceptual framework to talk about cross-disciplinarity. Moreover, this framework should (i) unambiguously distinguish the main types of cross-disciplinarity, (ii) provide an accurate definition for each type, and (iii) be neutral regarding the overall value of each individual type. While there is indeed a received framework that is shared by all three parties, (a) it is ambiguous, (b) contains at least one inaccurate definition and (c) presents one type of cross-disciplinarity as less valuable. In the two papers included in Part I, major steps are taken in the development of a useful shared conceptual framework for cross-disciplinarity that respects (i)-(iii) by resolving (a)-(c).

Paper 1 addresses the ambiguity of the received framework. It is first explained why the received framework is ambiguous. Next, it is argued that the use of an ambiguous framework for cross-disciplinarity by research policy makers and funding bodies causes three problems for cross-disciplinarity-supporting initiatives. The first problem is a decreased effectiveness of mission statements: policy makers and funders fail to clearly express their interest in and commitment to cross-disciplinarity. This makes it hard for researchers with cross-disciplinary aspirations to know which governments or funders would support their work. The second problem is a decreased effectiveness of initiatives that are set up to support cross-disciplinarity: policy makers and funders fail to clearly define the target audience of the cross-disciplinarity-supporting initiatives they set up. This unclarity impedes researchers in assessing whether their project proposal qualifies for funding. The third problem also is a decrease of the effectiveness of cross-disciplinarity-supporting initiatives, but here, for the reason that policy makers and funders fail to provide complete and coherent selection criteria. This impedes researchers in making sure that their proposal meets all (relevant) criteria. Next, the received framework is extended with a new umbrella term (as was done in Introduction) to solve the ambiguity. Finally, the paper shows how the use of the extended framework allows to avoid the three problems mentioned in this paragraph and to increase the effectiveness of mission statements and initiatives for the facilitation of cross-disciplinarity.

Paper 2 takes up on the issues of accurate definitions and neutrality. The received framework is closely associated with a received theory of cross-disciplinarity, which contains rich definitions of the different types of cross-disciplinarity and informs on their strengths and limitations. First, it is shown that the received theory is not neutral regarding the value of the different types of cross-disciplinarity because it presents multidisciplinarity as less valuable than interdisciplinarity. It is pointed out that the adoption of the received framework and associated theory (for example by research policy makers, funding bodies or researchers) thus supports the development of a preference for interdisciplinarity over multidisciplinarity. Next, it is argued that the theory only presents multidisciplinarity as less valuable because it is incomplete and thus, does not provide adequate support for a preference for interdisciplinarity. This means that having such a preference is unjustified. It is explained why the presence of an unjustified preference for interdisciplinarity (among research policy makers, funding bodies or researchers) is harmful for the practice of cross-disciplinarity. Finally, the imbalances in the received theory are analyzed and resolved. One of these imbalances is the contrast between the accuracy of the definition of ‘interdisciplinarity’ and the inaccuracy of the definition of ‘multidisciplinarity’. The received definition of multidisciplinarity is replaced with a new, more accurate one.

Part II focuses on facilitating communication within cross-disciplinary teams. As was mentioned above, terminological ambiguity constitutes an important challenge for this communication. However, not all ambiguous terms have the potential to cause misunderstandings. Only terms that play a key role in a given project (i.e. their interpretation affects the course of action) can cause communication problems because of their ambiguity. Such terms are called *problematically ambiguous terms* (PATs). Because cross-disciplinary projects differ with respect to their key terms and the disciplines (and hence, jargons) that are involved, PATs are project-specific.

The three papers of Part II contribute, directly or more indirectly, to the development of a tool for PAT resolution.

Paper 3 contains the theoretical outline of a new procedure that can be used by a given cross-disciplinary team to resolve the PATs of their project. Before putting forward the outline, it is ascertained that none of the existing procedures for the facilitation of cross-disciplinary communication can be used for PAT resolution. The new procedure is called 'SenseDisclosure' and consists of seven steps. In the outline, it is proposed that techniques from natural language processing (NLP) are incorporated to automate some of the steps. The most innovative proposal involves the use of word sense induction (WSI) techniques, i.e. techniques that allow to differentiate the senses, or meanings of a word based on the occurrences of that word in a corpus. The idea would be to use WSI techniques for the automated identification of terms that have different meanings across the jargons of the source disciplines of a project (and thus, are candidate PATs). The proposal is innovative because WSI techniques were originally developed in the context of machine translation and document retrieval, where it often suffices to make rough, or high-level, sense distinctions. However, in order to be a functional part of SenseDisclosure, the techniques also have to allow for the differentiation of more fine-grained senses, and this on the basis of a corpus that represents highly specialized disciplinary jargons (which has not been done before). Hence, before a WSI technique can be integrated into SenseDisclosure, further research and experiments are required.

Paper 4 reports on first steps in that research and experiment. It presents a pilot study that assesses whether WSI techniques meet an important precondition for becoming a functional part of SenseDisclosure. As a part of SenseDisclosure, the techniques should allow to determine, for any term that plays a key role in a given project and is part of multiple jargons (regardless of its compositionality and grammatical category), how many (fine-grained) senses it has across those jargons. In the pilot study, it is experimentally checked whether a WSI technique can be used to detect meanings of single word nouns (e.g. 'matrix' and 'model') when the number of meanings is already known. In the experiments, one implementation (viz. SenseClusters) of one WSI technique (viz. vector-based clustering) is applied to a corpus that was custom created to represent the jargons of the source discipline of a bone tissue engineering project that serves as a case study. While, on average, the results of the experiments were very good, much more research is required to assure that SenseClusters (or another implementation of a WSI technique) can be integrated into SenseDisclosure.[2]

In Paper 5, a more indirect contribution is made to the development of a PAT resolution procedure. When the experiments for Paper 4 were performed, it became clear that characterizations of the potential of WSI techniques in the literature are not always accurate, sometimes suggesting that the techniques allow to achieve more than actually is the case. In Paper 5, the limitations of WSI techniques are explored by means of a critical philosophical analysis of the concepts and methods that underlie the techniques. Thus, Paper 5 sheds some light on what can ultimately be expected from WSI techniques in the context of PAT resolution (and what not).

Part I and Part II are very different regarding their subject and methodology. Yet, their starting points are very similar. Instead of starting from an ongoing theoretical debate or from a gap in the literature, both parts are rooted in the challenges for the everyday practice of cross-disciplinarity. The way in which these challenges are addressed is also very similar. In both parts, a hands-on approach is taken. While the necessary attention is devoted to problem analyses, the main goal is always to present (part of) a practical solution. To this end, in Part I, not only the (general) literature on cross-disciplinarity is consulted, but also text material that is currently used in the allocation of funding for cross-disciplinary projects, as well as vision statements and research reports of major funding institutions. In Part II a practical solution is obtained by applying existing NLP technology in to new problem context, which required a custom corpus to be built. In sum, both parts aim at providing important first aid for cross-disciplinarity: they start from pressing problems, and present solutions.

[2] Only after Papers 3 and 4 were accepted for publication, I learned about the existence of another procedure that applies NLP techniques to PATs. Alessio Ferrari and colleagues developed a procedure that makes use of *word embeddings*, i.e. vector representations of words, to flag terms that are ambiguous in the context of a *requirement elicitation meeting*, i.e. a meeting where requirement analysts consult experts of a given domain to obtain the information software developers will need to build a system that can be used in that domain (Ferrari et al., 2018). Besides their methodology, also the functionalities of Ferrari et al.'s procedure differs from that of SenseDisclosure: (i) Ferrari et al.'s procedure is designed to facilitate one-off interactions between analysts and domain experts, whereas SenseDisclosure aims to facilitate interactions between members of a cross-disciplinary team, and (ii) Ferrari et al.'s procedure can only detect terms that are ambiguous across two different domains, whereas SenseClusters can be used to detect PATs across any number of disciplinary jargons.

Cross-disciplinarity of the dissertation

This dissertation draws from philosophy of science and computational linguistics. Roughly speaking, philosophy of science involves the study of the demarcation, foundation and methods of science, but also the study of scientific practice. Computational linguistics is about the modeling of language, either with the goal of generating linguistic theories or developing applications for the automated processing or generation of language. Application-oriented computational linguistics is commonly called ‘natural language processing’ (NLP). Because of the involvement of two disciplines, besides being *about* cross-disciplinarity, this dissertation also *is* cross-disciplinary. To be precise, because knowledge and techniques from the two disciplines are integrated, according to the conceptual framework, the second part is interdisciplinary. How this integration is achieved, is explained in detail in Conclusions and future research.

Part I
Facilitating communication about cross-disciplinarity

President "Bobby": *Mr. Gardner, do you agree with Ben, or do you think that we can stimulate growth through temporary incentives?*
Chance: *As long as the roots are not severed, all is well. And all will be well in the garden.*
President "Bobby": *In the garden.*
Chance: *Yes. In the garden, growth has it seasons. First comes spring and summer, but then we have fall and winter. And then we get spring and summer again.*
President "Bobby": *Spring and summer.*
Chance: *Yes.*
President "Bobby": *Then fall and winter.*
Chance: *Yes.*
Benjamin Rand: *I think what our insightful young friend is saying is that we welcome the inevitable seasons of nature, but we're upset by the seasons of our economy.*
Chance: *Yes! There will be growth in the spring!*
Benjamin Rand: *Hmm!*
Chance: *Hmm!*
President "Bobby": *Hm. Well, Mr. Gardner, I must admit that is one of the most refreshing and optimistic statements I've heard in a very, very long time. ... I admire your good, solid sense. That's precisely what we lack on Capitol Hill.*

— *Being There*

Prologue to Part I

Part I focuses on facilitating the communication between research policy makers, funding bodies and researchers about their needs and expectations with respect to cross-disciplinary research.

For this communication to be effective, all three parties should make use of the same conceptual framework for cross-disciplinarity. This framework should distinguish the main types of cross-disciplinarity in an unambiguous way, provide an accurate definition for each of the types and be neutral regarding the overall value of each individual type. While there is indeed a received conceptual framework that is shared by the three aforementioned parties, it does not meet the other desiderata. In the papers that constitute Part I, major steps are taken to transform the received framework into a new framework that does contribute to effective communication about cross-disciplinarity.

Paper 1 aims to solve the issue of ambiguity. The received framework is extended with an additional term (viz. 'cross-disciplinarity') to make sure that every included term only has one meaning. Moreover, it is shown how the use of the extended framework allows research policy makers and funding bodies to express their interest in cross-disciplinarity more precisely, and how it allows them to improve the set-up of initiatives for the targeted support of cross-disciplinary research.

Paper 2 takes up on the issues of inaccurate definitions and neutrality. The received framework is closely associated with a received theory of cross-disciplinarity, which contains elaborate definitions of the different types of cross-disciplinarity. Contrary to the included definition of interdisciplinarity, the definition of multidisciplinarity is inaccurate because it is incomplete. The incompleteness of this definition leads to multidisciplinarity being presented as a less valuable type of cross-disciplinarity. To make the theory more neutral, the received definition of multidisciplinarity is completed with the information that was missing, thereby creating a more accurate definition.

Paper 1
How increased conceptual clarity allows for more efficient supporting of cross-disciplinarity. Examples from Flanders, Belgium[1]

[1] This is a manuscript by myself and Prof. Dr. Erik Weber that has been submitted to *Minerva*.

Abstract

Statements about the importance of crossing disciplinary boundaries have become commonplace in mission statements and policy texts of research policy makers, research funding bodies, universities and other research-related institutions. In those statements, the use of terms like 'inter-', 'multi-' and 'transdisciplinarity' is highly common. Yet, the terms often remain undefined and (hence) are used interchangeably. It goes without saying that such a lack of conceptual clarity obfuscates the interest of research-related institutions regarding discipline-combining research, thereby rendering their statements uninformative for researchers who (aspire to) perform discipline-combining research. Moreover, and more importantly, the lack of conceptual clarity also plays in the set-up of initiatives that are taken by the institutions to (financially) support discipline-combining research. In that context, the conceptual unclarity causes (i) the scope of initiatives to be vague or ambiguous, and (ii) the selection criteria for the initiatives to be incomplete and/or incoherent. Both effects seriously impede the effectiveness of the initiatives. To enable more precise talk about discipline-combining research, and more effective initiatives for the supporting of discipline-combining research, this paper introduces an unambiguous conceptual framework. It also shows how the framework can be used to increase the accuracy of mission statements and policy texts and to improve the set-up of initiatives for the supporting of discipline-combining research. The latter is done by means of texts that are currently used by research-related institutions in Flanders, Belgium to communicate their interest in discipline-combining research or to allocate funding to discipline-combining projects via dedicated funding channels.

Keywords

ambiguity • cross-disciplinarity • interdisciplinarity • multidisciplinarity • research funding • research policy making

1.1 Introduction

1.1.1 A widespread interest for combining disciplines

The need for 'building bridges between scientific disciplines' is expressed in the mission statements and policy texts of many research-related institutions. For example, the report 'Facilitating Interdisciplinary Research' of the National Academies of the United States says that:

> Some of the most interesting scientific questions are found at the interfaces between disciplines and in the white spaces on organizational charts. Exploring such interfaces and interstices leads investigators beyond their own disciplines to invite the participation of researchers in adjacent or complementary fields [.] (The National Academies, 2005, p. 33)

The same idea is also present in texts of the main research-related institutions in Flanders (Belgium). These institutions fall into two categories, viz. institutions that do not, and institutions that do fund research. First, there are the institutions that are responsible for the organization and promotion of research, but do not distribute any funding. One of these institutions is the governmental advisory body called 'Flemish Council for Science and Innovation' ('Vlaamse Raad voor Wetenschap en Innovatie' – VRWI).[1] In their foresight report for 2025, they plead for the

> [r]estructuring of funding mechanisms of knowledge institutions in order to allow for the stimulation of interdisciplinary and *extra muros* collaborations. (VRWI, 2014, p. 119, italics in original, translated by the authors)[2]

Two other examples of research-related institutions that do not provide funding are the 'Royal Flemish Academy of Belgium for Science and the Arts' ('Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten' – KVAB) and its daughter academy 'The Young Academy' ('Jonge Academie' – JA), which are both responsible for the promotion of scholarship and science. In a report of the KVAB on researcher-driven science in Flanders, the following (radical) statement can be found:

[1] Since 2016, this body has become incorporated into the 'Flemish Advisory Council for Innovation and Enterprise' ('Vlaamse Adviesraad voor Innoveren en Ondernemen' – VARIO).

[2] (https://www.vlaanderen.be/publicaties/vrwi-toekomstverkenningen-2025-1)

> Interdisciplinary research should not be the exception, but as standard as mono-disciplinary research. (Monard et al., 2018, translation by the authors)

Second, there are the institutions that do distribute research funds. Flanders has six main research funding bodies. Five of them are universities (viz. University of Antwerp, the Free University of Brussels, University of Hasselt, Ghent University, and the Catholic University of Louvain), the sixth is the public research council (viz. Research Foundation Flanders – FWO). The Dutch and English names of these research-funding institutions, as well as their acronyms, are listed in Table 1.

Table 1 Overview of Flemish research funding bodies

Dutch name	English name	Acronym
Katholieke Universiteit Leuven	Catholic University of Leuven	KU Leuven
Universiteit Antwerpen	University of Antwerp	UAntwerpen
Universiteit Gent	Ghent University	UGent
Universiteit Hasselt	Hasselt University	UHasselt
Vrije Universiteit Brussel	Free University of Brussels	VUB
Fonds Wetenschappelijk Onderzoek – Vlaanderen	Research Foundation - Flanders	FWO

All of these funding bodies attach great importance to discipline-combining research. For example, UGent mentions 'openness' as one of its central values in its Research Policy Plan for 2017-2021 ('Beleidsplan Onderzoek 2017-2021'). This openness is described as multi-faceted, with one of the facets being:

> [openness] with respect to interdisciplinarity: the research and the research policy stand for cross-fertilization between different scientific fields. (UGent, 2017, p. 2, translated by the authors).

And, on the webpage that summarizes the research policy plan of KU Leuven it is stated that:

> KU Leuven attaches the highest importance to interdisciplinarity. Therefore, an active policy is pursued that provides expanding opportunities for interdisciplinary and transdisciplinary research to develop, including a novel funding channel for interdisciplinary networks (ID-N) and a custom valorization system for interdisciplinary initiatives. (KU Leuven, n.d., translated by the authors)

In sum, Flemish research institutions clearly express an interest in (enhancing) discipline-combining research.

1.1.2 Talking about ways of combining disciplines

A closer look at their mission and vision statements and policy texts reveals that Flemish research-related institutions acknowledge that the combining of disciplines can take different forms. For instance, the (internal) website of UGent states:

> Ghent University understands the term 'interdisciplinarity' as the various degrees and methods of collaboration outside of one's own discipline. This involves cross-, multi-, inter- and transdisciplinarity.[3]

And, in their vision statement on interdisciplinarity, the Young Academy writes:

> In this note, interdisciplinarity is defined as: "the intense collaboration across the boundaries of existing disciplines, both within and outside academia". The following gradations are distinguished on the basis of the intensity of the collaboration and the achieved result: multidisciplinarity [...], interdisciplinarity [...], transdisciplinarity." (Geris & Op de Beeck, 2015, p. 14, translated by the authors)

Moreover, in the policy plan of KU Leuven, it is stated that:

> In this policy vision we use interdisciplinarity as an umbrella term, knowing that this term covers variants and gradations of collaboration and integration. (2018, p. 49)

All three institutions mention different types of interdisciplinarity. However, it should be noted that they use 'interdisciplinarity' both as the umbrella term and as the name for a specific type of discipline-combining research, which entails an ambiguous use of the term.[4] Moreover, almost all funding bodies (KU Leuven being the exception) fail to define the concepts that are associated with the terms they use, and (hence) appear to use the terms that refer to different types of discipline-combining research interchangeably. For example, while some parts of the policy plan of UAntwerpen consistently mention 'multidisciplinarity', other parts are about 'interdisciplinarity' while the reason for the terminological switches is unclear (UAntwerpen, 2016). Hence, despite their interest in discipline-combining research and awareness of the existence of different types of that research, Flemish research-related institutions lack conceptual clarity when writing about the matter.

---

[3] (https://www.ugent.be/intranet/en/research/organisation/interdisciplinarity.htm, only accessible within the UGent domain)

[4] When echoing the (seeming) use of 'interdisciplinarity' as an umbrella term by institutions or authors, the term is put in italics. This is done because the word use does correspond with the meaning that is ascribed to 'interdisciplinarity' in the conceptual framework that is put forward in this paper (in Section 2).

1.1.3 Problems caused by a lack of conceptual clarity

The lack of conceptual clarity regarding discipline-combining research among Flemish research-related institutions causes three problems for the supporting of such research.

The first problem (P1) is that the potential impact of mission and vision statements and policy texts is limited. Institutions could use the texts as a means to inform the research community on what type(s) of discipline-combining research they are interested in and thus, to attract (teams of) researchers who share that interest. However, given the lack of conceptual clarity, institutions do not express their interest accurately and hence, the texts are uninformative to researchers who (aspire to) perform discipline-combining research. However, this is more of a missed opportunity than a genuine problem.

The other two problems are more serious. They both relate to initiatives that are taken by funding bodies to (financially) support discipline-combining research. Examples of such initiatives are the issuing of one-time calls for projects that involve multiple disciplines, the establishing of new institutes that harbor *interdisciplinary* networks or platforms, the opening of calls for projects to *interdisciplinary* proposals (and giving them an evaluation bonus so they can compete with disciplinary proposals) and the installation of funding channels that are dedicated to *interdisciplinary* projects (that issue calls for *interdisciplinary* projects regularly). For these initiatives to be effective, it is important that their set-up is well defined. This minimally requires (i) a clearly defined scope (i.e. it should be clear what type of project, collaboration or organization the initiative aims to support) and (ii) a complete and coherent set of selection criteria (i.e. a set of criteria that allows to determine whether a given project qualifies for support via the initiative). However, a lack of conceptual clarity causes scope definitions to be ambiguous and selection criteria to be incomplete and incoherent. An ill-defined scope leads to the problem (P2) that a project proposal that actually falls within the scope of an initiative might not seem to do so and hence, the team behind it might refrain from submitting it. The problem (P3) with incomplete and incoherent selection criteria is that they fail to fix the scope of the initiative (again possibly causing confusion) and can be unreasonably difficult to satisfy or even conflicting, which again, might discourage researchers from submitting a proposal.

Note that this paper focuses on problems (that are caused by a lack of conceptual clarity) for the future success of initiatives for the supporting of discipline-combining research. However, conceptual unclarity in the context of initiatives also has a negative impact on discipline-combining research that is already being financially supported; this was shown by Katrine Lindvig and Line Hillersdal (2019).

1.1.4 Goal and structure of the paper

As an answer to the lack of conceptual clarity among Flemish research-related institutions, this paper starts with proposing an unambiguous conceptual framework for discipline-combining research. The framework is an extended version of an existing one that is commonly accepted in the literature on *interdisciplinarity* (and that contains the terms that are used by the Flemish institutions).

The goal of this paper is to show how the framework can be used to solve the three problems presented above (P1-P3). To achieve this goal, the conceptual framework is applied to text material of Flemish research funding bodies (e.g. policy plans, selection criteria,...) that is currently used in the allocation of funding to discipline-combining projects. For reasons of brevity, only texts that are related to one specific initiative, viz. dedicated funding channels, are used.[5] Such channels exist at FWO, UGent and VUB. When no official English version of a relevant text is available, the original Dutch one is translated.

The remainder of this paper is structured as follows: Section 2 introduces the conceptual framework. Section 3 briefly shows how the framework can be used to solve P1 by resolving the uninformativeness of vision statements and policy texts. Section 4 provides a brief overview of the dedicated funding channels for discipline-combining projects that exist in Flanders. Section 4 shows how the framework can be used to solve P2 by clarifying the scope of initiatives. Section 5 does something similar for P3: it demonstrates how sets of selection criteria can be made complete and coherent by means of the framework. Section 6 presents general conclusions and contains directions for further research.

[5] The focus on dedicated funding channels does not imply that they are considered the best way to support discipline-combining research. A study of the German national research council ('Deutsche Forschungsgemeinschaft') even suggests that initiatives for the facilitation of discipline-combining research that revolve around the distribution of funds are redundant (DFG, 2016). This suggestion follows from their observation that, even when in direct competition, the funding success of discipline-combining proposals is not significantly lower than that of disciplinary proposals. Yet, their study has been contradicted by (Bromham et al., 2016).

1.2 The conceptual framework

1.2.1 General structure

The conceptual framework that is well-established in the literature on *interdisciplinarity* contains three terms, viz. 'inter-', 'multi-' and 'transdisciplinarity' (Bruce et al., 2004; Holbrook, 2013; Klein, 2014). Generally, their concepts are distinguished by a degree and scope of integration, where multidisciplinarity (MD) is considered non-integrative; interdisciplinarity (ID) integrative; and transdisciplinarity (TD) is taken to involve the integration of non-academic disciplines. In this framework, 'interdisciplinarity' is used both as the umbrella term and as the name for a type. To avoid this ambiguity, a fourth term is added to the original framework, viz. 'cross-disciplinarity'. Following the suggestion of Eigenbrode et al. 2007, this term is used as the umbrella term, meaning that cross- disciplinarity (CD) is considered to be the concept that encompasses ID, MD and TD.[6,7] Thus, in the conceptual framework, each term identifies only one concept (or, has one meaning) and the combination of the concepts roughly covers the diversity of CD. Yet, because it only consists of four concepts and a distinguishing principle (viz. integration, characterized in terms of degree and scope), the framework is highly abstract. In order to become truly useful, it has to be supplemented with more elaborate definitions. In Sections 2.2-2.4, paradigmatic definitions of the different types of CD are provided. Most of them find their origin in the work of Julie Klein (1990, 2004 and 2010). Klein is a philosopher of science who wrote much cited articles and books on *interdisciplinarity*, worked as an advisor on the subject of *interdisciplinarity* for research organizations (e.g. Bruun et al., 2005; NAS, 2005; Huutoniemi et al., 2010) and contributed to major compendia on *interdisciplinarity* (e.g. Weingart & Stehr, 2000; Frodeman et al., 2010 & 2017).

1.2.2 Integration as the crux of ID

As previously mentioned, the different types of CD are generally distinguished by means of integration. Or, as Julie Klein puts it:

[6] In this paper, both 'multidisciplinarity' and 'multidisciplinary' are abbreviated as 'MD'; 'interdisciplinarity and 'interdisciplinary' as 'ID'; 'transdisciplinarity' and 'transdisciplinary' as 'TD'; and 'cross-disciplinarity' and 'cross-disciplinary' as CD.

[7] Note that up to this point, the term 'discipline-combining research' was used to refer to cross-disciplinarity.

> [I]ntegration is the most common benchmark and, combined with degrees of disciplinary interaction, provides a comparative framework for understanding differences in types of interdisciplinary work. (Klein, 2010, p. 17)

Besides identifying integration as the distinguishing principle for different types of CD, Klein also indicates that it is "widely considered the crux of interdisciplinarity" (Klein, 2008, p. S119). She defines ID as follows:

> *Interdisciplinarity* integrates information, data, methods, tools, concepts or theories from two or more disciplines or bodies of knowledge to address a complex question, problem, topic, or theme. (2014, p. 13; italics in original)

The central role of integration in ID is also stressed by the National Academies. In their report titled 'Facilitating interdisciplinarity', they write that:

> Research is truly interdisciplinary when it is not just pasting two disciplines together to create one product but rather is an integration and synthesis of ideas and methods. (The National Academies 2005, p. 27)

It should be noted that, despite being a central notion, there is no commonly accepted general characterization of integration (O'Rourke et al., 2016). One reason for this is that studies of the phenomenon are divided: they are carried out by one discipline (e.g. see (König et al., 2013) for management science, (Czerniak et al., 2010) for educational science), or for one field (e.g. see (O'Malley, 2013) for integration in biology, (MacLeod and Nagatsu, 2018) for integration in sustainability sciences). Moreover, it can be argued that there are as many forms of integration as there are interdisciplinary projects (Klein, 2008), which means that there is no point in looking for a general characterization. Therefore, in this paper, integration is loosely defined as the fusing of elements, which previously were not combined, into a new whole.

1.2.3 Complementarity as the hallmark of MD

In contrast to ID, MD does not involve integration. It is defined by Klein as follows:

> Multidisciplinarity juxtaposes two or more disciplines or bodies of knowledge focused on a common problem, question, topic or theme. (2014, p. 12; italics in original)

MD can thus be understood as the simultaneous studying of a subject from different disciplinary angles (Holbrook, 2013). In an MD project, this is generally realized via distinct, but complementary disciplinary subprojects. Since the subprojects are (to some extent) mutually coordinated and information is exchanged between them, MD facilitates cross-fertilization, i.e. the re-thinking or enriching of traditional disciplinary perspectives, methods and concepts. In some MD projects, the cross-fertilization consists in the subprojects affecting each other's course or enabling each other's results. In other projects, the cross-fertilization will be limited to the exchange of ideas and perspectives (de Boer et al., 2006).

1.2.4 Extra-academic input in TD

Originally, TD was most commonly understood as referring to a type of CD that is more integrative than ID, and therefore, truly *transcends* disciplines (e.g. Klein, 1990; Stokols et al., 2003). However, in the past decades, an alternative definition has become predominant (Klein, 2008; Pohl et al., 2011). J. Britt Holbrook summarizes this definition as follows:

> [Transdisciplinarity stands for] the integration of one or more academic disciplines with extra-academic perspectives on a common (and usually a real-world, as opposed to a merely academic) problem. (2013, p. 1867)

She clarifies the difference with ID as follows:

> What distinguishes TD from ID, on the other hand, is not the idea of integration *per se*, but rather *what* ought to be integrated. In the case of ID, what ought to be integrated are various academic disciplinary approaches to a problem; in the case of TD, what ought to be integrated are academic and non-academic stakeholders' approaches to a problem. (2013, p. 1867).

The 'stakeholders' Holbrook mentions, can be practitioners, professionals, local policy makers, businesses etc.

1.3 Improving vision statements and policy texts

If research-related institutions used the framework when producing text material, they would be able to express themselves more accurately. For example, the previously quoted part of the Research Policy Plan for 2017-2021 of UGent could be reformulated as follows:

> [openness] with respect to cross-disciplinarity: the research and the research policy stand for the collaboration between different scientific fields. This collaboration can be multidisciplinary or interdisciplinary.[8]

And, the previously quoted advice of VRWI could be rephrased as follows:

> [r]estructuring of funding mechanisms of knowledge institutions in order to allow for the stimulation of interdisciplinary and transdisciplinary collaborations.[9]

These formulations are more informative as they give a better idea of what types of CD are of interest to the institutions, and what types are not (viz. TD in case of UGent, MD in case of VRWI).

1.4 Brief overview of dedicated funding channels for CD projects in Flanders

In Flanders, there are three funding channels that are dedicated to CD projects: one at FWO, one at UGent and one at VUB.[10]

At FWO, funding is distributed via expert panels. Every year, the total available budget is allocated to the panels based on a distribution key, leaving each panel with its own budget to distribute among the proposals they receive.[11] In 2010, the 'Interdisciplinary Expert Panel' was added to the list of the 30 previously existing disciplinary panels (e.g. 'History and Archeology', 'Biodiversity and Ecology', 'Cancer Research'). Like the disciplinary panels, the Interdisciplinary Expert Panel distributes doctoral and postdoctoral grants as well as project grants. Project grants are assigned to the principal investigator(s) and typically allow for the hiring of one PhD researcher or postdoctoral researcher, and sometimes more.

---

[8] Here, 'transdisciplinary' is omitted because UGent explicitly states to be interested in the combining of *scientific* fields. (UGent, 2017, p.2)

[9] Here, 'multidisciplinary' is omitted because the report shows that VRWI attaches great importance to integration.

[10] KU Leuven and UHasselt do not have an operational dedicated channel, but are currently (thinking about) introducing one. The channel of the former will be focused on supporting networks rather than projects. Note that the absence of dedicated channels does not mean that CD projects are not being funded. It only means that they are in direct competition with disciplinary projects to obtain funding.

[11] (https://www.fwo.be/nl/mandaten-financiering/verdeling-toekenning-mandaten-per-panel/)

Since 2016, UGent spends part of its special research fund ('Bijzonder Onderzoeksfonds'- BOF) on biannual calls for *interdisciplinary* research projects.[12,13] Project proposals are submitted by two promotors[14] (with different disciplinary backgrounds), and the funding they can obtain is the equivalent of two doctoral research grants.

VUB has the longest history of dedicated funding of CD. In 2002, it used its special research fund to introduce a program for *interdisciplinary* research, viz. the so-called 'Horizontal Research Actions' ('Horizontale Onderzoeksacties' – HOA) (VLIR, 2010). In 2011, after positive evaluations, it was replaced by a program called 'Interdisciplinary Research Program' (INTDI). Like at UGent, funding is assigned to the principal investigators of a project and the grant allows for the hiring of two doctoral researchers.

1.5 Clarifying the scope of Flemish dedicated funding channels

The names of all three dedicated funding channels (viz. 'Interdisciplinary Expert Panel', 'Calls for interdisciplinary research projects' and 'Interdisciplinary Research Program') suggest that the scope of the channels consists in projects that are ID. Yet, given the previously explained ambiguity of 'interdisciplinary', the names fail to clarify what type(s) of projects are being targeted: it could be CD projects, or ID projects.

In what follows, the conceptual framework is used to clarify the scope of each of the channels. This is done by comparing general characterizations of the scopes of these channels to the concepts that are included in the framework and determining which concept(s) best capture them.

1.5.1 The scope of the FWO channel

On the FWO website, the scope of the Interdisciplinary Expert Panel is characterized as follows:

[12] (http://www.secretariaat.rvb.ugent.be/jaarverslag/jaarverslag-2017/jaarverslag2017-volledig.pdf)

[13] Every Flemish university has a special research fund. This fund is made available by the Flemish government on the condition that the receiving universities increase the provided budget by 12 %, using their own means. How the funds are spend is decided by the universities.

[14] In Flanders, 'promotor' is used as a synonym for 'principal investigator', or 'supervisor'.

> Interdisciplinarity means that progress can be made in each of the disciplines (belonging to a different scientific area e.g. Biological Sciences, Humanities, Social Sciences, Medical Sciences, Science and Technology) involved by the execution of the collective project, in each of the disciplines separately and/or on a collective area. Projects where one of the disciplines is only instrumental for the other, where one discipline for example supplies tools for a subject of another discipline, are not eligible.[15]

The mentioning of "progress on a collective area" suggests that (researchers from different) disciplines are expected to work closely together. This counts as an indication of FWO being interested in ID projects. However, the absence of 'integration' or related terms such as 'synthesis' or 'interconnection', and the option of a project leading to progress in individual disciplines point towards an interest in MD projects. The scope of the FWO channel thus appears to consist in ID and MD projects.

1.5.2 The scope of the UGent channel

The latest call for *interdisciplinary* projects that was issued by UGent contains the following statement:

> The objective is to enable a consortium, consisting of two professors, to establish a joint interdisciplinary research project. Promoters, with a different field of research, work together on the same research topic, which is studied from different perspectives.[16]

The interest of UGent in "joint projects" that are carried out by researchers who "work together", would be best answered to by ID projects. Yet, since all references to integration are missing, and projects should enable 'the study of a topic from different perspectives', it seems that also the dedicated funding channel of UGent is open to both proposals for ID and MD projects.

1.5.3 The scope of the VUB channel

In their Strategic Research Policy Plan for 2013-2017 ('Strategisch beleidsplan onderzoek 2013-2017'), VUB indicates that the Interdisciplinary Research Program is open to:

---

[15](https://www.fwo.be/en/the-fwo/organisation/fwo-expertpanels/panels-fundamental-research/interdisciplinary-research/interdisciplinary-expertpanel/)

[16] (https://www.ugent.be/en/research/funding/bof/iop/iop.htm)

> [groups of o]ne or more teams, of which most members have a high level of performance, that profile themselves as a thematic platform with a total expertise on societal challenges.[17] (translated by the authors)

The mentioning of societal, and thus complex challenges might be taken as a cue for an interest in ID projects. Yet, this scope description is too concise to allow to determine the scope of the VUB channel. It could consist in ID and/or MD projects.

1.5.4 Conclusions about scopes

The above analysis shows that the scope of the three dedicated channels could consist in ID and/or MD projects. Thus, the problem appears to be more fundamental than the names of the channels being ambiguous: the type(s) of projects that are targeted is undecided. This means that funding bodies have to make a choice. In case a funder decides for its channel to be(come) dedicated to both ID and MD projects, the name of the channel and the characterization of its scope need to be changed to reflect this decision. In such a case, it is important that applicants are asked to indicate whether their proposal is ID or MD. After all, this information is prerequisite for a fair selection process: when a proposal is ID, it should involve integration, but if it is MD, it need not to meet this criterion. (Selection criteria for CD projects are further discussed in Section 5.) Alternatively, a funding body can decide to restrict the scope of its dedicated channel to ID projects. While this decision does not require the name to be adjusted, the characterization of the scope has to be expanded to include a mention of ‘integration’. Furthermore, in case of focusing the channel of ID projects, the funder might consider the introduction of a channel for MD projects.

Note that, because none of the characterizations of the scopes mentions practitioners, TD is not taken into account in the remainder of this paper.

1.6 Improving the selection criteria of Flemish dedicated funding channels

Sets of selection criteria fix the scope of a dedicated funding channel (i.e. one or more types of CD). This requires the criteria to be complete, i.e. to capture the necessary and

[17] (https://www.vub.ac.be/sites/vub/files/research/vub-beleidsplan-onderzoeksraad.pdf)

sufficient conditions for falling within the scope; and to be coherent, i.e. they should not contain criteria that fix something else than the scope. The framework can be used to improve sets of selection criteria by making sure that they are complete and coherent. To be precise, it is not the framework itself, but rather *essential criteria* for ID and MD projects, which can be derived from the framework, that can be used. However, since the scope of the Flemish dedicated funding channels could not be determined, it is unclear what set(s) of essential criteria should be used to improve the sets of selection criteria. Hence, to enable a demonstration of the potential of the framework, the selection criteria of UGent and VUB will be improved *as if* the scope of the channels is restricted to ID projects. As is explained in Section 6.4., FWO has a set of selection criteria for MD projects. These will be improved by means of the essential criteria for MD projects.

In what follows, first, sets of essential criteria are derived and illustrated, and then are used to improve the selection criteria of the dedicated funding channels.

1.6.1 Deriving essential criteria for ID projects

1.6.1.1 Goal criterion

Klein's definition of 'interdisciplinarity' mentions the addressing of a "complex, question, problem, topic or theme". The idea that ID revolves around the addressing of complex problems is widespread (e.g. NAS, 2005; Bammer, 2012). For example, Uskali Mäki writes:

> It is commonplace to define interdisciplinarity rather richly in end-means terms. Interdisciplinarity in these characterizations manifests the ambition to solve broad or complex problems by combining and integrating two or more disciplinary perspectives themselves alone too narrow or simple for the task [.] (2016, p. 331)

ID is thus considered to be characterized by a goal that transcends (possible goals of) the involved disciplines. This allows to derive a first criterion ID projects should meet: the project should have a clear goal that is shared by all disciplines involved and that cannot be reached individually by any of the disciplines, i.e. each discipline should make a non-redundant contribution. This is the *goal criterion*.

A good example of a project that meets this criterion, is the tuning of social robot 'Probo' for therapeutic purposes by a team of engineers and psychologists (of VUB).[18] In its basic form, Probo is a green, elephant-like stuffed animal, which can interact via a voice and lip synchronization mechanism, but also non-verbally by means of a touch screen and by imitating human facial expressions. The ultimate goal is to use Probo in therapy for children with an autism spectrum disorder. However, before this goal can

[18] (http://probo.vub.ac.be/)

attained, another, intermediate goal needs to be reached: the basic form of Probo needs to be optimized by making it *huggable* and *expressive*. Here, ‘huggability’ refers to being soft to the touch and being visually agreeable (also when carrying out its basic functionalities) (Goris et al., 2011). ‘Expressiveness’ is understood as being able to express basic emotions (happiness, sadness, anger, surprise, fear and disgust) in a way that is recognizable for children and adults (Saldien et al., 2010). The tuning of Probo is a good example of a goal that transcends the potential of the disciplines involved as the engineers are not trained to evaluate whether the psychological desiderata are met and serve the desired therapeutic purposes, whilst psychologists would not be able to carry out the required technical work.

1.6.1.2 Necessity criterion

As explained above, Klein’s definition (and that of others) states that ID stands for the combining of disciplines to address a complex problem. Since disciplines are combined *in order to* reach a certain goal, i.e. the addressing (or even solving) of complex problems, only disciplines that are necessary to reach the goal should be involved. This yields a second criterion for ID projects: all disciplines involved in the project should be necessary to achieve the common goal. This is the *necessity criterion*.

Note that from an institutional perspective, the second essential criterion ensures that a project is financially coherent, i.e. that the distribution of funding across different parts of a university structure is justified by the fact that all of these parts will devote energy (staff, operating costs, infrastructure) to the project.

As explained above, in the case of Probo, it is clear that the overarching goal requires input from engineers and psychologists. However, if it were claimed that the project also requires a biologist, the necessity criterion would not be respected unless further explanation and justification would be provided (e.g. Probo is to be equipped with locomotion skills and biological or organismal forms of locomotion will be used for inspiration).

1.6.1.3 Integration criterion

As has been stated repeatedly, integration is central to Klein’s definition of ID. The third criterion for ID projects that follows from the conceptual framework is an obvious one: the project should integrate (elements from) different disciplines. This is the *integration criterion*.

As indicated above, there is no general definition of integration. Knowledge of ID integration in terms of hard methodological or conceptual criteria is still lacking (MacLeod & Nagatsu, 2018). Therefore, the vagueness of the third essential criterion is unavoidable. To enable judgement on the integrative nature of a project, the project proposal should clarify (i) what would be integrated, (ii) how the integration would be carried out, and (iii) why integration is required.

For example, in the case of Probo, given the combination of psychological and technical desiderata, it is concepts and methods that need to be integrated. With respect to concepts, ‘huggable’ and the basic emotions need to get a technical translation to become compatible with the engineering concepts in which the design and functionalities of Probo are described. ‘Huggable’ is translated in terms of coating and type(s) of actuators, and ‘expressivity’ is translated in terms of the positions of facial parts. With respect to methods, integration is required to allow for the evaluation of the parts and functionalities of the robot. In the final phase of the research, i.e. when the prototype of the robot is tested, multiple runs are required to determine whether everything functions properly. At the same time, the huggability and expressivity of the robot should be measured. To do this, evaluation methods from engineering and psychology need to be combined. Here (in the final phase), the integration of methods is synchronous as different methods are applied to the same research subject, at the same time. However, in earlier phases of research (when models of alternative designs are explored), the methodological integration is more of a diachronic nature: designs are created by the engineers and then evaluated by psychologists, whose feedback is taken into account in the subsequent round of design. The outcome of integration is a social robot that can be used for therapeutic purposes.

1.6.2 Improving the selection criteria of UGent and VUB

The essential criteria for ID projects can now be used to improve the selection criteria of the dedicated funding channels of UGent and VUB. Yet, before doing so, two remarks need to be made. First, it should be noted that only selection criteria for the *interdisciplinarity* of projects can be improved, and not the more general criteria concerning the quality of projects (e.g. criteria related to the innovative nature of the project, international scientific recognition of the promotors) that might also be provided. Secondly, selection criteria are often complex and contain multiple requirements. To keep things simple, in what follows, ‘requirement’ and ‘criterion’ are used interchangeably.

1.6.2.1 UGent

The most recent call for *interdisciplinary* research projects at UGent included the following selection criteria:

1. The disciplines and the kinds of expertise that are brought together in the proposal are dissimilar from one another. For both disciplines, a broad definition rather than a narrow one must be given. To verify whether the research disciplines of the promotors are sufficiently dissimilar, in first

instance, the FWO discipline codes are compared. Secondly, the committee examines whether the codes correspond with the content of the project proposal.
2. The input of expertise, knowledge and methodology from both disciplines must be equally essential in order to carry out the proposed research. The research project can only be carried out in an integrated, concerted way and cannot be split up into 2 separated research lines. One discipline cannot be an auxiliary science for the other.
3. The results of the interdisciplinary research lead to new scientific insights in both disciplines or contribute to the development of a new field of study.[19]

To be complete, the set of selection criteria of UGent should contain all of the essential criteria. It can now be assessed whether this is indeed the case, and, if not, what essential criteria should be included to make the set complete. The goal criterion is indirectly contained in selection criterion 3, which states that the results of an ID project may consist in contributions to "the development of a new field of study". The ability to make such a contribution implies that the (problem-solving) potential of the disciplines involved is transcended. Note that selection criterion 3 also allows an ID project to yield "new scientific insights in both disciplines". This means that, with respect to the nature of the desired research results, UGent is too mild. The necessity criterion is contained in selection criterion 2, which requires that input "from both disciplines must be equally essential in order to carry out the proposed method". Finally, the integration criterion is also contained in selection criterion 2, which mentions a request for 'integration' ("the research project can only be carried out in an integrated, concerted way"). Yet, to ensure that the integrative character of proposals can be judged, calls for projects should explicitly ask applicants to clarify what, how and why they plan to integrate.

To be coherent, the set of selection criteria should not contain any criterion that captures something other than the scope of the channel. In other words, they should not include criteria that cannot be reduced to one of the essential criteria. In the case of UGent, there are three criteria that do not correspond directly with an essential criterion (henceforth 'additional criteria') and for which it should be checked whether they can be reduced to one. If this is not the case, they should be discarded. Yet, exceptions should be made for selection criteria that cannot be reduced to an essential criterion, but are derived from a central policy of the funding body (and transcends the funding channel).

The first additional criterion is selection criterion 1. It requires that disciplines are sufficiently dissimilar, where dissimilarity is expressed in terms of FWO discipline codes. This criterion can be considered an operationalization of the integration criterion: if you request the integration of different disciplines, a standard is required to determine what

[19] (https://www.ugent.be/en/research/funding/bof/iop/iop.htm)

counts as a discipline. This is what is done by stating that disciplines are distinguished in terms of their FWO codes. Hence, the criterion can stay. The second additional selection criterion is the exclusion of the option for disciplines to be "auxiliary sciences for each other" (as part of selection criterion 2). This can be considered an (indirect) operationalization of the goal criterion: if discipline A is an auxiliary science for B, than the goal of the project belongs to B. Hence, also this criterion can be retained. The last additional selection criterion states that both disciplines that are involved in a project must be equally necessary to carry out the proposed research (as part of selection criterion 2). The inclusion of this requirement is probably an effect of the policy that project grants are divided equally among the principle investigators of the project. Given this policy, equal roles of disciplines ensures the financial coherence of projects. However, this means that, if alternative distributions become allowed for, the equality requirement should be omitted (or adjusted).

1.6.2.2 VUB

According to the Strategic Research Policy Plan 2013-2017 of VUB, to qualify for funding via the interdisciplinary networks, a project should:

> [1.] Entail a **synergy** between expertises of different disciplines, each of which is necessary for common results (theory /concept / tool/ technique / information / data set),
> [2.] Entail a **broad** interdisciplinary collaboration, which, preferably combines expertise from two or more of the three main domains HW-BNTW-BMW,
> [3.] Result in innovative contributions to each of the disciplines involved.[20]
> (VUB, 2013, annex 4, p. 37, bold in original, translated by the authors).

By applying the essential selection criteria, it becomes clear that this set of selection criteria is complete: both the goal criterion and the necessity criterion are contained in selection criterion 1 and, while the integration criterion is not mentioned explicitly, it is implied by the results needing to be 'common'. For reasons of clarity, it would be good if VUB included an explicit integration requirement.

When it comes to the coherence of the VUB criteria, there are three additional criteria that need to be looked into. The first one is the request for "broad" collaborations that are not restricted to the alpha-, beta- or gamma-group. Since it is unclear to what policy the request could be related and since it cannot be reduced to an essential criterion, it should be discarded. The second additional criterion states that an ID project should result in "innovative contributions to each of the disciplines involved". Like the equality requirement of UGent, the inclusion of this criterion is probably the effect of a policy

[20] (https://www.vub.ac.be/sites/vub/files/research/vub-beleidsplan-onderzoeksraad.pdf)

regarding the distribution of means. Yet, it should be pointed out that by requiring both common results (as is done in selection criterion 1) and progress in individual disciplines (as is done by criterion 3), the selection criteria are extreme. They require projects to be simultaneously ID and MD (Lyall et al., 2011). Thus, VUB should consider revising their set of selection criteria and make them less demanding.

1.6.3 Deriving essential criteria for MD

1.6.3.1 Juxtaposition criterion

Klein's definition indicates that MD stands for different disciplines working on the same subject in parallel, or in juxtaposition. This yields a first criterion for MD projects: the project should place different disciplines next to each other. This is the *juxtaposition criterion*.

A good example of a project that meets this criterion is the Automatic Monitoring for Cyberspace Applications (AMiCA) project, which revolves around the automatic detection of potentially threatening situations on social networks by means of text- and image-analysis.[21] The focus of the project is on the detection of cyberbullying and exposure to suicidal and sexually transgressive behavior. It consists of three disciplinary subprojects that are carried out by (i) language technologists (from Antwerp University and UGent), (ii) communication scientists (from Antwerp University) and (iii) engineers (from KU Leuven and UGent) respectively.

1.6.3.2 Complementarity criterion

Klein's definition of MD is in line with definitions that state that MD is about disciplines working independently in order to obtain a broader analysis of a problem (e.g. Rosenfield, 1992; Stokols et al. 2008). In order for an MD project to allow for such a broad analysis, its disciplinary subprojects should be complementary: they should supplement or enable each other in the light of the problem around which the project revolves. This yields a second criterion for MD projects: the different disciplinary subprojects should enable and/or complement each other in the light of a given goal. This is the *complementarity criterion*.

AMiCA is a good example of a project with complementary subprojects. As explained above, the project revolves around the automatic detection of harmful situations (related to bullying, suicide and sexual transgression) on social media. In the context of this goal, communication scientists study risks that are associated with social media use; language

[21] (https://research.flw.ugent.be/en/projects/amica)

technologists focus on the detection of threats in written text; and engineers work on the analysis of photo and video material. The subprojects of AMiCA are not only supplementing each other, but also enabling each other: the communication scientists created a simulated social network to determine whether role playing (with 'bullies', 'victims' and 'bystanders') leads to better behavior in real social media settings (Van den Broeck et al., 2014). The data generated during those role plays was used by the language technologists to train their self-learning algorithms (Van Hee et al., 2018).

1.6.3.3 Disciplinary progression criterion

As indicated by Klein's definition, MD stands for the studying of the same subject from different disciplinary angles, which means that MD research yields disciplinary knowledge. This allows to formulate the final essential criterion for MD projects: a project should lead to progress in the disciplines that are responsible for the different subprojects. This is the *disciplinary progression criterion*.

In the AMiCA project, progress was made within all disciplines involved. For example, in language technology, new techniques were developed for the standardization of Dutch chat-language, for the analysis of the sentiments underlying Dutch texts, and for the profiling of authors (in terms of age and gender). In engineering science, one advancement was the development of novel techniques for the detection of nude in photos (since this forms a strong indication of sexually transgressive behavior). In communication science, one of the main achievements consisted in novel insight into the nature (and ratio) of threatening situations social media users encounter.

1.6.4 Improving the selection criteria of FWO

Besides a description of the scope of the 'Interdisciplinary Expert Panel', the FWO website features a decision tree that allows to determine whether a project is *interdisciplinary*. Three of the six paths in the tree lead to a project qualifying as *interdisciplinary*. Of these three paths, two are associated with a type of CD that leans towards ID (because they include a criterion that requires an integrated result), and one with a type that tends more to MD. It is the set of criteria associated with the latter path that is improved here by means of the essential criterion for MD projects.

In order for a project to qualify as MD on the FWO decision tree, its disciplines need to be "coordinate" instead of "subordinate" disciplines. The legend of the diagram specifies that a project contains coordinate disciplines when:

1. [Results of the project] can be published in quality journals of different disciplines, to be specified by the applicant.
2. Mutual interactive input is necessary from specialists of different disciplines.

3. Subjects should offer benefit to and mutual influence on each other.

Secondly, the different disciplines involved should have a “mutual interest”, which is defined as the “confrontation, comparison, interaction of existing knowledge of different disciplines”. Finally, there is a selection criterion stating that the expertise required for the project may not be present within one discipline. [22]

Now, the essential criteria for MD projects can be used to make sure that the set of criteria used by the FWO is complete. The juxtaposition criterion is covered by (i) the idea that “the expertise required for the project may not be present within one discipline”, which implies a requirement for the involvement of multiple disciplines, and (ii) the requirement for the “confrontation, comparison or interaction” between disciplines because the first two relations require disciplines to be put next to each other. The complementarity criterion is well-represented. It is captured by the requirement for “benefit and mutual influence” and counts as a precondition for meaningful confrontation, comparison or interaction. The third essential criterion, viz. the disciplinary progression criterion, is represented indirectly by the requirement for publications in journals “of different disciplines”. While all three essential criteria for MD projects are included in the selection criteria of the FWO, none of them is mentioned explicitly. Doing so would make the selection criteria more transparent.

Regarding the coherence of the FWO criteria, there are three additional criteria that need to be evaluated. The first additional criterion requires disciplines to be coordinate, which is to be interpreted as them being equally important for the execution of the project. However, since the FWO leaves the distribution of granted means to the principle investigator of the project, the request for equal roles for the disciplines does not seem to be the effect of a policy. Instead, its inclusion appears to be idiosyncratic and hence, the criterion should be discarded. The second additional criterion demands “mutual interactive input”, which can be interpreted as requesting cross-fertilization. This can be considered an operationalization of the compatibility criterion. Hence, the criterion can be kept. Finally, the FWO requires the involvement of disciplines to be necessary. While this is an essential criterion for ID projects, its inclusion in a set of selection criteria for MD projects is superfluous and should be removed.

1.6.5 Conclusions about selection criteria

The improving of the selection criteria confirmed that the scope of the three Flemish dedicated funding channels is unclear as the sets of selection criteria of the channels are a (problematic) mixture of good criteria for ID projects and good criteria for MD projects.

[22] (https://www.fwo.be/media/110097/schema_interdisciplinair_20110202.pdf)

For example, criterion 3 of VUB (“a project should result in innovative contributions to each of the disciplines involved”) is a good criterion for MD projects, but not for ID projects. On the other hand, the selection criteria of FWO cover most of the essential criteria for MD projects, but also include one for ID projects (viz. that the involvement of each of the involved disciplines is necessary). Yet, because of the singularity of both types of CD, it is a lot to ask for a project to meet all of the criteria for both types of CD simultaneously. Hence, again, the solution is to make sure that the selection criteria fix a scope that consists in ID projects, MD projects, or both.

1.7 Conclusions and future research

Research-related institutions often lack conceptual clarity when it comes to cross-disciplinarity. This lack causes expressions of interest in cross-disciplinarity that are contained in mission statements and policy texts to be vague or ambiguous, thereby rendering those texts uninformative. Moreover, the conceptual unclarity affects the quality of the set-up of initiatives that are taken by institutions for the supporting of cross-disciplinary research, thereby decreasing the effectiveness of those initiatives. This paper introduced an unambiguous conceptual framework and showed how it can be used to overcome the lack of conceptual unclarity and the problems it causes. This was done by applying the framework to texts of research-related institutions in Flanders, Belgium. First, it was shown how the framework can be used to make mission statements more accurate, and thus, more informative for researchers who (aspire) to perform CD research. Next, the conceptual framework was applied to the set-up of dedicated funding channels. It was shown how the framework can be used to clarify the scope of these funding channels. Finally, after having derived essential criteria for ID and MD projects from the conceptual framework, it was shown how these essential criteria can be used to make sure that the selection criteria of the channels are complete and coherent.

It should be noted that the lack of conceptual clarity regarding cross-disciplinarity and the problems (P1-P3) it causes for the supporting of cross-disciplinary research, are not specific to Flanders. For example, also the National Science Foundation (NSF) of the United States has a funding program that is dedicated to cross-disciplinary research (viz. The ‘cross-cutting and NSF-wide program’), the set-up of which suffers from a lack of conceptual clarity.[23] Hence, this paper addressed general problems for the supporting of cross-disciplinary research. Moreover, the conceptual framework and the essential

[23] For more info on the program, see https://www.nsf.gov/funding/pgm_list.jsp?type=xcut.

criteria for ID and MD projects that were put forward, are not tailored to the Flemish case. Hence, this paper not only addressed general problems, but also presented general tools that can be used to solve them. From this point of view, sections in which the tools were applied to texts from Flemish research-related institutions provided exemplars, i.e. demonstrations of how the tools can be applied to a given case.

With the paper being focused exclusively on ID and MD research, future research that immediately presents itself is the articulation of essential criteria for TD projects. Furthermore, the sets of (already articulated) essential criteria for ID and MD projects and the-to-be articulated criteria for TD projects should be adapted to become applicable to other types of initiatives, such as the founding of CD institutes and the funding of CD networks.

Funding

This research was supported by the Research Foundation – Flanders (FWO).

Acknowledgement

The authors would like to thank Prof. Dr. Dingmar van Eck for his feedback that helped to improve the paper significantly.

References

Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature*, *534*(7609), pp. 684-687.

Bruce, A., Lyall, C., Tait, J., & Williams, R. (2004). Interdisciplinary integration in Europe: the case of the Fifth Framework Programme. *Futures*, *36*(4), pp. 457-470.

Czerniak, C., Weber Jr., W.B., Sandmann, A., & Ahern, J. (2010). A Literature Review of Science and Mathematics integration. *School Science and Mathematics*, 99(8), 421-430.

de Boer, Y., de Gier, A., Verschuur, M. & de Wit, B., (2006). *Bruggen Bouwen. Onderzoekers over hun ervaringen met interdisciplinair onderzoek in Nederland*. RMNO – KNAW – NWO – COS. Retrieved from: https://www.rvo.nl/sites/default/files/bijlagen/Onderzoekers_over_interdisciplinair_onderzoek_Bruggen_bouwen_Ymkje_de_Boer_2006.pdf

Deutsche Forschungsgemeinschaft (2016). Crossing Borders – Interdiscipli nary Reviews and Their Effects. An exploration based on new proposals for DFG individual grants (2005 to 2010). DFG. Retrieved from: http://www.dfg.de/en/dfg_profile/facts_figures/evaluation_studies_monitoring/studies/report_interdisciplinarity/index.html.

Eigenbrode, S. D., O'Rourke, M., Wulfhorst, J. D., Althoff, D. M., Goldberg, C. S., Merrill, K., & Bosque-Pérez, N. A. (2007). Employing philosophical dialogue in collaborative science. *BioScience*, *57*(1), pp. 55-64.

Geris, L. & Op de Beeck, H. (Eds.) (2015). *Interdisciplinariteit in Vlaanderen.* Brussel: KVAB.

Goris, K., Saldien, J., Vanderborght, B., & Lefeber, D. (2011). 'How to Achieve the Huggable Behavior of the Social Robot Probo? A reflection on the actuators'. *Mechatronics,* 21, pp. 490-500.

Heckhausen, H. (1972). Discipline and interdisciplinarity. In Apostel, L., Berger, G., Briggs, A., & Michaud, G. (Eds.), *Interdisciplinarity: problems of teaching and research in universities*, pp. 83-90. Belgium: OECD.

Holbrook, J.B. (2013). 'What is Interdisciplinary Communication? Reflections on the Very Idea of Disciplinary Integration'. *Synthese*, 190, pp. 1865–1879.

Huutoniemi, K., Klein, J. T., Bruun, H., & Hukkinen, J. (2010). Analyzing interdisciplinarity: Typology and indicators. *Research Policy*, *39*(1), pp. 79-88.

Klein, J.T. (2008). 'Evaluation of Interdisciplinary and Transdisciplinary Research. A Literature Review', *American Journal of Preventive Medicine,* 35, pp. S116-S123.

Klein, J. T. (2010). A taxonomy of interdisciplinarity. In Frodeman, R., Klein, J.T. & Mitcham, C. (Eds.), *The Oxford handbook of interdisciplinarity*, pp. 15-30. United States: Oxford University Press.

Klein, J.T. (2014). 'Communication and Collaboration in Interdisciplinary Research', In O'Rourke et al. (Eds.), *Enhancing Communication & Collaboration in Interdisciplinary Research*, pp. 11-30. United States: Sage Publications.

König, B., Diehl, K., Tscherning, K., & Helming, K. (2013). A framework for structuring interdisciplinary research management. *Research Policy*, 42(1), 261-272.

KU Leuven (2018). Strategisch plan KU Leuven. Op kruispunten, voor een duurzame samenleving. Retrieved from: https://www.kuleuven.be/over-kuleuven/beleidsplannen/onderzoeksbeleid.

Lindvig, K., & Hillersdal, L. (2019). Strategically Unclear? Organising Interdisciplinarity in an Excellence Programme of Interdisciplinary Research in Denmark. *Minerva*, *57*(1), 23-46.

Lyall, C., Bruce, A., Marsden, W., & Meagher, L. (2011). Identifying key success factors in the quest for interdisciplinary knowledge. *Report to NERC*.

MacLeod, M., & Nagatsu, M. (2016). Model coupling in resource economics: Conditions for effective interdisciplinary collaboration. *Philosophy of science*, *83*(3), pp. 412-433.

Mäki, U. (2016). 'Philosophy of interdisciplinarity. What? Why? How?', *European Journal for Philosophy of Science,* 6, pp. 327–342.

Monard, E., Schoofs, L., Sevrin, A., Vandewalle, J., Van Dyck, D., & Wenmackers, S. (2018). *Onderzoeker-gedreven wetenschap. Analyse van de situatie in Vlaanderen*. Brussel: KVAB.

Pohl, C. (2011). What is progress in transdisciplinary research?. *Futures*, *43*(6), pp. 618-626.

Saldien, J., Goris, K., Vanderborght, B., Vanderfaeillie, J., & Lefeber, D. (2010). 'Expressing Emotions with the Social Robot Probo'. *International Journal of Social Robotics,* 2, pp. 377–389.

Stokols, D., Hall, K. L., Taylor, B. K., & Moser, R. P. (2008). The science of team science: overview of the field and introduction to the supplement. *American journal of preventive medicine*, *35*(2), pp. S77-S89.

The National Academies (2005). *Facilitating Interdisciplinary Research*. Washington: The National Academies Press.

Toulmin, S. (1972). *Human understanding: The collective use and evolution of concepts*. Princeton, NJ: Princeton University Press.

UAntwerpen (2016). *Beleidsnota 2016-2020*. Retrieved from: https://www.uantwerpen.be/nl/overuantwerpen/missie-en-visie/.

UGent (2017). *Beleidsplan Onderzoek 2017-2021*. Retrieved from: https://www.ugent.be/nl/univgent/waarvoor-staat-ugent/onderzoeksbeleid.

Van Den Broeck, E., Poels, K., Vandebosch, H., & Van Royen, K. (2014). Online perspective-taking as an intervention tool against cyberbullying. *Studies in Health Technology and Informatics*, *199*, pp. 113–117.

Van Hee, C., Jacobs, G., Emmery, C., Desmet, B., Lefever, E., Verhoeven, B., De Pauw, G., Daelemans, W., & Hoste, V. (2018). Automatic detection of cyberbullying in social media text. *PloS one*, *13*(10), e0203794.

Vlaamse Interuniversitaire Raad (2010). Beoordeling van de kwaliteit van het onderzoeks-management van de Vlaamse universiteiten. *Brussel: VLIR*. Retrieved from: http://www.vlir.be/media/docs/Onderzoeksbeleid/VLIREvaluatieonderzoeksmanagement2010.pdf.

VUB (2013), *Strategisch Beleidsplan Onderzoek 2013*-2017. Retrieved from: https://www.vub.ac.be/sites/vub/files/research/vub-beleidsplan-onderzoeksraad.pdf.

Paper 2
Putting multidisciplinarity (back) on the map[1]

[1] This is a manuscript that has been submitted to the *European Journal for Philosophy of Science*.

Abstract

This paper focuses on the hierarchical structure of dominant theories of cross-disciplinarity, which consists in the theories systematically representing multidisciplinarity as the 'lowest' or 'least interesting' type of cross-disciplinarity. It is shown that this representation of multidisciplinarity is an effect of the theories being incomplete, and thus, that their hierarchical structure is unfounded. It is argued that addressing the unfounded hierarchical theories is important because their adoption provides (unjustified) support for having a preference for interdisciplinarity over multidisciplinarity. The issue is all the more pressing because research policy makers and funders are among the adopters of the theories and this implies a risk of (funding) policies reflecting an unjustified preference for interdisciplinarity, which would be harmful for the practice of cross-disciplinarity. After the investigation of (i) what the hierarchical structure of dominant theories consists in, (ii) how it arises, and (iii) why and (iv) how it should be neutralized, the paper presents an alternative, well-founded theory of that is not hierarchical, and presents multidisciplinarity as a full-fledged type of cross-disciplinarity.

Keywords

cross-disciplinarity • interdisciplinarity • multidisciplinarity • research funding • research policy making

2.1 Introduction

Throughout the last decades there has been a significant increase in funding opportunities for research that combines different disciplines. Many research funders formally adjusted the scope of their schemes in order to include research that involves multiple disciplines. For example, the National Science Foundation (NSF)[1] of the United States, the European Research Council (ERC)[2] and the Deutsche Forschungsgemeinschaft (DFG)[3] now accept research proposals that go beyond the boundaries of a single discipline. Moreover, dedicated funding programmes and grants have been created. Some of these are connected to a certain research area, like the 'Dynamics of coupled natural and human systems programme' (CNH)[4] of the NSF and the 'National Health Institute grants'[5] (also from the United States). Others aim at supporting research that involves multiple disciplines "tout court", such as the 'Discovery Programme'[6] of the Australian research council, 'Sinergia'[7] of the Swiss National Science Foundation (SNSF) and the 'National Academies Keck Futures Initiative'[8] (NAKFI) in the United States. Many of these policy changes and initiatives were brought into existence based on official research reports on *interdisciplinarity.*[9] For example, the current setup of ERC grants is partially based on a study of the European Research Advisory Board on how to "avoid the introduction of barriers to interdisciplinarity" (EURAB, 2004); the Australian Council of Learnt Academies commissioned professor Gabriele Bammer, a public policy fellow at the Australian National University (ANU), to investigate "the potential of interdisciplinarity and how to support it" (Bammer, 2012); and for NAKFI, the National Academies wrote a report on "the state of interdisciplinary research and education in science and engineering" and recommends "ways to facilitate them" (the National Academies, 2005 p. x). Some other

[1] (https://www.nsf.gov/od/oia/additional_resources/interdisciplinary_research/support.jsp)
[2] (https://erc.europa.eu/sites/default/files/document/file/ERC_Grant_Schemes.PDF)
[3] (http://www.dfg.de/en/dfg_profile/mission/index.html)
[4] (https://www.nsf.gov/pubs/2018/nsf18503/nsf18503.htm)
[5] (https://www.nih.gov/news-events/nih-research-grants-digital-press-kit)
[6] (https://www.arc.gov.au/grants/discovery-program)
[7] (http://www.snf.ch/en/funding/programmes/sinergia/Pages/default.aspx)
[8] (https://www.keckfutures.org/)
[9] Below, '*interdisciplinarity*' is assigned an alternative, more precise meaning than the one assigned by research funders. Until this alternative meaning is introduced, occurrences of the term are put in italics to indicate that the way in which it is used, is not the definitive way.

important official studies on *interdisciplinarity* are those of The Academy of Finland (Bruun et al., 2005), of the Swiss Academies of Arts and Sciences (Pohl et al., 2011) and of the League of European Research Universities (LERU) (Wernli & Darbellay, 2016).

The majority of official reports on *interdisciplinarity* as well as formulations of initiatives for the facilitation of *interdisciplinary* research and calls for proposals related to those initiatives, make use of one and the same conceptual framework. This framework contains three concepts of *interdisciplinarity*, which are distinguished by means of the degree and scope of integration: multidisciplinarity is non-integrative, interdisciplinarity is integrative, and transdisciplinarity is hyper integrative, or involves the integration of non-academic disciplines.[10] The framework is discussed in detail in Section 2.

While research policy makers, funders and researchers use the conceptual framework to think and talk about research that involves multiple disciplines, it is not ideally suited to be used for this purpose. The reason is language-related: in the framework, the term 'interdisciplinarity' is used both to refer to "the general category of research that crosses disciplinary boundaries", and to refer to "a specific type of research across disciplinary boundaries that is integrative". It is easy to see how this double role gives rise to misunderstandings. For example, if a funder issues a call for *interdisciplinary* proposals, it remains unclear whether the call aims at multi- inter- and transdisciplinary projects, or focuses exclusively on interdisciplinary (i.e. the specific type) projects. Solving this problem is fairly simple: add an extra term to the framework so that all of its concepts are referred to by unique terms. Following the suggestion of Sanford D. Eigenbrode et al. (2007) and Michael O'Rourke and Stephen Crowley (2013), in this paper, 'cross-disciplinarity' (CD) is used to refer to the general category of research that crosses disciplinary boundaries, and 'interdisciplinarity' (ID) is kept as the name for a specific type of research that crosses disciplinary boundaries and is integrative.[11]

Using the conceptual framework introduced above, this paper addresses another, more profound issue. It concerns the theories of CD that have come to be associated with the framework. These theories are hierarchical in that they represent multidisciplinarity (MD) as the "lowest" or least interesting type of CD. Thus, the adoption of such a theory is supportive of a preference for ID over MD. The problem, however, is that the subordinate role these theories ascribe to MD is unfounded. Yet, research policy makers and funders have adopted the hierarchical theories and use them when organizing and managing CD research. This entails a risk of (funding) policies reflecting an unjustified preference for ID, which, in turn would affect the behavior of researchers and lead to a skewed and disrupted practice of CD. Indeed, in practice a preference for ID can be

---

[10] Christian Pohl et al. (2011) explain that the 'hyper integrative' definition is mostly used in the United States, while the 'integrating non-academic disciplines' definition is more common in Europe.

[11] In this paper, both 'multidisciplinarity' and 'multidisciplinary' are abbreviated as 'MD'; 'interdisciplinarity and 'interdisciplinary' are abbreviated as 'ID'; and 'CD' and 'cross-disciplinary' are abbreviated as CD.

observed among research policy makers, funders and researchers. Irrespective of there being a causal link between theory and practice (establishing such a link requires research that falls outside the scope of this paper), it is important to address the unfounded hierarchical character of theories of CD since it provides unjustified support for having a preference for ID.

The first goal of this paper is to provide more insight into why the use of hierarchical theories of CD is problematic. To achieve this goal, it is clarified how the framework is associated with theories of CD, what the hierarchical structure of these theories (i.e. their subordination of MD) consists in, how it arises, and why it is unfounded. Moreover, an overview is provided of the ways in which a preference for ID among research policy makers, funders and researchers could lead to a skewed practice of CD. The second goal is to introduce an alternative theory of CD that, though still associated with the conceptual framework, is well-founded and represents MD as a valuable and full-fledged type of CD. Given that the use of unfounded hierarchical theories by policy makers and funders poses a serious risk of (unintentionally) leading to a skewed practice of CD, such an alternative theory is much needed.

Before continuing, the paper's scope need to be further clarified. It is limited to the practice of CD in so far as it concerns CD *research projects* that are carried out in *academic settings*. Here, CD research projects are taken to be units of research that revolve around one or more problems or questions related to a certain subject, and are carried out by fixed teams of researchers with different disciplinary backgrounds. The focus on such projects means that CD science education is not covered, nor are other organizational forms of research such as institutes, consortia or centers. Secondly, from the focus on academic settings it follows that transdisciplinarity falls outside the scope of this paper as it is often taken to include extra-academic knowledge and actors.

The rest of this paper is structured as follows. Section 2 outlines the conceptual framework of CD, explains how it is associated with theories of CD, and argues that one kind of theory became dominant among research policy makers, funders and researchers. Section 3 clarifies that theories of the dominant kind are hierarchical because that they present MD as an unfavorable type of CD. It also shows that their hierarchical structure is unfounded because it is an effect of the theories being incomplete. Section 4 discusses the preference of research policy makers, funders and researchers for ID research over MD research. Moreover, it explores the harmful effects of a preference for ID on the practice of CD. Section 5 identifies elements that should be integrated into dominant theories of CD to make their representation of MD more realistic. Section 6 presents an alternative, more neutral theory of CD (in line with the scope of this paper, only covering MD and ID). Section 7 contains conclusions and makes suggestions for further research.

2.2 The conceptual framework

In her book 'Interdisciplinarity: History, Theory, and Practice', Julie Klein presented a minimal taxonomy of CD that captured the common denominator(s) of the then plethora of theories of CD (Klein, 1990). Her taxonomy, or conceptual framework, has come to play a central role in the literature on CD. It consists of three concepts i.e. 'multi-', 'inter-' and 'transdisciplinarity'. The concepts stand for different types of CD that are distinguished by means of integration. Or, as Julie Klein puts it:

> [I]ntegration is the most common benchmark and, combined with degrees of disciplinary interaction, provides a comparative framework for understanding differences in types of interdisciplinary work. (Klein, 2010, p. 17)

Only consisting of three concepts and a distinguishing principle, the framework is highly abstract. When scholars of CD (e.g. philosophers of science or scientists with a meta-interest in their work across disciplinary boundaries) or other people (e.g. board members of science funding organizations) use the framework, they further enrich it, thereby generating theories of CD. In general, these theories consist of definitions for the three types of CD. More often than not, they also include characterizations that further clarify the relation between the three types. These characterizations are often of an evaluative nature. Thus, while the conceptual framework is highly abstract and neutral regarding the value of the three types of CD, theories of CD are more detailed and normatively loaded.

Though there are different theories of CD, one kind of theory has clearly become dominant (Lattuca, 2001; Klein, 2010; Mobjörk, 2010; Holbrook, 2013). It includes definitions that reflect the following ideas:

- 'multidisciplinarity' refers to the collaboration of researchers with different backgrounds where their respective disciplines are juxtaposed instead of integrated (e.g. Holbrook, 2013);

- 'interdisciplinarity' stands for the collaboration of researchers with different disciplinary backgrounds where (elements of) the respective disciplines are integrated (e.g. Klein, 2014), and

- 'transdisciplinarity' either refers to a collaboration where the integration of (elements from) different disciplines is so extensive that the origin of the elements gets lost, or refers to a collaboration of researchers and non-academics, such as stakeholders and/or practitioners who integrate their knowledge and know-how (Pohl et al., 2011).

It should be noted that, even though the definitions above are more substantial than the concepts included in the framework, they are still rather imprecise as the two main defining notions, i.e. discipline and integration, remain vague. Scholars of CD have long recognized the importance of clarifying both notions, and have taken up the task of doing so.[12] However, a clarification of 'discipline' and 'integration' falls outside the scope of this paper. Its goal is not to put forward a fully operational theory of CD, but rather to show that the theories of CD that are currently used by research policy makers, funders and researchers are not ideally suited to serve as the basis for the organization of CD practice (given their unsubstantiated hierarchical structure) and to indicate how they should be adjusted.

2.3 The unfavorable representation of multidisciplinarity

Theories of CD of the dominant kind (henceforth 'dominant theories') are hierarchical as they represent MD as the lowest, and least interesting type of CD. To gain insight into what this unfavorable representation of MD consists in, how it arises, and why it is unfounded, definitions of 'ID' and 'MD' included in such theories (henceforth 'dominant definitions') are contrasted. The same is done for characterizations of ID and MD (henceforth 'dominant characterizations').

2.3.1 Contrasting dominant definitions

Consider the following dominant definitions of 'ID':

[12] With respect to 'integration', general characterizations such as provided by Holbrook (2013) are extremely scarce, but field-specific characterizations can be found. For example, biology-bound characterizations are provided in (Lionelli, 2013) and in (O'Rourke, et al., 2016) and natural resources management-related one in (Land & Water Australia, 2004). Furthermore, there are characterizations of integration of specific types of data. For example, MacLeod & Nagatsu (2018) characterize integration of models. Finally, alternatives for 'integration' have been put for forward, for example in (Grüne-Yanoff, 2015). Characterizations of 'discipline' can be found in (Darden & Maull, 1977; Turner, 2000; Lattuca, 2001; Lélé & Norgaard, 2005).

> Incorporation or integration of disciplinary perspectives into a larger, more holistic perspective is the chief distinguishing characteristic of interdisciplinary studies.[13] (Newell, 1992, pp. 212-213)

> Interdisciplinary research similarly approaches an issue from a range of disciplinary perspectives but in this case the contributions of the various disciplines are integrated to provide a holistic system outcome. (Bruce et al., 2004, p. 459)

> Interdisciplinarity is an interactive process in which researchers work jointly, each drawing from his or her own discipline-specific perspective, to address a common research problem. (Rosenfield, 1992 in Stokols et al., 2008, p. S79)

> Interdisciplinarity integrates information, data, methods, tools, concepts or theories from two or more disciplines or bodies of knowledge to address a complex question, problem, topic or theme. Work may occur individually or in teams, though, in the latter case, communication is essential to successful collaboration. (Klein, 2013, p. 13)

> Interdisciplinary research fosters linkages between disparate disciplines while some aspects of each discrete discipline are still recognizable and left intact (Strober, 2006). For example, researchers may integrate methods by mixing methods from various disciplines or create a shared language to approach a common problem, which can lead to a more holistic approach to problem-solving than a single discipline would allow (Buizer et al., 2015; Hickey and Nitschke, 2005). (Pischke et al., 2017, p. 1012)

These definitions provide information on six dimensions of ID. The first dimension is that of 'CD' and regards the number of disciplines that is involved. For example, the definitions by Ann Bruce et al. and Klein mention the involvement of "one or more disciplines". The second dimension is that of 'discipline combination' and is about the presence or degree of integration of disciplines. All five definitions cited above indicate somehow that ID involves integration. Thirdly, there is the dimension of 'research outcome', i.e. what the results of the research look like. This outcome is described as "holistic" by Bruce et al. (2004) and Erin C. Pischke (2017). The fourth dimension is that of 'research organisation' and regards the way in which research is carried out. Concerning this dimension, Patricia L. Rosenfield writes about "researchers work jointly" and Klein mentions "collaboration in teams". The fifth dimension is that of the 'use of disciplines' and is about the way in

[13] Strickly speaking, William Newell's definition should not be included in the list since it is one of interdisciplinary *studies*, which relates to education rather than research. However, given that Newell played an important role in establishing the framework and co-authored some important papers with Klein, his definition is nevertheless included.

which disciplines are deployed in the course of research. The definitions by Klein and Pischke et al. indicate that it is *parts* of disciplines that are being integrated, implying that disciplines are "broken up". The sixth and last dimension is that of 'motivation', i.e. why the research is taken up on. Three out of five definitions state that ID is problem driven (i.e. Stokols et al., 2008; Klein, 2013; Pischke et al., 2017).

Next, consider the following dominant definitions of 'MD':

> Multidisciplinary research involves low levels of collaboration, does not challenge the structure or functioning of academic communities and does not require any changes in the academic worldviews of the researchers themselves. (Bruce et al., 2004, p. 459)

> Multidisciplinarity is a sequential process whereby researchers in different disciplines work independently, each from his or her own discipline-specific perspective, with a goal of eventually combining efforts to address a common research problem. (Rosenfield, 1992 in Stokols et al., 2008, p. S79)

> A multidisciplinary approach to a problem would involve multiple disciplines each investigating the problem in its own way, with its own definition of the problem, according to its own standards, and arriving at its own independent solution. (Holbrook, 2013, p. 1867)

> Multidisciplinarity juxtaposes two or more disciplines or bodies of knowledge focused on a common problem, question, topic or theme. [C]ommunication occurs typically at the level of coordinating information, not collaborating. (Klein, 2013, p. 12)

> Multidisciplinarity is the placing side by side of insights from two or more disciplines without attempting to integrate them. (Repko et al., 2008, p. 13)

These definitions of 'MD' provide information on the same six dimensions as the ones of 'ID'. Two dimensions are even filled in similarly, i.e. that of 'CD' (e.g. 'two or more disciplines' – Klein, 2013) and that of 'motivation' (e.g. 'focused on a common problem' – Klein, 2013). Regarding the dimension of 'discipline combination', it is indicated that integration of disciplines is absent, but that instead, disciplines are juxtaposed. For example, Rosenfield mentions that "each researcher works from his or her own discipline-specific perspective" and Allen F. Repko writes that "there are no attempts at integration". The dimension of 'research outcome' is addressed by Holbrook. She explains that results are discipline-specific. When it comes to the dimension of 'research organization', it is stressed repeatedly that there is no collaboration among researchers with different disciplinary backgrounds (e.g. Bruce et al., 2004; Stokols et al., 2008; Klein, 2013). Finally, regarding the dimension of 'use of disciplines', it is mentioned that boundaries of disciplines are respected, which implies that disciplines are put to work in their entirety (e.g. Bruce et al., 2004; Stokols et al., 2008; Holbrook, 2013).

Despite covering the same dimensions, and two of them being filled in similarly, there is a striking difference between dominant definitions of 'ID' and 'MD'. While the first consist exclusively of positive elements (i.e. elements that indicate what ID *is*), the latter also contain negative elements (i.e. elements that indicate what MD *is not*). For example, where ID is defined in terms of "integration" and "joint work", MD is defined in terms of "no integration" and "no collaboration". An extreme example of a negative definition is the one by Pischke et al. (2017):

> Multidisciplinary [...] research involves researchers from more than one discipline who do not attempt to interconnect and integrate their research methods (Hickey and Nitschke, 2005; Strober, 2006). (Pischke et al., 2017)

While (sometimes) useful for distinguishing MD from ID, the inclusion of negative elements into definitions of 'MD' makes them pejorative and less informative.

A second difference between definitions of 'ID' and 'MD' is that, while the first distinguish ID from the alternatives of mono-disciplinarity (understood as collections of independent disciplinary research projects that revolve around the same subject) and of MD, the latter fail to do the same for MD. Definitions of 'MD' do point out where the difference lies with ID (i.e. 'no integration'), but do not make clear how the type of CD differs from mono-disciplinarity.

2.3.2 Contrasting dominant characterizations

Also in characterizations included in dominant theories of CD, there is a stark contrast between the ways in which ID and MD are approached. While ID is praised for allowing to generate "innovative solutions to complex, multi-dimensional, policy-related problems" (Lyall & Fletcher, 2013, p. 1) as well as for enabling "the advancement of knowledge and the yielding of comprehensive explanations" (Boix Mansilla et al., 2006), it is mostly the limitations of MD that are discussed:

> [MD] research is not usually conceptually pathbreaking but has shed light on different aspects of a particular problem, leading to immediate, but possibly short-lived solutions. (Rosenfield, 1992, p. 1351)

> [In MD,] researchers may contribute a few pieces to the jigsaw puzzle, but there is no improved understanding of the nature of the picture as a whole, and no fundamental change in perception, understanding or quality of knowledge-based outcomes. (Lyall et al., 2015, p. 13)

> Since no real cooperation among the disciplinary practitioners was assumed, MD projects did not result in changes or enrichment of the participating disciplines. Such disciplinary relations were thought to be transitory and limited. (Klein 1990 in Lattuca, 2001, p. 10).

> By requiring only close proximity, rather than integration, of different disciplines, multidisciplinarity lacks the transformative qualities in both knowledge and practice that are necessary to address complex environmental problems. (Clark & Wallace, 2015, p. 240)

> [Multidisciplinarity] can be compared to a fruit salad containing a variety of fruits, each fruit representing a discipline and each fruit being in close proximity to the others. The number of fruits used and the proportions of each in the salad may not be based on anything more than visual appeal. This is not so with [interdisciplinarity], however, which Moti Nissani (1995) compares to a "smoothie." The smoothie is "finely blended so that the distinctive flavor of each [fruit] is no longer recognizable, yielding instead the delectable experience of the smoothie (p. 125).[14] (Repko, 2008, p. 11)

2.3.3 What can be learned from the contrasts

The above contrasts show that the unfavorable representation of MD in dominant theories consists in portraying it as an unattractive and vague type of CD. The contrasts also provide insight into how this unfavorable representation arises. The depiction of MD as unattractive is caused by (i) dominant definitions being pejorative and (ii) characterizations failing to clarify the potential of MD. The image of MD being vague is an effect of dominant definitions (iii) containing negative elements and (iv) containing insufficient elements that allow to distinguish MD from mono-disciplinarity. Except for the pejorative phrasing, all causes involve a lack of information, i.e. on the potential of MD, on actual (positive) features of MD and on the difference between MD and mono-disciplinarity. Since the theories do provide the corresponding information on ID, rather than information missing, the theories are incomplete.[15] Because the unfavorable representation of MD is (primordially) caused by the incompleteness of the theories, it is unfounded.

[14] Repko writes about multi- and interdisciplinary *studies*. However, since in the 1995 paper of Nissani the metaphors apply both for CD education and research, in the quote above, 'multi- and interdisciplinary studies' can be replaced by 'multi- and interdisciplinarity'.

[15] Note that this summary captures the union of the limitations of dominant representations of multidisciplinarity. These limitations are not equally present in all dominant representations.

2.4 The preference for ID research

Research policy makers and funders have adopted dominant theories of CD. This clearly shows from their mission statements, policy notes and calls for projects. As stated repeatedly, the adoption of such theories provides unfounded support for a preference of ID over MD. This means that there is a risk of policy makers and funders developing an unjustified preference for ID and of organizing and managing CD research accordingly, which, in turn would (unwarrantedly) nudge researchers to prefer ID research over MD research. In the practice of CD such a preference can indeed be observed, both on the side of policy makers and funders and on the side of researchers. Even if this preference would turn out not to be caused by the adoption of dominant theories, it is definitely supported by it. This already makes it worthwhile to briefly look into the preference. Therefore, this section discusses manifestations of the preference and zooms in on its harmful effects on CD practice.

2.4.1 Manifestations of the preference

The preference for ID research (over MD research) on the side of research policy makers and funders is best illustrated by the inclusion of requests for integration and/or innovation in descriptions of CD programs and in calls for CD projects (Eigenbrode et al., 2007; Petts et al., 2008). For example, the NSF has eleven categories of funding opportunities, two of which target CD proposals. One CD category is labeled 'integrative activities' and targets ID and TD research proposals. The other one is called 'crosscutting and NSF-wide' and targets "activities sponsored by more than one NSF organization"[16]. While both the name and scope description of the second category suggest openness to all types of CD, the calls for proposals belonging to the category request 'connected', 'comprehensive' and/or 'integrated' approaches. Another can be found at the ERC. On their website, it is indicated that they welcome proposals that "cross boundaries between different fields of research, pioneering proposals addressing new and emerging fields of research or proposals introducing unconventional, innovative approaches and scientific inventions"[17]. The request for contributions to new and emerging fields or for the introduction of innovations, rules out MD and favors ID.

Among researchers, the preference for ID research is best illustrated by their tendency to present their (aspired) projects as interdisciplinary, even when they are in fact MD.

[16] https://www.nsf.gov/funding/pgm_list.jsp?type=xcut

[17] http://ec.europa.eu/research/participants/data/ref/h2020/other/guides_for_applicants/h2020-guide16-erc-stg-cog_en.pdf

The unjust use of the label 'interdisciplinary' by research groups has been observed on multiple occasions (e.g. by Kostoff, 2002; Mestenhauser, 2002; Clark & Wallace, 2015). Bruce et al. (2004) confirmed these observations in their study of projects that obtained funding via the European Union Fifth Framework Programme (FP5). For this study, they sent a survey to the coordinators of 754 FP5 projects that contained the question whether their project should be categorized as mono-, multi-, interdisciplinary. Of the 160 coordinators who responded, 66% classified their project as interdisciplinary and 22% considered their project to be multidisciplinary. Yet, it was found that many projects were categorized incorrectly:

> [W]e found [that] the degree of interdisciplinarity varied enormously among the projects rating themselves as interdisciplinary, with very few projects fully integrating disciplines. (2004, p. 462)

In many cases, instead of 'interdisciplinary', the label of 'multi-' or 'transdisciplinarity' would have been more appropriate:

> [A]mong the researchers and projects we studied, we found many examples of multidisciplinary research and also some examples of transdisciplinary research. (2004, p. 459)

Diana Rothen made a similar observation in the context of a study on ID research centres and programmes commissioned by the NSF:

> It was common to hear, for example, the mechanical engineer, atmospheric physicist, and public policy analyst describing themselves as "co-investigators on an interdisciplinary project" yet to observe them conducting their respective pieces of the research in near isolation from one another. (2004, p. 6)

This tendency towards opting for the label 'interdisciplinary' could be explained by a lack of awareness of the option of calling one's work 'MD', but only to some extent. For example, in the case of the survey of Bruce et al., coordinators were made aware of the option of labeling their work as MD. Moreover, in general, definitions of the tree types of CD are so common in publications and on webpages of governments and funders that it is hard to avoid learning about the existence of MD while searching for funding opportunities. It seems more likely that the tendency is an expression of a strategic preference for ID research, i.e. a preference that is adopted to maximize the chance of obtaining funding (given the current landscape created by research policy makers and funders).

2.4.2 Effects of a preference for ID

A preference for ID research can cause serious problems, both on the side of CD researchers and on that of research policy makers and funders. This is also hinted at by Bruce et al.:

> It would promote the overall quality of research and the effectiveness of research programmes if clearer distinctions were made between projects where interdisciplinary work is really valuable and where it is not so important for the outcomes. (2004, p. 460)

When it comes to research teams, it is those that (aspire to) work on a MD research project that are most at risk of being disadvantaged by a preference for ID. First and foremost, they might be impeded by the preference for ID research that is propagated by policy makers and funders. For example, upon perceiving that their MD project does not meet the expectations of funders (i.e. "CD projects must be interdisciplinary"), a research team might decide to cancel the project. If this scenario takes place repeatedly, it implies a serious impoverishment of the scientific landscape. Alternatively, a research team might decide to pitch their MD project as ID, thereby flirting with deontological boundaries as they misrepresent the CD nature of their project. Research teams are also at risk of being negatively affected by their own preference for ID research. For example, a team's wish for their project to be interdisciplinary might lower the accuracy of their conception of the project, e.g. their knowledge of how much integration the project requires, and what form of collaboration is required to achieve that level of integration. Having an inaccurate conception of one's project is not just unproductive, but counterproductive. A good understanding of the planning, the division of roles and coordination is crucial for successful CD collaboration (Enderby, 2002; Nancarrow et al., 2013).

Also research policy makers and funders are not immune to negative effects of a preference for ID research. First, research policy makers can be misled by researchers. For example, a common first step in evaluating CD project proposals is an assessment of their CD nature (where different types of CD are assessed by different criteria), but when researchers mislabel their projects, or coin them in misleading terms, it becomes hard to determine what evaluation criteria to use. Using the wrong set of criteria might lead to the rejection of a valuable project. Moreover, research policy makers and funders are vulnerable for the effects of their own preference for ID research. For example, while efforts to facilitate CD are usually aimed at "[pushing] fields forward and accelerating scientific discovery" (NSF) and to "address the grand challenges that society faces" (Gleed & Marchant, 2006, p. 5), a preoccupation with ID might delay, or even prevent from reaching those goals. After all, there is no information on desirable ratios of different types of CD in the light of the goals, and hence, it could be the case that the goals require more MD projects (see, for example, Evans & Macnaughton, 2004).

In short, a preference for ID research, or a neglect of MD research, undermines the possibility of a balanced practice of CD research. However, it should be noted that, if the preference for ID is a deliberate one, the negative effects listed above will be accepted as they are offset by other reasons for preferring ID. Yet, it is exactly in the selection of such reasons that research policy makers and funders might be misguided: if they adopt a dominant theory of CD, they adopt a theory in which ID is presented as better than MD, which could be a sufficient, or decisive reason to opt for ID. However, the representation of ID as superior is unfounded, and thus does not justify a preference for ID. This is why the adoption of dominant theories by research policy makers and funders poses the risk of causing an unjustified preference for ID to be reflected in research (funding) policies.

2.5 Collecting elements of a more neutral theory

In what follows, an alternative theory of CD is developed. It is still associated with the conceptual framework once introduced by Klein, but, in contrast to dominant theories, is well-founded and contains a more neutral representation of MD. Since the unfoundedness of dominant theories is an effect of their incompleteness, the alternative theory is developed by complementing them with the information they are missing. This means that (i) dominant characterizations of MD are enriched with elements that clarify the potential of MD (henceforth 'valuative elements'), (ii) dominant definitions of MD are complemented with elements that allow to distinguish it from mono-disciplinarity (henceforth 'non-mono-disciplinary elements'), and (iii) the negative elements of dominant definitions of 'MD' are replaced by positive ones. In what follows, the necessary valuative, non-mono-disciplinary and positive elements are identified by means of a critical analysis of elements included in dominant theories. Positive elements are mentioned in the subtitles, the others are discussed more extensively. Table 2 provides an overview of all collected elements.

2.5.1 Discipline combination: no integration, but juxtaposition

Every dominant definition somehow mentions that MD does not involve the integration of disciplines. While some straightforwardly state that "multidisciplinarity does not involve integration" (e.g. Repko et al., 2008; Pischke et al., 2017), others phrase it in terms of a lack of "synergy" (Okumus & van Niekerk, 2015) or "interconnection" (Chon-Torres, 2018). What happens to disciplines instead of being integrated is usually described as 'juxtaposition' (e.g. Miller, 1982; Klein 1990). Yet, in the literature, little can be found about what 'juxtaposition' means. One cue is Erich Jantsch's distinction between "a variety of disciplines" and "a juxtaposition of disciplines", where the latter is defined as "involving some sort of co-ordination among different disciplines" (1972, p. 107). Another one is Klein's comparison of "juxtaposition" to "assemblage" (2010, p. 17) Finally, Lisa Latucca picks up on the suggestion of one of her "informants on cross-disciplinarity" that juxtaposing stands for the combining of things that "typically don't go together" (2001, p. 74). All three cues imply that a juxtaposition presupposes a certain level of deliberate organization. Outside the literature on CD, a similar idea can be found. In dictionaries, 'juxtaposition' is defined as 'the placing of things next to each other to obtain a contrastive or comparative effect' (Merriam-Webster.com, n.d). From such a definition, it follows that there needs to be a certain relation between things that are juxtaposed, which makes putting them next to each other interesting. In the case of MD, these 'things' are disciplines and the relevant relations revolve around how the disciplines compare to each other in the light of the central research subject given their method(s), research subject(s), level(s) of description, etc. Possible relations include ones of conflict, support and competition.

This analysis points at two valuative elements. First, given that MD research involves the juxtaposing of disciplines, it allows to add (disciplinary) breath to a project without having to operate in the no man's land between disciplines. Or, as Klein puts it: "Juxtaposition fosters wider knowledge, information and methods" (2010, p. 17). A second valuative element is that MD allows to deepen the knowledge about how disciplines relate to each other in the light of a given subject. While the basic relation ought to be defined at the outset of an MD project, in the course of it, more detailed insights will be gained. What these insights are, depends on the type of relation. For example, when conflicting or competing disciplines are juxtaposed, one learns how their fruitfulness compares, and when they are supportive, one can learn about their interdependence.

The analysis also yields the following non-mono-disciplinary element: the disciplines involved in an MD project are carefully selected. This distinguishes MD from mono-disciplinarity since, in collections of independent projects, the relation between disciplines is completely irrelevant.

2.5.2 Research organization: no collaboration, but coordination

Besides informing on the results of a type of CD research (i.e. integrated or not), most dominant theories also include information on the process or organization of the type of CD research. For example, definitions of 'MD' often mention that "there is no collaboration" between researchers involved in the different disciplinary sub projects (e.g. Bruce et al. 2004), but that instead, their work is "coordinated" (Klein, 2013). A rare source on what CD coordination can consist in, is the paper 'Working alone together' by

Hille Bruns (2013). She defines ‘coordination’ as “the ongoing accomplishment of managing interdependencies in collective work” (2013, p. 36). In any MD project, the first interdependency that needs to be managed is the problem (e.g. Klein, 2013) or question (e.g. Cluck, 1980) that constitutes the core of the project. Since it has to be something that is recognized, or shared by all disciplines involved, it needs to be carefully negotiated. After that, there is at least one more interdependency that needs to be managed, i.e. timing. MD is often characterized as the “simultaneous” execution of different disciplinary sub projects (e.g. Jantsch, 1972; Marques, 2008). Strictly speaking ‘simultaneous’ stands for “happening at the same time”. However, a more broad interpretation of ‘simultaneous’ is also compatible with alternation (de Boer et al., 2014) or a succession (Cluck, 1980) of sub projects. Other potential interdependencies include terminology, infrastructure and evaluation standards (Brun, 2013).

Because of the simultaneity of its sub projects, an MD project allows to carry out research at reduced costs. For example, researchers responsible for different sub projects can go on joint expeditions or share expensive materials or infrastructure (Huutomieni et al. 2010, p. 83). This is an important valuative element. Note however, that MD projects that are set up exclusively with the goal of saving costs are edge cases of MD since they are defined by shared research conditions rather than by a shared a problem or question.

The above analysis also yields three non-mono-disciplinary elements. First, MD projects are characterized by a shared problem or question. Secondly, taking a MD approach entails strategic timing of the sub projects, i.e. in perfect parallel (i.e. in the exact same time period), in dialogue or in succession. Thirdly, given that the different sub projects are coordinated, they can be expected to be complementary. All three of these features are absent in collections of independent mono-disciplinary projects on the same subject.

2.5.3 Research outcome: no innovative solutions, but cumulation

Some dominant definitions and characterizations indicate that MD does not allow for solving complex issues (e.g. Rosenfield, 1992; Blevis & Stolterman, 2009; Clark & Wallace, 2015) nor for producing innovative solutions (e.g. Paletz et al., 2010). The reasoning behind this is that instead of the integrated resolutions that are required to settle complex issues, MD results in distinct and partial disciplinary solutions. Because of the nature of its output, MD is commonly described as “cumulative” (e.g. Huutoniemi et al., 2010; Zaman & Goschin, 2010) or “additive” (e.g. Klein, 1990).

This analysis points at three valuative elements. First, by being cumulative, MD projects allow to obtain broad knowledge on a given problem or question, making them valuable preliminary studies for large scale research projects (e.g. Repko, 2007), policy making, or ID projects. This idea that MD makes a great forerunner for ID is also expressed by the ZiF (i.e. the 'Center for Interdisciplinary Research' of the University of Bielefeld), which "considers sustained multidisciplinary cooperation crucial in order to eventually induce interdisciplinary" (Maasen, 2000, p. 177). The second valuative element is that, since the output of MD research is not integrated, it is versatile. The knowledge that is produced, can be integrated by different actors according to their needs and/or values. Thirdly, not all broad problems/questions are complex and (eventually) require an ID approach, some can be broken down into disciplinary sub problems/questions (e.g. Ross, 2009). In such cases, the additiveness of MD allows to resolve the problem by addressing individual sub problems/questions.

2.5.4 Use of disciplines. No challenging of disciplinary boundaries, but exchange between disciplines in their entirety

Though already implicit in 'juxtaposition' and 'no collaboration', many dominant definitions mention explicitly that MD research 'maintains disciplinary boundaries' or, that it 'does not challenge the structure of academic communities' (e.g. Bruce et al., 2004; Choi & Pak, 2006; Holbrook 2013), which means that disciplines are used in their entirety. Yet, the conservation of disciplinary boundaries does not mean that disciplinary sub projects are carried out in perfect isolation. For example, in their definition of 'MD', Gunther Tress et al. (2005) indicate that members of MD teams do exchange knowledge (but are quick to add that they do not create new, integrative knowledge), and Huutomieni et al. mention that knowledge is 'imported, exported or pooled across boundaries' (2010). Some more information on the nature of these 'exchanges' can be found in a report by Ymke De Boer et al. (2006) on CD in the Netherlands. The report distinguishes between five types of 'scientific collaboration'. Though the typology is not associated with the conceptual framework introduced by Klein, De Boer et al. indicate that, through the lens of framework, two types of their typology can be considered MD. In collaborations of the first type (i.e. "mutual knowledge exchange"), the exchange between disciplines is limited to developing understanding of each other's sub projects (and their results). In collaborations of the second type (i.e. "mutual influence"), the exchange is more extensive as some sub projects require input from other sub projects.

Discussions of the integrity of disciplines in MD point to three more valuative elements. First, since research is performed within disciplines, MD allows for disciplines to further develop and grow. Secondly, given its conservatism regarding disciplinary boundaries, MD involves a lower risk than integrative forms of CD (Lyall & Meagher, 2012). Finally, the exchange between disciplines fosters cross-fertilization of disciplines.

Since the exchange between disciplines can be very limited (or even absent), it does not provide an additional non-mono-disciplinary element.

Table 2   Overview of the collected elements

Dimension	Positive elements	Valuative elements	Non-mono-disciplinary elements
Discipline combination	• Juxtaposition of disciplines	• Broad knowledge obtained • Insight into the relation(s) between disciplines	• Careful selection of disciplines
Research organization	• Coordinated research	• Reduced research costs	• Addressing of a shared problem or question • Strategic timing of sub projects • Complementary sub project
Research outcome	• Cumulative outcome	• Preliminary study • Versatile research results • Non-complex problem-solving	
Use of disciplines	• Disciplines used in their entirety • Knowledge exchanged between disciplines	• Progress in disciplines • Low risk • Cross-fertilization	

2.6 An alternative, more neutral theory of CD

By integrating the positive and the non-mono-disciplinary elements obtained from the analysis into (a summary of) dominant definitions of 'MD', the following, alternative definition of 'MD' is obtained:

> Multidisciplinarity is a type of CD that involves the juxtaposition of strategically selected disciplines. In MD projects, this juxtaposition is realized via distinct, disciplinary sub projects. While these sub projects are largely independent, they revolve around a shared problem or a question. To be able to contribute to the solving of this problem, the sub projects are complementary and are carried out in coordination. Depending on how the disciplines responsible for the sub projects relate to each other in the light of the central problem or question (e.g. a relation of conflict, complementarity or supportiveness), the sub projects can be carried out in parallel, alteration or succession. Multidisciplinarity is cumulative and results in broad knowledge, which either constitutes an end point by 'offering a more

complete image' or 'sufficing to solve a broad, but modular question or problem', or, serves as a rich basis for further research or for policy making.

This alternative definition of 'MD' forms a more neutral complement for the dominant definitions of 'ID', which can be summarized as follows:

> Interdisciplinarity is a type of CD that involves the integration of strategically selected information, data, methods, tools, concepts or theories from two or more disciplines in order to address a complex question or problem. In an interdisciplinary research project, the integration is (typically) realized by a team of researchers with different disciplinary backgrounds who collaborate intensely to combine relevant elements of their respective disciplines. Interdisciplinarity allows to produce holistic solutions for complex problems. Since interdisciplinarity involves the re-organization of disciplinary elements around problems that, previously, could not be approached from a single discipline, it can also give rise to the emergence of new disciplines.

The scouting of the literature on CD also yielded valuative elements that allow to turn (a summary of) dominant characterizations of MD into the following alternative characterization:

> Multidisciplinarity yields broad knowledge that covers different disciplinary perspectives on a given problem or question. That knowledge might solve a problem or question (if it is a modular one), or may form the basis for another, large scale project. In the latter case, it is important to note that the units of disciplinary knowledge yielded by multidisciplinarity, are highly versatile as they can be integrated according to different interests, needs or values (for example of researchers or policy makers). Multidisciplinarity does not challenge disciplinary boundaries, which means that it is not more risky than mono-disciplinarity, and that it fosters the growth and development of disciplines. Finally, multidisciplinarity allows to gain insight into how disciplines relate to each other in the light of a given subject.

Again, the alternative characterization of MD better balances dominant characterizations of ID, which can be summarized as follows:

> Interdisciplinarity allows to forge break-troughs regarding problems that fall outside the scope of (collections of) individual disciplines. The novel combinations of disciplinary elements it generates, can result in the enriching of existing disciplines, or in the emergence of new disciplines. Interdisciplinarity thus not only makes maximal the use of knowledge and know-how available in existing disciplines, but also contributes to the restructuring of the scientific landscape in function of intriguing or pressing problems. Yet, the innovative character of interdisciplinarity makes a more risky type of CD.

2.7 Conclusions and further research

The literature on CD lacks consensus on how to define CD. Yet, there is a basic conceptual framework that has become widely accepted. The framework consists of three concepts i.e. 'multi-', 'inter-' and 'transdisciplinarity', which stand for different types of CD that are to be distinguished by means of a degree or scope of integration. Since the framework is highly abstract, it can only be used when complemented with a theory of CD. Roughly put, these theories consist of definitions of the three types of CD and characterizations, i.e. more loose remarks on how the different types relate, or compare to each other. While different theories have been developed, one kind of theory has clearly become dominant in the literature, but also among research policy makers, funders and researchers.

The first goal of this paper was to show that there is a problem with these theories, and to provide insight into that problem. In the context of this goal, it was shown that (i) dominant theories are unfoundedly hierarchic and that (ii) their adoption by research policy makers and funders poses a risk for the practice of CD. Regarding (i), it was clarified that the hierarchic structure lies in the representation of MD as the lowest type of CD. An analysis of dominant theories revealed that this representation is an effect of the theories being incomplete: while they provide certain information about ID, they do not provide the same information for MD, causing to the latter to be portrayed as vague and unattractive. It was thus concluded that the hierarchical structure of the theories is unfounded. Regarding (ii), it was argued that because of their (unfounded) hierarchical structure, the adoption of a dominant theory (unjustly) provides support for a preference for ID. It was also pointed out that in the case of research policy makers and funders, the adoption of dominant theories could lead to an unjust preference for ID being reflected in (funding) policies, which would seriously harm the practice of CD. Finally, it was noted that, though a causal link remains to be established, in practice, a preference for ID can indeed be observed. The manifestations and (potential) negative effects of this preference were discussed.

The second goal of the paper was to provide a solution for the problem. This was done by presenting an alternative, well-founded (and therefore less hierarchical) theory of CD. It was created by complementing dominant theories with the information (on MD) they were missing. This information consisted in (i) valuative elements (i.e. elements that indicate the value of MD), (ii) non-mono-disciplinary elements (i.e. elements that allow to differentiate MD from mono-disciplinarity), and (iii) positive elements (i.e. elements that state what MD is, instead of what it is not). The elements were derived from a critical analysis of the (broader) literature on CD.

Yet, with the alternative theory presented in this paper, the work has not been concluded. In order to become truly useful for research policy makers, funders and researchers, it should be extended to also cover transdisciplinarity. Additionally, it would be valuable to illustrate the elements constituting the alternative theory with examples from contemporary CD practice. This would show how the elements can be interpreted, and hence, make them less abstract.

Funding

This research was supported by the Fund for Scientific Research – Flanders (FWO).

Acknowledgements

The author wishes to thank Prof. Dr. Erik Weber and Prof. Dr. Michiru Nagatsu for their invaluable help throughout the research and writing process.

References

Bammer, G. (2012). Strengthening interdisciplinary research: What it is, what it does, how it does it and how it is supported. Australian Council of Learned Academies.

Blevis, E., & Stolterman, E. (2009). Transcending disciplinary boundaries in interaction design: breaking dominant boundaries of interaction. *Interactions*, 14, 48-51.

Boix Mansilla, V. B., Feller, I., & Gardner, H. (2006). Quality assessment in interdisciplinary research and education. *Research Evaluation*, *15*(1), 69-74.

Bruce, A., Lyall, C., Tait, J., & Williams, R. (2004). Interdisciplinary integration in Europe: the case of the Fifth Framework programme. *Futures*, *36*(4), 457-470.

Bruns, H. C. (2013). Working alone together: Coordination in collaboration across domains of expertise. *Academy of Management Journal*, *56*(1), 62-83.

Bruun, H., Hukkinen, J. I., Huutoniemi, K. I., & Klein, J. T. (2005). Promoting interdisciplinary research. Academy of Finland.

Choi, B. C., & Pak, A. W. (2006). Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. *Clinical and investigative medicine*, *29*(6), 351-364.

Chon-Torres, O. (2018). Disciplinary challenges of Today Science. *Biomed J Sci & Tech Res*, *3*(5), 1-2.

Clark, S. G., & Wallace, R. L. (2015). Integration and interdisciplinarity: concepts, frameworks, and education. *Policy Sciences*, *48*(2), 233-255.

Cluck, N. A. (1980). Reflections on the Interdisciplinary Approaches to the Humanities. *Liberal Education*, *66*(1), 67-77.

Darden, L.,& Maull, N. (1977). Interfield theories. *Philosophy of science*, *44*(1), 43-64.

De Boer, Y., de Gier, A., Verschuur, M., & De Wit, B. (2006) Bruggen Bouwen. Onderzoekers over hun ervaringen met interdisciplinair onderzoek in Nederland. RMNO – KNAW – NWO – COS.

Enderby, P. (2002). Teamworking in community rehabilitation. *Journal of clinical nursing*, *11*(3), 409-411.

Eigenbrode, S. D., O'Rourke, M., Wulfhorst, J. D., Althoff, D. M., Goldberg, C. S., Merrill, K., Morse, W., Nielsen-Pincus, M., Stephens, J., Winowiecki, L., & Bosque-Pérez, N. A. (2007). Employing philosophical dialogue in collaborative science. *AIBS Bulletin*, *57*(1), 55-64.

European Union Research Advisory Board (EURAB), 2004. Interdisciplinarity in Research. Retrieved from: ec.europa.eu/research/eurab/pdf/eurab_04_2009_interdisciplinarity_research_final.pdf.

Evans, H. M., & Macnaughton, J. (2004). Should medical humanities be a multidisciplinary or an interdisciplinary study?. *Medical Humanities*, 30, 1-4.

Gleed, A., & Marchant, D. (2016). Interdisciplinarity: Survey report for the Global Research Council 2016 annual meeting. Stockport, UK: DSJ research. Retrieved from: https://docplayer.net/98718536-Interdisciplinarity-survey-report-for-the-global-research-council-2016-annual-meeting-may-2016.html.

Grüne-Yanoff, T. (2016). Interdisciplinary success without integration. *European Journal for Philosophy of Science*, *6*(3), 343-360.

Holbrook, J. B. (2013). What is interdisciplinary communication? Reflections on the very idea of disciplinary integration. *Synthese*, *190*(11), 1865-1879.

Huutoniemi, K., Klein, J. T., Bruun, H., & Hukkinen, J. (2010). Analyzing interdisciplinarity: Typology and indicators. *Research Policy*, *39*(1), 79-88.

Jantsch, E. (1972). Towards interdisciplinarity and transdisciplinarity in Education and Innovation. In Apostel, L., Berger, G., Briggs, A. and Michaud, G. (Eds.), *Interdisciplinarity. Problems of Teaching and Research in Universities* (pp. 97-120), OECD.

Juxtaposition. (2019). In *Merriam-Webster.com*. Retrieved from: https://www.merriam-webster.com/dictionary/juxtaposition.

Klein, J. T. (1990). Interdisciplinarity: History, theory, and practice. Detroit: Wayne State University Press.

Klein, J. T. (2010). A taxonomy of interdisciplinarity. In Frodeman, R., Klein, J.T., & Dos Santos Pacheco, R. C. (Eds.), *The Oxford handbook of interdisciplinarity* (pp. 15-30). Oxford: Oxford University Press.

Klein, J. T. (2013). Communication and collaboration in interdisciplinary research. In O'Rourke, M., Crowley, S., Eigenbrode, S.D. & Wulforst, J.D. (Eds.), *Enhancing Communication & Collaboration in Crossdisciplinary Research* (pp. 11-30). SAGE Publications.

Klein, J. T. (2014). Interdisciplinarity and transdisciplinarity: keyword meanings for collaboration science and translational medicine. *J Transl Med Epi*, 2, 1024-1030.

Lélé, S., & Norgaard, R. B. (2005). Practicing interdisciplinarity. *BioScience*, *55*(11), 967-975.

Kostoff, R. N. (2002). Overcoming specialization. *BioScience*, *52*(10), 937-941.

Land and Water Australia. (2004). Integration Symposium 2004. Canberra : Land & Water Australia.

Lattuca, L. R. (2001). Creating interdisciplinarity: Interdisciplinary research and teaching among college and university faculty. Nashville: Vanderbilt University Press.

Leonelli, S. (2013). Why the current insistence on open access to scientific data? Big data, knowledge production, and the political economy of contemporary biology. *Bulletin of Science, Technology & Society*, *33*(1-2), 6-11.

Lyall, C., & Meagher, L. R. (2012). A masterclass in interdisciplinarity: Research into practice in training the next generation of interdisciplinary researchers. *Futures*, *44*(6), 608-617.

Lyall, C., & Fletcher, I. (2013). Experiments in interdisciplinary capacity-building: The successes and challenges of large-scale interdisciplinary investments. *Science and Public Policy*, *40*(1), 1-7.

Lyall, C., Bruce, A., Tait, J., & Meagher, L. (2015). Interdisciplinary research journeys: Practical strategies for capturing creativity. Bloomsbury Publishing.

Maasen, S. (2000). Inducing interdisciplinarity: Irresistible infliction? The example of a research group at the Center for Interdisciplinary Research (ZiF), Bielefeld, Germany. In

Weingart, P. & Stehr, N. (Eds.), *Practicing interdisciplinarity* (pp. 173-193), Toronto, Buffalo, London: University of Toronto Press.
MacLeod, M., & Nagatsu, M. (2018). What does interdisciplinarity look like in practice: Mapping interdisciplinarity and its limits in the environmental sciences. *Studies in History and Philosophy of Science Part A*, 67, 74-84.
Marques, J. M. M. C. (2008). Inter and multidisciplinarity in engineering education. *International Journal of Mechanics and Materials in Design*, *4*(2), 229-237.
Mestenhauser, J. A. (2002). In search of a comprehensive approach to international education: A systems perspective. Rockin'in Red Square: Critical approaches to international education in the age of cyberculture, 2, 165-213.
Miller, R. C. (1982). Varieties of interdisciplinary approaches in the social sciences: A 1981 overview. Issues in Interdisciplinary Studies: An Occasional Publication of the Association for Integrative Studies (pp. 1-37).
Mobjörk, M. (2010). Consulting versus participatory transdisciplinarity: A refined classification of transdisciplinary research. *Futures*, *42*(8), 866-873.
Nancarrow, S. A., Booth, A., Ariss, S., Smith, T., Enderby, P., & Roots, A. (2013). Ten principles of good interdisciplinary team work. *Human resources for Health*, *11*(1), 11-19.
National Academy of Sciences (NAS), Committee on Facilitating Interdisciplinary Research and Committee on Science Engineering and Public Policy, 2005. Facilitating Interdisciplinary Research. Washington (DC): National Academies Press.
Newell, W. H. (1992). Academic disciplines and undergraduate interdisciplinary education: Lessons from the School of Interdisciplinary Studies at Miami University, Ohio. European Journal of Education, *27*(3), 211-221.
Nissani, M. (1995). Fruits, salads, and smoothies: A working definition of interdisciplinarity. *The Journal of Educational Thought (JET)/Revue de la Pensée Educative*, *29*(2), 121-128.
Okumus, F., & van Niekerk, M. (2015). Multidisciplinarity, tourism. In J. Jafari & H. Xiao (Eds.), *Encyclopedia of Tourism* (pp. 1-2), Springer.
O’Rourke, M., & Crowley, S. J. (2013). Philosophical intervention and cross-disciplinary science: the story of the Toolbox Project. *Synthese*, *190*(11), 1937-1954.
O'Rourke, M., Crowley, S., & Gonnerman, C. (2016). On the nature of cross-disciplinary integration: A philosophical framework. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 56, 62-70.
Paletz, S., Smith-Doerr, L., & Vardi, I. (2010). National science foundation workshop report: Interdisciplinary collaboration in innovative science and engineering fields. Retrieved from:
https://www.researchgate.net/publication/272682934_National_Science_Foundation_Workshop_Report_Interdisciplinary_Collaboration_in_Innovative_Science_and_Engineering_Fields.
Petts, J., Owens, S., & Bulkeley, H. (2008). Crossing boundaries: Interdisciplinarity in the context of urban environments. *Geoforum*, *39*(2), 593-601.
Pischke, E. C., Knowlton, J. L., Phifer, C. C., Lopez, J. G., Propato, T. S., Eastmond, A., de Souza, T.M., Kuhlberg, M., Risso, V. P., Veron, S. R., Garcia, C., Chiappe, M., & Halvorsen, K. (2017). Barriers and Solutions to Conducting Large International, Interdisciplinary Research Projects. *Environmental management*, *60*(6), 1011-1021.
Pohl, C. (2011). What is progress in transdisciplinary research?. *Futures*, 43(6), 618-626.
Repko, A. (2007). Integrating interdisciplinarity: How the theories of common ground and cognitive interdisciplinarity are informing the debate on interdisciplinary integration. *Issues in Interdisciplinary Studies*, 25, 1-31.
Repko, A. F. (2008). Interdisciplinary research: Process and theory. USA: Sage.
Rosenfield, P. L. (1992). The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences. *Social science & medicine*, *35*(11), 1343-1357.

Ross, F. (2009). Degrees of disciplinarity in comparative politics: interdisciplinarity, multidisciplinarity and borrowing. *European Political Science*, *8*(1), 26-36.
Stokols, D., Hall, K. L., Taylor, B. K., & Moser, R. P. (2008). The science of team science: overview of the field and introduction to the supplement. *American journal of preventive medicine*, *35*(2), S77-S89.
Tress, G., Tress, B., & Fry, G. (2005). Clarifying integrative research concepts in landscape ecology. *Landscape Ecology*, *20*(4), 479-493.
Turner, R. K., Van Den Bergh, J. C., Söderqvist, T., Barendregt, A., Van Der Straaten, J., Maltby, E., & Van Ierland, E. C. (2000). Ecological-economic analysis of wetlands: scientific integration for management and policy. *Ecological economics*, *35*(1), 7-23.
Wernli, D., & Darbellay, F. (2016). Interdisciplinarity and the 21st century research intensive university, League of European Research Universities (LERU).
Zaman, G., & Goschin, Z. (2010). Multidisciplinarity, Interdisciplinarity and Transdisciplinarity: Theoretical Approaches and Implications for the Strategy of Post-Crisis Sustainable Development. *Theoretical & Applied Economics*, *17*(12), 5-20.

Epilogue to Part I

In Part I, the received conceptual framework for what is called ‘cross-disciplinarity’ in this dissertation was transformed so that it would better meet the desiderata of (i) unambiguously distinguishing the main types of cross-disciplinarity, (ii) providing an accurate definition for each type, and (iii) being neutral regarding the overall value of each individual type. To do so (a) the framework was extended with a new umbrella term (viz. ‘cross-disciplinarity’), (b) the definition of multidisciplinarity was completed and (c) it was argued that multidisciplinarity is not less valuable than interdisciplinarity.

This work facilitates communication about cross-disciplinarity, but its scope is limited in two important ways.

First, both papers are focused exclusively on research projects. Yet, they also bear relevance to other contexts in which cross-disciplinary research is performed, such as institutes, programmes, ateliers, etc. The actors that are active in those contexts also benefit from using a conceptual framework for cross-disciplinarity that is unambiguous and neutral, for example when interacting with research policy makers. Moreover, they can use the essential criteria for inter- and multidisciplinarity that can be derived from the framework to evaluate the cross-disciplinary nature of their work.

Second, because Papers 1 and 2 focused on strictly academic research projects, transdisciplinarity (which involves the integration of non-academic disciplines) fell outside their scopes. Hence, in retrospect, the aim of Part I is described more correctly as ‘the facilitation of communication about inter- and multidisciplinarity’.

Part II
Facilitating communication within cross-disciplinary teams

Then the bowspirit got mixed with the rudder sometimes:
A thing, as the Bellman remarked,
That frequently happens in tropical climes,
When a vessel is, so to speak, 'snarked'.

Lewis Caroll, *The Hunting of the Snark*

Prologue to Part II

Usually, a cross-disciplinary project is carried out by a team of researchers and each of the source disciplines of the project is represented by one or more team members with a background in that discipline (which does not exclude members having a background in multiple source disciplines). Part II focuses on facilitating *communication within cross-disciplinary teams*, i.e. communication between the members of a given cross-disciplinary team on the content and execution of the project they are working on. Because the team members often have different disciplinary backgrounds, this communication is also called 'cross-disciplinary communication'.

On this level of communication, misunderstandings may arise when a term that plays a central role in a given project, is part of the jargon of multiple source disciplines of that project, but has different meanings across those jargons. Therefore, in this dissertation, ambiguous key terms are called 'problematically ambiguous terms' (PATs). The papers that are included in Part II, provide the theoretical outline of a tool that can be used by the members of a cross-disciplinary team to resolve the PATs of their project. Moreover, first steps are taken in the development of the tool.

Paper 3 provides a theoretical outline of SenseDisclosure, which is a novel PAT resolution procedure. To ensure that the procedure is economical, thorough and precise, it is proposed that some steps of the procedure are automated by means of existing techniques from natural language processing. The most innovative proposal is the use of word sense induction (WSI) techniques for the identification of terms that are included in the jargon of multiple source disciplines of a project, and are polysemous.

Paper 4 presents a pilot study that assesses whether WSI techniques meet an important precondition to become a functional part of SenseDisclosure, i.e. whether they allow for the detection of the senses of a given term across disciplinary jargons. For the study, experiments were carried out by means of SenseClusters (an implementation of a WSI technique) for a cross-disciplinary project that revolves around bone tissue engineering and serves as a case study. The results of the experiments are promising as SenseClusters indeed meets the precondition, but more research is required before SenseDisclosure can be fully developed.

Paper 5 makes a more indirect contribution to the development of SenseDisclosure. It contains a critical philosophical analysis of the (methodological) assumptions that underlie WSI techniques. This analysis informs on the limitations of WSI techniques, thereby shedding light on what further research regarding the use of WSI techniques in the context of SenseDisclosure would be useful (and what research would not).

Paper 3 SenseDisclosure: a new procedure for dealing with problematically ambiguous terms in cross-disciplinary communication[1]

[1] This is (the manuscript version of) a paper that has been published in *Language Sciences*.

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 characterization 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 characterization 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.

Key Words

cross-disciplinarity • interdisciplinarity • interdisciplinary communication • natural language processing • terminological ambiguity • word sense induction

3.1 Introduction

Contemporary research projects are often characterized 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, 'multidisciplinarity' 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% to 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, 1997)), 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.

[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’ (2015, p. 534). Hodges (2008) makes a similar argument with respect to research in ecology.

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 characterization of a new PAT resolution procedure called ‘SenseDisclosure’ is presented. Its characterization incorporates multiple techniques from Natural Language Processing (henceforth ‘NLP’) to realize the necessary automations.[2] The new procedure is meant to be applicable to all kinds of cross-disciplinary projects (by an external facilitator).

The remainder of this paper is structured as follows. Section 2 contains an overview of existing CD communication procedures that also address PATs (ranging from communication guidelines to visual representation techniques). In Section 3, a new set of task and performance criteria for a (good) PAT resolution procedure is presented. In Section 4, these criteria are used to evaluate the capacity of the existing CD communication procedures for PAT resolution. Section 5 contains the theoretical characterization of the new PAT resolution procedure ‘SenseDisclosure’. The characterization incorporates some technical notions from the field of NLP, which are further explained in Section 6. In Section 7, the task and performance criteria are used for the evaluation of SenseDisclosure. Section 8 forms a small excursion into possible alternative applications of SenseDisclosure in several theoretical linguistic research fields. Section 9 contains conclusions.

3.2 Existing CD communication procedures that address PATs

Given the need of CD teams for a PAT resolution procedure, it seems probable that such procedures already exist. However, there is no systematic overview of such procedures.

[2] The theoretical characterisation of SenseDisclosure does not only serve a linguistic purpose, but also a philosophical one. More specifically, it corresponds to three objectives present in Mäki’s outline of a *philosophical study of interdisciplinarity* (2016): (i) it should yield information on ‘what happens (or fails to happen) when two or more disciplines are brought in contact with one another’, (ii) the philosophical study of interdisciplinarity should generate insights that are useful for the scientific community, and (iii) its methodology should extend the traditional philosophical methodology (Mäki, 2016, p. 336). The paper does so (i) by identifying a set of recurring communication problems between collaborating researchers with different backgrounds, and conceptualizing its underlying cause as ‘PATs’, (ii) by providing new task and performance criteria for the assessment of the capacity of a procedure to resolve PATs and by presenting SenseDisclosure, and (iii) by using techniques from the field of Natural Language Processing throughout the characterization of SenseDisclosure.

Therefore, a focused, small-scale literature search was performed. Academic research papers were selected via Google Scholar using search queries consisting of two components: one referring to the crossing of disciplinary boundaries (e.g. 'interdisciplinarity', 'cross-disciplinarity' and 'multidisciplinarity') and one referring to terminological ambiguity (e.g. 'different meaning' and 'ambiguous term'). Search results that did not mention the negative effect(s) of terminological ambiguity on CD communication were omitted. Based on the bibliographies of the resulting selection, additional papers were selected. The search generated more than twenty papers. These papers were then screened for CD communication procedures that, directly or indirectly, address PATs. This yielded two CD communication procedures that directly address PATs (henceforth 'direct procedures'), and six CD communication procedures that address PATs indirectly (henceforth 'indirect procedures'). The procedures are discussed below.

A first direct procedure is a technique called 'articulation' (Bracken and Oughton, 2006). When a CD team 'articulates' a PAT, team members with different backgrounds share their definition of the term, and explain why they use this definition in their particular discipline. The technique is nicely illustrated by the example of the articulation of the term 'catchment' by collaborating physical and social scientists discussed in the introduction.

A second direct procedure is a technique called 'registration' (Monteiro and Keating, 2009). The term is borrowed from imaging research, where it refers to the process of systematically linking up images of (a certain part of) the human body that are made at different times or from different angles. Registration in the context of a CD project is not related to images, but focusses on knowledge, i.e. visual and written information produced by the CD team in question. More specifically, weekly meetings are scheduled during which the relations between the different skills, models, concepts, etc. present in the CD team are clarified. According to Monteiro and Keating, such knowledge registration enables 'the identification of erroneous interpretations, illumination of unshared premises (that can then be mitigated) and development of shared understandings of what the common goals of the [CD] project are' (Monteiro and Keating, 2009, p. 20). This also encompasses the detection of PATs.

There are two types of indirect procedures: *relational indirect procedures*, which focus on the relational aspects of CD communication, i.e. the interaction between CD team members, and *content-oriented indirect procedures*, which deal with the content of CD communication, i.e. the knowledge that is exchanged between CD team members. Relational indirect procedures tend to revolve around general, high-level characteristics of communication. For example, when describing processes that are essential for fruitful CD communication, Thompson (2009) discusses so-called 'enabling processes', such as 'spending time together' and 'meaning negotiation', and 'expediting processes', such as 'shared humor and laughter' and 'reflexive talk' (Thompson, 2009, p. 284). Thompson explicitly states that (some of) these processes are also crucial for dealing with PATs:

> [...] people from different scholarly backgrounds assign qualitatively different meanings to the same term. Transcending such differences requires recognition, reflection, and negotiating of meaning, especially in interdisciplinary contexts. (Thompson, 2009, p. 286)

A similar approach can be found in Naiman (1999). Naiman argues that regular, informal communication is crucial for CD teams and stresses the importance of 'working towards consensus [and] making sure that everyone knows the guidelines for building a consensus' (1999, p. 294).

Content-oriented indirect procedures consist of two subtypes: those that deal with the exchange of 'tacit knowledge' and those that focus on the exchange of 'explicit knowledge'. A good example of the first subtype is the *Toolbox Project* developed by O'Rourke and Crowley (2013). The project revolves around so-called 'worldviews', i.e. the sets of metaphysical and epistemological assumptions that underlie disciplines. A worldview often consists of tacit knowledge because it is adopted unknowingly by the researchers working in a discipline. Yet, as a worldview greatly affects the research activities in the discipline, communication across disciplines with different (or even conflicting) worldviews is difficult. O'Rourke and Crowley compiled a 'toolbox' that consists of lists of carefully selected questions on metaphysical and epistemological topics. In a workshop organized for a given CD team, the toolbox is used to encourage the team members to reflect on their worldview, to make (parts of) their worldview explicit, and to discuss them with each other. By increasing mutual understanding, the CD communication within the team improves. An excerpt of the report of a workshop organized for a CD team of physical scientists shows how the application of the toolbox can also lead to the detection of PATs:

> [Person] 8: ... I mean, every time I hear the word 'diversity' I have to look to see who said it. Because if X said it, X means something completely different than if Y said it ... these terms are funny in terms of who's saying it and what the purposes are if they mean somewhat the same things, but not exactly the same things.
> [Person] 7: [overlap] 'Replication'.
> [Person] 8: 'Replication' is another one.
> [Person] 7: 'Representation'.
> [Person] 6: 'Model'.
> [Person] 1: 'Artificial', 'natural'. (O'Rourke & Crowley 2013: p. 1941)

The second subtype, focused on the exchange of explicit knowledge, mostly revolves around *conceptual models*. Conceptual models are 'typically drawn as diagrams with boxes and arrows that show the main elements and flows of material, information and causation that define a system' (Heemskerk et al., 2003, p. 2). For example, Heemskerk et al. recommend CD teams to create conceptual models of the system they are developing or the problem they are studying in order to improve the interaction between team members. They explicitly state that conceptual models allow to address communication problems that arise ‘from the use of jargon and terms that mean different things to different people’ (Heemskerk et al., 2003, p. 5), or, in other words, the problems caused by PATs. A similar argument can be found in (Pennington, 2008).

3.3 A new set of task and performance criteria

Most of the CD communication procedures discussed in Section 2 are not specifically developed to resolve PATs. This alone makes it necessary to evaluate their capacity for PAT resolution. The current literature does not provide such a systematic evaluation. Therefore, in this paper, the CD communication procedures are evaluated by means of a new set of *task* and *performance criteria*. Task criteria indicate what tasks a good PAT resolution procedure should be able to perform, and performance criteria indicate how those tasks should be performed. The criteria are introduced by means of the examples already discussed in the introduction.

Sometimes, the team of a CD project is unaware of the PATs hindering their CD communication. A good example of this is the case of the medical expert systems. Initially, the CD team was unaware of the fact that its knowledge engineers and medical specialists used different definitions of the terms ‘diagnosis’ and ‘outcome’. The (unnoticed) terminological ambiguity caused a misunderstanding that resulted in the waste of valuable time and energy. Hence, given a CD project, a good PAT resolution procedure satisfies the following task criterion:

(T1) the procedure should facilitate the detection of PATs in the CD communication.

In the case of the medical expert system project, a good PAT resolution procedure would (help the CD team to) flag ‘diagnosis’ and ‘outcome’ as PATs of the project.

However, this is not the only task. In some cases, CD teams are already aware of the problematic ambiguity of one or more terms in the context of their project. A good example of this is the case of the experts in chemical and electrical engineering who warn their fellows for the ambiguity of terms such as ‘phase’ and ‘transducer’ in projects that involve both disciplines (Ranade et al., 2011). However, such awareness alone clearly does not suffice to resolve PATs. Knowing that ‘phase’ might mean something else, but not knowing what exactly, does not help CD team members to communicate more efficiently. It is also necessary to have an overview of the concepts that might be invoked as the meaning of a PAT. Hence, given a CD project, a good PAT resolution procedure also satisfies the following task criterion:

(T2) the procedure should facilitate the identification of the concepts behind a PAT.

In the case of the collaboration between chemical and electrical engineers, a good PAT resolution procedure would help (the CD team) to discover the different concepts underlying 'phase' by providing pointers to the chemistry-related concept of 'aggregation state' (e.g. solid, liquid and gas) and the physics-related concept of 'relative displacement between two corresponding features of two waveforms having the same frequency'.

Last but not least, efficient CD communication also presupposes a new, shared interpretation standard with respect to the PATs. Without such a standard, in practice, CD teams still work 'in the dark' when it comes to the meaning of these terms: they might have detected PATs and identified their concepts, yet, they still would not know which concept is the correct one for each occurrence of a PAT. Think, for example, of the case of the physical and social scientists working on a project on sustainable land use. After they had identified the two concepts that were associated with 'catchment', they needed to agree on how they would use the term in the future. Hence, given a CD project, a good PAT resolution procedure also satisfies the following task criterion:

(T3) the procedure should facilitate the creation of a new, shared interpretation standard for a PAT.

In the case of the sustainable land use project, a good PAT resolution procedure would help the CD team to determine what 'catchment' concept best serves the needs of the project, and thus, should constitute the default interpretation of the term. In general, the default interpretation could be one of the concepts that was underlying a PAT, a combination of several of them, or a newly defined one. Alternatively, the PAT itself could be dropped, and new, unique terms could be generated for each concept that was underlying the PAT. The creation of an interpretation for a PAT may also include the selection of (unique) terms to refer to the concepts that are associated with the PAT, but are not included in its new, default interpretation. Note that, in principle, a new interpretation standard could be determined by the leader(s) of a CD team and imposed on the other team members. However, as an interpretation standard should meet the needs of the CD team as a whole, a new interpretation standard is best created by means of a process in which all CD team members participate.

Next to the task criteria above, three performance criteria are crucial.

The first performance criterion focusses on the costs of a good PAT resolution procedure, i.e. the amount of time and manpower the procedure requires. Members of CD teams already experience heavy workloads: on top of the work that comes with mono-disciplinary research, they also need to synchronize with other source disciplines (that have divergent interests and agendas). Hence, the cost of a good PAT resolution procedure should be as low as possible. In other words, given a CD project, a good PAT resolution procedure satisfies the following performance criterion:

(P1) the procedure should be economical (with respect to its time and manpower requirements).

The second performance criterion specifies the thoroughness of a PAT resolution procedure. More specifically, with respect to criterion T1, the procedure should not ‘simply’ lead to the detection of PATs, but to the detection of *all* PATs. Likewise, with respect to criterion T2, the procedure should identify *all* concepts of each PAT, and with respect to criterion T3, a new interpretation standard should be created for *all* detected PATs. In sum, given a CD project, a good PAT resolution procedure satisfies the following criterion:

(P2) the procedure should warrant a complete solution.

The third and final performance criterion specifies the accurate nature of a PAT resolution procedure. More specifically, with respect to criterion T1, the procedure should *only* detect those PATs that have a true potential to cause communication problems in the context of the given CD project. Similarly, with respect to criterion T2, the procedure should *only* identify those concepts of a PAT that play a role in the given CD project. Concerning criterion T3, it demands that the new, shared interpretation standards for PATs are project-specific, i.e. the standards *only* take into account the semantic needs that are at play in the given CD project. Simply adopting interpretation standards from similar CD projects should be avoided because there is no guarantee that these standards would fully apply. In sum, given a CD project, a good PAT resolution procedure satisfies the following criterion:

(P3) the procedure should warrant a precise solution.

Note that the criteria are defined as normative ideals (e.g. a phrase like ‘should warrant a complete solution’). In line with this approach, when the criteria are used to evaluate candidate procedures for PAT resolution, the procedures are analyzed with respect to their ‘ideal potential performance’, i.e. how they would perform if the input requested from the CD team or the external facilitator would be perfect.

3.4 Evaluation of the existing CD communication procedures that address PATs

In what follows, task criteria T1-T3 and performance criteria P1-P3 are used to evaluate the efficiency of the existing CD communication procedures with respect to PAT resolution.

Starting with the task criteria T1-T3, criterion T1 is only met by one direct tool, i.e. registration. When the members of a CD team systematically try to link the skills, models, concepts, etc. of the different disciplines involved, all sorts of disagreements, among which erroneous interpretations caused by terminological ambiguity, are discovered. Criteria T2 and T3 are met by both direct procedures, i.e. registration and articulation. In line with criterion T2, both techniques require the members of a CD team to explicate the role(s) of the different concepts of a PAT. As a direct result, in accordance with criterion T3, they also provide the team with the necessary information for creating a new interpretation standard for a PAT. The indirect procedures meet all three task criteria, but only to a certain extent. As the procedures enhance the relational or content dimension of CD communication, they increase the chance for a CD team to detect ambiguous terms (criterion T1) and to identify the different concepts underlying such terms (criterion T2). Hence, they might also facilitate the actual search for a new interpretation standard for the terms in the context of the project (criterion T3). However, instead of providing genuine *assistance* for PAT resolution, indirect procedures only create favorable *conditions* for addressing PATs, which can be considered a serious drawback.

Moving on to the performance criteria P1-P3, none of the existing CD communication procedures truly meets criterion P1. The procedures require CD team members to engage in group discussions, to draw diagrams, to reflect on their worldview, etc. Registration even requires weekly meetings where CD team members 'produce and present to others visual renderings of their part of the project, accomplishments, problems, and unmet challenges' (Monteiro and Keating, 2009, p. 2). In other words, the existing CD communication procedures require great efforts of a CD team, and hence, are not economical. The procedures also clearly do not meet criterion P2 (in multiple respects). More specifically, with respect to the detection of PATs (criterion T1), none of them provide CD teams with a standardized protocol that warrants the exhaustive detection of the PATs of their project. Indirect procedures score particularly badly. Whether or not a PAT is detected depends exclusively on the (contingent) flow of conversations between the CD team members. Consequently, because of the lack of systematicity in the detection of PATs, CD communication procedures also fail to ensure that new interpretation standards are created for all detected PATs (criterion T3). The procedures also cannot guarantee the identification of all concepts of detected PATs (criterion T2) as they rely exclusively on the CD team members for information on the concepts of terms. Yet, the ability of CD team members to (successfully) use certain terms does not guarantee that they are able identify all the concepts they associate with those terms. In contrast to criteria P1 and P2, all the existing procedures do meet criterion P3. Because the procedures are exclusively based on input provided by the team members of a given CD project and the goal is to improve the communication within that project, it is highly likely that the team members will not generate out of context information such as irrelevant PATs, concepts, etc. Hence, the procedures seem to be precise.

The evaluation shows that there is no existing CD communication procedure that sufficiently satisfies criteria T1-T3 and P1-P3. In other words, currently, despite the need for a good PAT resolution procedure, no such procedure is readily available. Also remark that the bad evaluation of the procedures with respect to criteria P1 and P2 is related to the following fact: when used for PAT resolution, the existing procedures require the CD team members to generate vast quantities of complex, interwoven linguistic data. They need to browse their jargon to find *all* PATs as well as identify *all* concepts related to those PATs. This is extremely costly (in the sense of criterion P1) and it is very unlikely that it leads to a complete solution (in the sense of criterion P2). Hence, the development of a new, truly effective PAT resolution procedure demands some level of automation. Given the linguistic nature of the data that needs to be processed, automating by means of techniques from the field of applied (computational) linguistics seems the best way to go.

3.5 SenseDisclosure: a new PAT resolution procedure

Keeping in mind the need for automation mentioned above as well as the new set of task and performance criteria, a theoretical characterization of a new PAT resolution procedure called ‘SenseDisclosure’ is presented. SenseDisclosure is meant to be applicable to all kinds of cross-disciplinary projects. Its application can be controlled by an *external facilitator*, i.e. a person who is an expert with respect to the application, general methodology and technical aspects of SenseDisclosure and uses this expertise to guide the CD team during their interactions with SenseDisclosure. Strictly speaking, the external facilitator does not belong to the CD team and is not necessarily familiar with (one of) its source disciplines.[3] SenseDisclosure is conceived as a procedure that consists of seven steps, some of which incorporate techniques from Natural Language Processing (NLP) to realize critical automations.[4] Note that, as these NLP techniques were not specifically developed for PAT resolution, some of them still require further research and development before they can be reliably integrated. Hence, at the moment SenseDisclosure remains a theoretical construct, its successful realization depending on whether the necessary research and development yields positive results. However, the already hi-end performance of current NLP techniques in combination with the rapid developments in the field form a strong indicator that such positive results are within reach. In what follows, the seven steps of the SenseDisclosure procedure are discussed. For the sake of clarity and readability, the discussion is kept brief and conceptual. Technical notions related to mentioned NLP techniques are explained in the next section. Also the way in which some of these techniques are still in need of further research and development is discussed in Section 6.

Step one: generate corpora

The first step is to generate a representative *corpus*, i.e. a preprocessed collection of characteristic texts, for the expert language of each source discipline of the project. A good way to do this is to first let the external facilitator ask the CD team members to identify the different source disciplines that are represented in the team and to provide a list of characteristic texts for each of these disciplines. Next, the collections of texts are compiled and *preprocessed* to prepare them for automatic processing. In it most basic form, preprocessing comes down to cutting text up into meaningful elements, and tagging those elements with grammatical or lexical information. Preprocessing can be carried out automatically by means of (modules of) existing NLP software packages such as the Natural Language Toolkit (NLTK) (Bird et al., 2009), Stanford CoreNLP (Manning et al., 2014), or the LeTs Preprocess Toolkit (van de Kauter et al., 2013).

Step two: create jargon sets

The second step is to create a *jargon set* for each source discipline, i.e. a set of terms that are highly characteristic for the discipline in question. This is done best by applying NLP techniques for *term extraction*, i.e. the identification of the terms that occur in a text or corpus, to the corpora that represent the source disciplines. There are several (modules

[3] The notion of external facilitator can also be found in related literature (Müller, 1998; Sinnet and Williams, 2011).

[4] Note that SenseDisclosure relies on techniques that only process *written* natural language. Extra-linguistic content in the corpora (e.g. mathematical expressions, tables, graphs and images) is not taken into account.

of) software packages that allow for automated term extraction, e.g. the Natural Language Toolkit (NLTK) (Bird et al., 2009), AlchemyAPI (n.d.) and TExSIS (Macken et al., 2013).

Step three: create the shared jargon set

The third step is to create the *shared jargon set*, i.e. the set of terms that are members of at least two different jargon sets. The members of a shared jargon set are thus always characteristic for *at least two different* source disciplines. Note that it might be argued that the property of being shared by 'at least two different source disciplines' is too broad and that it would be more efficient to take only those characteristic terms that are shared by *all* source disciplines. On first sight this might seem convincing, as it meets criterion P1. However, when also taking into account criterion P2, it becomes clear that such a definition is actually too restrictive. There might very well be terms that are both shared by some, but not all source disciplines and responsible for serious communication problems that hinder the CD team as a whole.

Step four: create the set of polysemous shared jargon terms

Step four is to create the set of polysemous shared jargon terms, i.e. the set of shared jargon terms that have at least two concepts, or senses. NLP techniques related to the problem of *word sense induction* show great potential for executing this step. 'Word sense induction' can be defined as 'automatically identifying the senses of words in texts without the need for handcrafted recourses or manually annotated data' (Van de Cruys and Apidianaki, 2011, p. 1476). Note that word sense induction techniques do not generate clear-cut characterizations for each concept of a (target) term, but rather a collection of 'concept indicative words', i.e. words that are strongly correlated with a concept of the (target) term. An interesting existing word sense induction software package is 'SenseClusters' (Purandare and Pedersen, 2004).

Step five: create the set of PATs

Step five is to create the set of PATs. The most effective way to do this is to consult the CD team members, present them with the set of polysemous shared jargon terms generated in the previous step, and ask them to identify the PATs.

Step six: generate word clouds

Step six is to generate word clouds for each concept of each PAT. Word clouds are visual representations of textual information, consisting of words that are organized around a central point. The font size of the displayed words is an indicator for their importance. Here, the word clouds are generated based on the concepts (viz. the collections of concept indicative words) that were identified by means of word sense induction techniques in step four. This can easily be done automatically, for example by means of the word cloud package in R, a platform for statistical computing and graphics (The R-project for Statistical Computing, n.d.).

Step seven: negotiate new interpretation standards

Step seven is to invite the CD team members to negotiate a new interpretation standard for each PAT, i.e. to transform each PAT into a term with just one (new) suitable meaning definition. The negotiation process is structured as follows. First, the team members familiarize themselves with the different concepts of each PAT by means of its word clouds. Next, for each PAT, they collectively provide exactly one meaning definition that is suitable for future CD communication. This can be done in three different ways: (i) they agree to use one of the original concepts, (ii) they create a new concept by strategically combining information from one or more of the original concepts, or (iii) they create a new concept by combining information from one or more of the original concepts *and* new information that they consider to be critical in the context of their CD project. Note that it is imperative for the CD team to also assign new, unambiguous terms to any type of conceptual information related to the PAT that is considered vital to future CD communication, but was left out of the new meaning definition.

3.6 SenseDisclosure: some technical details

In what follows, the notions related to NLP techniques mentioned in Section 5 are discussed in more detail. Also the way in which these techniques still require further research and development with respect to their application in SenseDisclosure is outlined. Note, however, that a complete technical characterization of SenseDisclosure falls outside the scope of this paper.

In step one, a representative corpus is compiled for each source discipline. The text material for each corpus is selected by those CD team members that have a background in the related source discipline. The initial selection might need to be adjusted by the external facilitator to ensure that the corpora provide a good starting point for SenseDisclosure. To be useful, a corpus has to be (i) representative for the expert language of the source discipline and (ii) contain explicit information on the concepts of the jargon terms of the source discipline. The first requirement ensures that the relevant set of terms is screened for PATs. It is met by including academic research papers in the corpora. The second requirement warrants that the concepts of PATs can be identified by means of NLP techniques. It is met by including pedagogical text material, such as intermediate textbooks and course material, which often contain explicit characterizations of central jargon terms. Once the final selection of texts for a corpus has been made, the texts have to be collected and preprocessed. The preprocessing involves breaking text up into sentences (sentence boundary detection), breaking sentences up into meaningful elements such as words, punctuation marks and numerals (tokenization), assigning the appropriate grammatical category to each token (part-of-speech tagging), bringing words back to their uninflected form (lemmatization), and detecting person names, organizations, locations, etc. (named entity recognition). As indicated before, all of this can be done automatically by means of dedicated software. Note, however, that the current software is not specifically developed to process highly specialized text such as scientific textbook material. Hence, further testing is required to verify whether it is already up for the task.

In step two, a jargon set is created for each source discipline by applying term extraction techniques to the corpora. As mentioned previously, this can also be done by means of software. The process starts with the listing of *candidate terms,* i.e. instantiations of term formation patterns (e.g. 'noun', 'adjective noun', 'adverb adjective noun') that occur in the corpus. Term formation patterns are language-specific. Hence, as the textbook material is in English, a set of patterns for the English language has to be used. Note, however, that, as the goal is to detect jargon terms, the set of formation patterns might need to be adjusted to make sure that it covers *scientific* English term formation patterns. Next, for each candidate term, its *termhood* is calculated. 'Termhood' is a measure for the domain specificity of a candidate term. More specifically, the termhood of a word indicates the extent to which the word is characteristic of a given corpus. Roughly speaking, it is determined by comparing the relative occurrence frequency of the term in the corpus to its relative occurrence frequency in a *general corpus*, usually created from newspaper articles (Kageura and Umino 1996). Imagine, for example, a corpus for experimental biology and the candidate terms 'model organism' and 'growth'. It is very likely that 'model organism' will have a high termhood as the relative occurrence frequency of this rather technical term will be (much) higher in the experimental biology corpus than in the general corpus. In contrast, 'growth' most likely will have a relatively low termhood as the occurrence frequency of this more generally used term will not be higher in the corpus than in the general corpus. After all, in the general corpus, on top of medical contexts (e.g. "new drug stops cancer cell growth"), the term will, for example, also be used in financial contexts (e.g. "the GDP growth rate has stagnated). Common functions for determining the termhood of a candidate term are 'C-value' and 'term frequency - inverse document frequency' (tf-idf) (Nakagawa and Mori, 2002). For candidate terms consisting of multiple words, a second value called 'unithood' is calculated. The *unithood* of a candidate term is a measure for its level of internal cohesiveness. For example, consider again the hypothetical corpus for experimental biology for which holds that the candidate term 'model organism' has a higher unithood than candidate terms like ‘main model’ and ‘living organism’. This would mean that the relative occurrence frequency of ‘model organism’ is higher than that of ‘model’ and ‘organism’ in combination with other words (such as ‘main’ and ‘living’ respectively). Common functions for determining the unithood of candidate terms are pointwise mutual information (PMI) and ‘log-likelihood’ (Korkontzelos et al., 2008). Candidate terms with a termhood (and, if applicable, unithood) above a predefined threshold become the members of the jargon set (of the given corpus).[5]

In step four, the set of polysemous shared jargon terms is created. As mentioned in Section 5 (step four), NLP techniques related to the problem of word sense induction are promising candidates for automating this step as they revolve around the automatic identification of the senses of words. There are several different types of word sense induction techniques: some are graph based (e.g. Biemann, 2006; Véronis, 2004), some use clustering (e.g. Pantel & Lin 2002, Schütze 1998; Purandare and Pedersen, 2004) and one even relies on a Bayesian framework (Brody & Lapata, 2009). However, all the techniques start from the so-called *distributional hypothesis*, which states that words occurring in similar contexts tend to be semantically similar (e.g. Harris, 1954). In line with this hypothesis, the techniques gather information on the sense(s) of words by looking at the contexts in which they are used. Note, however, that the techniques were originally developed in the context of information retrieval and machine translation and therefore primarily aimed at *differentiating* between the concepts of a given (target) term and not so much at generating detailed information about these concept(s). For example, when a query (in the case of information retrieval) or a word that needs to be translated (in the case of machine translation) is ambiguous, the techniques only try to determine what concept(s) might be addressed, and which ones are excluded. Hence, as the resolution of PATs does require detailed information about concepts, existing word sense induction techniques need to be further studied, developed and tested before one can be selected for integration in SenseDisclosure. Right now, the most promising technique seems to be ‘Sense Clusters’ (Purandare and Pedersen, 2004) as it is a very user-friendly software package and came out as a highly effective technique in a thorough evaluation by Agirre and Soroa (2007).

[5] It is not a problem to use a predefined threshold, as there are several strategies for determining a-priori threshold values. Unfortunately, a discussion of these strategies falls outside the scope of this paper.

3.7 Evaluation of SenseDisclosure

Given the theoretical characterization of SenseDisclosure in the previous section, it is possible to evaluate this new PAT resolution procedure with respect to task criteria T1-T3 and performance criteria P1-P3. However, note, that the evaluation remains conditional on the positive outcome of the necessary further research and development of specific NLP techniques (as explained in Section 5).

SenseDisclosure meets all three task criteria. Criterion T1, the facilitation of the detection of PATs in the CD communication, is covered by steps one to five. Criterion T2, the facilitation of the identification of the concepts of a PAT, is warranted by step six. Criterion T3, the facilitation of the creation of a new interpretation standard for a PAT, is covered by step seven.

SenseDisclosure also fully meets performance criteria P1 and P2, and P3 to a large degree. The procedure covers criterion P1, stating the need for a good PAT resolution procedure to be economical, because of several reasons. First of all, four of its steps can be automated by means of software, i.e. the generation of jargon sets (step two), the creation of the shared jargon set (step three), the creation of the set of shared polysemous jargon terms (step four), and the generation of word clouds (step six). Because these steps can be automated, they require almost no human effort. Moreover, their execution, as well as that of the non-automatable process of generating corpora (step one), can be controlled by an external facilitator. Furthermore, the two remaining steps that cannot be automated or outsourced are greatly facilitated. For the creation of the set of PATs (step five), the CD team members are provided with a set of polysemous shared jargon terms. This allows them to browse through the provided set of terms and indicate which of the terms are problematic, instead of having to reflect on their terminology by themselves, without any guidance (which can lead to a laborious identification and deliberation process with neither clear quality standards nor stop conditions). The creation of a new interpretation standard for each PAT (step seven) is supported by providing the CD team with word clouds (created in step six). The word clouds function as pointers to the different concepts of each PAT. In sum, the general effort required from the CD team members is minimal. SenseDisclosure also covers criterion P2, stressing the need for a complete solution, as it is essentially a well-characterized set of actions forming a partially automated procedure that warrants the detection of *all* PATs (criterion T1), the identification of *all* concepts of each PAT (criterion T2), and the creation of a new interpretation standard for *all* detected PATs (criterion T3). Finally, SenseDisclosure also largely covers criterion P3, stating the need for a PAT resolution procedure to be precise. Firstly, given a CD project, the methodology of the procedure rules out the detection of irrelevant PATs (criterion T1). The actual detection is carried out by the CD team members using a set of candidate PATs, i.e. the set of shared polysemous terms. Only they can identify those PATs that are able to cause communication problems in the context of the project. Secondly, the procedure also minimizes the chance that, for a detected PAT, concepts are identified (criterion T2) that are irrelevant to the CD project. Given a detected PAT, its concepts are identified automatically based on corpora that represent the jargons of the source disciplines of the given project. Hence, concepts that are completely irrelevant are excluded (e.g. the exclusion of the geography-related concept of the term ‘bank’, as in ‘a river bank’, in the context of a CD project with sociology and economy as its source disciplines). However, a more ‘fine-grained’ type of irrelevance cannot be fully excluded. Given a project, the procedure might identify concepts that are present in the corpora of its source disciplines, but not used in the context of the project. A project does not necessarily demand the usage of *all* the jargon of its source disciplines (e.g. a CD project that involves psychology and economy does not necessarily use both the psychology-related concept and the economy-related concept of the term ‘rational’). Finally, because the procedure relies on CD team members to create new interpretation standards (criterion T3), it is certain that during this creative process only semantic considerations that are relevant to their project are taken into account.

In sum, the evaluation of SenseDisclosure is very good. Even though the procedure might perform suboptimal when it comes the identification of relevant concepts of PATs (performance criterion P3 with respect to task criterion T2), its evaluation in total still surpasses the evaluation of the existing procedures (discussed in Section 4).

3.8 Alternative applications of SenseDisclosure

As SenseDisclosure incorporates multi-applicable NLP techniques (e.g. term extraction and word sense induction), the procedure also has potential for theoretical linguistic research. This section forms a small excursion into SenseDisclosure’s application possibilities in several theoretical linguistic research fields.

Once fully developed, the most straightforward alternative application of SenseDisclosure would be in the field of semasiology, i.e. the subdiscipline of lexicology that (roughly put) ‘studies the meaning(s) of words’ (Ifversen, 2011). Two examples of semasiological research are (i) an inquiry into the meaning of ‘riht’, i.e. a precursor word to the German word ‘Recht’ and the English word ‘right’ by Fruscione (2013) and (ii) a study of the different interpretations of the word ‘method’ in scientific research by Fishuk (2016). Phrased in the language of this paper, the main challenge of (i) is to collect concept indicative words in order to identify the concept of the term ‘riht’, whereas (ii) aims at the collection of concept indicative words in order to identify the different concepts of the polysemous term ‘method’. SenseDisclosure would be well equipped for both types of research: its core functionality is to disclose the concepts of a term by collecting concept indicative words, and the goal of this functionality is to determine whether a term is polysemous. Note that in this context some steps of the procedure are not used (e.g. the creation of a (shared) jargon set, the creation of the set of PATs and the creation of new interpretation standards). Those steps that are used, however, do not require major tweaking.

A second alternative application of SenseDisclosure would be in historical linguistics, more specifically in the study of semantic change. An often-cited example of semantic change is the shift in the meaning of the word ‘gay’. Initially, the word stood for ‘cheerful’ or ‘dapper’, but in the 1970’s, its meaning started to change. Nowadays, it is most commonly used to describe someone’s sexual orientation (Kulkarni et al., 2015). Again phrased in the language of this paper, the main challenge of studying the semantic change of the term ‘gay’ is detecting significant alterations in the collection(s) of concept indicative words, thereby gaining insight into how the concept(s) of ‘gay’ have changed over time. SenseDisclosure can be used for this task when the corpora used as input, are diachronic.

Finally, SenseDisclosure might also be used in morphological research. Morphology is the study of the structure of words (Bybee, 1985). It analyses, among other things, how the inflection of words, i.e. adding specific pre- or suffixes, changes their meaning. The different inflections of a word can be considered as different terms and SenseDisclosure can be used to generate word clouds for each of those terms. By comparing these word clouds, it should become possible to clarify in greater detail the way in which a specific inflection of a word alters its semantics.

3.9 Conclusions

The communication in cross-disciplinary (‘CD’) projects is often severely hindered by problematically ambiguous terms (‘PATs’) and CD teams are generally not trained to deal with them in an efficient and methodical way. Hence, CD teams are in need of an effective PAT resolution procedure. A small-scale literature search shows that there are several CD communication procedures that address PATs (directly or indirectly). However, the procedures were not designed primarily for resolving PATs. Hence, their capacity to do so is evaluated using a new set of three task criteria (T1-T3) and three performance criteria (P1-P3) for a good PAT resolution procedure. The evaluation shows that none of them sufficiently meets all six criteria, the main problem being that they rely too much on the input of CD team members. 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 characterization 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 characterization incorporates multiple techniques from Natural Language Processing (‘NLP’) to integrate 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 research and development yields positive results, which seems likely, SenseDisclosure can be a truly effective PAT resolution procedure. In other words, the procedure could really help CD teams to identify and resolve PATs quickly and efficiently, thereby effectively streamlining their project-related communication. More broadly speaking, such facilitation of CD communication enables scientific progress. In this respect, this paper can also be considered as a humble attempt to inspire the NLP community to perform the research and development that is still required for the successful realization of SenseDisclosure.

Funding

This research was supported by the Research Foundation – Flanders (FWO).

Acknowledgements

The author wishes to thank Prof. Dr. Erik Weber, Prof. Dr. Els Lefever and Prof. Dr. Stephan van der Waart van Gulik for their invaluable help throughout the research and writing processes.

References

Agirre, E., & Soroa, A. (2007). Semeval-2007 task 02: Evaluating word sense induction and discrimination systems. In *Proceedings of the 4th International Workshop on Semantic Evaluations I* (pp. 7-12). Association for Computational Linguistics.

Bieman, C. (2006). Chinese Whispers - An Efficient Graph Clustering Algorithm and Its Application to Natural Language Processing Problems. In *Proceedings of the TextGraphs: the First Workshop on Graph Based Methods for Natural Language Processing* (pp. 73-80). Stroudsburg: Association for Computational Linguistics.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. USA: O'Reilly Media.

Bordons, M., Morillo., F., & Gómez, I. (2004) Analysis of cross-disciplinary research through bibliometric procedures. In Moed, H.F., Glänzel, W., & Schmoch, U. (Eds.), *Handbook of quantitative science and technology research*. (pp. 437-456), Dordrecht: Springer Netherlands.

Bracken, L.J., & Oughton, E. A. (2006). 'What do you mean?' The Importance of Language in Developing Interdisciplinary Research. *Transactions of the Institute of British Geographers, New Series*, *31*(3), 371-382.

Brody, S., & Lapata, M. (2009). Bayesian Word Sense Induction. In *Proceedings of the 12th Conference of the EACL 2009*, Athens, Greece, 30 March – 4 April 2009, (pp. 103-111).

Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form* (Vol. 9). John Benjamins Publishing.

File, P.E. & Dugard, P. I. (1997). Misunderstanding between knowledge engineers and domain experts: a re-evaluation of the use of induction for medical experts. *International Journal of Medical Informatics*, *46*(2), 113-118.

Fischuk, N. (2016). SEMASIOLOGICAL CONTENT OF METHOD IN SCIENTIFIC RESEARCHES. *Scientific Messenger of LNU of Veterinary Medicine and Biotechnologies*, *18*(2 (69)), 165-168.

Francl, M. (2015). Chemical doublespeak. *Nature Chemistry*, 7, 533-534.

Fruscione, D. (2013). Riht in earlier Anglo-Saxon legislation: a semasiological approach. *Historical Research*, *86*(233), 498-504.

Halldin, S., Bynander, F., & van Groningen, E. (2015). Natural Disaster Science: A Nordic approach to Integrated Research on Disaster Risk. *Geografiska Annaler: Series A, Physical Geography*, *97*(1), 1-7.

Harris, Z. S. (1954). Distributional Structure. *Word*, *10*(23), 146-162.

Heemskerk, M., Wilson, K., & Pavao-Zuckerman, M. (2003) Conceptual Models as Procedures for Communication Across Disciplines. *Conservation Ecology*, 7(3), 8.

Hodges, K. E. (2008). Defining the Problem: terminology and progress in ecology. *Frontiers in Ecology and the Environment*, *6*(1), 35-42.

Ifversen, J. (2011). About key concepts and how to study them. *Contributions to the History of Concepts*, *6*(1), 65-88.

Kageura, K., & Umino, B. (1996). Methods of automatic term recognition: a review. *Terminology*, *3*(2), 259-289.

Kemp, C. (1996). The Use of Abstraction: Remarks on Interdisciplinary Efforts in Law and Philosophy. *Denv. UL Rev.*, *74*, 877.

Klein, J. T. (2010). A Taxonomy of Interdisciplinarity. In Frodeman, R., Klein, J.T., & Mitcham, C. (Eds.) *The Oxford Handbook of Interdisciplinarity* (pp. 15-30). Oxford: Oxford University Press.

Korkontzelos, I., Klapaftis, I., & Manandhar, S. (2008). Reviewing and evaluating automatic term recognition techniques. In *Proceedings of the 6th International Conference on Advances in Natural Language Processing*, Gothenburg, Sweden, 25-27 August 2008 (pp. 248-259). Berlin, Heidelberg: Springer.

Kulkarni, V., Al-Rfou, R., Perozzi, B., & Skiena, S. (2015). Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 625-635). International World Wide Web Conferences Steering Committee.

Macken, L., Lefever, E., & Hoste, V. (2013). TExSIS: Bilingual terminology extraction from parallel corpora using chunk-based alignment. *Terminology*, 19(1), 1-30.

Mäki, U. (2016). Philosophy of interdisciplinarity. What? Why? How? *European Journal for Philosophy of Science, 6*(3), 327-342.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Baltimore, Maryland, USA, 22-27 June 2014, (pp. 55-60). Stroudsburg: Association for Computational Linguistics.

Monteiro, M., & Keating, E. (2009). Managing Misunderstandings: The Role of Language in Interdisciplinary Scientific Collaboration. *Science Communication*, *31*(1), 6-28.
Müller, V. C. (1998). About differences and blind spots: A systemic *view* on an international, interdisciplinary research team. *Journal of Managerial Psychology 13*(3/4), 259-270.
Nakagawa, H., & Mori, T. (2002). A simple but powerful term extraction method. In *Proceedings COLING-02 on COMPUTERM 2002: second international workshop on computational terminology* (pp. 1-7). Stroudsburg: Association for Computational Linguistics.
Naiman, R. J. (1999). A perspective on Interdisciplinary Science. *Ecosystems*, *2*(4), 292-295.
O'Rourke, M., & Crowley, S. J. (2013). Philosophical intervention and cross-disciplinary science: the story of the Toolbox Project. *Synthese*, *190*(11), 1937-1954.
Pantel, P., & Lin, D. (2002). Discovering word senses from text. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 613-619). ACM.
Pennington, D. (2008). Cross-Disciplinary Collaboration and Learning. *Ecology and Society*, 13(2): 8.
Purandare, A., & Pedersen, T. (2004) Word sense discrimination by clustering contexts in vector and similarity spaces. In *Proceedings of the Conference on Computational Natural Language Learning*, Boston, Massachusetts, USA, 6-7 May 2004 (pp. 41-48).
Ranade, S. M., Salazar, H., & Rodriguez, L.A. (2011). Back to Basics: Relearning terms, concepts for process control. Retrieved from: https://www.controleng.com/articles/back-to-basics-relearning-terms-concepts-for-process-control/.
Rossetto, T. (2016). Semantic ruminations on ‘post-representational cartography’. *International Journal of CartographyI, 1*(2), 151-167.
Schütze, H. (1998). Automatic Word Sense Discrimination. *Computational Linguistics*, *24*(1), 97-123.
Sinnett, D., & Williams, K. (2011). *Best practice guidance for interdisciplinary research: Experiences from Bridging the Gaps*. Technical Report, University of the West of England, UK.
The R Project for Statistical Computing (n.d.). Retrieved from: https://www.r-project.org/.
Thompson, J. L. (2009). Building Collective Communication Competence in Interdisciplinary Research Teams. *Journal of Applied Communication Research*, *37*(3), 278-297.
Van de Kauter, M., Coorman, M., Lefever, E., Desmet, B., Macken, L., & Hoste, V. (2013). LeTs Preprocess: The multilingual LT3 linguistic preprocessing toolkit. *Computational Linguistics in the Netherlands Journal*, 3, 103-120.
Van de Cruys, T., & Apidianaki, M. (2011). Latent Semantic Word Sense Induction and Disambiguation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, Oregon, USA, 19-24 June 2011 (pp. 1476-1485).
Véronis, J. (2004). Hyperlex: Lexical Cartography for Information Retrieval. *Computer Speech and Language*, *18*(3), 223-252.
Watson Alchemy Language Service (n.d.). Retrieved from: http://www.ibm.com/watson/developercloud/alchemy-language/api/v1/.

Paper 4 Approaching terminological ambiguity in cross-disciplinary communication as a word sense induction task: a pilot study[1]

[1] This is (the manuscript version) of the paper by myself, Prof. Dr. Ted Pedersen and Prof. Dr. Els Lefever that has been published in *Language Resources and Evaluation*.

Abstract

Cross-disciplinary communication is often impeded by terminological ambiguity. Hence, cross-disciplinary teams would greatly benefit from using a language technology-based tool that allows for the (at least semi-) automated resolution of ambiguous terms. Although no such tool is readily available, an interesting theoretical outline of one does exist. The main obstacle for the concrete realization of this tool is the current lack of an effective method for the automatic detection of the different meanings of ambiguous terms across different disciplinary jargons. In this paper, we set up a pilot study to experimentally assess whether the word sense induction technique of 'context clustering', as implemented in the software package 'SenseClusters', might be a solution. More specifically, given several sets of sentences coming from a cross-disciplinary corpus containing a specific ambiguous term, we verify whether this technique can classify each sentence in accordance to the meaning of the ambiguous term in that sentence. For the experiments, we first compile a corpus that represents the disciplinary jargons involved in a project on Bone Tissue Engineering. Next, we conduct two series of experiments. The first series focuses on determining appropriate SenseClusters parameter settings using manually selected test data for the ambiguous target terms 'matrix' and 'model'. The second series evaluates the actual performance of SenseClusters using randomly selected test data for an extended set of target terms. We observe that SenseClusters can successfully classify sentences from a cross-disciplinary corpus according to the meaning of the ambiguous term they contain. Hence, we argue that this implementation of context clustering shows potential as a method for the automatic detection of the meanings of ambiguous terms in cross-disciplinary communication.

Keywords

cross-disciplinary communication • disambiguation • SenseClusters • terminological ambiguity • word sense induction

4.1 Introduction

Cross-disciplinarity involves the combining of knowledge from different disciplines to answer a research question or to solve a problem. Because of its wide potential, the phenomenon is receiving much attention, and 'interdisciplinarity', a specific type of cross-disciplinarity, even became a buzzword (Klein, 1996). Yet, cross-disciplinarity is no easy way to success. Cross-disciplinary (CD) projects are usually carried out by a team of researchers with a background in one or more of the disciplines from which knowledge is drawn, i.e. the so-called 'source disciplines' of the project. An often cited challenge for CD communication, and thus for CD collaboration, is the lack of a common language among the members of CD teams (e.g. Benda et al., 2002; Hall and O'Rourke, 2014; Thompson, 2009).

One aspect of the challenge is the need to deal with terminological ambiguity, i.e. terms having multiple meanings, while it is not always clear which meaning is addressed.[1] This phenomenon is particularly common in CD communication settings. Each discipline has its own jargon, and while some terms are part of multiple jargons, they do not necessarily have the same meaning across those jargons. Hence, whenever disciplinary jargons are combined, terminological ambiguity arises. The close link between cross-disciplinarity and terminological ambiguity is acknowledged in philosophy of science, among others, by O'Rourke and Crowley (2013, p. 1941). They name "[...] the false appearance of agreement that can arise when the same word is unknowingly used with different meanings" as one of the main problems for CD communication. Furthermore, even though reports on struggles, mistakes and failures are rare in scientific literature, there are several accounts in concrete scientific research of CD communication being impeded by terminological ambiguity. For example, de Boer et al. (2006) mention the case of a project on sustainable water use where collaborating environmental psychologists and hydrologists were delayed by a misunderstanding regarding the meaning of 'water consumption'. While all researchers involved were familiar with the term, the psychologists took its meaning to be 'water use by households', while the hydrologists used the term as a synonym for 'evaporation'. Other, yet similar misunderstandings have been reported in the context of CD projects in environmental studies (Heemskerk, 2003;

[1] Note that, in this paper, we remain agnostic with respect to the relation between ambiguity and related phenomena like polysemy, fuzziness, vagueness and generality.

Naiman, 1999), law (Vick, 2004), historical anthropology (Lutter, 2015) and geography (Bracken and Oughton, 2006).

At this point, it is important to note that CD communication can also benefit from a certain amount of terminological ambiguity (Francl, 2015). However, this paper focuses on 'problematically ambiguous terms' (PATs) to refer to terms that (have the potential to) cause problems for CD communication because they have multiple meanings across the jargons of the source disciplines of a CD project. It goes without saying that it is in the interest of a CD team to resolve PATs, preferably before they cause any problems. Unfortunately, CD team members generally are not trained to do so, and even if they were, it would still be an extremely time-consuming endeavor. Therefore, they would benefit from using a 'PAT resolution tool' that automates as many parts as possible of the process of PAT resolution, and provides support for those parts that cannot be automated.

4.1.1 A theoretical outline of a PAT resolution tool

To the best of our knowledge, no concrete PAT resolution tool exists, although a theoretical outline of one can be found in Mennes (2018). The tool characterized by Mennes is called 'SenseDisclosure' and is a procedure consisting of seven steps, which, in principle, can be carried out for any given CD project. The steps are so conceived that four of them could be carried out automatedly by means of language technology. The first step consists in generating, for each discipline of a given CD project, a corpus that represents its jargon. Step two is to create a 'jargon set' for each discipline by extracting the terms from its corpus. The third step is to create a 'shared jargon set', i.e. the intersection of all jargon sets. Step four is to check for each term in the shared jargon set if it has multiple meanings, and if so, what these meanings are. The fifth step is to identify the polysemous shared jargon terms that are likely to cause communication problems, i.e. PATs. Step six is to generate visual representations of the different meanings (identified in step four) of each identified PAT. These representations are then used in the seventh step: the negotiation of new interpretation standards for all identified PATs.

When it comes to the realization of SenseDisclosure as a concrete tool, six of its steps mentioned above can already be implemented rather easily. More specifically, steps two, three and six are fully automatable with the required software applications and techniques readily available (or easy to develop). For example, to automate step two, existing term extraction software might need to be fine-tuned to be able to detect scientific and technical terms. Steps one, five and seven are not automatable, but can be executed by human experts. Step four, however, remains until now a serious obstacle for the full realization of SenseDisclosure. However, Mennes suggests to use a 'word sense induction' (WSI) technique for the implementation of this step. Such techniques allow for the automated identification of meanings, (or senses) of a term based on texts in which the term is used. The main advantages of using a WSI technique is that it is unsupervised, and does not require predefined sense inventories.[2] It is, however, not possible to simply take some WSI technique of the shelve and use it. WSI is still an open problem (Escudero et al., 2000). Whenever one wants to use a WSI technique for a certain purpose, its suitability in terms of performance, accuracy, etc. should be tested extensively.

4.1.2 Contributing to the realization of SenseDisclosure: testing context clustering

In this paper, we experimentally assess, for one WSI technique, whether it shows potential for the automation of the task described as step four in SenseDisclosure, i.e. the detection of different meanings of polysemous terms in a CD corpus. The technique we put to the test is called 'context clustering'. Briefly put, given a term of interest (henceforth 'target term'), and a corpus in which the target term occurs, this technique first generates vectors that represent the contexts in which the target term occurs (henceforth 'context vectors'). A 'context' is taken to be the *n* words occurring before and after the target term (Van de Cruys and Apidianaki 2011).[3] Next, the context vectors are clustered using their 'closeness' to each other. The underlying idea is that the closer two context vectors are, the more similar their corresponding contexts are, and thus, the more similar the meaning of the target term in those contexts. Also, the words stored in the context vectors of a cluster can be interpreted as pointers to the components of the meaning present in the cluster. A promising, state-of-the-art implementation of context clustering is the software package called 'SenseClusters', developed by Ted Pedersen and his team at the University of Minnesota.[4] We will use this implementation to test the potential of context clustering for step four in SenseDisclosure, thereby, hopefully paving the way to a concrete realization of the PAT resolution tool.

Remark, however, that this research still takes on the form of a restricted pilot study. First of all, we exclusively focus on just one implementation of one WSI technique.

[2] There are many general concerns about the use of sense inventory based approaches. For example, there is the difficulty of demarcating the semantic information that should be included in a sense description, and that of distinguishing between closely related senses (Edmonds and Kilgarriff, 2002). Relying on sense inventories is particularly problematic in the context of CD projects. Such projects require the compilation of a custom inventory by selecting sense descriptions from existing 'disciplinary' sense inventories. Yet, it is unclear how one can ensure that relevant sense descriptions are selected from relevant sense inventories. Moreover, new sense descriptions would need to be developed for terms that are not included in existing inventories.

[3] In this paper, we use 'word' and 'term' interchangeably, though we use the latter especially when we want to stress that a lexical unit has a meaning.

[4] For more information, go to http://senseclusters.sourceforge.net.

Secondly, it is rational to restrict the task at hand, i.e. we will not work with an unknown number of senses, any type of term (in terms of length and part of speech), nor with corpora that are highly specialized and span different disciplines.[5,6] Instead we will focus on the detection of a known number of senses for a restricted set of target terms, each of which is a single word noun, in a CD corpus compiled out of beginning and intermediate textbooks. Since there is no general agreement in related literature on what a 'detected meaning' should look like, nor on how they should be evaluated, we define the experimental task as the classification of sentences according to the meaning of a target term they contain (i.e. 'same meaning, same class'). Last but not least, this is a pilot study because it relies on just one corpus. The corpus is created for the CD project called 'Prometheus' focusing on Bone Tissue Engineering (BTE), i.e. the improvement of skeletal repair by means of stem cell technology, biomaterials and therapy.[7]

The pilot study encompasses two series of experiments. The first series is aimed at determining appropriate SenseClusters parameter settings for the experimental task. The second series of experiments is set up to evaluate the performance of SenseClusters with respect to the experimental task.

The remainder of this paper is structured as follows. Section 2 presents an overview of related research on word sense induction techniques and their implementations. Section 3 zooms in on the data that are used in the pilot study experiments and the measures used for the evaluation of the experimental results. It also characterizes SenseClusters and gives a detailed description of its parameter space. Section 4 describes the set-up of the two series of experiments, reports on the results of the experiments and reflects on the meaning of the experimental results. Finally, Section 5 presents our conclusions and outlines some prospects for future research.

---

[5] The more specialized a corpus is, the less broad definitions it contains. This means that references to more general or high-level components of the meanings of terms will be scarce, and thus are less likely to be picked up by means of a context clustering technique.

[6] By spanning different disciplines, the corpus becomes highly variegated as one meaning (e.g. 'having the capacity to cause rotation') will often be referred to by different terms (e.g. 'couple' in kinematics and 'force' in kinetics). This poses a challenge for context clustering, as not only term ambiguity is present but also (latent) synonymy.

[7] For more information, go to https://www.mtm.kuleuven.be/prometheus.

4.2 Related Research

This research takes a distributional semantics approach, which means that it is based on the distributional hypothesis, i.e. the idea that words occurring in similar contexts tend to have similar meanings (Harris, 1954). This idea has inspired computational linguists to use statistical measures of word usage to define the meaning of a word (See (Turney and Pantel, 2010) for an overview), leading to 'distributional models of meaning'.

One particular way of extracting semantics from word usage are vector space models (VSMs). VSMs represent each document (or vector) in a collection as a point in a vector space. Points (or vectors) that are closer in the vector space are then semantically more similar than points that are far from each other in the space. VSMs were originally developed for the SMART information retrieval system (Salton, 1971). Documents were represented as a vector containing a bag of words, and the entire set of documents were represented in a term-document matrix, where the rows of the matrix correspond to the words and the columns correspond to the documents. This idea was implemented in information retrieval to compare a given query (represented as a 'bag of words') with a document collection where each document is also represented as a bag of words. The relevance of a given document for the query is then determined by the similarity of their vectors, based on the idea that each column vector in a term-document matrix captures the content or topic of the corresponding document.

Deerwester et al. (1990) were inspired by the work of Salton to measure word similarity by building word-context matrices to compare context vectors of words in a document collection. This way, the word is represented by a vector of co-occurring words in the context of the focus word. These contexts can be windows of words, but also richer representations such as grammatical dependencies (Padó and Lapata, 2007). Similar row vectors in the word-context matrix then indicate similar word meanings (Turney and Pantel, 2010).

These co-occurrence matrices are usually built as a matrix of frequencies, where the cells list the number of times the context word co-occurs with the focus word, resulting in sparse matrices. In a next step, these raw frequencies are then adjusted in order to give more weight to more informative context words. A well-known weighting technique is *tf-idf* (Spârck Jones, 1972), which gives a higher weights to elements that are frequent in the corresponding document (*tf* or term frequency), but rare in other documents in the corpus (*idf* or inverse document frequency). Another popular weighting technique is Pointwise Mutual Information (PMI) (Church and Hanks, 1989), a measure of association between words, which tells us how much more often than chance two words co-occur.

Converting a text corpus into a co-occurrence matrix allows mathematical processing of the resulting context vectors. In order to measure the similarity between vectors, distance metrics such as the cosine distance are used, viz. The cosine of the angle between the two context vectors. Also machine learning approaches, such as clustering techniques or classification algorithms, can use real-valued vectors as input.

Co-occurrence matrices have a long history of use in a variety of NLP tasks. One important NLP application where feature vector representations are used is Word Sense Disambiguation, which consists in assigning the correct sense to a word in a given context (Agirre and Edmonds, 2006; Lefever et al., 2011). In this paper, however, we want to discover different senses for a given PAT in the corpus at hand. This task is referred to as word sense induction or word sense discrimination (Pedersen, 2006) and is carried out by clustering similar contexts (represented as vectors) that contain a given target term. Note that a target term can be any word of interest, but in this paper, is taken to be a PAT. The resulting clusters then correspond to the various meanings of the target term. Word sense induction thus starts from the hypothesis that contexts for a given target term that share co-occurring words with that target term, will have similar vectors, and will therefore be realizations of the same meaning of the respective target term. This way, senses are considered as clusters of similar contexts of the target term. However, these first-order methods are potentially brittle because they require some of the exact same words to occur across a range of contexts in order to establish a sense. While this is reasonable in cases where the contexts are relatively consistent and share a common vocabulary, in most other cases it is too restrictive and results in many small clusters being discovered that tend to fragment the discovered senses.

An alternative is to consider methods based on second-order co-occurrences (Schütze, 1998), which replaces every content word in a context with a vector representing its first-order co-occurrences. The word vectors for a given context are then averaged together in order to create a single vector that becomes the representation of that context. This is a second-order representation since the word vectors encode the words that co-occur with them, and so by averaging these vectors together indirect relationships between words that occur with the words in the context are discovered. These create a kind of 'friend-of-a-friend' relation among words and result in a representation where contexts can be clustered together to discover a sense even if those contexts have no words in common.

Second-order co-occurrences are naturally represented in a word graph, where words that co-occur are connected by an edge. Words that do not directly co-occur may share a common neighbour in the graph, in which case those words are jointed via a second-order relation created by that shared neighbour. Chinese Whispers (Biemann, 2006) is an algorithm that efficiently finds clusters in these kinds of co-occurrence graphs. If the words in each context of a target term are connected directly or indirectly via edges in a graph, then this algorithm can discover senses by clustering the graph.

In contrast with these 'count-based methods', which compute how often a particular target term co-occurs with neighbouring words, more recent techniques, so-called 'predictive models', try to predict a word from its neighbours. The vectors resulting from mapping a vocabulary into dense vectors of real numbers are called word embeddings. The most successful implementation of word embeddings is Word2Vec (Mikolov et al., 2013a, b), which learns the word embedding vectors using neural network-based language model techniques. During training, the algorithm moves through the training corpus with a sliding window, and learns to predict the word in the middle of the window based on the sum of the vector representations of the context words (Baroni et al., 2014). As a result, these vectors are similar to the vectors of paradigmatic neighbours (or second-order associations) in the training corpus.

A different approach to learn word embeddings is presented by the GloVe algorithm (Pennington et al., 2014), which tries to combine the count models and local context window predictive models.

Although these predictive models have shown to be very successful for various NLP tasks, such as word sense disambiguation (Iacobacci et al., 2016) and sentiment analysis (Yu et al., 2017), the models are often considered a black-box approach and the embedding dimensions are difficult to interpret (Levy and Goldberg, 2014).

For the presented research, we use SenseClusters (Purandare and Pedersen, 2004; Pedersen, 2006), which supports both first-order and second-order co-occurrence representations for word sense discrimination. The tool will be discussed in detail in Section 3.4.

4.3 Materials and methods

Before proceeding to the discussion of the actual experiments, this section gives a detailed overview of the data sets, the SenseClusters parameters and the evaluation measures used in the experiments.

4.3.1 Data

The training and testing of the underlying algorithm of SenseClusters requires dedicated data sets. To generate these sets, we first compiled a corpus for the Bone Tissue Engineering (BTE) project 'Prometheus' that serves as a test case. Prometheus combines knowledge from ten source disciplines, where 'disciplines' refers to the highly specific subjects that are studied by the working groups constituting the CD team. Some examples are the study of hydrogels, growth plates and calcium phosphate based scaffolds. The corpus consists of ten subcorpora, each of them representing the jargon of one of the source disciplines. To ensure that the corpus contains plenty of references to (the components of) the meanings of the terms it contains, we composed the subcorpora from chapters of beginning and intermediate textbooks. Such books are pedagogical and therefore take a more gradual approach on defining terms, proceeding from high-level characterizations to more specialized definitions, thereby providing much of semantic information. To find relevant textbook chapters, we used the search engines 'Google Scholar' and 'Google Books'. Only handbooks from well-known scientific publishing houses were withheld. For two source disciplines, (i.e. in vivo micro-environment and BMP-cell based implants), we could not find enough textbook material, so we included (much cited) review articles. We scanned the selected textbook chapters and downloaded the review articles. The PDF files thus obtained were converted to TXT files by means of optical character recognition software.[8] The size of the resulting subcorpora varies between 200,000 and 330,000 words, constituting a total corpus of 2,700,000 words. Detailed information on the size of the corpora can be found in Table 3. For methodological reasons, it is of great importance that the contexts of the terms in the corpus are continuous and maximally informative on the meaning (components) of these terms.[9] Therefore, we minimized the noise in the contexts by removing information related to the structure of the texts (e.g. references to figures and pages, and enumerations), pieces of text interrupting the flow of the main text (e.g. footnotes and captions of figures and tables), and different sorts of formal expressions (e.g. variables and formulas). Removing these types of noise led to a small decrease in size of the corpora, as can also be seen in Table 3.

Table 3 Overview of the number of words per subcorpus before and after cleaning

Source Discipline	Before cleaning	After cleaning
Bioprocessing	305,511	300,491
BMP cell-based implants	207,198	203,581
Calcium phosphate based scaffolds	300,084	295,344
Cell manipulation	274,525	268,970
Growth plate biology	276,804	270,070
Hydrogels	306,580	299,354
Imaging	274,920	270,354
In vivo micro-environment	242,028	238,651
Mechanisms of action	272,344	259,203
Mathematical modeling	338,151	325,073
Total	2,785,067	2,744,691

[8] The subcorpora do not perfectly mirror the original texts as the accuracy of the recognition results was only sanity-checked.

[9] The underlying reason is that SenseClusters is based on the distributional hypothesis as mentioned earlier in Section 2. See also Subsection 3.4.

Next, the subcorpora were preprocessed by means of the LeTs Preprocess toolkit (Van de Kauter et al., 2013). The preprocessing involved sentence boundary detection (i.e. breaking up text into sentences), tokenization (i.e. breaking up sentences into meaningful elements such as words, punctuation marks and numerals), part-of-speech (POS) tagging (i.e. assigning the appropriate grammatical category to each token), lemmatization (i.e. reducing each full form to its uninflected form), and named entity recognition (i.e. detecting person names, organizations, locations, etc.). A manual check of the automatedly generated POS tags for a random sample of five thousand tokens revealed a very good performance of LeTs for this specialized corpus with an accuracy of 0.99.

4.3.1.1 The set of training data

As mentioned earlier the pilot study consists of two series of experiments. The same set of training data was used in both series and is based on the full BTE corpus. It consists of all full forms that occur in the corpus and that were labelled as noun (i.e.'NN', 'NNP', 'NNS'), verb (i.e. 'VB', 'VBP', 'VBN', 'VBG', 'VBZ') or adjective (i.e. 'JJ') by the LeTs POS tagger.

4.3.1.2 The sets of test data

The test data used in the experiments are more diverse. For the first series, sets of test data were created for the target terms 'matrix' and 'model'. Both terms were indicated as PATs in the context of the BTE project by the scientific coordinator of Prometheus (personal communication, January 16, 2014). We first identified the main meanings of both terms. For 'matrix', we observed that the term usually has the 'mathematical' meaning of "a rectangular array with prescribed numbers *n* of rows and *m* of columns" (Serre, 2010, p. 15) or the 'biological' meaning of "the material or tissue in which more specialized structures are embedded" (Harvey and Lund, 2007, p. 3). For 'model' we also observed both a mathematical and a biological meaning. The mathematical meaning is that of "a mathematical description of a system" (Nijhout et al., 2008), the biological meaning is that of "a non-human species that is extensively studied to understand particular biological phenomena" (Ankeny and Leonelli, 2011, p. 314). Based on the identified meanings, we created the test sets. Per target term, thirty sentences were manually selected from the BTE corpus: fifteen sentences in which the target term clearly has a mathematical meaning, and fifteen in which it has a biological meaning. Again, only the nouns, verbs and adjectives occurring in the sentences were included in the test sets. For example, consider the following two sentences selected for 'matrix', where the term respectively has a mathematical and a biological meaning.

(1) The approach transforms a matrix [X], containing measurements from n measured variables, into a matrix of mutually uncorrelated PCs, tk (where to n), which are transforms of the original data into a new basis defined by a set of orthogonal loading vectors, pk.

(2) The role of membrane receptors for steroid hormones is of particular significance in bone growth and development since chondrocytes in the endochondral lineage have the goal of mineralizing their extracellular matrix, a process that involves extracellular membrane organelles called matrix vesicles.

For the second series of experiments, we created sets of test data for the following seventeen target terms: 'analysis', 'cell', 'differentiation', 'fraction', 'function', 'gene', 'mass', 'matrix', 'model', 'mutation', 'pathway', 'phase', 'protein', 'reaction', 'receptor', 'sensitivity' and 'tissue'. These target terms were manually selected from the top of a list containing the terms of the corpus (automatically extracted by means of TExSIS) ordered on the basis of their domain-specificity scores (as predicted by TExSIS) (Macken et al., 2013). The selection was made with an eye on including target terms that were expected to have different amounts of meanings, and meanings of different granularities. For each of the target terms, we randomly selected fifty sentences from the corpus containing the target term in question, the only restriction being that the sentences contain at least twenty words that are nouns, verbs or adjectives. For example, the following two sentences are the first instances of the test set for 'matrix' and 'phase' respectively.

(3) Although remarkable progress has been made in developing biodegradable composite matrices as scaffolds for bone regeneration, these substrates do not provide signals to the migrating bone marrow stromal cells to differentiate into multiple mature phenotypes that are different from their tissue of origin.
(4) In the heavy reacting phase, mass balance of species j at steady state lead where Qh is the volumetric flow rate of the heavy (fluid) reacting phase, CHj is the concentration of species j in the heavy reacting phase, Jj is the mass transfer flux of species j from the heavy reacting phase, and a is the specific area of contact between the heavy and light reacting phases.

Since the selection was made randomly, we had no knowledge of the (number of) meanings included in the test sets. These only became clear after the gold standards were created.

4.3.2 Gold Standards

To allow for the evaluation of the experimental results, gold standards were created. Given a test set focused on a specific target term, a gold standard partitions the test set on the basis of the meaning of the target term in the sentences. With respect to the first series of experiments, the gold standards are known a priori as the sets of test data were manually created by selecting sentences on the basis of the meaning of the present target terms. For the second series of experiments, the gold standards had to be created a posteriori as the sentences of the test sets were randomly selected (i.e. without prior knowledge of the meanings of the target terms). Moreover, for some test sets, multiple gold standards were created, where the number of gold standards corresponds with the number of granularity levels at which different meanings of the target term can be distinguished. An overview of the gold standards, the meanings they contain, and the relative frequency of each of those meanings in the test set can be found in Table 4.

Table 4 Overview of the gold standards and related meanings per target term

Term	**No. of senses**	**Definition**	**Freq.**
Series 1			
Matrix	2	Rectangular array with *n* rows and *m* columns	0.5
		Material in which structures are embedded	0.5
Model	2	Mathematical description of a system	0.5
		Species that is studied extensively	0.5
Series 2			
Analysis	3	Investigating and cleaning data	0.34
		Dissecting something theoretical, abstract or mathematical	0.34
		Separating a physical whole into parts	0.32
Cell	1	Basic unit of organisms	1
Differentiation	2	Process of specialization	0.98
		Mathematical process of finding a derivative	0.02
Fraction	2	Ratio	0.7
		(Physical) part of something	0.3
	3	Specific, scientific ratio	0.52
		Ratio	0.18
		(Physical) part of something	0.3
Function	2	Mathematical relation between elements of sets	0.3
		Role of a (body) part	0.7
Gene	1	Segment of DNA	1
Mass	3	Quantity expressed in kg	0.46
		Body of matter	0.52
		Large number or amount	0.02
Matrix	2	Material in which structures are embedded	0.88
		Rectangular array with *n* rows and *m* columns	0.12
Model	2	Species that is studied extensively	0.3
		Representation of a system	0.7
	3	Species that is studied extensively	0.3
		Mathematical description of a system	0.54
		Explanatory tool	0.16
Mutation	2	Event of change in genetic structure	0.54
		Effect of change in genetic structure	0.46
Pathway	3	Linked series of reactions related to metabolism	0.54
		Process of signal transduction	0.34
		Chain of reactions related to gene expression	0.12
Phase	3	Homogenous part of a heterogeneous mixture	0.4
		(Part of a mixture in an) aggregation state	0.24
		Stage or period of time	0.36
Protein	1	Polymer of amino acids	1
Reaction	2	Response to an intervention	0.14
		Process in which substances change	0.86
	3	Response to an intervention	0.16
		Chemical process of change of substances	0.64
		Physiological process or change of substances	0.2
Receptor	1	Structure that registers stimuli	1
Sensitivity	2	Precision with which a measurement is made	0.7
		Degree of responsiveness to stimuli	0.3
Tissue	1	Aggregate of cells with similar structure or function	1

4.3.3 Evaluation measures

The result of each experiment is evaluated by means of a *global score*, which is a count of the true positives, i.e. sentences that are correctly classified as belonging to the given class (and thus, in accordance with the gold standard). Besides the global score, three evaluation measures were applied for each classification experiment: *accuracy*, *precision* and *recall*. There are four notions constitutive for these measures: (1) *true positives* (TP) (i.e. as explained above), (2) *true negatives* (TN) (i.e. sentences that were correctly classified as not belonging to the given class), (3) *false positives* (FP) (i.e. sentences that are wrongly classified as belonging to the given class), and (4) *false negatives* (FN) (i.e. sentences that are wrongly classified as not belonging to the given class). The evaluation measures are defined as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

Note that, while precision and recall are calculated per class, accuracy is calculated for all classes together.

4.3.4 SenseClusters: an overview

As mentioned above, we used SenseClusters for our experiments. This section provides a brief characterization of SenseClusters, with a special focus on its parameter space. SenseClusters is a state-of-the-art context clustering software package, which supports both 1storder and 2nd order clustering. It incorporates the clustering package ‘Cluto’ (Karypis, 2002) and provides an extensive parameter space capturing all relevant theoretical degrees of freedom in context clustering, thereby allowing us to extensively tune context clustering processes. What follows is an overview of the parameter space of SenseClusters (Purandare and Pedersen, 2004).

Context Vector Type. Default: 2nd order.

This parameter determines what information (from the training data) is stored in the context vectors for a target term. Where 1st order context vectors contain information on the words that occur in the contexts of the target term, 2nd order context vectors contain information about the words occurring in the contexts of the words that occur in the contexts of the target term. For a detailed description of both types, see Section 2.

Feature Type. Default: bigrams.

This parameter determines the features, i.e. the type of information stored in the dimensions of a context vector. 'Bigrams' are pairs of words that are ordered in correspondence to the sequence in which they occur in a given sentence, but are not necessarily adjacent (also known as 'skippy bigrams'). Alternative settings are 'co-occurrences', i.e. unordered bigrams and 'target co-occurrences', i.e. co-occurrences containing the target term. When 1st order clustering is applied, one also has the option of using 'unigrams', or single words (Pedersen et al., 2005).

Window Size. Default: 2.

This parameter specifies the maximum number of consecutive words in the training data from which bigram, co-occurrence or target co-occurrence features are selected. Thus, '2' is the smallest possible window size and does not allow for other words to intervene between the constituents of a feature in the training data.

Lower Frequency Cut-off. Default: 5.

This parameter specifies the minimum number of times a feature has to occur in the training data in order to serve as a dimension of the context vectors.

Precision. Default: 6.

This parameter captures the number of digits after the decimal point that are taken into account.

Clustering Space. Default: vector clustering space.

This parameter defines the dimensions of the space in which the context vectors are clustered. In a 'vector clustering space', the dimensions are the same as those of the context vectors. Alternatively, the clustering is performed in a 'similarity space', where the dimensions are defined by the similarity of the context vectors.

Clustering Method. Default: repeated bisections.

This parameter determines the way in which the context vectors are processed and clusters are generated. For our experiments, we applied the default, 'repeated bisections', which is a standard divisive 'hierarchical' clustering method that takes a 'top down' approach: at the start of the clustering process, all data are assigned to one and the same cluster, and at each stage of the process, the available clusters are divided into two clusters in a way that optimizes a given clustering criterion function (see below). The process is repeated until the requested number of clusters has been reached (or when the cluster stopping algorithm ends it). SenseClusters also supports a 'bottom-up' hierarchical clustering method, i.e. 'agglomerative clustering', where initially, all data are assigned to individual clusters, and at each stage, two clusters are combined or 'agglomerated' into a bigger cluster. There is also the option 'partitional biased agglomerative'. In this case extra processing is performed before the agglomerative clustering, i.e. a repeated bisections method is applied to generate preliminary clusters that are used as extra dimensions (or features) in the actual clustering process. The two other methods supported by SenseClusters are 'direct clustering' and 'graph based clustering'. The hallmark of the first method is that is does not span multiple stages: it is a partitional clustering method where all clusters are generated at once. When a graph based clustering method is used, the data are first cast into a graph, where each training sentence is a node that is connected with the nodes that represent the most similar data. Next, a partitional clustering method is applied.

Cluster Stopping. Default: set manually.

This parameter indicates the way in which the number of generated clusters is determined. When the cluster stopping is set manually, the number of clusters that are generated is predefined. Alternatively, several cluster stopping algorithms can be used.

Number of clusters. Default: 10.

This parameter is self-explanatory. We changed its value to the number of classes in the gold standard. Note that the clustering of context vectors coincides with the classification of sentences (as the vectors represent the sentences, and the clusters determine the classes). Hence, the number of requested clusters is also the number of classes the test sentences are divided into.

Clustering Criterion Function. Default: I2.

This parameter determines how the clusters should be evaluated, and thus, optimized. I2 is an *internal* type criterion function. It optimizes the quality of clustering results by maximizing the similarity of context vectors assigned to the same cluster. SenseClusters also offers clustering criterion functions of the 'external' and 'hybrid' type. External functions maximize the dissimilarity between vectors assigned to different clusters. Hybrid functions balance intra-cluster similarity and inter-cluster dissimilarity.

Similarity Measure. Default: cosine distance.

This parameter determines how the similarity of context vectors is calculated. SenseClusters supports the use of ‘Cosine Distance’ and ‘Extended Jaccard Coefficient’. Both measures determine the similarity of two context vectors based on their direction (and not their length).

Statistical Test of Association. Default: none.

This parameter indicates how the strength of association between words constituting a bigram or co-occurrence feature is measured. SenseClusters supports ‘Log-likelihood’, ‘chi-square’, ‘Dice’, ‘Phi’, ‘Odds ratio’, ‘Pointwise mutual information’ (PMI), ‘True mutual information’ (TMI), ‘T-score’, ‘Left Fishers’ and ‘Right Fishers’.

4.4 Experiments and discussion

The experiments are aimed at assessing whether SenseClusters is able to detect a known number of meanings of a term in a CD corpus. The experimental task consists of the classification of sentences containing a target term according to the meaning of those target terms. The experiments are carried out in two series. The first series is aimed at determining appropriate SenseClusters parameter settings for the experimental task. The second series is carried out to allow for an evaluation of SenseClusters’ performance of the experimental task.

4.4.1 Determining appropriate parameter settings

In the first series of experiments, we start with the default settings of Senseclusters and gradually determine optima for the different parameters. In each round of experiments, we built on previously established optima. The SenseClusters algorithm is trained on the set of training data, and tested on the test sets for ‘matrix’ and ‘model’ (consisting of manually selected sentences). As these test sets each contain two meanings, the number of clusters was set to ‘2’.

The parameter setting we varied in the first round of experiments, was the context vector type. Since this parameter is related to that of the context vector type, we co-varied the latter. We tested all combinations, except for the combination of first order context vectors and target co-occurrence features, which is not supported by the software, and the combination of second order context vectors and unigram features, which cannot be done. The results of the first round of experiments are shown in Table 5. The first two columns of the table contain the different parameter settings that were tested, i.e. context vector types and feature types. The other six columns contain different evaluations of the experimental results based on the test sets for ‘matrix’ and ‘model’, i.e. score, precision for sentences containing meaning 1 (P1) and meaning 2 (P2), recall for sentences containing meaning 1 (R1) and meaning 2 (R2) and accuracy.

Table 5 Results of the experimental round with varying vector types and feature types

Vector type	Feature type		Score	P1	P2	R1	R2	A
1st order	Unigram	Matrix	30	1	1	1	1	1
		Model	26	0.87	0.87	0.87	0.87	0.87
	Bigram	Matrix	18	0.56	1	1	0.2	0.6
		Model	21	0.75	0.67	0.6	0.8	0.7
	Co-oc	Matrix	18	0.56	1	0.33	0.2	0.6
		Model	21	0.75	0.67	0.27	0.8	0.7
2nd order	Bigram	Matrix	18	0.71	0.57	0.33	0.87	0.6
		Model	17	0.67	0.54	0.27	0.87	0.57
	Co-oc	Matrix	17	0.63	0.55	0.33	0.8	0.57
		Model	17	0.67	0.54	0.26	0.87	0.57
	Target co	Matrix	22	0.89	0.67	0.53	0.93	0.73
		Model	19	0.6	0.7	0.8	0.47	0.63

As can be seen in Table 5, the best results were obtained by means of 1st order vector representations and unigram features.[10] According to previous research by Pedersen, this indicates that the training data contain a high number of contexts in which the two meanings of interest of the target terms are present (Pedersen, 2015). Therefore, as this was not guaranteed in the second series of experiments, we shifted our attention to finding appropriate 2nd order clustering parameter settings. Hence, in the second round of experiments, the vector type parameter was kept at 2nd order, and different feature type and window size parameter settings were tested. The results of the second round are shown in Table 6.

According to our experiments, the window size that yields the best result varies per feature type. When bigram features were used, the best result was obtained for (sentences containing) ‘matrix’ with a window size of ‘6’. In combination with co-occurrence features, a window size of ‘3’ or ‘4’ yielded the best clustering result for ‘matrix’, whereas the best result for ‘model’ was achieved with a window size of ‘6’. When target co-occurrences were used, the best result was achieved with a window size of ‘3’ (for

[10] We define ‘best result’ as the highest accuracy.

‘matrix’), but a window size of ‘4’ yielded the best average result (i.e. the best result for ‘model’, and the second best result for ‘matrix’). In the third round of experiments we explored the effect of varying the lower frequency cut-off values. We learned that varying this value barely affects the clustering results. In so far as the results did vary, better results were obtained with a lower cut-off frequency of ‘0’. An exception to this rule was encountered when the feature type was set to ‘target co-occurrence’ and the window size to ‘3’. In that case, the best classification result was obtained (for ‘matrix’) with a lower cut-off frequency of ‘4’. Therefore, in the fourth round of experiments, we kept the lower cut-off frequency at ‘0’, except for the target co-occurrence case, and varied the statistical tests of association.[11] The results of the fourth experimental round are shown in Table 7.

For eight out of twenty four experiments, an accuracy of 0.9 or more was achieved for both ‘matrix’ and ‘model’. Perfect clustering results were obtained (for ‘matrix’) when a Dice or Pointwise Mutual Information (PMI) filter was applied in combination with bigram and co-occurrence features.

Table 6 Results of the experimental round with varying window size

Feature Type	Window size		Score	P1	P2	R1	R2	A
Bigram	3	Matrix	19	0.75	0.59	0.4	0.87	0.63
		Model	17	0.67	0.54	0.27	0.87	0.57
	4	Matrix	19	0.75	0.59	0.4	0.87	0.63
		Model	17	0.67	0.54	0.27	0.87	0.57
	5	Matrix	19	0.75	0.59	0.4	0.87	0.63
		Model	17	0.67	0.54	0.27	0.87	0.57
	6	Matrix	25	1	0.75	0.67	1	0.83
		Model	17	0.67	0.54	0.27	0.87	0.57
Co-oc	3	Matrix	17	0.56	0.58	0.67	0.47	0.57
		Model	17	0.67	0.54	0.27	0.87	0.57
	4	Matrix	17	0.56	0.58	0.67	0.47	0.57
		Model	17	0.67	0.54	0.27	0.87	0.57
	5	Matrix	16	0.55	0.53	0.33	0.73	0.53
		Model	23	0.9	0.7	0.6	0.93	0.77
	6	Matrix	16	0.56	0.52	0.33	0.73	0.53
		Model	26	0.79	1	1	0.73	0.87
Target co	3	Matrix	22	0.77	0.71	0.67	0.8	0.73
		Model	19	0.6	0.7	0.8	0.47	0.63
	4	Matrix	21	0.71	0.69	0.67	0.73	0.7
		Model	21	0.67	0.75	0.8	0.6	0.7
	5	Matrix	20	0.63	0.73	0.8	0.53	0.67
		Model	21	0.67	0.75	0.8	0.6	0.7
	6	Matrix	20	0.63	0.73	0.8	0.53	0.67
		Model	21	0.67	0.75	0.8	0.6	0.7

[11] Because the settings combination of feature type ‘co-occurrences’ and a window size of ‘6’ yielded better results than the settings combination of feature type ‘co-occurrence’ and a window size of ‘3’, we omitted the latter combination of parameter settings in the fourth round of experiments.

After four rounds of experiments, appropriate parameter settings were found both for 1st and 2nd order clustering. For 1st order clustering, the default settings generate very good results. For 2nd order clustering, the statistical test of association should be set to 'Dice' or 'PMI', the feature type to 'bigrams' or 'co-occurrences', the window size to '6', and the lower frequency cut-off value to '0'.

4.4.2 Evaluating the performance of SenseClusters

The second series of experiments is set up to evaluate SenseClusters' performance on the experimental task. The SenseClusters algorithm was trained on the training set and tested on the sets of randomly selected sentences for seventeen target terms. A fixed set of experimentally determined "appropriate" parameter settings was used. Even though 1st order clustering combined with unigram features generated great results, we opted for a set of parameters for 2nd order clustering, i.e. feature type set to 'co-occurrences', a window size of '6', a lower frequency cut-off of '0' and the statistical test of association set to 'Dice'. This choice was based on previous research into the performance of SenseClusters. In the context of a SemEval challenge, Pedersen showed that 2nd order clustering allows for the detection of more "indirect relationships" between terms, meaning that more numerous and subtle semantic connections are picked up on (Pedersen, 2013, p. 203). Moreover, it is the safest choice given that it is not known whether there are enough contexts in the training data to allow for the detection of the different meanings of the target terms by means of 1st order clustering.

Experiments were performed per test set. To determine into how many classes SenseClusters had to divide the test sentences, we looked at the number of meanings included in the gold standard for the test set. In case of multiple gold standards, an experiment was performed for each of them. An exception was made for test sets for which the gold standard contains one meaning (e.g. the ones for 'cell' and 'gene'). In those cases, we let SenseClusters divide the test sentences into two classes. The underlying idea is that setting the number of clusters to '1' would yield perfect, yet trivial results and would not allow us to evaluate the performance of the software package. However, by setting the cluster parameter to '2', it can be evaluated whether the (large) majority of test sentences are indeed assigned to the same class. The results of the second series of experiments are shown in Table 8. The first two columns of the table present the test set that was used. The other six columns contain different evaluations of the experimental results, i.e. score, precision, recall and accuracy.

Table 7 Results of the experimental round with varying tests of statistical association

Feature type	Window size	Statistical test of association		Score	P1	R1	P2	R2	A
Bigram	6	Log-likelihood	Matrix	18	0.56	0.8	0.93	0.27	0.6
			Model	17	0.67	0.54	0.27	0.87	0.57
		Chi-square	Matrix	22	0.65	1	1	0.47	0.73
			Model	22	0.82	0.68	0.6	0.87	0.73
		Dice	Matrix	30	1	1	1	1	1
			Model	29	1	0.94	0.93	1	0.97
		Phi	Matrix	22	0.65	1	1	0.47	0.73
			Model	22	0.82	0.68	0.6	0.87	0.73
		PMI	Matrix	30	1	1	1	1	1
			Model	28	0.88	1	1	0.87	0.93
		TMI	Matrix	18	0.56	0.8	0.93	0.27	0.6
			Model	17	0.67	0.54	0.27	0.87	0.57
Co-oc	6	Log-likelihood	Matrix	16	0.56	0.52	0.33	0.73	0.53
			Model	15	0.5	0.5	0.8	0.2	0.5
		Chi-square	Matrix	20	0.6	1	1	0.33	0.67
			Model	20	0.78	0.62	0.57	0.87	0.67
		Dice	Matrix	30	1	1	1	1	1
			Model	29	1	0.94	0.93	1	0.97
		Phi	Matrix	20	0.6	1	1	0.33	0.67
			Model	20	0.78	0.62	0.47	0.87	0.67
		PMI	Matrix	30	1	1	1	1	1
			Model	28	0.88	1	1	0.87	0.93
		TMI	Matrix	16	0.56	0.52	0.33	0.73	0.53
			Model	15	0.5	0.5	0.8	0.2	0.5
Target co	3	Log-likelihood	Matrix	20	1	0.6	0.33	1	0.67
			Model	18	0.59	0.62	0.67	0.53	0.6
		Chi-square	Matrix	20	1	0.6	0.33	1	0.67
			Model	16	0.53	0.53	0.53	0.53	0.53
		Dice	Matrix	22	1	0.65	0.47	1	0.73
			Model	17	0.55	0.63	0.8	0.33	0.57
		Phi	Matrix	20	1	0.6	0.33	1	0.67
			Model	16	0.53	0.53	0.53	0.53	0.53
		PMI	Matrix	20	0.86	0.61	0.4	0.93	0.67
			Model	22	0.89	0.67	0.53	0.93	0.73
		TMI	Matrix	20	1	0.6	0.33	1	0.67
			Model	18	0.59	0.62	0.67	0.53	0.6
	4	Log-likelihood	Matrix	20	1	0.6	0.33	1	0.67
			Model	20	0.69	0.64	0.6	0.73	0.67
		Chi-square	Matrix	20	1	0.6	0.33	1	0.67
			Model	17	0.56	0.57	0.6	0.53	0.57
		Dice	Matrix	22	1	0.65	0.47	1	0.73
			Model	17	0.55	0.63	0.8	0.33	0.57
		Phi	Matrix	20	1	0.6	0.33	1	0.67
			Model	20	0.69	0.65	0.6	0.73	0.67
		PMI	Matrix	24	0.8	0.8	0.8	0.8	0.8
			Model	20	0.86	0.61	0.4	0.93	0.67
		TMI	Matrix	20	1	0.6	0.33	1	0.67
			Model	21	0.75	0.67	0.6	0.8	0.7

While in the first series of experiments, the accuracy of the results obtained with a fixed set of parameter settings was rather steady, the accuracy varies greatly in the second series, with the lowest accuracy being 0.54 and the highest 0.94. However, the overall results are satisfying. First of all, it is noteworthy that the experiments based on randomly selected test sentences containing two meanings for the target terms ‘matrix’ and ‘model’ yielded similar results to those obtained for the manually selected test sentences. This indicates that SenseClusters does not seem to require carefully curated test data. Secondly, nine out of twenty one experiments yielded results with an accuracy of 0.8 or more. The best results, with an accuracy of 0.9 or higher, were obtained for experiments based on test sets that contain two meanings for a target term. It is remarkable that very good results were obtained when the two meanings are *disparate*, as is the case for ‘differentiation’, ‘function’, ‘matrix’, ‘model’ and ‘sensitivity’, and when they are *related*, as holds for ‘fraction’. Note that the fact that these experiments are the most successful, could be an effect of the parameter settings being determined in experiments with a similar set-up. Another assuring observation is that the accuracy of the results for experiments based on test sets that contain *fuzzy* meanings of a target term, i.e. meanings that are not clearly defined, are relatively steady and lie between 0.64 and 0.68. Fuzzy meanings can be found in the test sets for the terms ‘analysis’, ‘mutation’, ‘reaction’, ‘mass’ and ‘phase’. For example, in the case of ‘analysis’, the “separating of a physical whole into parts” often entails “the investigation and cleansing of data”, blurring the boundary between these meanings. Consider the following test sentence, in which it is not clear whether ‘analysis’ refers to the study of something physical or the study of data:

(5) The specific objectives of this research included the evaluation of friction coefficient evolution using an oscillating pin-on-plate device, determination of the effect of normal load and lubrication fluid on friction and wear of PVA hydrogels, and analysis of predominant wear mechanisms.

More interesting observations regarding the performance of SenseClusters can be made when looking at some of the less favorable results. First of all, inferior results were obtained for experiments based on a test set containing only one meaning of a target term, i.e. the test sets for ‘cell’, ‘protein’, ‘receptor’ and ‘tissue’. One hypothesis is that although the terms only have one meaning, it is of a *general* or unspecific nature. For example, to know what is referred to with ‘receptor’, a type of stimulus should be specified as well as the way in which that stimulus is transmitted and received. As the details of a general term can be further characterized in many different ways, general terms can be expected to occur in extremely diverse contexts. Such a high variety of contexts makes it unlikely for context clustering software such as SenseClusters to detect only one meaning. Secondly, there are experiments based on sets of test data in which one meaning of the target term has a relative frequency of less than 0.12, as is the case for ‘differentiation’, ‘mass’ and ‘pathway’. Even though the accuracy of the overall results for these experiments is acceptable, the precision and recall for such underrepresented meanings was equal to 0. This means that SenseClusters can handle a kind of "class imbalance", but fails to detect meanings that are poorly represented. Thirdly, there is an observable decline with respect to the quality of the experimental results when the sentences of a test set that were first divided into two classes, are re-divided into three classes, and thus three meanings. This is the case for experiments performed for the terms 'fraction', 'matrix' and 'model' ('reaction' being an exception). The decline is most likely explained by the fact that the third meaning is obtained by splitting one of the original meanings of the target term into two more fine-grained meanings. For example, for 'matrix', "material in which structures are embedded" is split into "mold or breeding ground" and "extracellular matrix". In other words, SenseClusters is probably better at detecting coarse-grained meanings than fine-grained meanings.

Table 8 Results of the second series of experiments

Term	No. of senses	S	P1	R1	P2	R2	P3	R3	A
Analysis	3	20	0.5	0.11	0.56	0.5	0.4	0.8	0.64
Cell	1	28	1	0.56					0.56
Differentiation	2	47	0.97	0.95	0	0			0.94
Fraction	2	46	0.92	0.7	0.92	0.8			0.92
	3	27	0.76	0.93	0.25	0.13	0.92	0.8	0.84
Function	2	46	0.88	0.88	0.94	0.94			0.92
Gene	1	47	1	0.94					0.94
Mass	3	26	0.52	0.61	0.57	0.46	0	0	0.68
Matrix	2	45	1	0.89	0.55	1			0.9
	3	12	0.45	0.36	0	0	0.54	0.76	0.66
Model	2	42	1	0.47	0.81	1			0.84
	3	30	1	0.47	0.89	0.63	0.25	0.75	0.73
Mutation	2	17	0.91	0.41	0.58	0.96			0.66
Pathway	3	27	0.62	0.7	0.47	0.47	0	0	0.69
Phase	3	11	0.88	0.4	0.33	0.75	0.64	0.5	0.68
Protein	1	28	1	0.56					0.56
Reaction	2	16	0	0	0.83	0.79			0.68
	3	17	0.33	0.86	0.78	0.54	0.22	0.2	0.68
Receptor	1	31	1	0.62					0.62
Sensitivity	2	42	0.89	0.89	0.73	0.73			0.84
Tissue	1	41	1	0.82					0.82

In sum, based on the pilot study experiments discussed above, we can conclude that SenseClusters succeeds at the experimental task: SenseClusters is able to classify sentences that are derived from a CD corpus in accordance with the meaning of some polysemous target term they contain. Hence, this implementation of the WSI technique of 'context clustering' has clear potential to carry out step four in the PAT resolution tool SenseDisclosure. Yet, SenseClusters' take on the experimental task is not an unqualified success. It succeeds well for sentences containing a target term that has two distinct (though not necessarily disparate) meanings, but sputters when the term has fuzzy or general meanings. Moreover, SenseClusters generates better results for sentences containing a target term with a coarse-grained meaning, than for those sentences containing a target term with a fine-grained meaning. Finally, two qualifying remarks regarding the evaluation of the performance of SenseClusters are in place. First of all, the experiments of the second series only allow for an evaluation of the performance of SenseClusters insofar as the parameter settings determined in the first series are indeed appropriate for the experimental task. Secondly, not all entries of the gold standards are equally solid. Since the test sentences for the second series of experiments were randomly selected, some of them did not contain enough meaningful context words to determine the meaning of the target term. Consider, for example, the following test sentence for target term 'phase':

(6) Under many implementations, the "phase" information is directly proportional to displacements, while the "magnitude" image is familiar to clinicians and scientists interested to observe tissue morphology, typically with excellent soft tissue contrast compared to other techniques like optical imaging or ultrasound.

The sentence does not contain enough information to decide with certainty whether the term 'phase' refers to "a period of time" or to "a part of a mixture". Yet, in the relevant gold standard, the sentence was assigned to the class associated with the latter meaning. The deciding factor here was the occurrence of 'tissue morphology', instead of, for example, 'tissue development'. A cursory analysis shows that just a little more than half of the sentences about which we were unsure when creating the gold standards were classified wrongly by SenseClusters. Most likely, this problem would be solved by increasing the context size.

4.4.3 Excursion: putting the classes to work

The cautiously established potential of SenseClusters for the implementation of the fourth step in SenseDisclosure makes it possible to also tentatively explore its potential for the implementation of the sixth step. To recap: after the *detection* of the meanings of polysemous shared jargon terms in the second step, the sixth step deals with the generation of the *visual representations* of the meanings of each polysemous shared jargon term that was identified as a PAT. In this brief excursion, we test whether the classes of test sentences generated by SenseClusters can be used to visualize the meaning of the

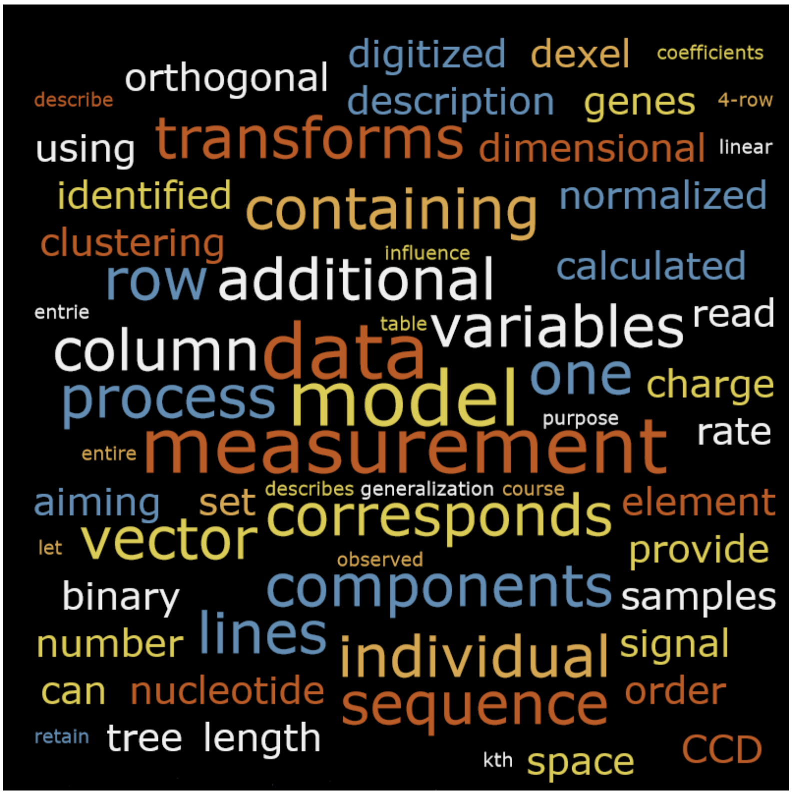

Figure 1 *Matrix* sense 1

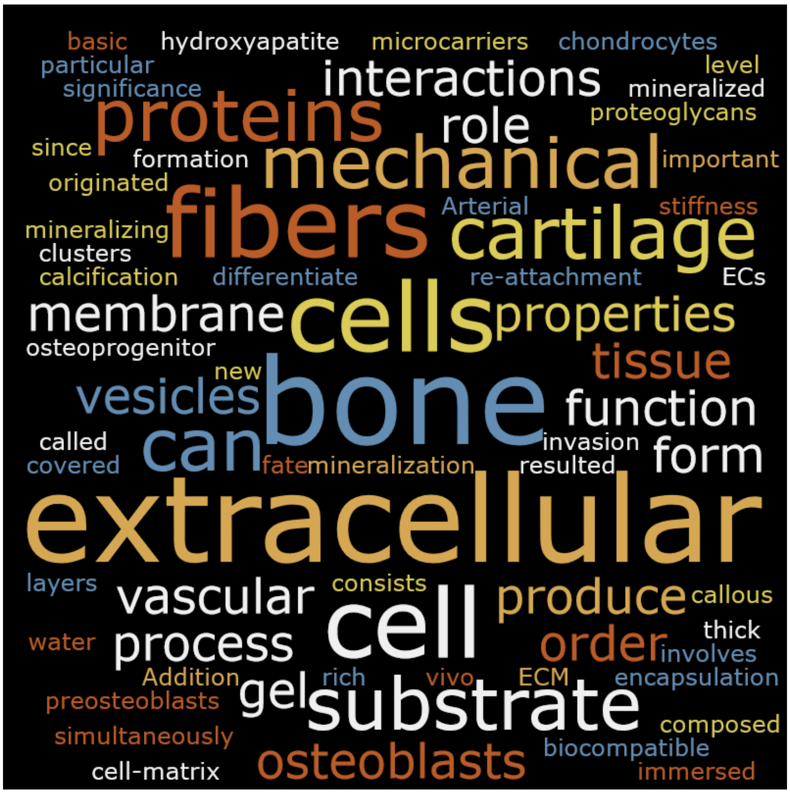

Figure 2 *Matrix* sense 2

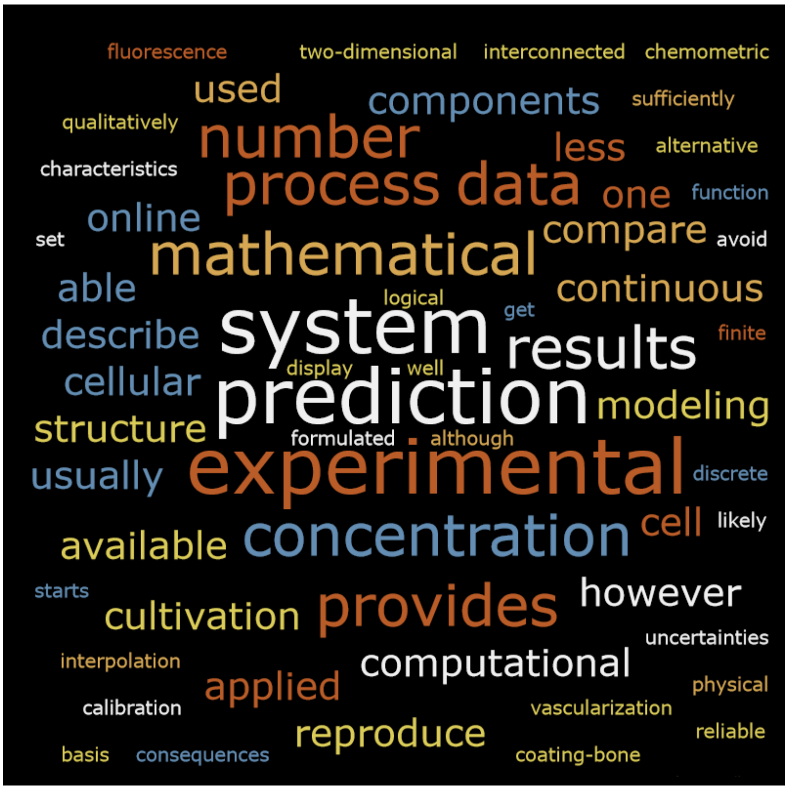

Figure 3 *Model* sense 1

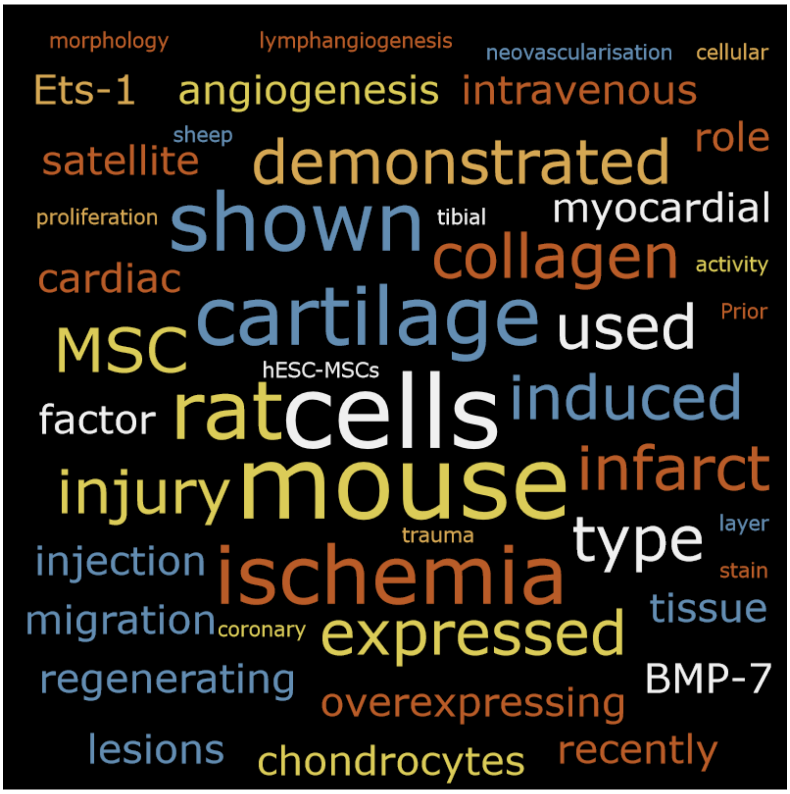

Figure 4 *Model* sense 2

target terms corresponding to those classes. The type of visual representation we need are ‘word clouds’, i.e. groups of words organized around a central point, where the font size of the displayed words indicates their importance or weight. We created word clouds for the classes of sentences generated for ‘matrix’ and ‘model’ in the first series of experiments by means of an online word cloud generator.[12]

Figures 1 and 2 visualize the two meanings of the term ‘matrix’, where the first word cloud corresponds with the mathematical sense of the term, i.e. “rectangular array with rows and columns” and the latter corresponds with the biological sense of the term, i.e. “material or tissue in which more specialized structures are embedded”.

Figures 3 and 4 visualize the two meanings of ‘model’. The first word cloud corresponds with the mathematical meaning of the term, i.e. “a mathematical description of a system”, the latter with its biological meaning, i.e. “a non-human species that is extensively studied to understand particular biological phenomena”. For Bone Tissue Engineering (BTE) laymen, the word clouds seem to give a good picture of both meanings of the target terms. However, a conclusive evaluation of the usefulness of the word clouds requires further research. In follow-up research, the CD team members working on the BTE project will be consulted and it will be investigated whether the word clouds are truly helpful in informing them on the different meanings of the PATs.

4.5 Conclusions and further research

This paper addresses the issue of problematically ambiguous terms (PATs) in cross-disciplinary (CD) communication settings. PATs are terms that have different meanings across the jargons of the source disciplines of a CD project and therefore (potentially) impede the communication between the CD team members. In that context, CD teams are requiring a tool assisting them in the resolution of these problematically ambiguous terms. The main obstacle for the realization of such a PAT resolution tool is the implementation of an effective method for the automatic detection of the different meanings of ambiguous terms across different disciplinary jargons. In this study, we therefore set up a pilot study to experimentally assess whether the word sense induction (WSI) technique of ‘context clustering’, as implemented in the software package ‘SenseClusters’, might be a solution.

The pilot study consists of two series of experiments revolving around the experimental task of classifying sentences derived from a CD corpus according to the

[12] https://www.wordclouds.com

meaning of a target term they contain. The first series of experiments was carried out to determine appropriate SenseClusters parameter settings for the experimental task. In this series, the algorithm underlying SenseClusters was trained on a corpus that represents the source disciplines of a CD project on Bone Tissue Engineering (BTE). It was tested using manually selected sentences containing the target term 'matrix' or 'model'. We observed that, for 1st order clustering, the default parameter settings of SenseClusters work best. For 2nd order clustering, the feature type should be changed to 'bigrams' or 'co-occurrences', the window size to '6', the lower cut-off frequency to '0', and the statistical test of association to 'Pointwise Mutual Information' (PMI) or 'Dice'. In the second series of experiments, we evaluated SenseClusters' performance on the experimental task using one of the experimentally determined sets of parameter settings. Again, SenseClusters was trained on the BTE corpus, but was tested on sets of randomly selected sentences, which were created for seventeen different target terms. We observed that SenseClusters indeed is able to perform the experimental task very well. The best results were obtained for test sets containing only two meanings of the target terms in question, especially when the meanings are not fuzzy or general. Hence, we conclude that this implementation of context clustering shows potential as a method for the automatic detection of the meanings of ambiguous terms in cross-disciplinary communication. As a bonus, we briefly explored whether SenseClusters could also be used for the automation of step six in the PAT resolution tool 'SenseDisclosure', i.e. the visualization of the different meanings of PATs. At first sight, also these results are promising. However, as the research presented in this paper constitutes a restricted pilot study, further research is critical.

A first interesting research path preserves the current set-up and methodology revolving around SenseClusters, but varies experimentally with respect to the following aspects: (1) the usage of alternative test data focused on target terms with more than two meanings, (2) the usage of automatic cluster stopping algorithms (as the number of meanings of a PAT generally is not known), (3) the usage of larger contexts (as the pilot study experiments indicated this might improve the results), (4) the inclusion of lemma's and part of speech (POS) tags of context words in the context vectors, and (5) the inclusion of multi-word compound target terms. Secondly, independent, complementary studies are required. To gain more insight in the potential of WSI techniques for PAT resolution, these studies should be carried out with different implementations of context clustering, but also with implementations of other WSI techniques, such as graph based clustering and predictive models. Finally, experiments should be performed based on CD corpora for fields other than BTE.

Funding

The work presented in this paper was carried out in the context of a PhD fellowship funded by the Research Foundation—Flanders (FWO).

Acknowledgements

We thank Prof. Dr. Liesbet Geris for sharing her cross-disciplinary experiences as the Scientific Coordinator of Prometheus and providing us with the necessary information for the corpus compilation. We also want to thank Prof. Dr. Stephan van der Waart van Gulik for his constructive feedback which helped to improve the paper significantly.

References

Agirre, E., & Edmonds, P. (2006). Word sense disambiguation: Algorithms and applications. Berlin: Springer.

Ankeny, R. A., & Leonelli, S. (2011). What's so special about model organisms? Studies in History and Philosophy of Science Part A, *42*(2), 313–323.

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd annual meeting of the chapter of the association for computational linguistics*, Baltimore (pp. 238–247). Maryland, USA: ACL.

Benda, L., Poff, L., Tague, C., Palmer, M., Pizzuto, J., Cooper, S., Stanley, E., & Moglen, G.(2002). How to avoid train wrecks when using science in environmental problem solving. *BioScience*, *52*(12), 1127–1139.

Biemann, C. (2006). Chinese whispers: An efficient graph clustering algorithm and its application to natural language processing problems. In *Proceedings of the first workshop on graph based methods for natural language processing*, New York City (pp. 73–80).

Bracken, L. J., & Oughton, E. A. (2006). 'What do you mean?' The importance of language in developing interdisciplinary research. *Transactions of the Institute of British Geographers*, *31*(3), 371–382.

Church, K., & Hanks, P. (1989). Word association norms, mutual information, and lexicography. In *Proceedings of the 27th annual conference of the association of computational linguistics*, Vancouver, British Columbia (pp. 76–83).

de Boer, Y., de Gier, A., Verschuur, M., & de Wit, B. (2006). Bruggen bouwen. Onderzoekers over hun ervaringen met interdisciplinair onderzoek in Nederland. RMNO, KNAW, NWO & COS. Retrieved from: https://www.knaw.nl/shared/resources/actueel/publicaties/pdf/Bruggen_Bouwen_Onderzoekers_over_interdisciplinair_onderzoek_2006.pdf/view .

Deerwester, S., Dumais, S., Landauer, T., Furnas, G., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407.

Edmonds, P., & Kilgarriff, A. (2002). Introduction to the special issue on evaluating word sense disambiguation systems. *Natural Language Engineering*, *8*(4), 279–291.

Escudero, G., Màrquez, L., & Rigau, G. (2000). Boosting applied to word sense disambiguation. In R. López de Mántaras & E. Plaza (Eds.), *Machine learning: ECML 2000* (pp. 129–141). Berlin: Springer.

Francl, M. (2015). Chemical doublespeak. *Nature Chemistry*, *7*(7), 533.

Hall, T. E., & O'Rourke, M. (2014). Responding to communication challenges in transdisciplinary sustainability science. In Huutoniemi, K. & Tapio, P. (Eds.), *Transdisciplinary Sustainability Studies* (pp. 135–155). Routledge.

Harris, Z. (1954). Distributional structure. *Word*, *10*(23), 146–162.

Harvey, R., & Lund, V. (2007). Biofilms and chronic rhinosinusitis: systematic review of evidence, current concepts and directions for research. *Rhinology*, *45*(1), 3–13.

Heemskerk, M. (2003). Conceptual models as tools for communication across disciplines. *Conservation Ecology*, *7*(3), 1-13.

Iacobacci, I., Pilehvar, M., & Navigli, R. (2016). Embeddings for word sense disambiguation: An evaluation study. In *Proceedings of the 54th annual meeting of the association for computational linguistics* (pp. 897–907). Berlin, Germany: ACL.

Karypis, G. (2002). Cluto-a clustering toolkit. Tech. rep., Minnesota Univ Minneapolis Dept of Computer Science.

Klein, J. T. (1996). Crossing boundaries: Knowledge, disciplinarities, and interdisciplinarities. Charlottesville: University of Virginia Press.

Lefever, E., Hoste, V., & De Cock, M. (2011). ParaSense or how to use parallel corpora for word sense disambiguation. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp 317–322) Portland, Oregon, USA: Association for Computational Linguistics.

Levy, O., & Goldberg, Y. (2014). Dependency-based word embeddings. In *Proceedings of the 54th Annual meeting of the association for computational linguistics* (pp. 302–308) Baltimore, Maryland, USA: ACL.

Lutter, C. (2015). Comparative approaches to visions of community. *History and Anthropology*, *26*(1), 129–143.

Macken, L., Lefever, E., & Hoste, V. (2013). Texsis: Bilingual terminology extraction from parallel corpora using chunk-based alignment. *Terminology International Journal of Theoretical and Applied Issues in Specialized Communication*, *19*(1), 1–30.

Mennes, J. (2018). SenseDisclosure. A new procedure for dealing with problematically ambiguous terms in cross-disciplinary communication. *Language Sciences*, 69, 57–67.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. In *Proceedings of the international conference on learning representations* (ICLR).

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing sysems (pp. 3111–3119).

Naiman, R. (1999). A perspective on interdisciplinary science. *Ecosystems*, *2*(4), 292–295.

Nijhout, H., Reed, M., & Ulrich, C. (2008). Mathematical models of folate-mediated one-carbon metabolism. *Vitamins & Hormones*, 79, 45–82.

O'Rourke, M., & Crowley, S. J. (2013). Philosophical intervention and cross-disciplinary science: The story of the toolbox project. *Synthese*, 190, 1–18.

Padó, S., & Lapata, M. (2007). Dependency-based construction of semantic space models. *Computational Linguistics*, *33*(2), 161–199.

Pedersen, T. (2006). Unsupervised corpus-based methods for WSD. In Agirre, E. & Edmonds, P. (Eds.), *Word sense disambiguation: Algorithms and applications*, Springer, pp 133–166.

Pedersen, T. (2013). Duluth: Word sense induction applied to web page clustering. In *Second joint conference on lexical and computational semantics (* SEM), Volume 2: Proceedings of the seventh international workshop on semantic evaluation (SemEval 2013)* (pp. 202–206).

Pedersen, T. (2015). Duluth: Word sense discrimination in the service of lexicography. In *Proceedings of the 9th international workshop on semantic evaluation* (SemEval 2015) (pp. 438–442).
Pedersen, T., Purandare, A., & Kulkarni, A. (2005). Name discrimination by clustering similar contexts. In *Proceedings of the sixth international conference on intelligent text processing and computational linguistics*, Mexico City (pp. 220–231).
Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. In Empirical methods in natural language processing (EMNLP) (pp. 1532–1543).
Purandare, A., & Pedersen, T. (2004). Word sense discrimination by clustering contexts in vector and similarity spaces. In *Proceedings of the conference on computational natural language learning*, Boston, MA (pp. 41–48).
Salton, G. (1971). The SMART retrieval system: Experiments in automatic document processing. Upper Saddle River, NJ: Prentice-Hall.
Schütze, H. (1998). Automatic word sense discrimination. *Computational Linguistics*, *24*(1), 97–123.
Serre, D. (2010). Matrices: Theory and applications. Graduate texts in mathematics. (2nd ed.). Springer-Verlag New York.
Spârck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, *28*(1), 11–21.
Thompson, J. (2009). Building collective communication competence in interdisciplinary research teams. *Journal of Applied Communication Research*, *37*(3), 278–297.
Turney, P., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37, 141–188.
Van de Kauter, M., Coorman, G., Lefever, E., Desmet, B., Macken, L., & Hoste, V. (2013). Lets preprocess: The multilingual LT3 linguistic preprocessing toolkit. *Computational Linguistics in the Netherlands Journal*, 3, 103–120.
Van de Cruys, T., & Apidianaki, M. (2011). Latent semantic word sense induction and disambiguation. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies, association for computational linguistics*, Portland, Oregon, USA (pp. 1476–1485).
Vick, D. W. (2004). Interdisciplinarity and the discipline of law. *Journal of Law and Society*, *31*(2), 163–193.
Yu, L. C., Wang, J., Lai, K., & Zhang, X. (2017). Refining word embeddings for sentiment analysis. In Empirical methods in natural language processing (EMNLP) (pp. 545–550).

Paper 5
A critical analysis and explication of word sense disambiguation as approached by natural language processing[1]

[1] This is a manuscript by myself and Dr. Stephan van der Waart van Gulik that has been submitted to *Linguistics and Philosophy*

Abstract

This paper presents a critical analysis and explication of the approach to the problem of word sense disambiguation by the field of natural language processing. The paper is motivated by the observation that, despite the fact that the approach is very interesting and promising, the actual research has some problems with regard to its conceptual frameworks and methods. The methodology behind the analysis and explication is philosophical, focussing on the identification of conceptual unclarities and methodological incoherencies as well as their (practical) implications. The methodology helps to pinpoint some structural and rather complex confusion in the research with respect to both a central concept and an important set of methods. First, the characterization of the problem of word sense disambiguation as a computational task allows for different interpretations. Depending on whether a rich or less rich interpretation is adopted, the methods developed for performing the task have or do not have certain problematic limitations that need to be solved. Secondly, the applicability of methods for the inductive discrimination of word senses is overestimated. The acknowledgement of this confusion makes it possible to understand the sometimes erratic course of the research and underlines the need for one, commonly shared, more refined and integrated conceptual framework and methodology. This paper may be interpreted as a humble onset of some of the necessary groundwork for developing these.

Keywords

machine learning • natural language processing • sense repository • word sense disambiguation • word sense induction

5.1 Introduction

In natural languages, many words are 'ambiguous': they have more than one meaning or 'sense'. When an ambiguous word is communicated, either through text or verbally, it is often challenging to understand which sense is actually used. A classic example is the English word 'bank'. This word is ambiguous as it has several different senses, two of which are (roughly put) 'the terrain alongside the bed of a river, creek or stream' and 'an organization where money can be deposited'. Now consider the following two sentences.

a. "He was at the bank."
b. "He was at the bank as he wanted to close his personal account."

In sentence (a), without any extra information, it is unclear which sense is used by the occurrence of 'bank'. The person referred to could very well be at (as in 'near') the river bank or at (as in 'in') the building of a financial organization. In sentence (b), however, extra information is provided, more specifically with respect to a financial transaction, i.e. 'closing a personal account', thereby making it highly likely that the occurrence of 'bank' in that sentence uses the sense 'an organization where money can be deposited'.

The problem of how to correctly disambiguate ambiguous words is studied in different research fields. In fields such as artificial intelligence and psycholinguistics, a 'theoretical' or 'scientific' approach is taken as the problem is considered to be part of a larger set of fundamental research problems concerning the 'deep' relation between semantics and cognition (Ide & Véronis, 1998). By contrast, in natural language processing, the problem is approached as a problem that can be solved by (rather pragmatic) computational means.

This paper focuses exclusively on the approach taken by natural language processing, presenting a critical analysis and explication of the conceptual frameworks and methods developed within this approach. The paper is also restricted to those methods that exclusively work with 'mono-lingual corpora', i.e. large bodies of text written exclusively in one language.[1] The paper is motivated by the observation that, despite the fact that the approach is very interesting and promising, the actual research regularly suffers from a

[1] Another popular set of method uses 'cross-lingual' corpora, i.e. large bodies of text written in one language and translations of those texts into one or more other languages. These methods are often used in the context of machine translation and multilingual information retrieval (e.g. Gale & Church, 1993; Lefever et al., 2011).

lack of conceptual clarity and methodological systematicity.[2] The methodology behind the analysis and explication is philosophical, focusing on the identification of conceptual unclarities and methodological incoherencies as well as their (practical) implications. This methodology helps to pinpoint some structural and rather complex confusion in the research with respect to both a central concept and an important set of methods. The acknowledgement of this confusion makes it possible to understand the sometimes erratic course of the research and underlines the need for one, commonly shared, more refined and integrated conceptual framework and methodology. This paper may be interpreted as a humble onset of some of the necessary groundwork for developing these.

It is also important to note that the paper contains almost no (formal) technicalities, despite the fact that many of the discussed methods are essentially algorithms and thus rather technical. This is intentionally so and for two reasons. First of all, many (if not all) of the technical details of the methods are irrelevant for (understanding) the actual analysis and explication. Secondly, a (potentially) broad and heterogeneous reader audience is assumed, thus making it critical to avoid diving too deep into a lot of specialist knowledge, especially when there is only little to gain by doing so. Therefore, where it might be helpful, technical elements are sketched out in an informal, conceptual manner. Readers acquainted with technical details can of course easily skip these sketches. Readers who are not, but are (becoming) interested, are always provided with references for further in-depth reading.

The paper is structured as follows. Section 2 discusses the way in which the problem of word sense disambiguation is approached in the field of natural language processing, i.e. its history, key concepts and methods. In Section 3 several severe limitations of word sense disambiguation methods that rely on pre-existing sense repositories are presented. Section 4 introduces an attempt by the research community to overcome these limitations. The attempt is based on the integration of methods for solving yet another problem, i.e. the problem of word sense induction. In Section 5, these methods are discussed in more detail. Section 6 explains how the attempt introduced in Section 4 is based on structural confusion within the research community both with respect to the computational characterization of the problem of word sense disambiguation and word sense induction methods. The confusion is responsible for two types of problems for the attempt: (i) a serious limitation that needs to be solved when adopting a specific interpretation of the aforementioned computational characterization and (ii) how to deal with certain inherent limits of word sense induction methods. Finally, Section 7 summarizes the most important findings.

[2] This is also the personal experience of one of the authors of this paper, Julie Mennes. While working in the field, she observed a lack of conceptual clarity in some reported results in the literature and experienced difficulties when performing computational experiments that were set-up and directed using those same reported results.

5.2 Word sense disambiguation: history, key concepts and methods

Word sense disambiguation (WSD) is an important research problem in the field of natural language processing (NLP) (e.g. Agirre & Edmonds, 2007). The history of WSD is discussed in great detail by Nancy Ide and Jean Véronis (1998) and Mark Stevenson and Yorick Wilks (2003). They locate the origin of WSD in the first explorations of machine translation in the 1950's. Almost immediately it became clear that machine translation can only work if it is possible to solve another automation problem, i.e. given an ambiguous word and a context in which it occurs, how to automatically determine which sense of the word is used. Today, WSD is studied by a large community of NLP researchers.

The community is highly organized and interactive. This is nicely illustrated by the periodically organized event 'SemEval' (e.g. Nasiruddin, 2013). SemEval was originally known as 'Senseval' and called into life by the 'Association for Computational Linguistics Special Interest Group on the LEXicon' (ACL-SIGLEX) in 1998 to systematically compare different methods developed for solving WSD. The name change happened in 2007 and indicated the inclusion of NLP methods not related to WSD (e.g. Kapetanious et al., 2013). Until 2016, every edition of SemEval included different WSD-related challenges. A typical example of such a challenge is 'lexical sample', where different WSD research teams try to outperform each other with respect to the automatic disambiguation of specific occurrences of a sample set of words in a given text (Martínez et al., 2008). For each challenge, the organizers provided the necessary data and evaluated the performance of the methods of the competing teams using a fixed evaluation methodology.

Not surprisingly, in the community, WSD is characterized as a computational task:

> (WSD-T) 'given an ambiguous word and a text, label each occurrence of the word in the text with the sense it is using'.

In this paper, a method that performs WSD-T is called a 'WSD method'.

Note that a WSD method always requires some kind of repository of senses from which sense labels can be derived. Originally, four types of pre-existing sense repositories were used.

- Type 1: 'machine-readable dictionaries', i.e. alphabetically ordered lists of words with characterizations of their sense(s) stored in a data format that can be queried by software applications (e.g. Diekema, 2003).
- Type 2: 'thesauri', i.e. formal data structures consisting of words and their semantic relations to other words, such as synonymy, antonymy, hyponymy and meronymy (e.g. Abuzir & Vandamme, 2002).

- Type 3: ‘computational lexica’, i.e. combinations of machine-readable dictionaries and thesauri (e.g. Navigli, 2009; Dias-Da-Silva, 2010).
- Type 4: ‘sense-annotated corpora’, i.e. texts in which occurrences of ambiguous words are labelled (or ‘annotated’) with the sense they are using (e.g. Vial et al., 2017).

The actual WSD methods can be divided into two types: ‘knowledge-based’ and ‘supervised’ methods.

Knowledge-based methods are methods that rely primarily on repositories of type 1-3 above (e.g. Agirre & Edmonds, 2007). There are three main types of knowledge-based methods.[3]

- Type 1: methods that use some formal notion of overlap between the context of an occurrence of an ambiguous word and its entries in dictionaries, thesauri or lexica. Famous examples are the LESK algorithm (e.g. Lesk, 1986; Ide & Véronis, 1998) and its variations (e.g. Mihalcea, 2007).
- Type 2: methods based on similarity measures applied to semantic networks (e.g. Rada et al., 1989; Budanitsky & Hirst, 2001).
- Type 3: methods that generate selectional preferences that can be used to constrain the possible senses of an occurrence of an ambiguous word in a given context (e.g. Agirre & Martínez, 2001; Brockman & Lapata, 2003).

The LESK algorithm is arguably the most classic knowledge-based method. The main idea behind its simplest version goes as follows: ‘given an occurrence of an ambiguous word in a text and a set of sense characterizations provided by a repository, label the occurrence with a unique sense by identifying the sense characterization that has a maximal word overlap with the context of the occurrence’ (e.g. Mihalcea, 2007), where the notion of ‘context’ for some occurrence of a word is defined either quantitatively as the *n* words before and after the occurrence, with ‘*n*’ standing for some strategically chosen natural number, or qualitatively as the sentence or paragraph in which the occurrence resides. Note that this definition of ‘context’ is also exclusively used throughout this paper.

Supervised methods are essentially supervised machine learning algorithms that use sense-annotated corpora as training data, thus exclusively relying on repositories of type 4 (e.g. Màrquez et al., 2007). In order to explain conceptually how such algorithms go

[3] Two other types of WSD methods that also sometimes are considered to be knowledge-based are handcrafted disambiguation rules (e.g. Weiss, 1973) and general heuristics such as the ‘one sense per discourse’-rule (e.g. Gale et al., 1992a) (e.g. Ide & Véronis, 1998; Mihalcea, 2007; Navigli, 2009).

about, a small digression into machine learning and some of its basic concepts is required. The digression is also critical for later sections.

Machine learning is a field of research that is focused on the study and development of algorithms that allow computers to learn how to perform certain tasks without making use of preprogrammed instructions, only relying on patterns in the data that are provided for a given task. Examples of such tasks are classification, regression, clustering, anomaly detection, rule extraction and prediction (e.g. Bishop, 2006; Witten et al., 2011; Bonacorso, 2017).

The learning process always involves some kind of data-processing model. During a 'training phase' the model is (iteratively) adjusted to try and meet a set of predefined desiderata. Depending on the type of learning, the desiderata are characterized differently.

A first type of learning is called 'supervised learning'. Classic examples are classification and regression. In this case, the characterization of the desiderata is content-related: input-output data couples are provided that need to be paired correctly. More specifically, throughout the training phase, the model is tuned in such a way that, if all goes well, when presented with the input of any of these couples at the end of the phase, it correctly generates the corresponding output. Note that the data couples often presuppose strategic data preparation. The training phase normally is followed by a 'testing phase'. In this phase, new input-output data couples are used to test whether the tuned model is able to effectively generalize the learned task over new input data, i.e. whether it is able to correctly predict related output data. If so, the model can be considered trustworthy enough to be used with new, unknown, yet similar data.

A second type of learning is called 'unsupervised learning'. A good example is clustering. In this case, the characterization of the desiderata necessarily is of a more formal nature, as the model is supposed to tune itself to (latent) statistical patterns in the input data, thereby (hopefully) generating interesting insights in those same patterns. Hence, there are no preconceived input-output data pairings, only desiderata that revolve around performance and output properties, like, for example, in the case of a clustering task, a predefined number of clusters to search for or criteria for checking whether clusters are sufficiently compact and well-separated. This also means that unsupervised learning does not require a testing phase similar to the one in supervised learning, during which the generalization abilities of a model are tested.[4] An illustrative conceptual example of unsupervised learning is presented in Section 5.

[4] For the sake of clarity and completeness: there do exist cases of unsupervised learning in which a 'test phase' is integrated, yet in these cases, the test phase is used to test the accuracy of the trained data-processing model. For example, in the case of clustering, a test phase can be implemented in order to verify whether certain pre-analyzed data, known to 'belong together', are also effectively clustered together.

Returning to the original discussion above, concerning supervised machine learning algorithms that use sense-annotated corpora as training data: consider the following conceptual and (thus) very simplified example of a 'two-layered feedforward backpropagation artificial neural network' trained as a classifier that can label the occurrences of an ambiguous word with their senses (e.g. McClelland & Rumelhart, 1989; Bishop, 1995; Aggarwal, 2018).

In this example, the data-processing model is a computational network consisting of a layer of input nodes, a layer of output nodes and a weighted connection between every input and output node. The data presented to the model during the training phase are based on the provided sense-annotated corpus. More specifically, the data are the result of the corpus being strategically transformed by the application of a preprocessing protocol and a subsequent vector transformation protocol. Note that both protocols also play a vital role in the unsupervised machine learning example in Section 5.

The preprocessing protocol prepares the corpus for automatic processing. Basic processes of this protocol are (i) 'sentence boundary detection', i.e. breaking up text into sentences, (ii) 'tokenization', i.e. breaking sentences up into meaningful elements, (iii) 'part-of-speech tagging', i.e. labeling tokens to indicate the grammatical category to which they belong, (iv) 'lemmatization', i.e. reducing words to their uninflected form, and (v) 'named entity recognition', i.e. detecting proper names (e.g. Hardeniya et al., 2016). In this example, all processes except named entity recognition are carried out.

The vector transformation protocol transforms the output of the preprocessing protocol into a specific set of vectors. In this example, it generates a set of input-output data couples consisting of vector pairs. More specifically, for each annotated occurrence of the ambiguous word in the corpus, a paired input vector and output vector are created. The input vector encodes information about features of the context of the occurrence.[5] In general, a 'feature' is a piece of information or measurable property of the context and can be lexical or syntactical (Thanaki, 2017). Each dimension of the input vector corresponds with exactly one feature from the set of all features observed in the context of at least one occurrence of the ambiguous word in the corpus. Note that all input vectors thus have the same length and that this length also determines the exact number of input nodes of the data-processing model. In this example, only lexical features of the type 'unigram', i.e. individual words, are used.[6] If a word is present in the context of the

[5] To be precise, an input vector can be either a so-called 'first order vector', i.e. a vector that encodes information about the context of an occurrence, or a 'second order vector', i.e. a vector that encodes information about the contexts of the words present in the context of an occurrence (Schütze, 1998; Purandare & Pedersen, 2004).

[6] Alternatively, a lexical feature can be (i) a 'bag of words', i.e. an unordered collection of *all* the words present in the context of a given occurrence (so possibly containing some words more than once), (ii) an 'n-gram', i.e. a set of *n* contiguous words ordered in correspondence to the sequence in which they are present in the context of a given occurrence, where '*n*' stands for some strategically chosen natural number, (iii) a 'k-skip n-gram', i.e.

occurrence, the value of its corresponding dimension in the related input vector is '1', and otherwise '0'.[7] The output vector encodes the sense annotation of the occurrence. Each dimension of the output vector corresponds with exactly one sense from the set of all different senses that are used to annotate the occurrences of the ambiguous word in the corpus. Consequently, also here, all the output vectors have the same length and this length determines the exact number of output nodes of the data-processing model. The annotation of the occurrence with a specific sense is represented in the related output vector by setting the value of the dimension corresponding with this sense to '1' and the value of each other dimension to '0'.

During the training phase, the data-processing model is tuned using training data. In this example, this comes down to the following process.

1. A training set of paired input vectors and output vectors is created by randomly selecting 70 percent of the total set of pairs.
2. The weights of the connections in the network are randomly set to values within the real interval [-0.5, 0.5].
3. Next, in turn, for each pair of input and output vectors in the training set, the following process is performed.
   a. The given input vector is 'propagated' through the network: the weights of the connections form a matrix that is multiplied by the input vector, generating a response vector that, after some extra processing of its dimensional values (using a squashing function mapping values lower than or equal to 0 to '0' and otherwise to '1'), can be interpreted as 'the network trying to state which sense is used'.
4. The response vector is compared to the given output vector, and by using a so-called 'backpropagation learning rule', very specific error corrective feedback is sent back to each weight in the network, i.e. proportional to its share in the error of its corresponding output dimension, thereby pushing each weight slightly more into a direction that should lead to a better overall result.
   b. The process in step 3 is repeated until either a number of predefined iterations is reached or the average response error is below a predefined threshold.

If all goes well, the connection weights converge to settings that ensure that for all (or at least for some predefined minimum of) input vectors the right output vector is generated.

---

a variation of an n-gram where the words are not adjacent, but are maximally *k* positions apart from each other in the context of a given occurrence, where '*k*' refers to a strategically chosen natural number, or (iv) an 'unordered bigram', i.e. a set of two contiguous words (regardless of the order in which they occur) (e.g. Thanaki, 2017; Beslow, 2018).

[7] Note that it is also possible to have non-binary values for the dimensions of the vectors, for example when they represent features that are based on measurable properties that are continuous, like relative frequencies of some kind.

When this is the case, the tuning of the data-processing model is finished, i.e. the network has been trained. Next is the test phase during which the remaining 30 percent of paired input vectors and output vectors is presented to the network in order to check whether it is able to effectively generalize the learned classification task. If so, the network can be truly considered a trustworthy and effective classifier. More specifically, a classifier that uses the senses with which the given corpus is annotated to label any occurrence of the ambiguous word, even previously unseen ones.[8]

5.3 Severe limitations of word sense disambiguation methods using pre-existing sense repositories

Unfortunately, as is already widely acknowledged by the community of NLP researchers, WSD methods using pre-existing sense repositories cannot offer perfect solutions, as each of these repository types invokes one or more severe limitations, some of which are type-specific.

The use of machine-readable dictionaries limits WSD methods in two type-specific ways. First, they often contain inconsistencies. Secondly, these dictionaries generally do not provide much useful information about word co-occurrence, information that can facilitate WSD methods significantly. Ide and Véronis illustrate this with the words 'ash', 'tobacco', 'cigarette' and 'tray', contrasting dictionaries with sense-annotated corpora: while these words frequently co-occur in corpora, they are only linked indirectly in dictionaries via their respective characterizations (1998). Using thesauri also limits WSD methods in a type-specific way, as the upper levels of the concept hierarchies in thesauri are too broad to be useful. This is also mentioned by Ide and Véronis (1998).

Two other limitations are invoked when using either machine-readable dictionaries, thesauri or computational lexica. One has to do with the fact that, in practice, because of the continuously evolving character of natural languages, these three repository types will always lag behind events, not capturing extremely new or very rare senses of words, thus remaining incomplete (e.g. Pantel & Lin, 2002). This fact is commonly known as the 'knowledge acquisition bottleneck' and is first described by William Gale, Kenneth Church and David Yarowsky (1992b). The effects of the bottleneck are very striking, for example, when dealing with rather specific domains. Such domains often invoke highly specialized

[8] An interesting practical research variation on this conceptual example can be found in (Towell & Voorhees, 1998).

and thus rare senses, see also Subsection 6.2. The other limitation follows from the fact that all three types of repositories contain senses with different degrees of granularity. For example, consider the sense 'the building of a financial organization' of the ambiguous word 'bank' as discussed in the introduction. This sense of 'bank' is already pretty fine-grained. A more coarse-grained sense variant is, for example, 'a financial organization' (so without any reference to a building). When such information is present, it is often extremely hard, if not impossible for a WSD method to systematically determine the correct sense with an appropriate degree of granularity (Brody & Lapata, 2009).

Note that the last two limitations are also (indirectly) invoked when using sense-annotated corpora, as the set of sense labels used for annotating these corpora is derived from machine-readable dictionaries, thesauri or computational lexica.

5.4 Overcoming the limitations by using 'word sense induction'-based repositories?

Because of the limitations that the four types of pre-existing repositories invoke in WSD methods, leading researchers started searching for alternative methods that would not have to use these repositories. One approach that was considered very promising is based on a set of related methods developed in the context of solving a problem known as 'word sense induction' (WSI).[9] The problem of WSI can be understood best when characterized as a computational task:

> (WSI-T) 'given an ambiguous word and a text, determine which occurrences of the word in the text use a similar sense, thereby inductively discriminating the different senses of the word used in the text'.

In this paper, a method that performs WSI-T is called a 'WSI method'.

The enthusiasm of the researchers with respect to WSI methods seems to have been based on the intuition that these methods create a new kind of sense repository, a repository very much tailored to the textual context in question and thus not invoking any of the limitations of the original, pre-existing repositories when used by a WSD method. For example, the knowledge acquisition bottleneck is avoided, as even very rare

[9] In the literature the problem goes by different names. Besides 'word sense induction' (e.g. Van de Cruys & Apidianaki, 2001; Brody & Lapata, 2009; Nasiruddin, 2013), the problem is also referred to as 'sense discrimination' (e.g. Schütze, 1998), 'unsupervised word sense induction' (e.g. Denkowski, 2009), 'unsupervised lexical acquisition' (e.g. Widdows & Dorow, 2002).

senses can be inductively discriminated, and inconsistencies cannot exist (rather trivially), as the new type of repository does not incorporate any notion of negation.

The following quotes illustrate how the researchers present and discuss the new approach.

> [W]e present a unified model for the automatic induction of word senses from text, and the subsequent disambiguation of particular word instances using the automatically extracted inventory. (Van de Cruys & Apidianaki, 2011, p. 1476)

> Although the role of WSI, in a disambiguating context is to build a sense inventory that can be used subsequently for WSD, therefore, WSI can be considered as part of WSD. (Nasiruddin, 2013, p. 5)

> Word-sense induction is closely related to the word-sense disambiguation (WSD) task where the goal is to choose which meaning of a word among [those] provided in the sense inventory was used in the context. The sense inventory may be obtained by a WSI system [.] (Bartunov et al., 2016, p. 2)

> Unsupervised WSD approaches rely neither on hand-annotated sense-labeled corpora, nor on handcrafted lexical resources. Instead, they automatically induce a sense inventory from raw corpora. (Pelevina et al., 2017, pp. 2-3)

Note that, as some of the quotes clearly illustrate, both the conceptual distinction between WSI and WSD and the naming of methods following the new approach vary a lot throughout the research literature. Moreover, and arguably more importantly, also the way in which these new methods are characterized varies a lot. However, the different characterizations do share an important core idea about the (combined) problem that needs to be solved and its operational characterization as a computational task. The problem is essentially that of `induction-based word sense disambiguation' (IWSD) and its operationalization as a computational task comes down to the execution of a simple sequence of the two computational tasks discussed earlier:

> (IWSD-T) ‘first perform WSI-T, then use its sense repository output for the execution of WSD-T’.

In this paper, a method that performs IWSD-T is called an ‘IWSD method’.

Unfortunately, the aforementioned intuition, ultimately responsible for the development of IWSD methods, is based on structural and rather complex confusion within the research community both with respect to WSD-T and WSI methods. First, with respect to WSD-T, the intuition presupposes a very specific interpretation. However, there also exists another interpretation, one that historically seems to be preferred by the community. From the point of view of this second interpretation, using sense repositories created by a WSI methods introduces a serious new limitation. Secondly, with respect to WSI methods, there seems to be a tendency within the community to overestimate their potential in relation to the knowledge acquisition bottleneck. In order to understand the exact nature of all this confusion, first the roots and some operational characteristics of WSI methods need to be discussed.

5.5 Word sense induction: roots, key concepts and methods

WSI methods are rooted conceptually in the so-called ‘distributional hypothesis’ (Cecchini et al., 2017). While the hypothesis is commonly ascribed to Zelig Harris (1954), the formulation by John Rupert Firth is the most well-known: “[Y]ou shall know a word by the company it keeps [.]” (1957, p. 17).[10] Put differently, context is key: words with similar meanings occur in similar contexts (e.g. Denkowksi, 2009). For example, assuming that in everyday language the words ‘chair’ and ‘seat’ often occur in similar contexts, i.e. are ‘frequently accompanied by the same words’, and more so than ‘chair’ and ‘prime number’, it can be inferred that in everyday language the word 'chair' has a meaning more similar to the meaning of 'seat' than to 'prime number'. Checking whether the respective sets of occurrences of two words *A* and *B* are ‘frequently accompanied by the same words’ always comes down to performing some kind of comparative analysis of the ‘distributional patterns’ generated by the words in the contexts of the occurrences of *A* and *B* respectively (hence, ‘*distributional* hypothesis’). It is easy to see that a variation of this analysis can also be performed for a set of occurrences of one word, thereby generating insight into those subsets of occurrences that are accompanied by the same words and thus contain occurrences that are semantically more similar (to each other than to other occurrences). This last variation is the practical starting point of WSI-methods.

---

[10] Note that actually neither Harris nor Firth were trying to provide foundations for a computational approach to word sense disambiguation. Harris’ claim that there is a correlation between semantic similarity and distributional similarity (of morphemes) is part of his idea that it is possible to develop a complete linguistic theory that is exclusively ‘language-internal’, i.e. a theory that is exclusively based on facts about certain distributional patterns in the syntactic structures of language (Sahlgren, 2008). Such a theory can only be developed if also the phenomenon ‘meaning’ can somehow be explicated in terms of syntactic notions. Firth, on the other hand, worked on a ‘polysystemic’ theory of meaning that would encompass all levels of meaning, including ‘situational meaning’, ‘syntactical meaning’ and ‘lexical meaning’ (Oyelaran, 1967). His statement about ‘words and the company they keep’ is part of an argument for the study of lexical meaning on the basis of so-called ‘collocations’, i.e. words that often co-occur.

In practice, the WSI methods are all unsupervised machine learning algorithms. Three types of algorithms can be distinguished: ‘clustering algorithms’, ‘graph-based algorithms’ and ‘Bayesian algorithms’ (Nasiruddin, 2013). As the first type is both popular and explicative, what follows is a conceptual and thus (again) very simplified example of a so-called ‘agglomerative’ clustering algorithm used to generate a sense repository for the senses used by the occurrences of one ambiguous word in a given corpus (e.g. Day & Edelsbrünner, 1984; Duda et al., 2001; Hastie et al., 2009).[11]

In this example, the data-processing model is essentially an algorithm that iteratively analyzes distributional patterns present in data residing in a multidimensional space. As the type of learning is unsupervised, there is only a training phase and input data, i.e. vectors residing in the (same) multidimensional space. The vectors are the result of the corpus being strategically transformed by applying the two protocols already discussed in Section 2, albeit slightly modified, i.e. first the preprocessing protocol and then only the ‘input vector’-related part of the vector transformation protocol (using unigram features). Note that, as a consequence, each dimension of the space corresponds to a unique word from the corpus.

Throughout the training phase, the (latent) patterns present in the distribution are explored in order to define and iteratively refine a set of vector clusters. In this example, this comes down to the following process.

1. A metric is chosen for measuring the distance between vectors, in this case, the standard Euclidean distance function.
2. A metric is chosen for measuring the distance between two vector clusters, in this case, a function returning the maximum Euclidean distance that exists between the vectors of each cluster.[12]
3. The distances between all the given input vectors are inventoried.
4. The two closest vectors are merged into a new cluster and all remaining vectors are reinterpreted as clusters with only one vector.
5. The distances between all the clusters are compared and the two closest clusters are merged into one new cluster.

[11] The example follows an approach known as ‘local’ or ‘token-based’, in contrast to another approach called ‘global’ or ‘type-based’, see (Van de Cruys & Apidianaki, 2001; Pedersen, 2007). The (technical) details of both approaches fall outside the scope of this paper. However, it is useful to remark that the global or type-based approach is known to be helpful when mapping the output of WSI methods to pre-existing sense repositories like those discussed in Section 2, see also (Pedersen, 2007). The (dis-) advantages of such mappings are discussed in Subsection 6.1.

[12] Cluster algorithms that use this cluster distance metric perform what is called ‘complete-linkage clustering’. Other well-known variations are ‘single-linkage clustering’, where the minimum distance is used, and ‘average-linkage clustering’, based on the mean distance (e.g. Hastie et al., 2009).

6. The process in step 5 is repeated until some stop criterion is satisfied. In this case, the criterion is that, for each cluster, the distances between its vectors are minimal.[13]

As mentioned earlier, the resulting set of vector clusters can be interpreted as a sense repository for the senses used by the occurrences of one ambiguous word in a given corpus.

5.6 New (and old) limitations caused by using 'word sense induction'-based repositories: always a matter of interpretation?

It is now possible to explicate the confusion concerning WSD-T and WSI methods and the (interpretation-dependent) limitations of IWSD methods (in Subsection 6.1 and 6.2 respectively).

5.6.1 A (too) poor notion of sense?

First of all, it is crucial to keep in mind that WSI methods revolve exclusively around the gradual notion of 'similarity'. WSI methods only generate insight into the level of similarity between the senses used by occurrences of a given ambiguous word, as is clearly illustrated in Section 5. WSI methods are not intrinsically concerned with the question whether some of those used senses should be considered as 'one and the same'. Of course, it can be argued that with the use of some realistic threshold at least some levels of similarity could be re-interpreted as a sufficient indication of 'sameness'.[14]

Yet, even if this threshold strategy is acceptable, WSI methods used as part of IWSD methods still do not provide what some might argue is needed for the succeeding WSD methods, i.e. informative sense labels. This is where the confusion regarding the interpretation of WSD-T resides. In itself, the characterization of WSD-T does not specify what sense labels should incorporate, i.e. what semantic information they should contain.

[13] Another classic stop criterion is that the distance between the vectors of different clusters is maximal (e.g. El Sayed et al., 2008).
[14] It is, however, far from trivial to provide a convincing theoretical framework for setting up such 'realistic' thresholds, let alone in every possible context.

However, when considering the originally used pre-existing sense repositories, all four types provide some kind of 'deep' or 'rich' semantic information per sense, like, for example, a definition or a node in some ontological graph. So, there seems to be at least a strong historical preference in the research community to use informative sense labels. In IWSD methods, however, there are no informative sense labels. The WSI methods that they use only discriminate 'sense options', generating knowledge that there is, for example, 'some sense option *x*' that differs from other distinguishable sense options, thereby warranting a unique label for identifying *x*, yet not linking any extra information about *x* to that label, i.e. what it is about, for example, by means of some definition or ontological implementation. Note that all this is also explicitly acknowledged by several leading NLP researchers such as Hinrich Schütze and Ted Pedersen.

> While WSD is typically identified as a sense labeling task, that is, the explicit assignment of a sense label to a target word, unsupervised WSD performs word sense discrimination, that is, it aims to divide 'the occurrences of a word into a number of classes by determining for any two occurrences whether they belong to the same sense or not'. (Schütze, 1998, p. 97)

> Distributional approaches do not assign meanings to words, but rather allow us to discriminate among the meanings of a word by identifying clusters of similar contexts, where each cluster shows that word being used in a particular meaning. This is quite distinct from the traditional task of word sense disambiguation, which classifies words relative to existing senses. (Pedersen, 2007, p. 134)

The strong historical preference in the research community to use informative sense labels naturally leads to a 'rich' interpretation of WSD-T, i.e. an interpretation in which WSD-T demands informative sense labels. When not following this preference, a 'poor' interpretation of WSD-T follows naturally, i.e. an interpretation in which WSD-T simply requires non-informative sense option labels. Hence, from the point of view of the rich interpretation, IWSD methods suffer from a serious limitation: they do not 'deliver' what is originally asked for by WSD-T. From the perspective of the poor interpretation, however, there is no problem to be found.

When adopting the rich interpretation and still wanting to use IWSD methods, it is necessary to try and overcome the new limitation. Two strategies easily come to mind. One is to extend IWSD methods with some kind of 'sense information extraction protocol' that, for each sense option in the generated repository, can identify the words that are highly characteristic for its related 'distributional pattern'. Highly characteristic words can be interpreted together as (crude) sense information. However, even if this part of the protocol can be automated, which is possible for the output of many concrete WSI methods, the words it identifies as characteristic for a particular sense option still need to be manually interpreted by a representative of the relevant language group as a kind of 'judge'. He or she needs to link the words to a specific existing sense. Apart from the problematic fact that this necessity opens up the (final step of the) protocol to highly subjective decisions, it also means that it is far from scalable. In sum: this strategy is not very promising.

Another strategy is to extend IWSD methods with a protocol that systematically maps the WSI-generated sense options to entries of a pre-existing repository. This way, the final sense labels actually are informative, i.e. they refer to 'rich' or 'deep' semantic information. Unfortunately, apart from the challenge of setting up a mapping protocol that works systematically and flawlessly (to standards that are themselves hard to define), this strategy also reintroduces all the old limitations invoked by using pre-existing sense repositories discussed in Section 3. Not only that, even new limitations can arise. For example, if the number of WSI-generated sense options for some word outnumber its related sense entries in the provided pre-existing repository, it is unclear how to interpret this situation: either the WSI method created some incorrect sense options, or the WSI method identified new, potentially more refined sense options not yet present in the pre-existing repository, in which case it is not clear how to proceed, at least not in an automated manner. Again, taking all this into consideration, also this strategy is rather problematic.

5.6.2 Corpus dependency

As already discussed in Section 4, part of the initial enthusiasm with respect to IWSD methods is related to the intuition that these methods "[...] have the potential to overcome the knowledge acquisition bottleneck[.]" (Navigli, 2009, p. 10:26), because all the senses present in the given corpus are inventoried using the integrated WSI-method, even if they are rare or highly domain-specific. However, this intuition indicates that there is also some confusion with respect to WSI methods. Overcoming the knowledge acquisition bottleneck (for some language of interest) means generating a sense repository that captures all possible senses of all words (in that language). As the characterization of WSI-T in Section 4 clearly indicates, this task is always performed on the basis of a text (or 'corpus'). Hence, if a WSI method is used to try and overcome the knowledge acquisition bottleneck (for some language), it should be applied to a corpus that contains all senses of all words (in that language). This, of course, is practically impossible because of (the combination of) two reasons. First, as is illustrated by the following quotes, the WSD community estimates the likelihood of creating a 'total corpus' (for some language) as very low.

> [I]t does not seem reasonable to think that the training material is large and representative enough to cover 'all' potential types of examples. (Escudero et al., 2000, p. 173)

> [I]t seems theoretically impossible to define a training corpus and sense inventory that is truly generic (i.e. general enough to represent any possible domain). (Buitelaar et al., 2007, p. 276)

Secondly, even if such a total corpus could be created, applying a WSI method to this corpus would come with an unrealistically high computational cost.

A WSI method can only overcome the bottleneck 'locally', i.e. by generating a sense option repository that is complete with respect to a given corpus. Hence, also IWSD methods are inherently limited as they can only correctly disambiguate a new word occurrence if its sense is present in the corpus to which the integrated WSI method is applied for the generation of the sense option repository.

This limitation needs to be taken into consideration when using an IWSD method. More specifically, a type of 'domain verification protocol' is needed that can check whether a previously unseen text presented to the IWSD method for disambiguation belongs to the same 'domain' as the corpus to which the integrated WSI method is applied.[15] Existing methods for text categorization can be used to develop such a protocol. One common method involves a supervised machine learning algorithm that classifies texts into predefined domain categories (Özgür, 2002), both coarse-grained, such as 'natural science', 'applied science', 'commerce', 'life', and more fine-grained, such as 'mathematics', 'biology', 'agriculture', 'medicine', 'finance', 'industry' (Sharoff, 2004).[16]

5.7 Summary of the main findings

The way in which NLP approaches the problem of WSD, i.e. how to correctly disambiguate ambiguous words, is very interesting and promising. However, the research also regularly suffers from a lack of conceptual clarity and methodological systematicity, which explains the sometimes erratic course of the research.

In the NLP community, WSD is characterized as a computational task (WSD-T) 'given an ambiguous word and a text, label each occurrence of the word in the text with the sense it is using'. A method that performs WSD-T is a 'WSD method'. An initial problem

[15] David Lee defines 'domain' as 'broad subject field' (2001).

[16] Note that the development history of text categorization methods resembles the one of word sense disambiguation methods. Text categorization methods that use supervised machine learning algorithms require labeled training data. The creation of these data is labor-intensive. Therefore, methods that use unsupervised machine learning algorithms were developed (e.g. Steinbach et al., 1999). Yet, these methods only allow for the domain-classification of texts, and do not generate any insight into what these domains are about (Dharmadhikari et al., 2011).

widely acknowledged by the research community consists of the severe limitations of WSD methods that rely on pre-existing sense repositories like machine-readable dictionaries and thesauri. Examples of these limitations are inconsistency, incompleteness (also known as the ‘knowledge acquisition bottleneck’) and the lack of useful information about word co-occurrence.

To overcome these limitations, leading researchers started developing a new type of methods, i.e. IWSD methods. These methods integrate both a WSI method and a WSD method. The WSI method inductively discriminates the different senses of a word used in a provided text. The intuition behind this new approach is that a WSI method can create a new kind of sense repository that does not invoke any of the limitations of the original, pre-existing repositories when used by a WSD method.

However, the intuition is based on structural and rather complex confusion within the research community both with respect to WSD-T and WSI methods.

First, with respect to WSD-T, the characterization of WSD-T in itself does not specify what sense labels should incorporate, i.e. what semantic information they should contain. Yet, there is a strong historical preference to use informative sense labels, which naturally leads to a ‘rich’ interpretation of WSD-T, i.e. an interpretation in which WSD-T demands informative sense labels. When not following this preference, a ‘poor’ interpretation of WSD-T comes naturally, i.e. an interpretation in which WSD-T simply requires non-informative sense option labels. In IWSD methods there are no informative sense labels. The WSI methods that they use only discriminate ‘sense options’. Hence, from the point of view of the rich interpretation, IWSD methods suffer from a serious limitation: they do not ‘deliver’ what is originally asked for by WSD-T. From the perspective of the poor interpretation, there is no problem to be found. When adopting the rich interpretation and still wanting to use IWSD methods, it is necessary to try and overcome the new limitation. One obvious strategy is to use some kind of sense information extraction protocol, but this is not scalable. Another evident strategy is to use a protocol that systematically maps the WSI-generated sense options to entries of a pre-existing repository. But such a protocol is difficult to develop, would re-introduce the old limitations invoked by using pre-existing sense repositories and even introduces new limitations.

Secondly, with respect to the confusion concerning WSI methods, there seems to be a tendency to overestimate their potential in relation to the knowledge acquisition bottleneck. A WSI method cannot overcome the bottleneck simultaneously across all possible domains (of a language), but only ‘locally’, i.e. by generating a sense option repository that is (only) complete with respect to the given corpus. This limitation needs to be taken into consideration when using an IWSD method. More specifically, a domain verification protocol is needed that can check whether a previously unseen text presented to the IWSD method for disambiguation belongs to the same domain as the corpus to which the integrated WSI method is applied.

Funding

This research was supported by the Research Foundation – Flanders (FWO).

References

Abuzir, Y., & Vandamme, F. (2002). ThesWB: Thesaurus construction WorkBench. In Vidal, T., & Liberatore, P. (Eds.), *Starting Artificial Intelligence Researchers Symposium (STAIRS) Proceedings* (pp. 95-104). The Netherlands, Amsterdam: IOS Press.

Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*. Springer.

Agirre, E., & Edmonds, P. (2007). Introduction. In Agirre, E., & Edmonds, P. (Eds.), Text, Speech and Language Technology (Vol. 33), *Word Sense Disambiguation: Algorithms and Applications* (pp. 1-29). Springer.

Agirre, E., & Martínez, D. (2001). Learning Class-to-class Selectional Preferences. In *Proceedings of the 2001 Workshop on Computational Natural Language Learning-Volume 7* (p. 3). Association for Computational Linguistics.

Bartunov, S., Kondrashkin, D., Osokin, A., & Vetrov, D. (2016). Breaking Sticks and Ambiguities with Adaptive Skip-gram. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS)* (pp. 130-138).

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. New York: Oxford University Press.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag.

Beslow, T. (2018). *Applied Natural Language Processing with Python. Implementing Machine Learning and Deep Learning Algorithms for Natural Language Processing*. San Francisco, California: Apress.

Bonaccorso, G. (2017). *Machine Learning Algorithms. A Reference Guide to Popular Algorithms for Data Science and Machine Learning*. Birmingham, Mumbai: Packt Publishing.

Brody, S., & Lapata, M. (2009). Bayesian Word Sense Induction. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 103-111). Association for Computational Linguistics.

Budanitsky, A., & Hirst, G. (2001). Semantic Distance in WordNet: An Experimental, Application-oriented Evaluation of Five Measures. In *Workshop on WordNet and Other Lexical Resources* (Vol. 2, pp. 2-2).

Buitelaar, P., Magnini, B., Strapparava, C.,& Vossen, P. (2007). Domain-Specific WSD. In Agirre, E., & Edmonds, P. (Eds.), Text, Speech and Language Technology (Vol. 33), *Word Sense Disambiguation: Algorithms and Applications* (pp. 275-299). Springer.

Brockmann, C., & Lapata, M. (2003). Evaluating and Combining Approaches to Selectional Preference Acquisition. In *Proceedings of the 10th Conference on European Chapter of the Association for Computational Linguistics* (Vol 1, pp. 27-34). Association for Computational Linguistics.

Cecchini, F. M., Riedl, M., & Biemann, C. (2017). Using Pseudowords for Algorithm comparison: an Evaluation Framework for Graph-based Word Sense Induction. In *Proceedings of the 21st Nordic Conference on Computational Linguistics (NoDaLiDa)* (No. 131, pp. 105-114). Linköping University Electronic Press.

Day, W. H. E., & Edelsbrünner, H. (1984). Efficient Algorithms for Agglomerative Hierarchical Clustering Methods. *Journal of Classification*, *1*(1), 7-24.

Denkowski, M. (2009). A Survey of Techniques for Unsupervised Word Sense Induction. *Language & Statistics II Literature Review*, 1-18.
Dharmadhikari, S. C., Ingle, M., & Kulkarni, P. (2011). A Comparative Analysis of Supervised Multi-label Text Classification Methods. *Advanced Computing: An International Journal*, *2*(6), 161-169.
Dias-Da-Silva, B. C. (2010). Brazilian Portuguese WordNet: a Computational-linguistic Exercise of Encoding Bilingual Relational Lexicons. *International Journal of Computational Linguistics and Applications*, *1*(1-2), 137-150.
Diekema, A. R. (2003). Translation Events in Cross-language Information Retrieval: Lexical Ambiguity, Lexical Holes, Vocabulary Mismatch, and Correct Translations. (Doctoral dissertation, Syracuse University). Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.75.141.
Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2nd ed.). New York: John Wiley.
El Sayed, A., Velcin, J., & Zighed, D. (2008). Word Clustering with Validity Indices. In *Conference of the Canadian Society for Computational Studies of Intelligence* (pp. 259-270). Berlin, Heidelberg: Springer.
Escudero, G., Màrquez, L., & Rigau, G. (2000). An empirical study of the domain dependence of supervised word sense disambiguation systems. In *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics* (Volume 13, pp. 172-180). Association for Computational Linguistics.
Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*, [special volume of the Philological Society]. Oxford: Basil Blackwell.
Gale, W. A., Church, K. W., & Yarowsky, D. (1992a). One Sense per Discourse. In *Proceedings of the workshop on Speech and Natural Language* (pp. 233-237). Association for Computational Linguistics.
Gale, W. A., Church, K. W., & Yarowsky, D. (1992b). A Method for Disambiguating Word Senses in a Large Corpus. *Computers and the Humanities*, *26*(5-6), 415-439.
Gale, W. A., & Church, K. W. (1993). A Program for Aligning Sentences in Bilingual Corpora. Computational Linguistics, 19(1), 75-102.
Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., & Mathur, I. (2016). Natural Language Processing: Python and NLTK. Learn to Build Expert NLP and Machine Learning Projects Using NLTK and Other Python Libraries. Birmingham, Mumbai: Packt.
Harris, Z. S. (1954). Distributional structure. Word, 10(2-3), 146-162.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Data Mining, Inference and Prediction (2nd edition). Springer Series in Statistics. New York: Springer.
Ide, N., & Véronis, J. (1998). Introduction to the special issue on word sense disambiguation: the state of the art. Computational Linguistics, 24(1), 2-40.
Kapetanios, E., Tatar, D., & Sacarea, C. (2013). Natural Language Processing: Semantic Aspects. CRC Press.
Lefever, E., Hoste, V., & De Cock, M. (2011). Parasense Or How to Use Parallel Corpora for Word Sense Disambiguation. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2 (pp. 317-322). Association for Computational Linguistics.
Lesk, M. (1986). Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. In Proceedings of the 5th Annual International Conference on Systems Documentation (pp. 24-26). ACM.
Màrquez, L., Escudero, G., Martínez, D., & Rigau, G. (2007). Supervised Corpus-Based Methods for WSD. In Agirre, E., & Edmonds, P. (Eds.), Text, Speech and Language Technology (Vol. 33), Word Sense Disambiguation: Algorithms and Applications (pp. 167-217). Springer.

Martínez, D., de Lacalle, O. L., & Agirre, E. (2008). On the Use of Automatically Acquired Examples for All-nouns Word Sense Disambiguation. Journal of Artificial Intelligence Research, 33, 79-107.

McLelland, J.L., & Rumelhart, D. E. (1989). Explorations in Parallel Distributed Processing. A Handbook of Models, Programs and Exercises. Cambridge, London: MIT Press.

Mihalcea, R. (2007). Knowledge-based methods for WSD. In Agirre, E., & Edmonds, P. (Eds.), Text, Speech and Language Technology (Vol. 33), *Word Sense Disambiguation: Algorithms and Applications* (pp. 107-133). Springer.

Nasiruddin, M. (2013). A state of the art of word sense induction: A way towards word sense disambiguation for under-resourced languages. *arXiv preprint arXiv:1310.1425*.

Navigli, R. (2009). Word Sense Disambiguation: A survey. *ACM Computing Surveys (CSUR)*, *41*(2), 1-69.

Oyelaran, O. O. (1967). Aspects of Linguistic Theory in Firthian Linguistics. *Word*, *23*(1-3), 428-452.

Özgür, A. (2004). *Supervised and Unsupervised Machine Learning Techniques for Text Document Categorization* (Unpublished Master's Thesis). Boğaziçi University.

Pantel, P., & Lin, D. (2002). Discovering Word Senses From Text. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 613-619). ACM.

Pedersen, T. (2007). Unsupervised Corpus-based Methods for WSD. In Agirre, E., & Edmonds, P. (Eds.), Text, Speech and Language Technology (Vol. 33), *Word Sense Disambiguation: Algorithms and Applications* (pp. 133-167). Springer.

Pelevina, M., Arefyev, N., Biemann, C., & Panchenko, A. (2017). Making sense of word embeddings. *arXiv preprint arXiv:1708.03390*.

Purandare, A., & Pedersen, T. (2004). Word Sense Discrimination by Clustering Contexts in Vector and Similarity Spaces. In *Proceedings of the 8th Conference on Computational Natural Language Learning (CoNLL-2004) at HLT-NAACL 2004*.

Rada, R., Mili, H., Bicknell, E., & Blettner, M. (1989). Development and Application of a Metric on Semantic Nets. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(1), 17-30.

Sahlgren, M. (2008). The Distributional Hypothesis. *Italian Journal of Linguistics*, *20*(1), 33–53.

Sharoff, S. (2004). Towards Basic Categories for Describing Properties of Texts in a Corpus. In Lino, M. T., Xavier, M. F., Ferreira, F., Costa, R., Silva, R., Pereira, C., Carvalho, F., Lopes, M., Catarino, M., & Barros, S. (Eds.), *LREC04: Proceedings of Language Resources and Evaluation Conference* (Volume V, pp. 1743–1746).

Steinbach, M., Karypis, M. S. G., & Kumar, V. (2000). A Comparison of Document Clustering Techniques. In *TextMining Workshop at KDD2000*.

Stevenson, M., & Wilks, Y. (2003). Word Sense Disambiguation. In Mitkov, R. (Ed.), *The Oxford Handbook of Comp. Linguistics* (pp. 249-265). Oxford University Press.

Thanaki, J. (2017). *Python Natural Language Processing. Explore NLP with Machine Learning and Deep Learning Techniques.* Birmingham, Mumbai: Packt Publishing.

Towell, G., & Voorhees, E. (1998). Disambiguating Highly Ambiguous Words. *Computational Linguistics*, *24*(1), 125-145.

Van de Cruys, T., & Apidianaki, M. (2011). Latent Semantic Word Sense Induction and Disambiguation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (Vol 1, pp. 1476-1485). Association for Computational Linguistics.

Vial, L., Lecouteux, B., & Schwab, D. (2017). *UFSAC: Unification of Sense Annotated Corpora and Tools* (Doctoral dissertation, UGA-Université Grenoble Alpes). Retrieved from: https://hal.archives-ouvertes.fr/hal-01680739

Weiss, S. F. (1973). Learning to Disambiguate. *Information Storage and Retrieval*, *9*(1), 33-41.

Widdows, D., & Dorow, B. (2002). A Graph Model for Unsupervised Lexical Acquisition. In *Proceedings of the 19th International Conference on Computational Linguistics* (Vol 1, pp. 1-7). Association for Computational Linguistics.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining. Practical Machine Learning Tools and Techniques* (3th ed.). United States: Morgan Kaufman.

Wu, C. (2009). Corpus-based Research. In Halliday, M.A.K. & Webster, J.J. (Eds.), *Continuum Companion to Systemic Functional Linguistics* (pp. 128-142), London, New York: Continuum International Publishing Group.

Yarowsky, D. (1993, March). One Sense Per Collocation. In *Proceedings of the Workshop on Human Language Technology* (pp. 266-271). Association for Computational Linguistics.

Epilogue to Part II

In Part II, a PAT resolution procedure called 'SenseDisclosure' was outlined and elaborate proposals were made regarding its implementation. One of the proposals is to use WSI techniques for the automation of a crucial step of the procedure. The possibility of using WSI techniques for this purpose was substantiated experimentally and the potential and limitations of WSI techniques were further explored by means of a philosophical analysis. What follows are some reflections on the research that is contained in Part II.

First, while transdisciplinarity fell out of the scope of Part I, it was covered in Part II. After all, the design of SenseDisclosure is such that, once fully developed, it could be used by any cross-disciplinary team, including transdisciplinary ones. The only precondition is that sufficient text material is available for each source discipline (which might not always be the case for non-academic disciplines).

Secondly, the philosophical analysis of WSI techniques that is carried out in Paper 5 shows that some parts of Paper 4 should be re-phrased. More specifically, Paper 5 shows that, given their roots in the structuralist distributional hypothesis, WSI techniques can be used to *discriminate* the different meanings (or senses) of a term, but not to gain insight into what these senses *are*. Thus, the clusters that are generated by means of the techniques are *correlated* with a sense of a term, but cannot be equated with it, and, the (most significant) words that occur in a cluster are, at best, pointers to (the components of) senses. However, in Papers 3 and 4 this limitation of WSI techniques is not always clearly reflected. Consider the following fragments:

> Here, the word clouds are generated based on the concepts (viz. the collections of concept indicative words) that were identified by means of word sense induction in step four. (Section 3.5)

> [T]he sixth step deals with the generation of the visual representations of the meanings of each polysemous shared jargon term that was identified as a PAT. In this brief excursion, we test whether the classes of test sentences generated by SenseClusters can be used to visualize the meaning of the target terms corresponding to those classes. (Section 4.4.3)

To be completely in line with the results of the critical analysis in Paper 5, and thus, more correct, the fragments should be rephrased as follows:

> Here, the word clouds are generated based on the *clusters of co-occurring terms* that were identified by means of word sense induction in step four.
>
> [T]he sixth step deals with the generation of the visual representations *for* the meanings of each polysemous shared jargon term that was identified as a PAT. In this brief excursion, we test whether the classes of test sentences generated by SenseClusters can be used to create visual *pointers to* the meaning of the target terms that *are correlated* with those classes.

Finally, something more needs to be said about the workings and potential of SenseDisclosure. SenseDisclosure is designed to enable PAT resolution by means of three functionalities, i.e. the facilitation of (i) the detection of PATs, (ii) the discrimination of the different concepts behind a PAT, and (iii) the creation of a shared interpretation standard for a PAT. Yet, it is highly likely that SenseDisclosure also contributes to PAT resolution in another, unintended way. When a team decides to use SenseDisclosure, this creates an opportunity for the members to ask each other questions that otherwise might be taboo, e.g. clarificatory questions about key terms of the project. In other words, SenseDisclosure might not only work as a semi-automated tool, but also as a conversation catalyst.

General conclusions and future research

Conclusions

The conclusions of the presented research have been discussed in detail in the conclusions of the included papers. Below the main findings are summarized:

- The received framework for discipline-combining research is ambiguous because 'interdisciplinarity' serves both as the umbrella term and as the name for a specific type of discipline-combining research. The proposed resolution for this ambiguity is the introduction of a new umbrella term, for example 'cross-disciplinarity'.
- The effectiveness of (funding) initiatives that are set up to support cross-disciplinary research can be increased by using the unambiguous framework for cross-disciplinarity when defining the target audience and the selection criteria of the initiative.
- If research policy makers, funding bodies and researchers were to develop an unjustified preference for interdisciplinarity over multidisciplinarity, this would be harmful for the practice of cross-disciplinarity.
- Contrary to what is suggested by the received theory of cross-disciplinarity, multidisciplinarity is a full-fledged and valuable type of cross-disciplinarity.
- Cross-disciplinary teams require a (semi-)automated procedure in order to systematically detect and resolve problematically ambiguous terms (PATs).
- SenseClusters, an implementation of a word sense induction (WSI) technique, shows great potential for automating a crucial step of a PAT resolution procedure, i.e. the detection of terms that have multiple meanings (or, are polysemous) across different disciplinary jargons.
- WSI techniques cannot be used to gain insight into what the different meanings of a term are, but do allow to discriminate senses and to find pointers to what these senses are.

As was indicated in the introduction, some of these findings were obtained by means of an interdisciplinary approach that combines philosophy of science and computational linguistics. Now that all the research has been presented, it can be explained how knowledge and techniques from the two disciplines were integrated.

In high-level terms, the integration can be summarized as the application of natural language processing (NLP) techniques (i.e. term extraction techniques and WSI techniques) to a philosophy of science research problem (viz. communication across disciplines). This application led to an alternative use of the techniques and necessitated a rephrasing of the research problem. In practice, the integration was more complex and took place in different stages of the research reported on in Part II.

In the first stage, the central research problem of Part II, i.e. PATs, was conceptualized. This phase was purely philosophical. It started with the reading of publications by philosophers of science and philosophically inclined scientists on differences between disciplinary “languages” and “jargons” and how these differences impede cross-disciplinary collaboration (e.g. Wear, 1999; Bracken and Oughton, 2006; Choi & Pak, 2007; Monteiro & Keating, 2009). By analyzing some concrete examples of misunderstandings caused by language-related differences, it became clear that many of them resulted from terminological ambiguity (e.g. File & Dugard, 1997; Bracken & Oughton, 2006; Ranade et al., 2011). Based on that observation, the concept of PAT was created. It was defined as ‘a term that can cause communication problems because it plays a key role in a given project and is ambiguous’.

The second stage of research revolved around the search for an effective tool for PAT resolution. Initially, this phase was also purely philosophical as it began with an evaluation of the potential of existing tools for the facilitation of cross-disciplinary communication (on a conceptual level). For this evaluation new sets of task and performance criteria for a good PAT resolution tool were developed. None of the existing tools met all criteria. Thus, in order to provide an effective PAT resolution tool, it would have to be newly created. It also became clear that in order for the new tool to meet all task and performance criteria, it would have to be (semi-) automated. Methods that allow for the necessary automation (of dealing with terms) cannot be provided by philosophy of science, but can be found in the field of NLP. Yet, the application of NLP techniques comes with some prerequisites. For example, written text is required that can serve as input and, if NLP techniques are to be used to resolve PATs, the notion of PAT needs to be operationalized in linguistic terms. In other words, in the outline of the new PAT resolution procedure called ‘SenseDisclosure’, the philosophical desiderata had to be integrated with these NLP prerequisites. To accommodate the requirement for written text, the first step of SenseDisclosure was dedicated to the compilation of a corpus that represents the jargons of the source disciplines of a given project. Note that this step would never have been included in the outline if the PAT resolution procedure were strictly philosophical. Regarding the operationalization of the notion PAT, PATs were redefined as a subset of the ‘terms that occur in at least two of the corpora that represent the jargon of one of the disciplines involved and are polysemous’.

In the third stage of research, it was assessed whether SenseClusters, i.e. a software package that implements the word sense induction (WSI) technique 'context clustering', shows potential for becoming a functional part of SenseDisclosure. As a functional part, SenseClusters would have to determine whether a term is polysemous across disciplinary jargons. The assessment of the potential of SenseClusters consisted in two series of experiments that were set up to determine whether the software package could meet a fundamental precondition for being able to detect terms that are polysemous across jargons. While these experiments were of a computational linguistic nature, the overarching philosophical research problem did affect the methodology of the experiments. For example, the output of SenseClusters was evaluated in a slightly unorthodox way (from an NLP perspective). At the core of SenseClusters, there is an unsupervised machine learning algorithm. The output of such algorithms is generally evaluated by checking whether test data (i.e. new data that had not been used for the training of the algorithm) are processed correctly (i.e. as prescribed by the gold standard). In one series of experiments, these test data were hand-picked (while they usually are randomly selected). Moreover, in both series of experiments, the gold standard was designed by one person (while usually, multiple people are consulted). These deviations from the standard evaluation practice were justified because the goal of the experiments was not to determine exactly how well SenseClusters performs a certain task, but to establish that SenseClusters can perform a certain task.

In the fourth and final research phase, WSI techniques were subjected to a critical analysis that shed light on the limits of these techniques. While the analysis was philosophical, the computational linguistic literature on WSI did not merely serve as a case study in the development or illustration of a philosophical theory. Instead, philosophical techniques were used to explicate some central (methodological) assumptions underlying WSI techniques. The results of the analysis cannot be plainly categorized as philosophical or linguistic, but instead, are hybrid.

In sum, throughout three research stages, knowledge and techniques from philosophy of science and computational linguistics (NLP to be precise) were integrated. In some stages the former discipline was predominant (e.g. in the characterization of a PAT resolution tool), in other the latter took the lead (e.g. in assessing the potential of WSI techniques). Throughout the three stages, integration was achieved by means of iteration, which is considered a hallmark of integration (see Klein, 2008; Bammer, 2013). For example, the notion of PATs is first defined in philosophical terms, and later further operationalized in NLP terms. On a higher level, another example of iteration can be found. A philosophical analysis of the needs of cross-disciplinary teams resulted in the proposal to use NLP techniques as part of a PAT resolution procedure. The first step in the development of this procedure consisted in NLP experiments aimed at assessing the feasibility of the proposed use of the techniques. These experiments revealed the need for a philosophical analysis of the limitations of the techniques. The results of this analysis, in turn, had consequences for the outlined PAT resolution procedure.

Future research

Cross-disciplinarity may already be heavily promoted but there still is a lot of work to be done with respect to the facilitation of cross-disciplinary research. The following ideas for future research emerge directly from this dissertation.

First, concepts of transdisciplinarity need to be streamlined and further developed. Currently, there are two competing concepts of transdisciplinarity. One concept is defined as ‘involving more extensive integration than interdisciplinarity’ (up to the point that the source disciplines can no longer be identified); the other as ‘involving input from non-academics’ (Pohl et al., 2011). In order to overcome the theoretical divide, it needs to be assessed whether both concepts are useful. If so, a new term should be assigned to one of them (i.e. to avoid that ‘transdisciplinarity’ identifies two concepts and thus, is ambiguous).[1] Once a decision is made about what concept(s) of transdisciplinarity are useful, and each concept is identified by a unique term, a rich definition needs to be developed for each concept to complete the theory of cross-disciplinarity.

Another direction for future research that follows from this dissertation is the further development of SenseDisclosure, which requires more research. First, more experiments with SenseClusters need to be performed in order to assess whether, besides meeting the necessary precondition, it also allows to detect terms that are polysemous across disciplinary jargons (i.e. step four of SenseDisclosure). In these experiments, the cluster stopping function of SenseClusters needs to be tested as it needs to be checked whether it allows to detect the number of meanings a term has across disciplinary jargons. Moreover, it should be experimentally determined whether the inclusion of lemma’s and part-of-speech information in the context vectors allows to improve the performance of SenseClusters. Second, the potential of other NLP techniques, such as graph based clustering, Bayesian word sense induction or word embeddings needs to be explored. Third, on a more theoretical level, more research needs to be done into the causes of terminological ambiguity in cross-disciplinary teams: compared to differences between disciplinary jargons, to what extent do differences between jargons and everyday

[1] In this dissertation, it was assumed that transdisciplinarity, defined as ‘involving non-academic input’, constitutes the most useful complement to multi- and interdisciplinarity because transdisciplinarity as ‘highly integrative’ could be considered an edge case of interdisciplinarity.

language create PATs? In case everyday language turns out to play an important role, a way should be found to also represent it in a corpus (one way could be to use a general corpus such as the Corpus of Contemporary American English). Finally, more insight is required into which type of concept representation is most useful in the process of negotiating a new, shared interpretation standard for a PAT. Are these indeed word clouds (i.e. collections of words that frequently co-occur and are associated with a concept), or is something more required (e.g. insight into the semantic relations between co-occurring words)? How could these representations be created?

Besides research that follows from this dissertation, there are also some gaps and missing links in the literature on cross-disciplinarity that need to be taken care of.

First, a concept that is at the heart of the research literature on cross-disciplinarity, viz. integration, needs to be further analyzed. Because the application contexts of the concept are highly diverse (ranging from practical to theoretical contexts and from descriptive to normative contexts), its analysis requires a pluralistic approach. Instead of looking for one general definition, it is more appropriate to develop sets of definitions, or typologies. Some important results have already been achieved in this direction. For example, O'Rourke et al. (2016) provided a list of dimensions on which types of integration can be distinguished. The main dimensions (for some of which, also subdimensions are defined), are the following: (i) input (i.e. what is being integrated), (ii) output (i.e. the result of integration), (iii) process (i.e. the combinational changes that are made to the input), (iv) parameters (i.e. the scale of integration, the commensurability of the input, the comprehensiveness of the output). MacLeod and Nagatsu (2018) used most of these dimensions in their typology of model integration in environmental sciences. To get more insight into the forms integration can take, more dimensions should be defined (e.g. the motivation for integration, the degree to which the input of the source disciplines is balanced, the time and budget that are required to achieve the integration) and more typologies should be created (e.g., for other fields, for other input or output categories, etc.).

Second, the issue of what the building blocks of cross-disciplinarity are needs to be addressed. The term 'cross-disciplinarity' suggests that these blocks are disciplines, but this raises the question of how 'discipline' should be defined. Yet, disciplines are not the only candidate building blocks. Some researchers have argued that cross-disciplinarity is about combining knowledge from different 'scientific communities' (Lélé & Norgaard, 2005) or 'fields' (Huutoniemi et al., 2010). Both notions need to be further characterized and operationalized, and more research needs to be done into the advantage(s) of considering them as the building blocks for cross-disciplinarity.

Third, there is an implicit assumption in the literature on cross-disciplinarity that needs to be addressed. This assumption is that the cross-disciplinary nature of a project can always be captured by a single label (i.e. 'inter-', 'multi-' or 'transdisciplinary'). The set of decision rules that is used to determine what label is most appropriate appears to be the following:

1. If a project involves more than one discipline, it is cross-disciplinary (O'Rourke and Crowley, 2013).
2. If a cross-disciplinary project involves knowledge or techniques from a non-academic discipline, it is transdisciplinary (Defila and Di Giulio, 1999).
3. If a cross-disciplinary (but non-transdisciplinary) project is not split up into discrete work packages, or if it involves a significant amount of knowledge integration, it is interdisciplinary (Huutoniemi et al., 2010; Eigenbrode et al., 2007).
4. If a project is cross-disciplinary but not transdisciplinary nor interdisciplinary, it is multidisciplinary.

Yet, projects almost never integrally belong to one and the same type of cross-disciplinarity. They consist of different phases (which are often further split up into work packages) that are not necessarily all cross-disciplinary, let alone of the same type of cross-disciplinarity. Hence, more research is required into the possible combinations of (mono-disciplinarity and) different types of cross-disciplinarity.

Last, but perhaps most importantly, in order to enable *strategic* facilitation of cross-disciplinarity, more insight is required into the reason(s) why cross-disciplinary research is performed. There already exist some inventories of motivations for cross-disciplinarity (e.g. Barry, 2008; Huutoniemi et al., 2010), but more needs to be done. For example, insight is required into which parties (e.g. policy makers, funding bodies, researchers) invoke which reasons, and how these parties influence each other (e.g. do policy makers listen to researchers, or do researchers try to fulfill the wishes of policy makers?).

References

Apostel, L., Berger, G., Briggs, A., & Michaud, G. (1972). *Interdisciplinarity. Problems of Teaching and Research in Universities*. Paris: Organization for Economic Cooperation and Development.

Bammer, G. (2013). *Disciplining interdisciplinarity: Integration and implementation sciences for researching complex real-world problems*. Canberra: ANU E Press.

Barlow, J., Ewers, R. M., Anderson, L., Aragao, L. E., Baker, T. R., Boyd, E., Feldpaush, T. R., Gloor, E., Hall, A., Malhi, Y., Milliken, W., Mulligan, M., Parry, L., Pennington, T., Peres, C. A., Phillips, O. L., Roman-Cuesta, R. M., Tobias, J. A., & Gardner, T. A. (2011). Using learning networks to understand complex systems: a case study of biological, geophysical and social research in the Amazon. *Biological Reviews*, *86*(2), 457-474.

Barry, A., Born, G., & Weszkalnys, G. (2008). Logics of interdisciplinarity. *Economy and Society*, *37*(1), 20-49.

Benda, L. E., Poff, L. N., Tague, C., Palmer, M. A., Pizzuto, J., Cooper, S., Stanley, E., & Moglen, G. (2002). How to avoid train wrecks when using science in environmental problem solving. *BioScience*, *52*(12), 1127-1136.

Bracken, L. J., & Oughton, E. A. (2006). 'What do you mean?' The importance of language in developing interdisciplinary research. *Transactions of the Institute of British Geographers*, *31*(3), 371-382.

Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature*, *534*(7609), 684-687.

Bruun, H., Hukkinen, J. I., Huutoniemi, K. I., & Thompson Klein, J. (2005). *Promoting interdisciplinary research: The case of the Academy of Finland*. Helsinki: Academy of Finland. Retrieved from: https://core.ac.uk/download/pdf/157585961.pdf.

Choi, B. C., & Pak, A. W. (2007). Multidisciplinarity, interdisciplinarity, and transdisciplinarity in health research, services, education and policy: 2. Promotors, barriers, and strategies of enhancement. *Clinical & Investigative Medicine*, *30*(6), 224-232.

Defila, R., & Di Giulio, A. (1999). Evaluating transdisciplinary research. *Panorama: Swiss Priority Programme Environment*, *1*, 29, 4-27. Retrieved from: www.ikaoe.unibe.ch/forschung/ip/Specialissue.Pano.1.99.pdf.

Duget, P. (1972). Approach to the problems. In Apostel, L., Berger, G., Briggs, A., & Michaud, G. (Eds.), *Interdisciplinarity. Problems of Teaching and Research in Universities* (pp. 11-23). OECD.

Eigenbrode, S. D., O'Rourke, M., Wulfhorst, J. D., Althoff, D. M., Goldberg, C. S., Merrill, K., Morse, W., Nielsen-Pincus, M., Stephens, J. Winowiecki, L., & Bosque-Pérez, N. A. (2007). Employing philosophical dialogue in collaborative science. *BioScience*, *57*(1), 55-64.

Ferrari, A., Esuli, A., & Gnesi, S. (2018). Identification of cross-domain ambiguity with language models. In *2018 5th International Workshop on Artificial Intelligence for Requirements Engineering (AIRE)* (pp. 31-38). IEEE.

File, P. E., & Dugard, P. I. (1997). Misunderstanding between knowledge engineers and domain experts: A re-evaluation of the use of induction for medical expert systems. *International Journal of Medical Informatics*, *46*(2), 113-118.

Holbrook, J. B. (2013). What is interdisciplinary communication? Reflections on the very idea of disciplinary integration. *Synthese*, *190*(11), 1865-1879.

Huutoniemi, K., Klein, J. T., Bruun, H., & Hukkinen, J. (2010). Analyzing interdisciplinarity: Typology and indicators. *Research Policy*, *39*(1), 79-88.

Irani, Z. (2018, January 24). The University of the Future will be Interdisciplinary. *The Guardian*. Retrieved from: https://www.theguardian.com/higher-education-network/2018/jan/24/the-university-of-the-future-will-be-interdisciplinary.

Jantch, E. (1972). Towards Interdisciplinarity and Transdisciplinarity in Education and Innovation. In Apostel, L., Berger, G., Briggs, A., & Michaud, G. (Eds.), *Interdisciplinarity. Problems of Teaching and Research in Universities* (pp. 97-121). OECD.

Klein, J.T. (1990). *Interdisciplinarity. History, Theory, and Practice*. Detroit: Wayne State University Press.

Klein, J. T. (2008). Evaluation of interdisciplinary and transdisciplinary research: a literature review. *American Journal of Preventive Medicine*, *35*(2), S116-S123.

Klein, J. T. (2010). A taxonomy of interdisciplinarity. In Frodeman, R., Klein, J.T., & Mitcham, C. (Eds.), *The Oxford handbook of interdisciplinarity* (pp. 15-30). Oxford: Oxford University Press.

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and language*, *81*(1-3), 205-223.

Kockelmans, J.J. (1979). Science and discipline: some historical and critical reflections, and Why interdisciplinarity? In Kockelmans, J.J. (Ed.), *Interdisciplinarity and higher education* (pp. 123-60). University Park: Pennsylvania University Press.

Kostoff, R. N. (2002). Overcoming specialization. *BioScience*, *52*(10), 937-941.

Lattuca, L. R. (2001). *Creating interdisciplinarity: Interdisciplinary research and teaching among college and university faculty*. Nashville: Vanderbilt University Press.

Lawrence, R. & Després, C. (Eds.) (2004). Transdisciplinarity [Special Issue]. Futures. *36*(4).

Lélé, S., & Norgaard, R. B. (2005). Practicing interdisciplinarity. *BioScience*, *55*(11), 967-975.

Lowe, P., Phillipson, J., & Wilkinson, K. (2013). Why social scientists should engage with natural scientists. *Contemporary Social Science*, *8*(3), 207-222.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language*, *32*(5), 692-715.

MacLeod, M. (2018). What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice. *Synthese*, *195*(2), 697-720.

MacLeod, M., & Nagatsu, M. (2018). What does interdisciplinarity look like in practice: Mapping interdisciplinarity and its limits in the environmental sciences. *Studies in History and Philosophy of Science Part A*, *67*, 74-84.

Mayville, W. (1978). Interdisciplinarity: The Mutable Paradigm. AAHE-ERIC / Higher Education Research Report No. 9. Washington D.C.: AAHE.

MIT (n.d.) Innovation and Collaboration. Retrieved from: http://web.mit.edu/mit2030/themes/innovation-collaboration.html.

Mobjörk, M. (2010). Consulting versus participatory transdisciplinarity: A refined classification of transdisciplinary research. *Futures*, *42*(8), 866-873.

Monteiro, M., & Keating, E. (2009). Managing misunderstandings: The role of language in interdisciplinary scientific collaboration. *Science Communication*, *31*(1), 6-28.

Morse, W. C., Nielsen-Pincus, M., Force, J. E., & Wulfhorst, J. D. (2007). Bridges and barriers to developing and conducting interdisciplinary graduate-student team research. *Ecology & Society*, *12*(2), 1-14.

National Academy of Sciences (NAS), Committee on Facilitating Interdisciplinary Research and Committee on Science Engineering and Public Policy (2005). *Facilitating Interdisciplinary Research*. Washington D.C.: National Academies Press.

NSF (n.d.) Introduction to Interdisciplinary Research. Retrieved from: https://www.nsf.gov/od/oia/additional_resources/interdisciplinary_research/

O'Rourke, M., & Crowley, S. J. (2013). Philosophical intervention and cross-disciplinary science: the story of the Toolbox Project. *Synthese*, *190*(11), 1937-1954.

O'Rourke, M., Crowley, S., & Gonnerman, C. (2016). On the nature of cross-disciplinary integration: A philosophical framework. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *56*, 62-70.

Pohl, C. (2011). What is progress in transdisciplinary research?. *Futures*, *43*(6), 618-626.

Ranade SM, Salazar H and Rodriguez LA (2011) Back to Basics: Relearning terms, concepts for process control. Retrieved from: http://www.plantengineering.com.

Stember, M. (1991). Advancing the social sciences through the interdisciplinary enterprise. *The Social Science Journal*, 28, 1–14.

Tuggy, D. (1993). Ambiguity, polysemy, and vagueness. *Cognitive Linguistics (includes Cognitive Linguistic Bibliography)*, *4*(3), 273-290.

Wear, D. N. (1999). Challenges to interdisciplinary discourse. *Ecosystems*, *2*(4), 299-301.

Weingart, P. (2010). A Short History of Knowledge Formations. In Frodeman, R., Klein, J. T., & Mitcham, C. (Eds.), *The Oxford Handbook of Interdisciplinarity* (pp. 3-15). Oxford: Oxford University Press.

Wilthagen, T., Aarts, E. H., & Valcke, P. (2018). A Time for Interdisciplinarity-An Essay on the Added Value of Collaboration for Science, University, and Society. Retrieved from: https://ssrn.com/abstract=3289727.

Zhang, Q. (1998). Fuzziness-vagueness-generality-ambiguity. *Journal of Pragmatics*, *29*(1), 13-31.

Summary

Loosely speaking, cross-disciplinarity involves the combining of knowledge or techniques from different disciplines. Based on the kind of disciplines from which knowledge is drawn (i.e. only academic disciplines, or also practical fields) and the way in which knowledge is combined (i.e. by means of integration or not), three types of cross-disciplinarity are distinguished: 'interdisciplinarity' stands for the integration of knowledge from academic disciplines; 'multidisciplinarity' for the combining of knowledge from academic disciplines, but not by means of integration; and 'transdisciplinarity' stands for the integration of knowledge from academic and non-academic disciplines.

In the past decades, cross-disciplinary research has become more and more popular among research policy makers, funding bodies and researchers. An important motor behind this increased popularity is a firm belief in the innovative potential that combinations of disciplines offers for achieving scientific breakthroughs and addressing complex (societal) problems.

While the (potential) results of cross-disciplinary research may be much desired, they are not easily obtained. Compared to (mono-)disciplinary research, cross-disciplinary research comes with some extra challenges. Most challenges result from (the interplay of) two factors: (i) the context in which cross-disciplinarity is to be performed and disseminated is structured around the needs of individual disciplines, and (ii) cross-disciplinarity involves the combining of knowledge from different disciplines, and thus, dealing with, differences between disciplines regarding i.a. epistemology, metaphysics, methodology and jargon.

Both factors affect a crucial aspect of the practice of cross-disciplinarity, viz. communication. First, the discipline-centeredness of research contexts affects communication *about* cross-disciplinarity. This communication takes place between research policy makers, funding bodies and researchers: via mission statements, calls for projects, project proposals and reports they coordinate their needs and expectations regarding cross-disciplinary research. Second, the differences between disciplines affect communication *within* cross-disciplinary teams. This takes place between the members of

a cross-disciplinary team and concerns the content and execution of the project the team is working on.

This dissertation aims to facilitate cross-disciplinarity by providing theoretical analyses and practical tools that can be used to overcome challenges for cross-disciplinarity. The focus is on two specific challenges, i.e. one for the first and one for the second level of communication.

On the level of communication *about* cross-disciplinarity, the challenge is the current lack of a shared conceptual framework that unambiguously distinguishes the main types of cross-disciplinarity, contains an accurate definition of each type, and is neutral regarding the overall value of the types. Currently, there is a received framework that is used by research policy makers, funding bodies and researchers, but it does not meet the other desiderata. To facilitate communication about cross-disciplinarity, major steps are taken to transform the received conceptual framework into a new framework that does meet the necessary requirements.

On the level of communication *within* cross-disciplinary teams, the focus is on 'problematically ambiguous terms' (PATs), i.e. terms that cause serious communication problems for a given cross-disciplinary team because they occur in the jargon of multiple source disciplines of the project, but have different meanings across those jargons. To facilitate communication within cross-disciplinary teams, a semi-automated PATs resolution procedure is outlined that makes use of existing techniques from computational linguistics. Moreover, important steps towards its development are taken.

Because this dissertation combines knowledge and techniques from philosophy of science and computational linguistics, besides being about cross-disciplinarity, it also is cross-disciplinary itself.

Samenvatting

Cross-disciplinariteit staat voor het combineren van kennis en technieken die afkomstig zijn uit verschillende disciplines. Op basis van het soort disciplines dat kennis aanlevert (i.e. enkel academische disciplines of ook praktijkvelden) en de manier waarop kennis wordt gecombineerd (i.e. al dan niet door middel van integratie), worden er drie types cross-disciplinariteit onderscheiden: ‘interdisciplinariteit’ staat voor het integreren van kennis uit academische disciplines; ‘multidisciplinariteit’ voor het combineren van kennis uit academische disciplines, maar niet door middel van integratie; en ‘transdisciplinariteit’ staat voor het integreren van kennis uit academische en niet-academische disciplines.

De afgelopen decennia is cross-disciplinair onderzoek steeds populairder geworden, zowel bij beleidsmakers als bij onderzoekers. Een belangrijke motor voor deze toenemende populariteit is de idee dat het combineren van disciplines innovatief potentieel biedt voor het genereren van wetenschappelijke doorbraken en het aanpakken van complexe maatschappelijke problemen.

Hoewel de (potentiële) resultaten van cross-disciplinair onderzoek zeer gewenst zijn, worden ze niet makkelijk verkregen. In vergelijking met disciplinair onderzoek gaat cross-disciplinair onderzoek immers gepaard met een aantal extra uitdagingen. Deze uitdagingen komen voornamelijk voort uit (de wisselwerking tussen) twee factoren: (i) de context waarin cross-disciplinair onderzoek wordt uitgevoerd en verspreid (bv. universiteiten en tijdschriften) is hoofdzakelijk gericht op het ondersteunen van onderzoek in individuele disciplines, en (ii) cross-disciplinariteit houdt het combineren van kennis uit verschillende disciplines in, en dus ook het omgaan met verschillen tussen disciplines op hete vlak van o.a. kennistheorie, metafysica, methodologie, wetenschappelijk taalgebruik.

Beide factoren bemoeilijken een cruciaal aspect van cross-disciplinariteit, namelijk communicatie. De discipline-gerichtheid van de onderzoekscontext bemoeilijkt communicatie *over* cross-disciplinariteit. Deze communicatie vindt plaats tussen beleidsmakers, wetenschap financierende instellingen en onderzoekers: door middel van visieteksten, projectoproepen, projectaanvragen en rapporten coördineren ze hun wensen en verwachtingen omtrent cross-disciplinair onderzoek. Verschillen tussen disciplines bemoeilijken de communicatie *in* cross-disciplinaire teams. Deze

communicatie vindt plaats tussen de leden van cross-disciplinaire teams en gaat over de inhoud en uitvoering van de projecten waar de teams aan werken.

Dit proefschrift heeft als doel cross-disciplinariteit te faciliteren door theoretische analyses en praktische instrumenten aan te leveren die gebruikt kunnen worden om uitdagingen voor cross-disciplinariteit het hoofd te bieden. De focus ligt hierbij op twee specifieke uitdagingen die spelen op respectievelijk het eerste en het tweede communicatieniveau.

Op het niveau van communicatie *over* cross-disciplinariteit bestaat de uitdaging in het gebrek aan een gedeeld conceptueel kader dat de belangrijkste types van cross-disciplinariteit op een eenduidige manier onderscheidt, deze types van accurate definities voorziet, en bovendien neutraal is ten aanzien van de waarde van de types. Er is momenteel wel een gevestigd conceptueel kader dat wordt gebruikt door beleidsmakers, wetenschap financierende instellingen en onderzoekers, maar het voldoet niet aan de andere vereisten. Om de communicatie over cross-disciplinariteit te faciliteren, worden belangrijke stappen ondernomen om het gevestigde kader te transformeren tot een nieuw kader dat wel aan de nodige eisen voldoet.

Op het niveau van communicatie *in* cross-disciplinaire teams ligt de focus op 'problematisch ambigue termen' (PATs), i.e. termen die communicatieproblemen kunnen veroorzaken binnen een cross-disciplinair team omdat ze deel uitmaken van het jargon van verschillende brondisciplines van het project waar het team aan werkt, maar verschillende betekenissen hebben in die jargons. Om de communicatie in cross-disciplinaire teams te vergemakkelijken, wordt een nieuwe, semi-geautomatiseerde procedure voorgesteld om PATs op te lossen. De procedure maakt gebruik van bestaande technieken uit de computationele linguïstiek. Bovendien worden enkele belangrijke stappen genomen in de ontwikkeling van de procedure.

Doordat dit proefschrift kennis en technieken uit de wetenschapsfilosofie en de computationele linguïstiek combineert, gaat het niet alleen over cross-disciplinariteit, maar is het ook zelf cross-disciplinair.