Bridging Remote Cultures

Cross-lingual Concept Mapping Based on the Information Receiver's Prior-knowledge

Kano Glückstad, Fumiko

COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN
SOLBJERG PLADS 3
DK-2000 FREDERIKSBERG
DANMARK

www.cbs.dk

ISSN 0906-6934

Print ISBN: 978-87-92842-88-6
Online ISBN: 978-87-92842-89-3

Bridging Remote Cultures

PhD Series 31-2012

Copenhagen Business School
HANDELSHØJSKOLEN

Bridging Remote Cultures: Cross-lingual concept mapping based on the information receiver's prior-knowledge

Fumiko Kano Glückstad

LIMAC PhD School
Programme in Language and Culture

PhD Series 31-2012

PhD dissertation

Bridging Remote Cultures: Cross-lingual concept mapping based on the information receiver's prior-knowledge

By Fumiko Kano Glückstad

LIMAC Graduate School
Department of International Business Communication
Copenhagen Business School

Fumiko Kano Glückstad
Bridging Remote Cultures: Cross-lingual concept mapping based on the information receiver's prior-knowledge

1st edition 2012
PhD Series 31.2012

ISSN 0906-6934

Print ISBN: 978-87-92842-88-6
Online ISBN: 978-87-92842-89-3

LIMAC PhD School is a cross disciplinary PhD School connected to research communities within the areas of Languages, Law, Informatics, Operations Management, Accounting, Communication and Cultural Studies.

Preface

My original motivation for undertaking this PhD research has inherently been rooted in my professional experiences as a freelance researcher for several consultant companies undertaking Japanese governmental research projects. More specifically, I have been involved in a plurality of projects where I reported on governmental policies and social frameworks in the Scandinavian countries for the Japanese central government. In the process of these undertakings, I have frequently faced situations where I had difficulties in translating a Japanese term written in an assignment to an ‘appropriate’ keyword expressed in an original Scandinavian language, which should be used for searching relevant documents, and in identifying an ‘appropriate’ translation in Japanese which can conceptually be understood by a Japanese audience. Since these international reports inherently address a wide audience in Japan, my humbly translated and communicated Japanese terms would potentially be disseminated nationwide. As a consequence, a translator or a communicator has in general – whether conscious about it or not – a substantial responsibility for introducing foreign concepts in a precise and unambiguous way. When considering this immense responsibility as a communicator, I realized a particularly important fact, that the number of people who have in-depth knowledge about the two cultures, i.e. Scandinavia and Japan, is in fact very limited. The number of Japanese residing in Denmark is only about 500-700 including temporal Japanese business expats. This inherently means that the domestic Japanese residing in Japan have no idea whether a given Japanese translation of Danish concepts is in fact appropriately reflecting the original meaning of the source concepts at hand. They must solely rely on Japanese translations that are provided as stimuli, and from these stimuli, they simply have to create their own picture about these Danish concepts, terms or meanings. These problems would at first glance seem to be regarded as an issue of proper translation between Danish and Japanese. However, the real underlying issue is that a third language - English - is in most cases needed in the process of translation between rare language combinations such as the case of Danish and Japanese. Thus, a major challenge is the lack of linguistic – and human – resources to directly bridge groups of remote languages. For example, in most business- or diplomatic situations, a human face-to-face or a written communication usually relies on English as intermediate or pivot language. The involvement of English as pivot means that inherently a third language culture – like the British or the American – which is obviously not a native culture of any of the two communicating parties is suddenly indirectly inherited and playing in during the communication process. This has the potential impact of amplifying semantic inconsistencies that already inherently exist between the parties of the two remote cultures undertaking a communication. When using English as pivot for expressing even a similar concept existing in two remote cultures the outcome can be rather different and hence, even in this "trivial" situation there is a potential risk of creating misunderstandings between the two parties. My proposed approach to resolve this remote culture communication challenge has been to apply Terminological Ontology (TO) proposed by Prof. Bodil Nistrup Madsen's research group at CBS. I have considered the communicator- and receiver's conceptual knowledge rooted in their respective cultures as TOs that consist of both original expressions and English translated expressions of concepts and their characteristic features (definitions). My idea has been to map these culturally-specific TOs using the English translations originating from the respective countries for bridging domain knowledge existing in remote language cultures.

My original PhD project plan was rather focused on technical aspects of TO mapping and my plan was to implement this framework in collaboration with Prof. Kentaro Inui's research group at Tohoku University in Japan. However, this stay and plan was abruptly terminated due to the Great Tohoku Earthquake that struck the region I was temporarily residing, on 11 March 2011. Hence, my PhD project and plan was forced to be strongly revised due to this sudden termination of my foreign research stay at Tohoku University. Surprisingly, this unexpected *force majour* did not completely ruin my research undertakings – as everyone would have expected, even myself at that point - but instead forced me to replan, refocus and "boot strap" a new research plan with success, into publishing a plurality of full peer-reviewed conference papers during the past 14 months. More specifically, the quake and its consequences somehow forced me to completely reconsider my project from several different and new angles and I eventually decided to apply a cognitive approach based on my earlier educational- and professional experiences. In the aftermath of the quake, I have published in total 8 papers, among these, 6 peer-reviewed papers have received overwhelmingly positive reviews. One has been selected for the best paper award among student authors and another has been nominated to be among the last four finalists for a best paper award, despite the fact that they were based on my very quick and preliminary empirical studies. Due to these successful appearances at the international stage, I have had opportunities to discuss my research ideas with several influential professors in diverse research domains, among others, computational linguistics, terminology, artificial intelligence, cognitive science and computer science such as machine learning. Through these interactions, I realized that people who have shown their interest and understanding in my ideas know the difficulties of bridging remote cultures because they - themselves - experienced the multi-cultural and multilingual environment in their respective international activities. This implies, I think, that my unexpected success after the sudden termination last year is in fact inherently rooted in my 15 years of multi-cultural experiences both privately and professionally. The many encouragements from international scientists indicate that the new way I have started to look at the problems of communication and translation, could potentially be a "game changer" playing into a diversity of domains of research, not only for machine translation and cross lingual information retrieval but also for multi-cultural domain knowledge integration and transfer, and to various human aspects such as modality research and so-called social signal processing involved in any translation and communication process. The key to future international research undertakings is to further facilitate the understanding of inter-cultural and inter-lingual communication processes from both a human- and a linguistic aspect. Here, in particular, I believe that my newest ideas based on so-called cognitive ontology mapping could, in the future, be excitingly integrated into the framework of Theory of Communicative Supertypes proposed by Prof. Per Durst-Andersen at CBS. The integration of the cognitive ontology mapping with the Communicative Supertypes has the potential to uniquely explain how linguistic supertypes influence the formation of cognitive ontologies of culturally-specific domain knowledge which may affect any cross-cultural communication process.

I would like to express my thanks to: first of all, the CBS Research Dean, Alan Irwin and the Head of Department of IBC, Alex Klinge who generously granted me a half-year extension of my PhD project due to the force majour caused by the Great Tohoku Earthquake on 11 March, 2011. This extension was perhaps the most important and exciting period to harness my new research ideas and to materialize them. Thanks go to my formal supervisors, Hanne Erdman Thomsen and Daniel Hardt, who have both always been critical to the settings of my empirical studies which have been necessary and enabled me to critically and carefully rethink the entire empirical workflow; and Dr. Morten Mørup from DTU Informatics who supported me with technical and mathematical implementations of cognitive models, which harnessed my basic mathematical understandings of the applied models and formalized the empirical studies. I also appreciate positive and valuable feedback from Prof. Jerker Denrell from Saïd Business School, University of Oxford, Per Durst-Andersen from IBC, the Terminology and Artificial Intelligence community, the Culture & Computing community, who all encouraged me to believe and continue with my ideas and for the completion of this research.

I would also like to express my appreciation and respect to people in Sendai, Japan – most importantly, Prof. Kentaro Inui's research group at Tohoku University, who, already just after the quake, struggled and fought together to recover from the psychological and physical damages caused by the disaster taking the life of more than 20.000 people. Their messages posted on the internet encouraged and kept me strong for the past one year in the aftermath of all this. I will also never forget the people in our common shelter in Sendai, who let us bring a lot of rice-balls etc. when we were leaving, despite the shortage of food in the period following the disaster. When I left Sendai and Japan days later with our daughter, I felt a strong guilt- and almost a feeling of being irresponsible "egoist", as Japanese. My mind was somehow split in two and I felt a tremendous sadness. However, the memory of the rice-balls from the shelter and the subsequent messages from students in Prof. Inui's research group made me think that "they are fighting with enormous courage and are never giving up, so I should at least do the same back in Denmark".

Finally, I thank my relatives – my Japanese parents who cleaned up the mess which I left in Sendai and my family here in Denmark. There have been many misunderstandings and challenges in our cross-cultural communication during the aftermath of the disaster, which ironically inspired my latest research. Finally, I have to show my deep respect for our daughter, Mie, who has grown up in a positive- and healthy way despite her "up-front experiences" with one of the largest disasters in the Japanese history and all the troubles that followed.

Basho's road
throughout Tohoku
hope remains

Haiku writer, S. DeGuire, paying homage to Matsuo Basho, the 17th century haiku virtuoso whose masterwork "Oku no Hosomichi" (The Narrow Road to the Deep North) was inspired by his treks in the Tohoku region, ravaged by the 3-11 earthquake, tsunami and devastating nuclear aftermath.

Contents

List of figures

List of tables

List of abbreviations

BMG	Bayesian Model of Generalization
CAOS	Computer Aided Ontology Structuring prototype
CLIR	Cross Lingual Information Retrieval
COM	Culturally-specific Ontology Mapping Cognitive Ontology Mapping
CRICED	Center for Research on International Cooperation in Educational Development
CRP	Chinese Restaurant Process
CSC	Culturally-specific Concept
IND	Immigration and Naturalization Office
IRM	Infinite Relational Model
ISCED	International Standard Classification for EDucation
KYOTO	Knowledge-Yielding Ontologies for Transition-based ORganization
MEXT	the Japanese Ministry of Education, Culture, Sports, Science and Technology
MONNET	Multilingual ONtologies for NETworked Knowledge
MT	Machine Translation
RTC	Relevance Theory of Communication
SC	Source Culture
SL	Source Language
TC	Target Culture
TL	Target Language
TO	Terminological Ontology (refers to method)
TOs	Terminological Ontologies (refer to constructed ontologies)
UIS	UNESCO Institute for Statistics
UNESCO	United Nations Educational, Scientific and Cultural Organization
XBRL	eXtensible Business Reporting Language

Chapter 1.

Introduction

1. Motivation: real-life problems

1.1. Human-based communication

During recent years, I have been involved in a plurality of projects where I reported on governmental policies and social frameworks in the Scandinavian countries for the Japanese central government. In the process of these undertakings, I have frequently faced situations where I had difficulties in: **i)** translating a Japanese term written in an assignment to an 'appropriate' keyword expressed in an original Scandinavian language, which should be used for searching relevant documents; and **ii)** identifying an 'appropriate' translation in Japanese which can conceptually be understood by a Japanese audience. Since these international reports inherently address a wide audience in Japan, my humbly translated and communicated Japanese terms would potentially be disseminated nation-wide. As a consequence, a translator or a communicator has in general – whether conscious about it or not – a substantial responsibility for introducing foreign concepts in a precise and unambiguous way.

One of my Japanese acquaintances who has been living in Denmark for more than 40 years formulated this difficult mission of undertaking such translation tasks in the following way: "*Once I deeply understood the two cultures (Denmark and Japan) and the cultural differences/nuances of conceptual meanings existing in the two countries, it became impossible for me to translate culturally-specific terms into the other language. Existing language resources (dictionaries etc.) are in this context useless*". What my acquaintance meant by this is that it is virtually impossible to precisely translate or convey the meaning of a culturally-specific term if no exact equivalent term exists in the Target Language (TL) culture. Despite this inherent frustration, communicators or translators are still required to convey such culturally-specific terms into a TL in an optimal manner such that a TL audience can instantly infer the original meaning of a given Source Language (SL) concept. - Is there some way to solve this inherently difficult challenge which even skilled and experienced translators cannot easily solve?

Figure 1-1: How to introduce a new Japanese delikattes to Danes

This grand challenge can also be identified in a more familiar and daily life scenario e.g. in the international marketing practices. As an example, consider a situation where a Dane travelling to Japan has been visiting a Japanese café and enjoyed a traditional Japanese dessert which tasted new and fantastic! The Dane believes that people from Denmark must also like this kind of dessert so that he/she is tempted to introduce this new type of dessert concept by opening a new café in the centre of Copenhagen. In such a situation, one must naturally apply the Danish people´s general prior knowledge about desserts and how Danes are expected to compare this new Japanese delikattes with the desert concepts they are already familiar with. From such an initial “guesstimate”, it should be identified how this new Japanese concept can be categorized and positively accepted by a general Danish mind-set.

The situation could be more complicated in cross-cultural communication scenarios in particular between remote cultures (e.g. between Scandinavians and Asians) both of whom use a non-English language. Such a situation implies that these two parties from two remote countries most likely communicate with each other in their second language - English. The situation is identical in written communication scenarios. If a Dane wants to know a specific Japanese concept written in Japanese, he/she has in most of the cases to rely on Japanese-English linguistic resources (e.g. dictionaries), since no highly reliable linguistic resource directly linking between Japanese and Danish exists. Hence, for any pattern of communication between remote cultures, use of English as an intermediate language is virtually unavoidable. This interference of English might create a further cultural bias for two non-native English speakers. In other words, the involvement of English as intermediate language in communication will most likely amplify semantic ambiguities between the two parties.

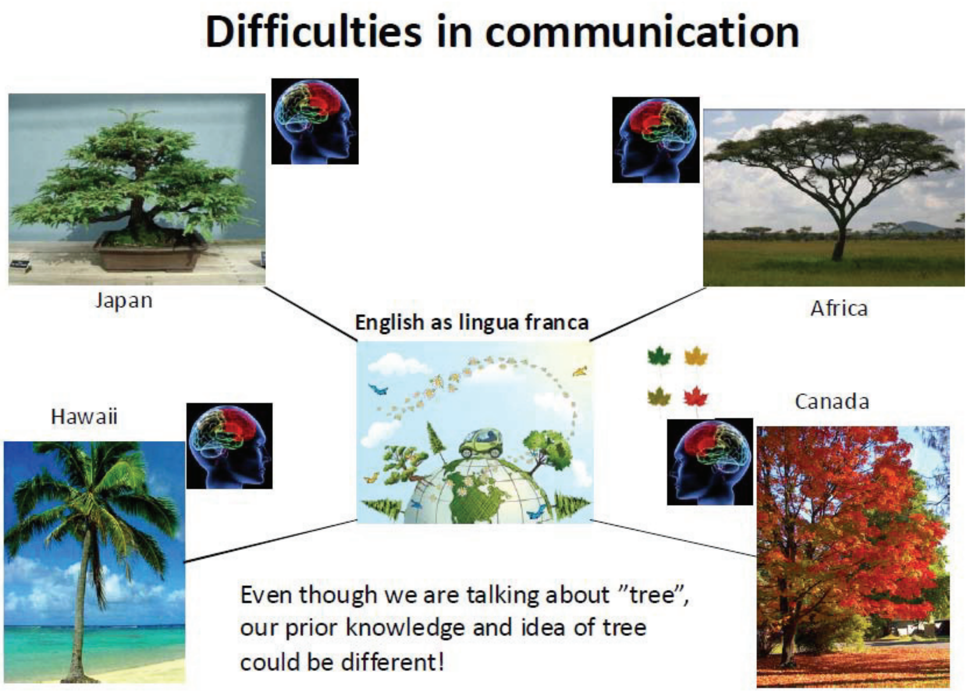


Figure 1-2: Picturing the difficulties in communication

Considering any of the above situations, the question is: how to convey meanings of culturally-specific concepts existing in one culture to a person coming from another (remote) culture, and how to achieve a successful knowledge transfer. For addressing this issue, my thesis applies the theoretical framework based on Sperber & Wilson´s Relevance Theory of Communication which considers communication as an inferential process. More specifically, a translated term of an SL concept is considered as a new object that appears to a TL reader´s taxonomically organized prior knowledge. Hence a translator should identify a stimulus (that is translation) optimally relevant to a concept existing in the TL reader´s taxonomy that is rooted in his/her culture. Such an optimally relevant stimulus (translation) would form a solid common ground between an SL communicator and a TL reader.

1.2. Machine-based communication

The fast-paced globalization with its internet revolution has brought about new possibilities of direct access to local first-hand information. This movement is further facilitated by the recent development of the Semantic Web in the future. The role of ontology in a multilingual setting is considered as an emerging challenge for Semantic Web development. For example, one of the interesting Japanese-based projects, the so-called Language Grid (Ishida, 2006), is a complete infrastructure in itself that combines language resources such as bilingual dictionaries, text corpora and natural language processing tools. Its backbone is based on an ontology called the Language Service Ontology and the Semantic Web Service Technology. These technologies enable one to combine bilingual dictionaries, e.g. Japanese-English and English-Danish dictionaries, text corpora and natural language processing tools from all over the world. Thus translations between languages, e.g. Japanese-Danish, can be realized through a *pivot* translation using English as an intermediate language. However, this *pivot* approach focuses on combining the existing language resources rather than carefully considering how to optimally convey meanings of culturally-specific concepts existing in one culture to a population from another culture, and how to achieve a successful knowledge transfer.

JP	English translation of JP	DK (based on TO mapping)
専門学校 (Syn. 専修学校専門課程) *Senmon-gakko* *(Syn. Sensyu-gakko senmon-katei)*	Specialized training collage specialist degree	Erhvervsakademiuddannelse Syn: kort videregående uddannelse
高等専修学校 *Koto-sensyu-gakko*	Specialized training school upper secondary course	Erhvervsuddannelse

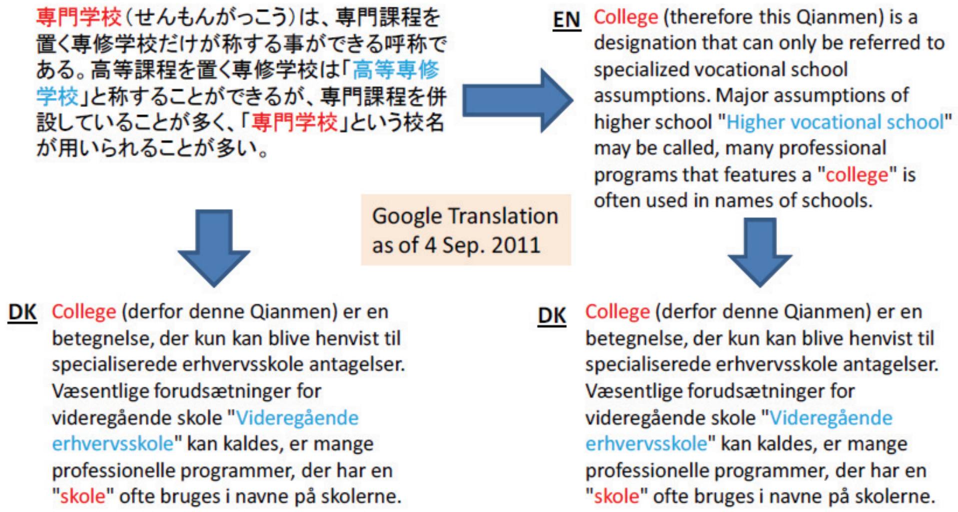


Figure 1-3: Google pivot translation example

The well-known Google translation is another interesting case. When translating the following Japanese text into Danish, the Google solution obviously employs the pivot translation approach. Hence culturally-specific terms such as *Senmon-gakko* (official English translation: “Specialized training college, specialist degree course”) which, according to my bi-cultural knowledge, possibly corresponds to the Danish term “Erhvervsakademiuddannelse” (synonym: “kort videregående uddannelse”) is translated into “college” or “school” by the Google translation. In the same way, *Koto Sensyu Gakko* (official English translation: “Specialized training school, upper secondary course”) which corresponds to the Danish term “Erhvervsuddannelse” is translated into “Videregående erhvervsskole” via the pivot translation “higher vocational school”. A critical problem lying here is that a solely statistical solution considers any kind of English expression as one large bag of words (as English corpus) which contains not only expressions of concepts originating from English-speaking countries but also English translations of concepts originating from other non-English speaking countries. When considering the obvious goal of translation, that is, to convey the original meaning of a source concept to an audience in a target culture, to achieve the most successful cross-cultural knowledge transfer and, to share the common knowledge at the maximum degree between the two parties, a critical question is how well such a pivot translation can convey an original conceptual meaning of an SL word into a TL translation.

This problem could be well-explained by the following example: Imagine a situation where a Japanese diplomatic officer receives a business card from a Danish *konsulent* at *Økonomistyrelsen*. The Japanese officer needs to understand the definition of three words *konsulent, økonomi* and *styrelsen*, in order to figure out at which level this Danish person is positioned and in which kind of organization, by reflecting on his/her current conceptual understandings of the Japanese political system. First of all, a term like *Økonomistyrelsen*[1] consists of several words. In addition, each Danish word has more than one sense (meaning). In the case of *konsulent, økonomi* and *styrelsen*, each word respectively has one, five and four senses according to the Danish WordNet. When checking up with the available Danish-English dictionaries, there are several English translation options: ***konsulent*** can potentially be translated as *consultant*, *advisor*, *advisory officer*, *reader*, or *counselor;* ***økonomi*** into *finance*, *economy* or *economics; and* ***styrelsen*** into *management*, *executive committee*, *council*, *board*, *steering committee*, *government agency*, *executive agency*, *state agency*, and *administration*. In addition to this, there are several Japanese translation options for each English translation candidate when looking them up in English-Japanese dictionaries. Since a term consists of several words and a word often carries several meanings, the dictionary-based transitive translation approach using English as a pivot language simply amplifies the probability of selecting an inappropriate sense in a TL. Thus the problem of word sense ambiguity becomes especially serious in the process of pivot translations.

Another point is that a source concept often carries so-called equivalent expressions in several languages, i.e. an original expression in a source language and its official translation in English. A problem is that these lexically expressed multi-word expressions in different languages are sometimes semantically inconsistent. A typical example is illustrated by the case where the formal English name of the Danish authority *Økonomistyrelsen* is *The Danish Agency for Governmental Management*. When readers see these expressions without knowing the synonymous relationship between them, it is impossible for them to judge that these expressions refer to the same concept. This example shows that it is a major challenge to find an appropriate translation

[1] Terminologist categorizes this type of term – i.e. organization - as an individual concept.

e.g. Japanese, that can optimally convey the original meaning of a specific Danish concept to a Japanese audience. One obvious point is that it is impossible to translate such a term without knowing its original conceptual meaning – that is, the definition of the concept under consideration. In the case of the Japanese translation of the Danish term *Økonomistyrelsen*, what readers need to know in a cross-cultural communication context is the level at which the Danish organization is situated, information about whether the organization is part of the ministry, and what kind of authorization the organization has, etc. Hence, an ideal Japanese translation should reflect on these pieces of knowledge to the maximum extent, contrasted to the Japanese conceptual structure in question.

These examples indicate that simple statistically-based or electronic dictionary-based pivot translation schemes seem to have severe inherent problems conveying original conceptual meaning of SL words into TL translations.

1.3. Is there any solution to these challenges ?

When considering research domains such as Machine Translation (MT) and multi-linguality of linguistic resources, it is, in my eyes, crucial to realize the aforementioned fundamental issues existing in real-life cross-cultural communication. That is, to focus on how well the meaning of an SL concept is optimally conveyed to a TL audience and how to achieve a successful knowledge transfer or cross-cultural communication. Surprisingly, I have often realized that researchers involved in MT research or research on multilingual linguistic resources do not have a multilingual/multi-cultural background themselves. Hence, the solutions they propose are often inherently depending on straight forward translation based on computational techniques without taking careful note of the existing differences in conceptual structures possessed by people from different cultures. Another challenge arising from this could be the difficulties in discovering the "real issue" lying behind the translation between e.g. a Danish and a Japanese, since there is no obvious way of understanding the "true meaning" of concepts rooted in the two cultures, in particular, if the source culture is remote to the target audience. For example, consider the situation where a wrong Japanese translation (国民学校: citizen school) of a Danish SL concept (folkeskole) is provided to a Japanese TL reader. How many Japanese living in Japan will potentially speculate on the true meaning of this Danish concept and start to investigate the real meaning of this specific Danish concept? From this point, the framework I propose in this thesis is strongly influenced by my (Danish-Japanese) bi-cultural intuition and self-analysis of my own cognitive processes involved when translating a specific Danish concept for a Japanese audience.

A critically important point is what defines a successful cross-cultural communication and what the objective of translation required in a successful cross-cultural communication is. The Relevance Theory of Communication (Sperber & Wilson) states that *communication is principally an inferential process.* This means that a communicator has to produce a *stimulus* from which the audience can infer what set of thoughts or assumptions the communicator intends to convey. When it comes to translation, a successful translator should create *stimulus* (that is translation) that is optimally relevant to the audience in the target culture (Gutt, 1990). This inferential process can be further divided into the issues related to the translator who translates SL concepts to TL translations and the reader who is supposed to infer the meaning of the given translations. Assuming that the translator´s mother tongue is a TL (which is often the case), the first thing the translator has to do is to comprehend a given SL text by inferring the context of text and meanings of terms described therein. In this process, the translator is supposed to generalize meanings of newly emerged terms in the text and to categorize them in accordance with his/her prior knowledge about the specific domain in terms of both the source culture and the target culture. Subsequently, he/she has to identify the best possible translation which can optimally convey the original meanings of SL into TL. According to (Murphy, 2004: Chapter 7 & 8), in such situations, various cognitive processes that have been investigated in the history of concept learning research are involved, among others, similarity based categorization, knowledge effects (the effect of prior knowledge), and taxonomic organization of categories. From the given translation that is identified by the translator as stimuli, the TL reader is subsequently supposed to infer the original meaning of given SL concepts. Again in this process, various cognitive processes such as similarity based categorization, knowledge effects (the effect of prior knowledge), and the taxonomic organization of categories are involved.

Consequently, a key research question that should be emphasized in this thesis is: **how should background knowledge possessed by an SL communicator and a TL reader be represented and linked in**

light of various cognitive processes involved in cross-cultural communication. My hypothesis starts by the assumption that Terminological Ontology (TO) (Madsen et al., 2004) could potentially be used as prior-knowledge of the culturally-specific domain knowledge, which is required for the inference process in a cross-cultural human communication scenario. Secondly, I consider several cognitive models originating from the categorization and concept learning research as potential ontology mapping algorithms, which reflect the inference process involved in cross-cultural communication. Hence, in this thesis, the Relevance Theory of Communication (Sperber & Wilson, 1986) and the categorization research that has been performed by diverse cognitive scientists (Murphy, 2004) is contrasted to a simulative framework, Culturally-specific Ontology Mapping (COM) applying the cognitive models to data-sets obtained from TO. The eventual aim of this work is to propose and assess whether the simulative framework, COM, is a technically feasible solution for identifying potential translation candidates by optimizing relevance between concepts in two remote languages such as a set of European and Asian languages, and for facilitating common understanding between information senders and receivers. However, in this thesis, I focus on identification of optimal components suitable for the COM framework that realizes the cognitive simulation based on the Relevance Theory of Communication (Sperber & Wilson, 1986), as well as, similarity based categorization, knowledge effects (the effect of prior knowledge), and taxonomic organization of categories in cognitive sciences (Murphy, 2004).

2. Proposal of a framework: COM

The uniqueness of the COM approach is to apply mathematical models derived from cognitive science to the issue having been considered as one of the most challenging topics within the research domain of Semantic-Web and multilinguality, that is, linking of culturally-specific concepts that are semantically inconsistent. The theoretical approach taken in this thesis is based on **1)** Sperber & Wilson´s Relevance Theory of Communication which considers communication as an inferential process and **2)** the Knowledge Effects proposed in the area of concept learning and categorization research in the cognitive science. In the Relevance Theory of Communication (Sperber & Wilson, 1986), a reader’s prior-knowledge is considered as *context*, and a deduction process of new information P and the context C is called *contextualization*. In here, new knowledge generated from the union of P and C is called *contextual effect* which is the necessary condition for *relevance*. The context is selected from a set of background assumptions that are organized in an individual's *encyclopaedic memory*. In the Knowledge Effects and Taxonomic Organization of Categories in the cognitive sciences, people *use their prior-knowledge to reason about a new object in order to decide what category it is, or in order to learn a new category* (Murphy, 2004: 60-61). Such a new category is stored in a person's mind as a concept and a plurality of such concepts are most likely connected in hierarchical networks in the brain. These hierarchical network connections are used for making inductive inferences and categorization judgments (Murphy, 2004: Chapter 7). The cognitive theories imply that an information receiver's prior-knowledge is used for his/her inferential process of learning new information and such prior-knowledge is organized in a taxonomic hierarchy in his/her memory. Supported by these ground theories, I will, at the second-half of a series of empirical studies in this thesis, employ a novel mathematical cognitive model called the Bayesian Model of Generalization, originally proposed by the cognitive scientists (Tenenbaum & Griffiths, 2001). This model is based on the traditional theories of the cognitive sciences, among others, Shepard's Universal Law of Generalization (Shepard, 1987) and Tversky's Contrast Model (Tversky, 1977) and addresses crucial questions of learning processes suggested by (Chomsky, 1986). In order to apply the Bayesian Model of Generalization, I consider a scenario of bi-directional communication: **a)** from an SL communicator to a TL audience: a translated term of an SL concept as a new object that appears to a TL audience´s taxonomically organized prior-knowledge; and **b)** an SL communicator identifies a stimulus (that is translation) optimally relevant to a concept existing in the TL audience´s taxonomy.

My hypothesis starts from the viewpoint that Terminological Ontology (TO), the method that has been introduced by (Madsen et al., 2004) at CBS, contributes to identify an optimally relevant translation by assisting one to systematically organize conceptual features from domain knowledge (corresponding to the taxonomic organizations). Such systematically organized features can be used for linking an SL concept with a TL concept based on a plurality of cognitive models as algorithms of aligning two culturally-dependent taxonomies.

TO is a domain-specific ontology used for knowledge sharing, which normally is applied in terminology work within the domain of language for special purposes. The unique points of TO that differentiate it from other types of ontologies are its feature specifications and subdivision criteria. A feature specification consists of a feature dimension and its value. Thus, a representation of a whole concept is a feature structure, i.e. a set of feature specifications corresponding to the unique set of characteristics that constitutes that particular concept. Terminological ontologists argue that concepts are defined in a language-dependent context, and therefore, TO is language- or culturally dependent. TO is developed within a knowledge sharing community, then dynamically updated and validated. If it is necessary to share knowledge with other communities, TOs developed in different communities should be compared, aligned and merged as needed. Hence, as a starting point of my work in this thesis, two culturally-dependent TOs are developed and aligned for identifying corresponding concepts existing in the two cultures under consideration, which would potentially form the optimal common ground between the two parties: the SL communicator and the TL audience.

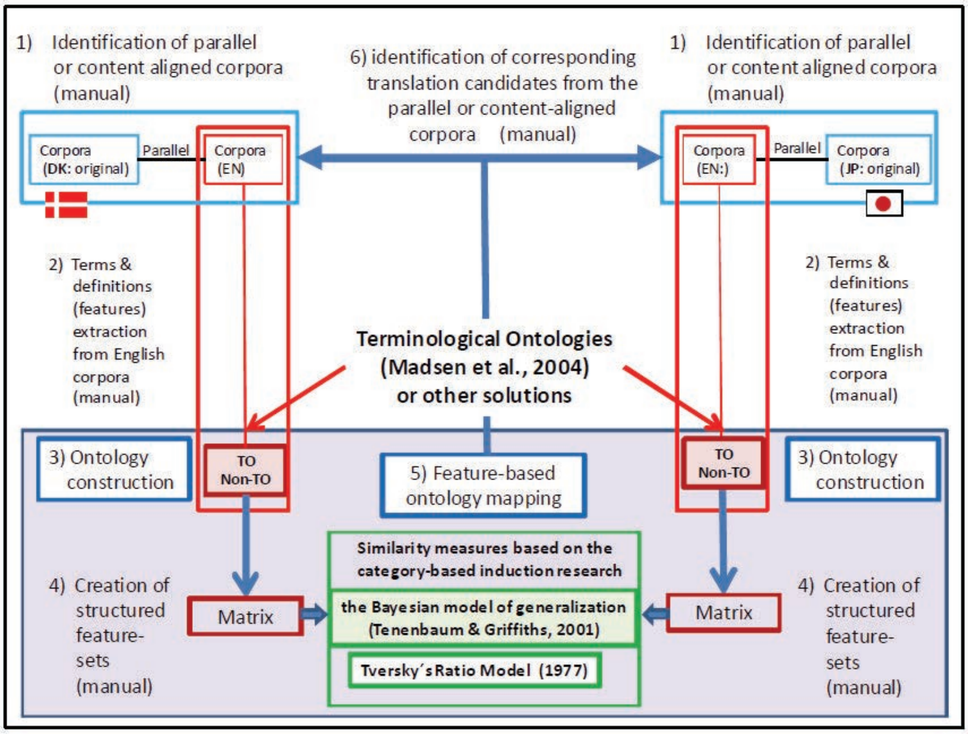


Figure 1-4: Culturally-specific Ontology Mapping Framework

To summarize, the initial framework of COM proposed in this thesis is assumed to consist of the following modules: **1)** identification of domain specific content-aligned (or parallel) corpora, consisting of SL-English and English-TL language combinations; **2)** terms and features extractions from content-aligned English corpora which describe domain specific terms and their definitions in English for the respective domain-knowledge in the SL- and TL cultures; **3)** construction of ontologies for the respective domain-knowledge in the SL- and TL countries based on the information extracted in 2); **4)** creation of feature structures for each concept for the respective domain-knowledge in the two countries, and standardization of feature labels used in the two country-specific ontologies; **5)** alignment of the two structured feature sets based on feature-based ontology mapping algorithms applying cognitive models, i.e. Tversky´s contrast model (Tversky, 1977) and eventually the Bayesian Model of Generalization (Tenenbaum & Griffiths, 2001); and **6)** identification of corresponding translation candidates from the content-aligned corpora consisting of SL-English and English-TL language combinations. For convenience, the initial COM framework is schematically illustrated in Figure 1-4. With this framework as starting point, I will further investigate **how the background knowledge possessed by an SL communicator and a TL reader should be represented and linked in light of various cognitive processes involved in cross-cultural communication.**

Since the entire framework covers a broad range of research domains, the specific scope and limitations of this thesis is described in the next section.

3. Research questions and scope limitations

As outlined in the previous section, a key research question in this thesis is identification of optimal components suitable for the COM framework that realizes the cognitive simulation based on the Relevance Theory of Communication (Sperber & Wilson, 1986), as well as, similarity based categorization, knowledge effects (the effect of prior knowledge), and taxonomic organization of categories in cognitive sciences (Murphy, 2004). More specifically, my thesis will attempt to address the question: **How should the background knowledge possessed by an SL communicator and a TL reader be repre-**

sented and linked in light of various cognitive processes involved in cross-cultural communication?

As described in the previous section, the eventual aim of this work is to propose a simulative framework for identifying potential translation candidates by optimizing *relevance* between concepts in two remote languages such as a set of European and Asian languages. The framework could be considered as a complementary approach to the MONNET project[2] and the KYOTO project[3] which have been both involving a number of outstanding researchers from several research institutions across a plurality of countries. Compared to these major international projects, this thesis is based on a "one-man" PhD project. Hence, there are obvious limitations in many of the considered aspects when proposing such an ambitious framework covering a broad range of research domains.

Consequently, I have decided to focus my research on a few but important issues that clearly differentiate my work from other major projects. First of all, the main focus of my undertakings is to align Culturally-specific Concepts (CSCs) that are respectively rooted in two remote languages or cultures. The idea of focusing on these CSCs is brought about by the following reasons: When focusing on problems of cross-cultural communication occurring at the lexical level: **1)** often, it is impossible to identify a semantically 100% equivalent CSC for a remote culture; and **2)** prior-knowledge of CSCs that is deeply rooted in one's own cultural background such as traditions, social systems etc. inherently influences as cultural bias on any cross-cultural communication scenario. Hence, a question might immediately arise: How can we possibly improve the transfer of original conceptual meanings of SL concepts to a TL audience when there are no 100% equivalent concepts existing between the two remote cultures under consideration and when both a TL audience and an SL communicator process inference based on their own culturally-rooted prior-knowledge? In other words: **How should the background knowledge possessed by an SL communicator and a TL reader be represented and linked in light of various cognitive processes involved in cross-cultural communication?**

As mentioned above, my hypothesis starts by the assumption that TO (Madsen et al., 2004) could potentially be used as prior-knowledge of the culturally-specific domain knowledge, which is re-

[2] http://www.monnet-project.eu/

[3] http://www.kyoto-project.eu/

quired for the inference process in cross-cultural human communication scenario. Accordingly my main research question can be divided into two dimensions: **a)** what algorithms are suitable for mapping TOs ? ; and **b)** within the defined framework of COM, what are the advantages or disadvantages of using TO compared with other approaches for systematically representing and organizing conceptual features as culturally-dependent prior-knowledge? In order to answer the first sub-question posed in a), I consider several cognitive models originating from the categorization and concept learning research as potential ontology mapping algorithms, which reflect the inference process involved in cross-cultural communication. I have implemented several preliminary studies of mapping two culturally-dependent TOs. The empirical studies started from a very primitive and manual feature-mapping approach, gradually applying a very basic cognitive model, and eventually identifying the algorithm that has been well-recognized as a novel cognitive model of concept learning in the cognitive science as the best possible algorithm for the COM framework. The gradual progress of my empirical studies and application of the relevant theories which support the eventual selection of the best possible algorithm will be carefully explained in this thesis. The explanation arrived at, is expected to automatically answer the sub-question posed in a). For addressing the second sub-question posed in b), I have implemented the comparative analysis of mapping two types of input data: **i)** feature-sets created by TOs and **ii)** feature-sets directly created from corpora without TO in combination with a novel unsupervised learning method called the Infinitive Relational Model (Kemp, et al. 2006). This enables me to answer what are potentially the advantages or disadvantages of using TO. These attempts from the two dimensions, will hopefully answer the key question: **How should the semantic knowledge possessed by an SL communicator and a TL reader be represented and linked in light of various cognitive processes involved in cross-cultural communication?**

The eventual aim of the COM framework could be to develop an automatic solution to handle each module. However, at this moment, my project is limited to the identification of optimal components that simultaneously support human cognitive process into the framework of COM. More specifically, my project is limited to the scope of:

- Systematically and manually extracting necessary information (terms, definitions, features, dimensions, relations) from text corpora and manually or semi-automatically develop ontologies, corresponding to modules 2 and 3 in the COM framework outlined on pages 21-24.

- From the constructed ontologies, manually create matrices consisting of terms and features for the respective culturally-dependent ontologies, corresponding to module 4 in the COM framework outlined in the aforementioned pages.

- Employing cognitive models as algorithms to computationally map domain-specific language-dependent structured-datasets, corresponding to module 5 in the COM framework outlined in the aforementioned pages.

My focus will especially be on the last point above, on the employment of cognitive models and its meanings/intensions behind the selection of the cognitive models. At the end of this thesis, I aim at arguing whether translation candidates obtained from different approaches taken under the COM framework convey original conceptual meanings of SL words to TL readers from the cognitive aspects of learning a new foreign concept. Strictly speaking, for assessing such effects, a thorough human-based assessment is required. However, the cognitive aspects mentioned in this thesis are only used as a theoretical apparatus behind the COM. Ideally, the empirical analysis for assessing the translation output should be included in the framework of this research. Considering the issues described below, the assessment of translation output from a cognitive aspect will obviously be considered as out-of-scope for this thesis work.

First of all, this research has been pursued as a “one-man project” for the limited period available for a PhD-project. Secondly, identifying people who have in-depth knowledge of two remote cultures and countries are obviously very limited. Thirdly, even though I manage to identify subjects having in-depth knowledge of the two remote cultures, additional “time factors”, such as the generation gap existing in conceptualization of CSCs, should be taken into consideration. Thus, a thorough human-based assessment method should be process oriented – for example, on how people induce knowledge or imagine associative words from a given translated word, e.g. the feature listing method (Rosch, 1978), instead of comparing mapping results against a sort of gold-standard mapping, that has no guarantee of being correct. Finally there are other issues to be considered, e.g. whether subjects should have bi-cultural or mono-cultural backgrounds and whether subjects should assess English translations or only local expressions of original concepts. These issues have to be taken into careful consideration and any human assessment methodologies should be carefully planned. Hence the human assessment itself should naturally be positioned as further research to be undertaken after the successful completion of this particular project. These important considerations will be further discussed in the final part of the thesis.

4. The outline of the chapter organization

The thesis is organized in the following way: in the next chapter, I will try to position the COM framework in comparison with other multilingual ontology frameworks, such as the KYOTO project and the MONNET project. In this comparison, I will analyze each of the projects from the viewpoint of whether the styles of cross-cultural communication achieved by these multilingual ontology solutions are bidirectional (inter-operable) or mono-directional (assimilative); and whether concepts existing in two remote cultures are linked in the form of symmetric or asymmetric relations. From these viewpoints, it will be discussed in the next chapter how the COM framework can be uniquely positioned. Chapter 3 first introduces the two basic theories: the Relevance Theory of Communication (Sperber & Wilson, 1986); and the Knowledge Approach (Murphy, 2004) proposed in the area of concept learning and categorization research in the cognitive sciences that support the uniqueness of the COM framework. After reviewing these important elements, I will debate how these theories are interrelated, which form the basis of the COM framework. Chapter 4 summarizes the basic methodological elements: construction of TOs (Madsen, et al., 2004); application of a novel unsupervised learning algorithm - the Infinite Relational Model (Kemps, et al., 2006); and the basic ontology mapping algorithm – Tversky's Contrast Model (Tversky, 1977); and the Bayesian Model of Generalization that is identified as the most appropriate algorithm for the COM framework proposed by cognitive scientists (Tenenbaum & Griffiths, 2001). Chapter 5 consists of seven published papers I have authored during this PhD project. Each paper presents a small progress in the series of empirical studies. Chapter 6 is discussing the results obtained from the empirical studies in contrast to the applied theories and human assessment, followed by future directions to be taken for the next research undertakings. After discussing the future prospects, the thesis is finally concluded by Chapter 7.

Chapter 2.

Relevant works

In this Chapter, the COM framework is compared with other multilingual ontology frameworks, such as the KYOTO project and the MONNET project. The comparison is based on a number of dimensions defined in (Cimiano et al., 2010). These dimensions are used in categorizing different types of *ontology localizations.* The term "ontology localization" is originally defined in (Suárez-Figueroa & Gómez-Pérez, 2008) as *the adaptation of an ontology to a particular language and culture*. This has been slightly re-formulated by (Cimiano et al, 2010) as *the process of adapting a given ontology to the needs to a certain community, which can be characterized by a common language, a common culture or a certain geo-political environment.* Cimiano et al. (2010) view the ontology localization problems from two different aspects: a *lexical layer* and a *conceptual layer*. The lexical layer of an ontology refers to labels of concepts, properties and other elements used to describe concepts. On the other hand, the conceptual layer implies a conceptual structure itself that may need to be adapted due to *a different cultural or geo-political context*. Cimiano et al. (2010) emphasize that *the adaptation of the conceptualization layer will be primarily driven by the inexistence of conceptual equivalents (or concepts with the same granularity level) in the target community whenever the final purpose of the ontology is to be equally valid in the source and target communities.* Accordingly, they describe how the localization of the different layers (lexical and conceptual) interact with each other, and introduce different dimensions characterizing the localization process based on the outline defined in (Espinoza et al., 2009).

***International (standardized) domain vs. culturally influenced domain*:** First of all, ontologies can be categorized based on characteristics of domains. Some domains are internationally standardized as they are identified in e.g. the technical domain or the financial report domain. On the other hand, some domains are culturally dependent. For example, domains such as related to taxation systems, legal systems, political systems, social welfare systems and educational systems are all culturally and geo-politically influenced. An ontology developed for a culturally influenced domain is referred to as a *culturally-dependent ontology* which is one of the main focuses in this thesis.

Functional vs. documental localization: Based on the so-called functionalist theories to translation by (Nord, 1997), Cimiano et al., (2010) argue that an *ontology might be localized with different goals in mind*. Their idea of *functional localization* is that, if a domain is culturally influenced, e.g. a different geo-political reality for a target community, the original source ontology has to be adapted to the target community by maintaining similar functions in both communities. It means that it is necessary to change the conceptual structure in a source community to fulfill the requirements of a target community. Hence, functional localization implies the creation of a new ontology on the basis of the original one. Contrary, documental localization means that the original ontology can be supported by members of another linguistic community. Hence, what is required in this type of localization is to document the meaning of the original ontology for a different language community.

***Interoperable vs. independent ontology*:** Interoperable ontology means that a new target ontology and the original source ontology are to a certain extent interoperable with each other. Thus, any changes to the conceptual structure are restricted in order to ensure a certain degree of interoperability. On the other hand, independent ontology means that the target ontology corresponds to the source ontology only in a functional manner so that significant changes to the conceptual structure of the target ontology are acceptable in order to meet the needs and capture the specificities of the target community.

In addition to these dimensions proposed by (Cimiano, et al., 2010), the following two major relevant works are analyzed from the viewpoint of whether the styles of cross-cultural communication achieved by these multilingual ontology solutions are bi-directional (interactive) or mono-directional (assimilative); and whether concepts existing in two cultures are linked in the form of symmetric or asymmetric relations. The analysis leads to an idea of how the COM framework can be uniquely positioned compared aforementioned major multilingual-ontology projects: MONNET and KYOTO.

1. MONNET

The FP 7 European Union-funded project, MONNET, on Multilingual ONtologies for NETworked Knowledge is a project that offers solutions for a cross-language information access problem based on MT and Semantic Web technology. The project is targeted for business and governmental users. Hence the main project employs two specific sectors consisting of several use cases.

The first project area is targeted for cross-lingual information access problems in the financial sector. One of the use cases employs XBRL (eXtensible Business Reporting Language) taxonomies that are an XML-based open standard for identifying and communicating complex financial information in corporate business reports. Its aim is to localize ontologies for supporting not only the translation of the central elements of business reports in Dutch, English, German and Spanish, but also the extraction, integration and presentation of financial data available in various types of documents in various languages, in order to contribute to the effective sharing of financial and business knowledge across Europe (Declerck et al. 2010). Hence the starting point of this use case is that the language-independent universal ontology is localized into several languages. In other words, this use case specifically clarifies that it is not the aim to restructure or to align the XBRL taxonomies for different countries or legislations. In the case that some concept is not directly translatable into a target language in a one-to-one fashion, paraphrases will be used instead. However, automatic generation of appropriate paraphrases is not within the scope of their localization component (Declerck et al. 2010). Hence the case is obviously dealing with an *international (standardized) domain*, and *documental localization*. Since this case does not create a new ontology, interoperability is obviously not required.

The second project area is targeted for cross-lingual information access problems in the public sector. One of the use cases deals with a localization of a Customer Service Guide (*Klantdienstwijzer* in Dutch) of the Immigration and Naturalization Office (IND) that is an application providing information and advice about rights and obligations for foreign citizens who want to live/work/come to the Netherlands. It means that this use case deals only with translation of Dutch legal concepts into other languages, i.e. there is no inherent need of aligning legal concepts across different countries/cultures. Though the ontology is in the *culturally influenced domain*, the localization of this application has a rather standardized aspect, and is considered as *documental localization*.

2. KYOTO

Another FP 7 European Union-funded project, KYOTO, on Knowledge-Yielding Ontologies for Transition-based Organization, pursues the idea of helping domain specific communities to model terms and concepts in their domain and to use this knowledge to apply text mining on documents (Cuadros et al. 2010). The KYOTO project aims at constructing a system that facilitates the exchange of information across cultures, domains and languages. According to (Vossen et al., 2008), the KYOTO project is based on *the assumption that language reflects culture and that the linguistic encoding of knowledge and information is therefore culturally biased.* Hence, the KYOTO platform establishes semantic interoperability across languages for a specific domain by creating language-specific WordNets that are inter-connected through a shared ontology. WordNet (Miller, 1995; Fellbaum, 1998) *is a large electronic lexical resource for English, organized as a semantic network (an acyclic graph).* WordNet is structured in the way that words and short phrases that possess the same conceptual meaning are grouped as a synonym set – called a *synset.* Synsets are interlinked with labeled arcs depending on different types of semantic and lexical relations such as synonymous relations, super- or subordinate relations, part- or whole relations and so on. These relations enable one to measure semantic relatedness of two synsets in the WordNet based on a distance of links that connect between the two synsets. Following this English WordNet, similar semantic networks have in recent years been developed in many other languages among others Dutch, Spanish, German, Italian, Japanese and Chinese. The KYOTO platform connects these language-specific wordnets by anchoring them to a generic language-independent wordnet called *Interlingua* (Vossen, 2004) and to a shared upper ontology such as the so-called DOLCE (Masalo et al., 2003) and SUMO (Niles & Pease, 2001). In practice, Vossen et al. (2008) developed the system for the domain of environment. Thus, culturally-specific concepts in the environmental domain are first connected to a language-specific domain-specific wordnet, and then further mapped to the shared ontology. The most notable point in the KYOTO framework is that it separates the linguistically and culturally-specific vocabularies from the shared ontology while using the shared ontology as a point of interface for the concepts used by the various communities (Vossen and Rigau, 2010). Cimiano et al. (2010) categorizes the KYOTO framework based on its dimensions as follows: the KYOTO project that deals with the *culturally influenced* domain; is considered as *functional localization* of a general lexicon to different target languages; and maintains a considerable degree of *interoperability through a mediator-ontology*, while each wordnet and its language-specific vocabularies in a target culture are considered as *independent*.

3. Culturally-specific Ontology Mapping

Terminological Ontology (TO) is a domain-specific ontology used for knowledge sharing (Madsen et al. 2004), which normally is applied to terminology work, cf. for example (ISO, 2000). The unique points of the TO method that differentiate it from other types of ontologies are primarily its feature specifications and subdivision criteria (Madsen et al., 2004). A feature specification consists of a feature dimension and its value. Thus, the representation of a whole concept is a feature structure, i.e. a set of feature specifications corresponding to the unique set of characteristics that constitutes a particular concept (Madsen et al., 2004). What it basically means is that a hierarchical structure of concepts - i.e. a TO - is shaped based on the feature specifications and subdivision criteria. Terminological ontologists argue that concepts are defined in a language-dependent context, and therefore a TO is language dependent. A TO is developed within a knowledge sharing community, then dynamically updated and validated. If it is necessary to share knowledge with other communities, TOs developed in different communities should be compared, aligned and merged, as needed. While the aforementioned two mainstream projects; MONNET and KYOTO, both deal with complex ontologies involving huge data-sets, TOs usually handle smaller amounts of concepts. Based on the dimensions defined by (Cimiano et al., 2010)[4], TOs could be categorized as follows:

1. The TO method handles a *culturally influenced domain*;
2. Since a TO is independently developed within a knowledge sharing community, it is considered as an *independent ontology*; and
3. Since concepts are defined in a language-dependent context, a conceptual structure is completely adapted to a target community. It means that the TO method is considered as a *functional localization*.

This categorization of the TO is indeed identical to the KYOTO framework. One notable point is that TOs that are developed in different language communities should directly be compared, aligned and merged as needed, while the KYOTO framework employs a shared ontology to anchor culturally-specific ontologies and maintains a considerable degree of *interoperability through a mediator-ontology.* Therefore, in my opinion, the KYOTO framework should rather be categorized as an *interoperable ontology* compared with the TO that maintains 100% independency. Another notable point that should be emphasized here is that a TO could potentially be a suitable tool for simulating cognitive processes explaining a real-life cross-cultural communication scenario. Since a TO maintains a hierarchical structure constructed based on conceptual features possessed by each concept, this enables one to compute feature-based semantic relatedness between two concepts, while the KYOTO framework allows one to compute semantic relatedness based on the distance-based measure. This difference will result in an argument that the feature-based measure enables one to compute either a symmetric- or an asymmetric relationship between two concepts, while the distance-based measure is only limited to compute a symmetric relationship between two concepts. This argument leads to one of the most important keywords in this thesis, namely *asymmetric* coordination in communication originally suggested in Sperber & Wilson´s Relevance Theory of Communication (1986), which will be carefully reviewed in the next chapter.

Before entering the next chapter, Figure 2-1 illustrates the summary of comparisons based on the dimensions proposed by (Cimiano, et al.,

[4] Strictly speaking, TO does not involve the ontology localization process described in (Cimiano et al., 2010), because TO is independently developed for an individual knowledge community. Thus, the dimensions of (Cimiano et al., 2010) are applied here purely for the purpose of comparison.

2010). In Figure 2-1, the culturally influenced domain is further divided into an "assimilative setting" and an "interactive setting". If an ontology for a culturally influenced domain requires documental localization such as the use case of the Dutch immigration service of the MONNET project, a communication style could rather be considered as an *assimilative setting*. On the other hand, if e.g. the Swedish governmental office and the Japanese governmental office are exchanging information in order to establish a pension welfare treaty, the communication style would rather be considered as an *interactive setting*. The *interactive setting* can be further divided into a *symmetric relation* and an *asymmetric relation*. In the case of the KYOTO framework, culturally dependent ontologies are linked via an intermediate common ontology and semantic relevance is measured based on the distance-based similarity measure. Hence the similarity relations between two concepts are symmetric. On the other hand, the COM framework employs a feature-based measure that computes an *asymmetric similarity relation* between two concepts. Based on these additional dimensions, Figure 2-1 clearly illustrates the position of the COM framework in comparison with the MONNET project and the KYOTO project.

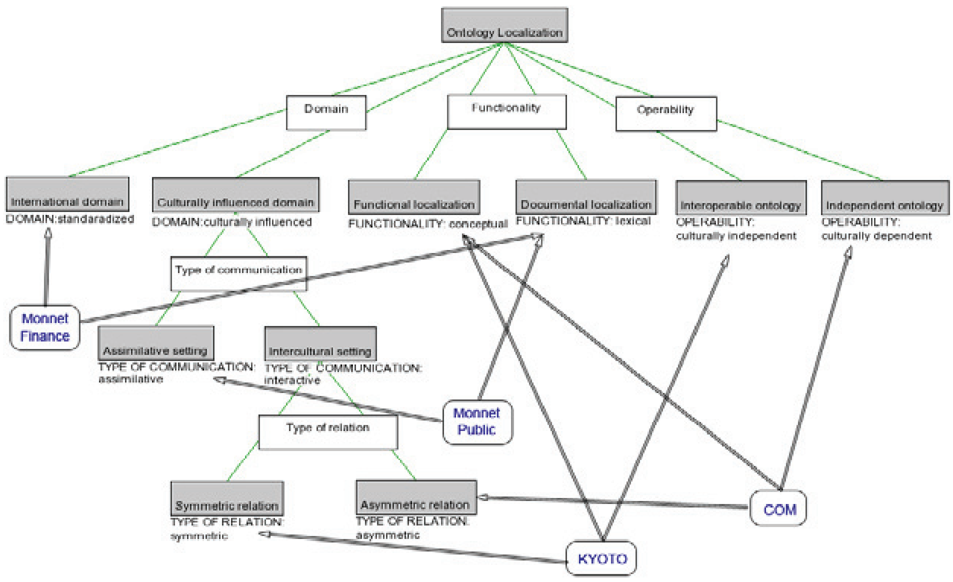


Figure 2-1: Dimensions of ontology localization

Chapter 3.

Theories

In this chapter, the following theories that support the COM framework are first reviewed: **1)** the Relevance Theory of Communication; and **2)** the Knowledge Effects proposed in the area of concept learning and categorization research in the cognitive sciences. These two theories are derived from different research domains, respectively, the communication and pragmatics studies, and the concept learning and categorization studies in cognitive sciences. Despite the fact, that these theories originate from different research domains, they directly or indirectly deal with issues of representing prior knowledge and the issue on how knowledge possessed by individuals is linked up in human communication. The most important keywords are *asymmetric* coordination in communication and taxonomic organization of concepts based on features. In the sections that follow, these theories are first reviewed independently. Finally, in the last section, it is carefully considered how they can be inter-related and applied to the COM framework.

1. The Relevance Theory of Communication

The aim of Sperber & Wilson's work (1986) is *to identify underlying mechanisms, rooted in human psychology, which explain how humans communicate with one another* (Sperber & Wilson, 1986:32). They consider Grice's theory as point of departure for their theoretical approach. Grice's analysis (1957), in their point of view, employs a model, the so-called *code model*, as the basic explanation of communication. In the code model, communication is achieved by: **1)** a message originating from an information source is encoded by an encoder into a signal; **2)** the signal is transferred via a channel, which may potentially pose a risk of destroying or distorting the signal; and **3)** the signal is finally decoded by a decoder into a message that will be received at its destination (Shannon and Weaver's diagram, 1949). When it comes to human communication, a language is considered as *a code which pairs phonetic and semantic representations of sentences* (Sperber & Wilson, 1986:9). Sperber & Wilson, however, argue that *there is a gap between the semantic representation of sentences and the thoughts actually communicated by utterances. This gap is filled not by more coding, but by inference* (Sperber & Wilson, 1986:9). In an inferential process, input data is *a set of premises* which logically result in *a set of conclusions* as output that are justified by the premises. A key point here is that the set of premises constitutes a *context*, which is a subset of the audience's assumptions about the world (*aka* prior knowledge). On the other hand, the framework of the code model requires *mutual knowledge*, in order to communicate a message by encoding and decoding. According to Sperber & Wilson, these two types of processes constitute communication in the following way: *communication requires some degree of co-ordination between communicator and audience on the choice of a code and a context. The notion of mutual knowledge is used to explain how this co-ordination can be achieved: given enough mutual knowledge, communicator and audience can make "symmetrical" choices of code and context* (Sperber & Wilson, 1986: 43). However, the *symmetrical* choice of code and context is not sufficient for establishing optimal communication between communicator and audience. Sperber & Wilson emphasizes that, although *all humans live in the physical world*, mental representations are constructed differently due to differences in our close environment and our different cognitive abilities. Because people use different languages and have mastered different concepts, the way they construct representations and make inference is also dissimilar. Sperber & Wilson call this narrower physical environment a *cognitive environment*. Especially, they focus on a human´s conceptual cognitive abilities involved in communication and emphasize that *manifest facts* are important elements for conceptual cognition. In their definition, *a fact is manifest to an individual at a given time if and only if he is capable at that time of representing it mentally and accepting its representation as true or probably true; and a cognitive environment of an individual is a set of facts that are manifest to him* (Sperber & Wilson, 1986: 39). As described above, communication is achieved by coordination between communicator and audience on the choice of a code and a context. A cognitive environment that is shared by two people is the set of all facts that are manifest to both communicator and reader. However, according to Sperber & Wilson, a realistic notion of *mutual manifestness* is not strong enough to support the *symmetrical* choice of code and context. Since an individual possesses a total cognitive environment that is the set of facts based on his/her perceptual ability, inferential ability, actual awareness of facts, knowledge he/she has acquired and so on, it is much easier to achieve "*asymmetric*" coordination between communicator and audience. This is in a way obvious because *human beings somehow manage to communicate in situations where a great deal can be assumed about what is manifest to themselves and others, but nothing can be assumed to be truly mutually known or assumed* (Sperber & Wilson, 1986:45). Sperber & Wilson (1986: 49) also use the word *ostension* referring to behaviour of initiating and realizing a manifest situation when humans have an intention to make something manifest. Based on this definition, they define human communication as ostensive-inferential communication in the following manner:

> *The communicator produces a stimulus which makes it mutually manifest to communicator and audience that the communicator intends, by means of this stimulus, to make manifest or more manifest to the audience a set of assumptions* (Sperber & Wilson, 1986:63).

Based on their definition of human communication that is assumed to be the asymmetrical coordination between communicator and audience, Sperber & Wilson (1986) argue that the core ability of humans to *perform spontaneous demonstrative inference is a set of deductive rules,* that is, *a set of computations which take account of the semantic properties of assumptions.* What is referred to in their argument about "the semantic properties of assumption" is the semantic properties of *a structured set of concepts* (Sperber & Wilson, 1986: 85). Concepts are defined here as *psychological objects considered at a fairly abstract level*. They assume that each concept possesses a label that function as an address in memory under which different types of information can be stored and retrieved. At the same time this label may also appear as an element of a logical form. The information that may be stored in memory at a certain conceptual address can be distinguished into the following three types: a logical entry; an encyclopaedic entry; and a lexical entry. Accordingly, *a conceptual address is a point of access to the logical, encyclopaedic and linguistic information which may be* *needed in the processing of logical forms containing that address* (Sperber & Wilson, 1986: 86).

The logical entry for each concept consists of a set of deductive rules. These deductive rules describe a set of premises and conclusions *in which there is a specified occurrence of that concept, and yield only conclusions from which that occurrence has been removed* (Sperber & Wilson, 1986:86). These elimination rules are defined based on standard logics, for example (Sperber & Wilson, 1986: 86-87):

And-elimination

(a) *Input*: (*P* and *Q*)
    Output: *P*

(b) *Input*: (*P* and *Q*)
    Output: *Q*

Modus ponendo ponens

Input: (i) *P*
         (ii) (If *P* then *Q*)
Output: *Q*

Modus tollendo ponens

(c) *Input*: (i) (*P* or *Q*)
             (ii) (not *P*)
    Output: *Q*

(d) *Input*: (i) (*P* or *Q*)
             (ii) (not Q)
    Output: *P*

The encyclopaedic entry consists of information about extension and/or denotation of concepts, i.e. information about the objects, events and/or characteristics that describe a specific concept. In other words, the encyclopaedic entry is the organization of conceptual information in human memory. It means that the encyclopaedic entries are typically organized differently in an individual's memory since new information is added to them constantly in his/her life. Finally, the lexical entry refers to information about a natural-language expression for a specific concept. The lexical entries are assumed to contain for example syntactic and phonological information.

Based on these definitions, Sperber & Wilson (1986) outline a formal deduction system for modeling *the system used by human beings* *in spontaneous inference, and in normal utterance comprehension in particular* (Sperber & Wilson, 1986: 94). The device they propose is defined as follows:

> *The device we envisage is an automaton with a memory and the ability to read, write and erase logical forms, compare their formal properties, store them in memory, and access the deductive rules contained in the logical entries for concepts. Deductions proceed as follows. A set of assumptions which will constitute the axioms, or initial theses, of the deduction are placed in the memory of the device. It reads each of these assumptions, accesses the logical entries of each of its constituent concepts, applied any rule whose structural description is satisfied by that assumption, and writes the resulting assumption down in its memory as a derived thesis. Where a rule provides descriptions of two input assumptions, the device checks to see whether it has in memory an appropriate pair of assumptions; if so, it writes the output assumption down in its memory as a derived thesis. The process applied to all initial and derived theses until no further deductions are possible* (Sperber & Wilson, 1986: 95).

Their hypothesis is that a human deductive device has access only to elimination rules and results only in non-trivial conclusions. The non-trivial logical implication is defined as follows:

> *A set of assumptions P logically and non-trivially implies an assumption Q if and only if, when P is the set of initial thesis in a derivation involving only elimination rules, Q belongs to the set of final theses* (Sperber & Wilson, 1986: 97).

They claim that, in normal circumstances, only non-trivial implications play a role in the comprehension process. They further suggest that the deductive device enables one to realize the improvements generated by new information to a stock of factual assumptions (an existing representation of the world). Such a process is a so-called *contextual implication* of *P* in *C* where *P* may be thought of as new information, and *C* as old information in the memory. In here, they call *a deduction based on the union of P and C as premises a contextualization of P in the context C* (Sperber & Wilson, 1986: 107). The contextual implication of *P* in *C* can be defined as (Sperber & Wilson, 1986: 107-108):

> ***A set of assumptions P contextually implies an assumption Q in the context C if and only if***
> *(i) The Union of P and C non-trivially implies Q,*
> *(ii) P does not non-trivially imply Q, and*
> *(iii) C does not non-trivially imply Q*

Accordingly, this can be interpreted in the following way: *A contextual implication is new information in the sense that it could not have been derived from C, the stock of existing assumptions, alone; however, it is not just new information, the newly presented information alone. It is a synthesis of old and new information, a result of interaction between the two* (Sperber & Wilson, 1986: 108). Such new information generated based on the union of new information *P* and old information *C* is called *contextual effects* in (Sperber & Wilson, 1986). They argue that *having contextual effects is a necessary condition for relevance* (Sperber & Wilson, 1986: 119). Their assumption is that people have intuition of relevance and they are consistently able to distinguish relevant from irrelevant information. Their claim is:

> *An assumption which has no contextual effect in a given context is irrelevant in that context. In other words, having some contextual effect in a context is a necessary condition for relevance* (Sperber & Wilson, 1986: 121).

In other words, the term, *relevance* is defined as follows:

> *An assumption is relevant in a context if and only if it has some contextual effect in that context* (Sperber & Wilson, 1986: 122).

There are two important dimensions in terms of *relevance* raised by Sperber & Wilson (1986). The first dimension is how to assess the degree of relevance. They argue that there are two factors to be taken into account for assessing the degree of relevance. The first factor is obviously the contextual effects. Secondly, the contextual effects are the results of a mental process. The processing effort involved in achieving a contextual effect should therefore be considered as the second factor.

> *Extent condition 1: an assumption is relevant in a context to the extent that its contextual effects in this context are large.*
> *Extent condition 2: an assumption is relevant in a context to the extent that the effort required to process it in this context is small* (Sperber & Wilson, 1986: 125).

In order to measure these effects and efforts, Sperber & Wilson emphasize the *comparative notion of relevance* instead of a quantitative. They claim that the comparative judgments are intuitive, since contextual effects and processing effort are non-representational dimensions of mental processes. Their speculation is that such an internal psychological mechanism of contextual effects and mental effort could be monitored by physico-chemical parameters. However, no concrete empirical approach has been proposed in (Sperber & Wilson, 1986).

Another dimension is related to how contexts are chosen. It has been suggested by Sperber & Wilson (1986) that *the context used to process new assumptions is, essentially, a subset of the individual's old assumptions, with which the new assumptions combine to yield a variety of contextual effects* (Sperber & Wilson, 1986: 132). In order to process new information, it has to be combined with *an adequately selected set of background assumptions*. Such background assumptions are selected from diverse sources such as long-term memory, short-term memory and perception, in combination with each item of new information. This requires *the organization of the individual's encyclopaedic memory.* For example, *it is generally agreed that encyclopaedic information in long-term memory is organized into chunks of some kind. Such chunks have been discussed in the literature under such names as "schema", "frame", "scenario" and "prototype"* (Sperber & Wilson, 1986: 138). It means that the encyclopaedic entries they have introduced are considered as organized chunks of a certain size. Smaller pieces of chunks can be grouped into larger chunks and so on. It suggests that the context is not determined at the first place. In the process of comprehension, people try to select a context which will maximize relevance. In other words, there are different sizes of contexts existing in the human memory. A larger context might contain one or several smaller contexts. A smaller context might be included in a larger context. As a result the set of accessible contexts is partly organized as inclusion relations where smaller contexts are included in larger contexts. This order of inclusion corresponds to the order of accessibility, according to Sperber & Wilson (1986). The minimal context can be immediately accessed, while contexts including the minimal context and the one-step extension as subpart can be accessed for example in two steps. This indicates that contexts that require several access steps require extra processing effort in the human mind. The final important remark on *relevance* is that an individual automatically aims at maximal relevance in terms of the effects and applied effort. In order to achieve a maximal relevance, humans, so to speak, have to identify the best possible context that can optimally process an assumption. Accordingly, Sperber & Wilson (1986) define *relevance* for an individual as follows:

> ***Extent condition1****: an assumption is relevant to an individual to the extent that the contextual effects achieved when it is optimally processed are large.*
> ***Extent condition2****: an assumption is relevant to an individual to the extent that the effort required to process it optimally is small* (Sperber & Wilson, 1986: 144-145).

2. Knowledge Effects: Category Learning, Taxonomic Organization of Categories and Category-based Inference

"The Big book of Concepts" written by Gregory L. Murphy (2004: 1) starts with the following statement: *Concepts are the glue that holds our mental world together. .... If we have formed a concept (mental representation) corresponding to that category (class of objects in the world), then the concept will help us understand and respond appropriately to a new entity in that category. Concepts are a kind of mental glue, then, in that they tie our past experiences to our present interactions with the world, and because the concepts themselves are connected to our knowledge structures.*

This statement inherently reflects that the study of concepts is highly connected to the study of communication that has just been dealt with in the previous section. However, in the study of concepts, the focus is rather on a representation of conceptual knowledge – relationships between groups of objects and features possessed by these objects. The representation of conceptual knowledge enables humans to learn new concepts and to make category-based induction which forms the basis for cross-cultural communication. Therefore, I will first review Murphy´s "The Big book of Concepts" describing the basic elements that have been discussed in the history of concept studies and issues relevant to the knowledge representation and communication.

Three major approaches

Within the cognitive sciences discipline, there have been two major views on concepts: the so-called prototype view and the exemplar view. The original idea of **the prototype view** is rooted in (Rosch and Mervis, 1975). In Rosch and Mervis's paradigm, a concept is represented as features that are typically identified among the category members. Among these features, the ones that appear frequently in a category are considered more important than the remaining in the category. Assuming that a feature list (a set of features) is a concept representation, the categorization process of a new object is the computation of similarity of the new item measured up against the existing feature list. It means, that if a feature that is commonly possessed by both a new object and the representation, the feature in question receives "credit". On the contrary, if a feature is possessed only by the representation or by the object, the feature in question loses "credit" (Tversky, 1977). In the prototypical view, all features identified in the new object are contrasted with features that are possessed by the representation. Afterwards, the weights of features that are present in both the object and the representation are added up, and all the weights of object's features that do not exist in the category representation are subtracted. Based on this computation, the object is judged to be in the category or not, according to an arbitrarily defined categorization criterion (Murphy, 2004: 41-48).

The exemplar view has first been proposed by (Medin and Schaffer, 1978). This view does not consider an individual's concept representation as a definition that refers to all category members of that representation nor as a list of features that are possessed to greater or lesser degrees in the category members of that representation. In this view, an individual's concept representation is considered as the set of category members which the person remembers. Again, similarity plays a key role in the exemplar theory. Assuming that an individual's concept of dogs is the set of dogs that the person remembers, a new object, say another animal, observed by this person, should be weighted on the basis of how similar his/her memory of dogs is to the new given animal object. A key aspect in this view is therefore to assess *how similar the object is to each memory.* (Murphy, 2004: 49-60).

A new theory called **the knowledge approach** has more recently appeared as a reaction to the prototype- and exemplar approaches. This new approach is in some sense based on the two aforementioned views. The idea of the knowledge approach is that when people learn a new concept, for example related to animals, the new information about animals is integrated with one's prior knowledge about biology, animal behavior or other relevant domain knowledge. The relation between the new concept and the prior knowledge is bi-directional: i.e. the learning process of the new concept can be influenced by the prior knowledge, while the new information of this concept may also influence the prior knowledge. The knowledge approach considers concepts as part of our general knowledge of the world. It means that concepts should be consistent with whatever else we know (Keil 1989; Murphy and Medin 1985). In order to maintain such consistency, people use their prior knowledge to reason about a new object and decide what category it is or to learn a new category. Unlike the prototype- and exemplar view, the knowledge view claims that *people do not rely on simple observations or feature learnings in order to learn new concepts. They pay attention to the features that their prior knowledge says are the important ones* (Murphy, 2004: 60-64).

Category learning

A diverse range of studies based on these three major views have been conducted towards various cognitive phenomena identified in our daily life: among others, category learning and category construction are some of the basic cognitive processes that are commonly involved in our daily life. In (Murphy, 2004), the cognitive processes involved in these phenomena are broadly covered from aspects of the three major approaches: the prototype approach; the exemplar approach; and the knowledge approach.

One interesting debate in the category learning research is centered on identifying which kind of features that will most likely influence a human's categorization process. Murphy (2004) introduces some examples from literatures, among others: Nosofsky (1988)'s reasoning indicating that *when one item occurs more than others, it is generally considered to be more typical and is categorized faster than less frequent items;* and in (Rosch and Mervis, 1975), *when a feature occurs more often in a category, items that possess it are learned faster and have greater typicality.* Opposed to these arguments of typicality and feature-frequency relations, some systematic exceptions are identified as the *inverse base-rate effect* (Medin and Edelson, 1988) or the *apparent base-rate neglect* (Gluck and Bower, 1988). As an example of these studies, a fictional scenario based on Gluck and Bower (1988) is illustrated as follows: In a category of a rare disease, a symptom "1" occurs 60 % of the entire cases, while in a category of a common disease, the same symptom "1" occurs only 20 % of the entire cases. Assuming that the common disease occurrence is three times higher than the rare disease, the absolute number of occurrences for symptom "1" must be equal in the common- and rare diseases. However, in their studies, subjects judged that the feature of symptom "1" was to be in the rare disease category and ignored the absolute number of category frequencies. Kruschke (1996) points out that *the base-rate neglect phenomenon occurs because of attention shifting from the common category to the most predictive features of the rare category* (Murphy, 2004: 115-118).

Another interesting issue within the category learning research, is the feature correlation in concepts. Rosch et al. (1976) argues that *categories contain clusters of correlated attributes that are fairly distinct from other clusters. The prototype structure of concepts is one way of representing these clusters of correlated features.* However, from another viewpoint, Murphy (2004) indicates that *if two features are both quite frequent within a category, a "correlated" link would probably not be necessary, because one can infer that the features will be co-occuring.* To investigate such *feature-feature correlations* within a category, there have been a number of experiments using subjects for making categorization judgments. However, these investigations have shown little evidence of such feature-feature correlations within a category. Murphy and Wisnienski (1989), based on their experiment, argue that subjects are not good at learning feature-feature correlations, and they are instead trying to learn how each feature is related to a specific category. In the view of Murphy and Wisnienski (1989), subjects use their prior knowledge when learning categories. Therefore, *people are primarily learning how features are related to the different categories, not how they are correlated within the categories.* In other words, Murphy indicates that the learning of feature-category relations implicitly results in feature-feature correlations within a category. Murphy concludes with his noteworthy remarks (Murphy, 2004: 118-125):

> *There are so many possible pairs of correlated features in natural categories that it is difficult to see how one could consider them all. One of the benefits of categories is that they allow us to reduce the information structure of the correlations in the environment. Instead of storing the facts that wings go with feathers, and feathers go with nests, and nests go with beaks, and wings go with beaks, and wings go with nets, and feathers go with beaks, and so on, we can store the facts that wings, feathers, nests and beaks are all properties of birds, which implicitly represents the feature-to-feature correlations categories* (Murphy, 2004: 125).

Category construction is another type of category learning, which is specified only for the learning of a new category without any feedback and therefore is sometimes referred to as unsupervised learning. The process of category construction is described in Murphy (2004: 126-127):

> *This process is probably most important in childhood, when children notice categories around them but may not yet have words to describe them or may not have received any feedback telling them that the things are indeed in the same category. ... As adults, we already have concepts for most of the things we encounter, and so we may not notice this process working unless we travel, encounter a recent invention, or otherwise expose ourselves to new kinds of entities*[5].

In the area of category construction research, one of the fascinating debates is what people focus on when categorizing objects. Medin, Wattenmaker, and Hampson (1987) experimented with subjects to divide artificial items consisting of four feature-dimensions with feature-values falling into two categories. Their results indicate that people use the so-called *uni-dimentional strategy*; that is, if pictures of objects are given as stimuli, subjects would first choose one dimension (e.g. size or shape etc) and subsequently divide the items up into categories that differ in that dimension (large vs. small, or square vs. round, etc.). A number of studies (Medin et al., 1987; Regehr and Brooks, 1995; Thompson, 1994; Ward et al., 1991) show the extremely strong bias towards uni-dimensional categories. However, Murphy

[5] Figure 1-1 describing the situation where a new Japanese delikattes is introduced to the Danish mindset, is a good example.

(2004: 128-129) reflects on *why people make categories that are so different from those of everyday life*. He points out the example of feature-feature correlations discussed in the category learning paragraph and emphasizes that multi-dimensionality of concepts gives much stronger induction power compared to that of the uni-dimensional. He argues that when several categorical features are related, the cluster of these features form a so-called family resemblance structure. *This can happen either through the relations of prior knowledge* (Ahn, 1990; Kaplan and Murphy, 1999; Spalding and Murphy, 1996) *or through induction that relate the features* (Lassaline and Murphy, 1996). Murphy concludes as follows:

> *When one notices a new animal, for example, one is likely to use prior knowledge to relate its physical characteristics to its habitat and behaviors. Furthermore, if one is trying to predict something about it (whether it will attack, whether it will fly away, or whether it will try to eat your garbage), one will notice relations between the features (things with sharp teeth and claws tend to attack, things with wings tend to fly away). Thus, the family resemblance structure of real categories is probably in part due to such attempts to understand and make predictions (inductions) about the world. Finally, when the categories are very highly structured (i.e., without any exceptions in the critical features), subjects can notice the structure. For any real categories that do have such strong structure, then unsupervised learning may be fairly easy* (Murphy, 2004: 133).

Knowledge effects

A key point emphasized in Murphy (2004: 141-197) is that many of the previous empirical works in the prototype- and exemplar views are based on artificial stimuli not on categories used in real-life. The reasons for this tendency are according to Murphy, first, that researchers often have a tendency to intentionally avoid using materials which subjects are already familiar with or have pre-experimental knowledge about. This is according to Murphy driven by the desire to obtain "clean" results solely reflecting on the learning process. Secondly, researchers often tend to believe that the discovery of general principles based on the use of simple stimuli can be applied to many different domains. Contrary to this, *the knowledge approach argues that what one knows outside of the concept influences how it is learned and then* *used. The concept is formed within the constraints of one's understanding of the world.* Hence, according to this viewpoint it is vital to examine *the effects of background knowledge on concept learning and use* (Murphy, 2004: 141-142). Murphy (2004) explains the knowledge effects by using an exemplar situation of visiting a Zoo. His argument is that, even before visiting the Zoo, people will have prior knowledge of the animal- or say the sub-leveled mammals domain which can be used for learning a new object within the most related domain. Hence he defines the term, *knowledge effect, as influences of prior knowledge of real objects and events that people bring to the category-learning situation* (Murphy, 2004: 146). Murphy (2004) argues that there are two main differences when comparing traditional experiments using artificial stimuli and the knowledge approach based on natural categories. The first difference is that features themselves are different. Secondly, the features of natural categories seem to be related to each other, while the features of artificial categories do not necessarily. Murphy and Allopenna (1994) shows that *when the features of a category formed a consistent set, the category was much easier to learn than when they were inconsistent or simply neutral.* Their findings suggest that *knowledge helps learning because it relates the features in the category, rather than through the properties of the features themselves* (Murphy, 2004: 151). However, in a more realistic scenario, people must also learn about properties that have not yet been connected with their background knowledge. Kaplan and Murphy's (2000) work shows that when thematic features are present in categories, background knowledge is helpful even though the knowledge is incomplete or imperfect. Subjects are also able to explain away or ignore features that are inconsistent and still be able to use the most accurate knowledge. Based on a series of studies, Murphy (2004) concludes that:

> *Background knowledge affects not only initial acquisition of a concept but also later categorization judgments. People tend to positively categorize items that are consistent with their knowledge and exclude items that are inconsistent, sometimes even overruling purely empirical sources of information.... It seems very likely that how and when knowledge influences categorization will depend on the nature of the category and whether other cues are readily available* (Murphy, 2004: 173).

However, one key question arises: where in the first place are the features of natural categories identified? One of the arguments have been that knowledge plays an important role, in other words, one's prior domain knowledge contributes to generate a set of properties that can be used for learning a new category e.g. by judging similarities between them (Goodman, 1965; Murphy and Medin, 1985; Markman and Wisniewski, 1997). An interesting finding expressed by Wisniewski and Medin's (1994) experiment where two groups (with and without prior knowledge, respectively) divide stimuli into two categories, demonstrates that *the knowledge subjects were likely to make hierarchical features of, those were the features that had both an abstract and a concrete component.* Wisniewski and Medin (1994) point out that people who is possessing prior knowledge also have strong expectations about properties of a specific category. However, Murphy argues that *those expectations are flexible and they change with actual experience with the category. Thus bottom-up and top-down processes combine to determine what features are associated with a category. Hence it is difficult to model concept learning when the features themselves are changing over the course of learning. .... To sum up, then, most work on concept learning has not said very much about where the object's features come from* (Murphy, 2004: 178-179).

Taxonomic organization of categories

The issues discussed up to now in (Murphy, 2004) have mainly focused on how people categorize objects and how they are learning new concepts based on features. Murphy (2004: 199-242) also deals with the issue of how all the categories that are learned are actually stored in our minds. The main focus is whether humans store concepts in a hierarchically organized structure. Murphy (2004) hypothesizes that a category structure is stored in the long-term memory in the form of a so-called categorical network. The categories located at the higher level in this hierarchy are called superordinate and the ones that are located at the lower level are referred to as subordinate categories. Category members are connected by the *set inclusion* relation. This relation is also referred to as an "is-a" relation in (Collins and Quillian, 1969). Murphy describes three basic principles for forming a hierarchical structure of categories. The first principle is that *any category can have only one immediate superordinate*. The second principle is that the "is-a" relation can only hold asymmetrically, so to speak. As an example of this asymmetry, all dogs are animals, but all animals are not necessarily dogs so that the "is-a" relation obviously only makes sense being applied as a uni-directional relation from dogs to animals. The third principle is that category relations are transitive so that properties of a superordinate category are inherited by sub-category members. For example, bulldogs are dogs and dogs are mammals. It means that bulldogs are mammals. If the superordinate category is, “mammals have blood”, then this particular feature of “having blood” is inherited by the sub-category, “dogs” and therefore inherited by “bulldogs”, as well. Murphy emphasizes that these hierarchical descriptions enable people to generalize the properties possessed by mammals in many categories under “mammal” in the hierarchy. This ability is important for us *to immediately access knowledge about new entities that one hasn’t had direct experience with* (Murphy, 2004: 202). Whether people actually organize their concepts in hierarchies that have these properties are further discussed from a psychological point of view in (Murphy, 2004: 199-203).

According to Murphy (2004), there are two possible ways to consider how hierarchies are mentally represented. The first possibility is that people’s concepts are pre-stored and structured in memory in the form of hierarchical networks based on “is-a” links. The view of this hierarchical network is assumed to provide a kind of cognitive economy. If the feature “breathes” is represented at the animal node, it is not necessary to store the same feature “breathes” at the mammals-, reptiles-, dogs-, or bulldogs nodes, because hierarchies automatically provide for the inheritance of properties. The second possibility is that the hierarchical structure of concepts is generated from a reasoning process instead of being explicitly pre-stored in memory. The idea is that a small group of objects X is included in a medium group of objects Y and the medium group Y is included in a larger group of objects Z. If all the items in the larger group Z have six fingers, it is obvious that all the items in the small group X also possess six fingers. The argument is that without storing the hierarchical links in the memory, information about category inclusion can be computed based on the properties that are known of a category (Rips, Shoben, and Smith, 1973). In this view of computed models, the properties that are generally true of a category are also true for its subordinates in a hierarchy. It means that more specific categories inherently have the same features as more general categories but with one or more additional features. There have been several studies comparing validities of the two views: pre-stored view and computed view. Collins and Quillian (1969) hypothesized that subjects respond faster at verifying “a pine is an evergreen” than on “a pine is a plant,” because the former involves fewer "is-a" links than the latter. Hence, their argument is that the more "is-a" links there are involved to be traversed in order to verify a sentence, the longer time it takes for us to verify that it is true. However, several subsequent studies are to some degree in opposition with Collins and Quillian's conclusions. E.g. Rips, Schoben and Smith (1973) added a new factor to Collins and Quillian's framework, namely *typicality*. For example, they compared the response time for the following two statements: "a robin is a bird" and "an ostrich is a bird." According to the superordinate and subordinate "is-a" link, both *robin* and *ostrich* is linked to bird with only one step. However, subjects responded faster to the "is-a" link "a robin is a bird". This result indicates inconsistencies with Collins and Quillian's conclusion that support the pre-stored view. Another influential study that also challenges Collins and Quillian's conclusions has been proposed by Sloman (1998). In this study, subjects were tested based on simple one-step logical inferences such as:

Birds have an ulnar artery.
Therefore robins/penguins have an ulnar artery.

Subjects were asked to rate the conditional probabilities that the conclusion was true provided that the premise was true. Two of Sloman´s important findings were that, first, subjects did not rate the probabilities of the conclusion category as "1", and secondly, subjects rated the probabilities higher for arguments with typical items (such as robins) than arguments with atypical items (such as penguins). Sloman (1998: 28) argues that *Inclusion relations are part of the set of rules that we can apply given the right conditions, not the set of relations that comprise memory structure.* Sloman emphasizes that subjects compute the similarity of one category to its superordinate concepts, and based on the similarity they draw inferences. As Sloman (1998) demonstrated, Murphy (2004: 209) summarizes that *the feature model, which say that hierarchies are computed rather than pre-stored, have the edge in this battle.* However, Murphy also points out that *a list of features may also be insufficient to account for all the data. At least some category relations may be explicitly learned and represented.* Accordingly, a possible conclusion indicated by Murphy (2004: 209) is that *it seems most likely that some category relations are directly learned and stored, whereas others are computed as needed. Although perhaps inelegant, this kind of mixed model may be the best explanation of the jumbled set of facts that people know about the world.*

Category-based inference

Murphy (2004: 241) argues that *one of the major uses of categories is to make predictions about novel items.* He emphasizes that category-based inferences are very common in communication. By referring something to a simple category name, speakers inherently assume that listeners will retrieve information about the category and use it to comprehend the meaning. Murphy specifies that the inferential process employed in this course of action is *category-based induction*. One reason the term *induction* is used for referring to this process is that it involves drawing an uncertain inference to the category as a whole. Especially, people are attempting to draw conclusions about one category or a new encountered item based on knowledge of another category. According to Murphy, categories whose members are extremely similar would have stronger inductive power than those that are less similar. Thus, more specific categories usually allow stronger induction than general categories.

According to Murphy (2004: 245), the modern study of category-based induction originated from (Rips, 1975). In Rips´ study, similarities were used for measuring the category structure of birds and mammals. One unique approach Rips applied in his study was to use so-called *blank predicates*. In these blank predicates, unfamiliar properties which people usually do not know about are specified. Murphy uses an example of the predicates, "(members of the category: birds) have sesamoid bones." Since most people do not know what the "sesamoid bones" are, people cannot answer this question based on their prior knowledge. On the other hand, people can use their prior knowledge directly to the predicates included "(members of the category: birds) can fly." In Rips´ study, the single predicate, whether members of a category would catch an unknown disease, was given to subjects. Here, the bird category "ducks" was used as a given category and "robins" and "geese" were respectively used as target categories. Subjects were asked to come up with the percentage of members of each category that would potentially catch this unknown disease. Accordingly, Rips found that there are primarily two things governing that subjects tend to extend this blank predicate to a target category. Those are the *typicality* of the given category, and the *similarity* of the given category to the target category. An example of the typicality case is that inductions of a given category of robins were stronger than that of eagles. For the similarity case, an induction from gees to ducks was stronger than from gees to cardinals. Murphy emphasizes that (2004: 246-247):

The similarity of the given and target items is not a question of what category or categories they belong to. Indeed, the given and target items could be individual objects not known to be in any category. However, the typicality question is obviously a matter of categorization.

Murphy (2004: 247) also points out an important notion of Rips' model that *the induction from X to Y is not in general the same as Y to X (see also Tversky, 1977). For example, the induction from a robin to a goose is more certain than the induction from a goose to a robin. Although the similarity between the two items is the same in both directions, robin is much more typical than goose is, and so it provides a better basis for induction. This predication was confirmed by Osherson et al. (1990), as well as by Rips.*

As pointed out by Murphy, the studies conducted by (Osherson et al., 1990) support the asymmetric view of similarities. Osherson et al. proposed arguments applying the blank predicates. They have proposed a model of reasoning processes using these arguments consisting of two components: the similarity of the premise categories to the conclusion category; and a component called *coverage*. To be more specific about the coverage, *when the premises are distributed across a certain category, subjects can become more confident that the property is true of the entire category.* Murphy (2004: 248-250) presents a number of interesting phenomena of inductive reasoning demonstrated by Osherson et al. (1990) as follows:

Premise typicality: *In general, the more typical the item or items in the premise, the stronger the argument is. So (1) is stronger than (2):*

(1) <u>Robins have a high potassium concentration in their blood</u>
All birds have a high potassium concentration in their blood

(2) <u>Penguins have a high potassium concentration in their blood</u>
All birds have a high potassium concentration in their blood

Premise diversity: *The more diverse the premise categories, the stronger the argument. So (4) is stronger than (3):*

(3) Rhinoceroses require Vitamin K
Hippopotamuses require Vitamin K
Humans require Vitamin K

(4) Bats require Vitamin K
Hippopotamuses require Vitamin K
Humans require Vitamin K

Premise monotonicity: *If you add more categories (all of them being at the same level) to the premises, the argument gets stronger.*

(5) Foxes have sesamoid bones
Pigs have sesamoid bones
Gorillas have sesamoid bones

(6) Foxes have sesamoid bones
Pigs have sesamoid bones
Wolves have sesamoid bones
Gorillas have sesamoid bones

Murphy emphasizes that the studies of Osherson et al. (1990) have even extended into related areas such as psychology of reasoning. Such extended works put emphasis on the influence of the similarity-based approach adopted for categorical inductions. However, Murphy also points out some limitations of the Osherson et al. (1990) model. One critique raised by Murphy is based on Heit and Rubinstein's study (1994) concluding that *the Osherson et al. model relies on only one similarity measure to predict induction, and so the fact that items are similar or dissimilar in different respects is not relevant.* Heit and Rubinstein (1994: 420) point out that *prior knowledge could be used dynamically to focus on certain features when similarity is evaluated. In this conception, inductive reasoning is an active process in which people identify features in the premise and the conclusion categories that are relevant to the property being inferred.* Murphy (2004: 253-254) concludes by citing Heit and Rubinsterin's words that

Rather than judging overall similarity, as in the Osherson et al. model, people may more specifically focus on the particularly relevant way in which the categories are related. One way to think of this is that the similarity computation is based not on the concepts' total properties but on the properties that are specifically relevant to the predicate under question (Heit and Rubinstein, 1994: 421)

3. Why and how these theories support the COM framework?

As outlined in the previous sections, the Relevance Theory of Communication and the Knowledge Effects are substantially consistent from a plurality of aspects. In my view, the Relevance Theory of Communication (Sperber & Wilson, 1986) forms a larger framework (RTC framework) that accommodates a communication scenario consisting of an information sender, an information receiver, and an inferential process that bridges the two parties. However, there are a number of critical issues on whether Sperber & Wilson's idea is straightforwardly applicable. When considering that the scope of this thesis to apply the COM framework to the mapping of domain-specific concepts, it raises the question whether Sperber & Wilson's idea, that a concept consisting of a logical entry, an encyclopaedic entry and a lexical entry, is directly applicable. Another question is, while the recent empirical studies introduced in (Murphy, 2004) emphasize the inductive power of categories, Sperber & Wilson's RTC framework employs a deductive system. Hence, I propose to replace these critical elements within Sperber & Wilson's RTC framework with elements based on the Knowledge Effects that empirically explain phenomena of human concept learning and categorization. The result of the integration of these two theories becomes plausibly consistent with the COM framework that aims for achieving a successful cultural knowledge transfer.

First, I clarify which elements in Sperber & Wilson´s RTC framework that are, respectively, consistent or inconsistent with the COM framework. One of the directly applicable elements of the RTC framework to the COM framework is the *asymmetric* coordination between communicator and audience on the choice of code and context. As described in the previous section, the asymmetric co-ordination of code and context is built upon the symmetric co-ordination. If a cognitive environment is shared by two parties, the set of all facts is manifest to both communicator and audience and therefore this may possibly generate a common ground based on the symmetric coordination. However, from a realistic viewpoint, people use different languages and are mastering different concepts so that the way people construct mental representations and make inference are inherently different. Thus, it seems much more realistic and easier to achieve the asymmetric coordination. When looking into the fine-grained level, i.e. at the level where an audience learns an individual concept from a given stimulus, the asymmetric relation between a communicator and an audience also exists. Rips (1975) and Osherson et al. (1990) both support this view by claiming that induction from X to Y is not in general the same as from Y to X. Thus the similarity of a given category to a target category is uni-directional. In fact, the model of computing similarities based on categorical features, originally proposed by Tversky (1977), enables one to compute such asymmetric similarities.

Hence, my speculation is that the RTC element of the asymmetric coordination should be maintained and thereby the asymmetric similarity computation algorithms originally derived from the prototypical approach should be possible to integrate into the COM framework.

The second key component in the RTC framework that in my opinion seems applicable to the COM framework is the general idea of the comprehension process or the so-called *contextual implications* of new information P in a context C. This contextual implication component seems applicable to the COM framework when considering new information P as a new stimulus provided by a communicator and the context C as the prior knowledge of an information receiver. The combination of P and C generates new knowledge Q for the information receiver. When contrasting this to the category learning, Murphy (2004) argues that *if we have formed a concept (mental representation) corresponding to that category (class of objects in the world), then the concept will help us understand and respond appropriately to a new entity in that category. Concepts are a kind of mental glue, then, in that they tie our past experiences to our present interactions with the world, and because the concepts themselves are connected to our knowledge structures.* From Murphy's argument, our knowledge structures that are formed based on our past experiences is in fact the knowledge that corresponds to the context C in the RTC framework; a new entity in that category is the new information P; and the union of P and C is playing the role of "glue" generating a new form of concept

Q. Thus, Sperber & Wilson's idea of contextual implication consistently fits with the Knowledge Effects that considers *influence of prior knowledge of real objects and events that people bring to the category-learning situation* (Murphy, 2004: 146). In here, the new knowledge generated based on the union of a new object P and the prior knowledge C is called *contextual effects* in (Sperber & Wilson, 1986). This also leads to the two important dimensions of *relevance* raised by Sperber & Wilson (1986: 125): *an assumption (new information) is relevant in a context to the extent that its contextual effects in this context are large; and an assumption is relevant in a context to the extent that the effort required to process it in this context is small*. These dimensions will also be applicable to the Knowledge Effects, when it comes to applying the Bayesian Model of Generalization (Tenenbaum & Griffiths, 2001) at a later stage in this thesis.

On the other hand, there are also some critical and inconsistent components at the fine-grained level in the RTC framework which are not directly applicable to the COM framework. Those issues are related to the way Sperber & Wilson view a structured set of concepts that are stored in the human memory and the deductive system they propose based on the elimination rules.

The largest difference between Sperber & Wilson's approach (1986) and Murphy's approach (2004) is that Sperber & Wilson employs the *deduction*, while Murphy employs the *induction* as the inferential process of knowledge comprehension. In my eyes, this major difference seems to be caused by the way they view a structured set of concepts that are stored in the human memory.

Before discussing how the conceptual knowledge should be stored, I need to clarify the scope of the COM framework. As described in Chapter 2, the aim of the Culturally-specific Ontology Mapping is to identify corresponding concepts existing in two remote cultures. The scope of the concept mapping is limited to culturally-specific domain knowledge at the level of individual lexical items – not the entire utterance. Hence general vocabularies such as verbs and adjectives etc. are here considered out of scope. Thus, the approach to be taken is the terminological approach and not the syntactic or lexical semantic approach. The terminological approach considers a concept consisting of several feature specifications, which possesses one or several lexical expressions, as starting point. The approach is indeed reasonably consistent with the way concepts are represented by features and with the basic principles for forming a hierarchical structure of categories described in Murphy (2004). On the other hand, Sperber & Wilson

(1986) considers an assumption as a set of structured concepts and emphasizes logical functions of concepts. Thus they distinguish logical entry and encyclopaedic entry, which are the information attached to a concept. More specifically, Sperber & Wilson argue that encyclopaedic entries typically vary across speakers and they are open-ended. On the other hand, *logical entries are small, finite and relatively constant across speakers and times….. the content of an assumption is constrained by the logical entries of the concepts it contains, while the context in which it is processed is, at least in part, determined by their encyclopaedic entries* (Sperber & Wilson, 1986: 88-90). This implies that the encyclopaedic entries in a way correspond to the concept representation in Murphy (2004). However, the way the context is processed is highly influenced by the lexical semantic view in Sperber & Wilson (1986). The general lexical semantics considers first a lexical item that possesses several meanings as starting point. The view Sperber & Wilson take is rooted in the lexical semantic approach and is implied from their considerations that *concepts have both logical and lexical entries that provide a point of contact between input and central processes, between the linguistic input system and the deductive rules of the central conceptual system. Recovery of the content of an utterance involves the ability to identify the individual words it contains, to recover the associated concepts, and to apply the deductive rules attached to their logical entries. … We assume, then, that the "meaning" of a word is provided by the associated concept…. This allows us to maintain a somewhat ecumenical view of lexical semantics* (Sperber & Wilson, 1986:90). This difference in views of the terminological approach by Murphy (2004), and the lexical semantic approach by Sperber & Wilson (1986), results in different inferential approaches, deductions and inductions. Sperber & Wilson's RTC framework employs the *deductive* system based on the elimination rules as inferential algorithm. On the other hand, Murphy (2004) argues that the inferential process involved in communication is a category-based *induction*. For example, for the RTC framework, the elimination rule for generating a conclusion from a premise can be illustrated in the following way:

Mother-elimination rule
Premise: (X-mother-Y)
Conclusion: (X-female parent-Y)

Here, it is considered that *the meaning of a word is provided by a definition which expresses the individually necessary and jointly sufficient conditions for the word to apply. For instance, the definition of "mother" could be "female parent". If this is so, it can be represented by assigning "mother" as the lexical entry for the concept "female parent" the elimination rule in* the above logical expression (Sperber & Wilson, 1986: 90). On the contrary, Murphy (2004: 241) argues that *one of the major uses of categories is to make predictions about novel items.* He emphasizes that speakers can assume that listeners will, by referring to something with a simple category name, retrieve information about the category and use it to comprehend the meaning. The reason the term *induction* is used for referring to this process is that it involves drawing an uncertain inference of the listeners to the category as a whole. Especially, people are attempting to draw conclusions about one category or a new encountered item based on their prior knowledge of another category. In this inductive process, people use similarity judgment. According to Heit and Rubinstein (1994: 420), *prior knowledge could be used dynamically to focus on certain features when similarity is evaluated. In this conception, inductive reasoning is an active process in which people identify features in the premise and the conclusion categories that are relevant to the property being inferred.* This statement indicates that, if prior knowledge is organized in a hierarchical structure based on the terminological approach, such prior knowledge could effectively be used for the inductive reasoning as pointed out by Heit and Rubinstein.

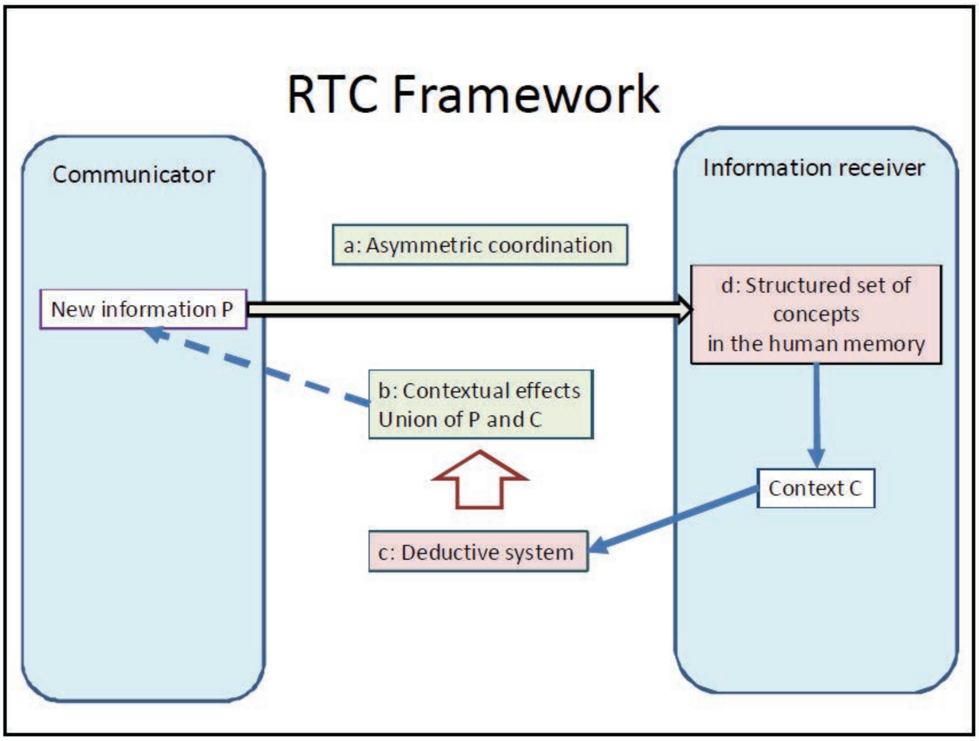


Figure 3-1: RTC Framework

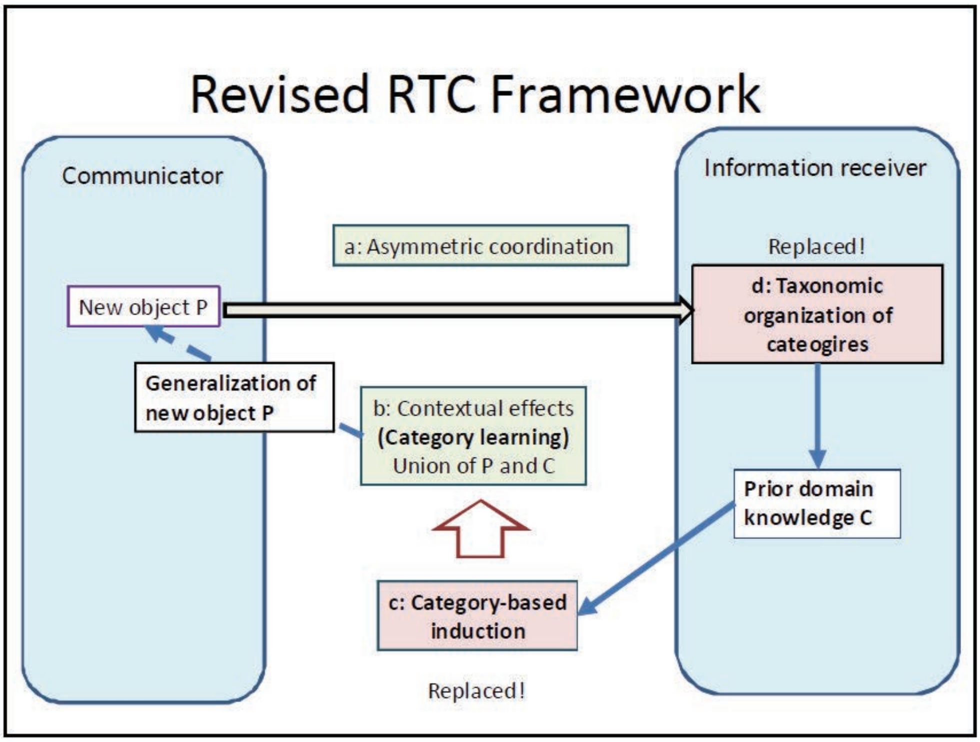


Figure 3-2: Revised RTC Framework

To summarize, while Figure 3-1 illustrates the original RTC framework, Figure 3-2 proposes a revised version of the RTC framework that forms the basis of the COM framework.

Hence, now it is possible to draw a more concrete picture of the COM framework that is an integration of the selected elements of the Relevance Theory of Communication and the replaced elements of the Knowledge Effects. First of all, the COM framework involves the two parties; one being the communicator who is conveying meanings of an SL concept to an information receiver who is receiving a TL stimulus that is supposed to be a translation of the SL concept in question. The COM framework only deals with domain-specific knowledge that is culturally-rooted in a specific country, e.g. the educational system, social system, legal system, traditional events etc. For convenience, the COM framework assumes that the average population in a specific country has general knowledge e.g. about the educational system in his/her country, as domain knowledge. Such country-specific knowledge is in most cases officially translated into English. Hence the English expression of each country-specific concept which also possesses an original local expression is considered as input data in this scenario. By identifying a country- and domain-specific corpus officially written in English, it is possible to manually extract English expressions of concepts and their definitions. These English expressions and their features that are identified in their definitions can be used for constructing a hierarchical structure of categories based on the basic principles described in Murphy (2004). These basic principles of forming a hierarchical structure of categories are, as a starting point, assumed to be consistent with the principles of Terminological Ontology whose methodological basis is laid out in the next chapter. Although some principles of Terminological Ontology may interfere with the natural hierarchical formation of categories, the assumption here is that the Terminological Ontology might still be a useful method to be employed in the framework. Terminological ontologies are constructed both for the SL- and TL domain knowledge, respectively considered as the communicator- and the information receiver's prior knowledge.

Contrasting to the Relevance Theory of Communication, such prior knowledge is considered as *context C*. We can draw two types of scenarios where a communicator, based on his/her prior knowledge, is going to identify an appropriate translation from new objects existing in a TL information receiver's cultural domain; and where an information receiver, based on his/her prior knowledge, is going to generalize the meanings of original concepts from stimuli given by a communicator. The former could be considered as the SL-oriented communication and the latter as the TL-oriented communication. It means that, e.g. in case of the TL-oriented communication, if the information receiver has his/her prior knowledge about the educational system in his/her country, this knowledge is considered as *context C*. The information receiver is supposed to have no knowledge about the educational system in the SL culture. The SL communicator is now providing a stimulus that is a TL translation of an SL educational concept. This TL translation appears as *a new encountered information P* to the information receiver's *context C*. The union of *P* and *C* is supposed to generate the contextual effect according to the Relevance Theory of Communication. In other words, the union of *P* and *C* implies the information receiver's assumption *Q* about the new information *P*, which is the communicator's intention. Here, if a cognitive environment is shared by two people, the set of all facts is manifest to both communicator and audience and therefore this may possibly generate a common ground based on the symmetric coordination. However, in a realistic scenario, the two parties use different languages and are mastering different concepts so that the way people construct mental representations and perform inference are inherently different. Thus, and as stated previously, it is most realistic and easiest to achieve the asymmetric coordination. In order for the TL information receiver to infer and generalize the original meaning of the SL concept, the category-based inductions, i.e. feature-based similarity measures, are applied as algorithms. For example, the model of computing similarities based on features proposed by Tversky (1977) enables one to compute such asymmetric similarities. To re-emphasize, this asymmetric similarity algorithm explains the views of Rips (1975) and Osherson et al. (1990) that induction from *X* to *Y* is not in general the same as from *Y* to *X*. Thus, again, the similarity of a given category to a target category is uni-directional.

The above is somehow forming the "ideal picture" of a cognitive framework consisting of at least four elements required for the COM framework. These four key elements are: **a)** the asymmetric coordination; **b)** the contextual effects generated based on the union of a new object *P* and prior knowledge *C*; **c)** the taxonomic organization of categories; and **d)** the category-based induction. These elements are integrated into the COM framework as shown in Figures 3-3 and 3-5 below. These figures respectively depict the SL-oriented communication and the TL-oriented communication described above. Figures 3-4 and 3-6, respectively illustrate how the two hierarchically structured concept systems are mapped depending on the communication patterns. To fulfill the requirements for the elements **a)** and **b)**, methods for representing element **c):** the taxonomic organization of categories, should be identified. In addition, algorithms for performing element **d):** category-based induction, should be identified. Accordingly, these identified methodological approaches will be summarized in the next chapter.

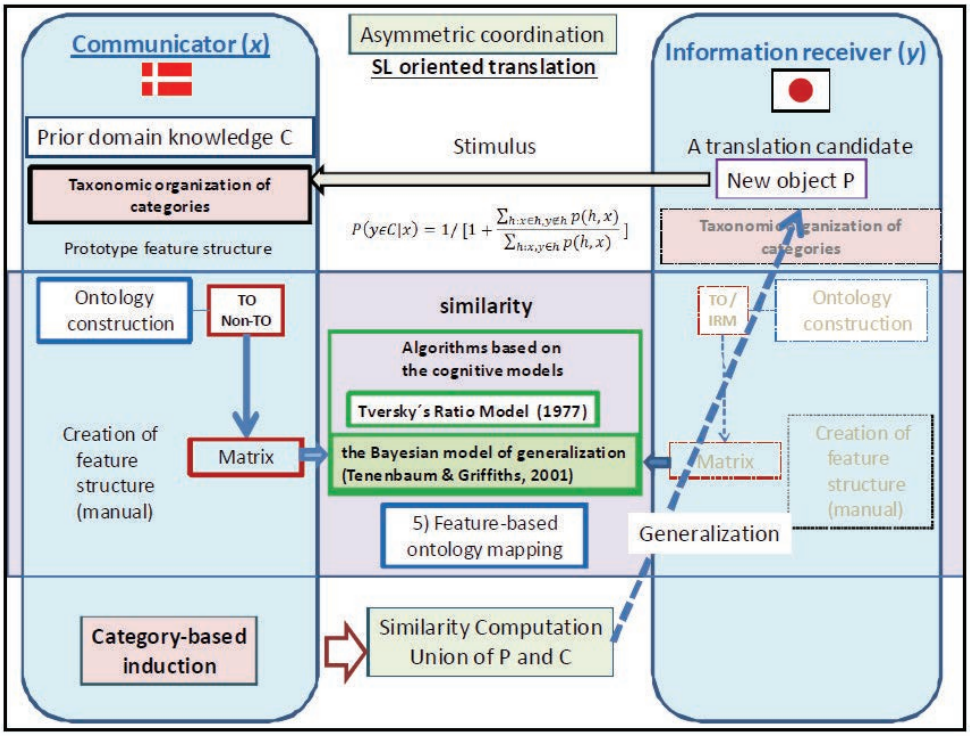


Figure 3-3: Four elements from the Revised RTC framework integrated into the COM framework (SL-oriented)

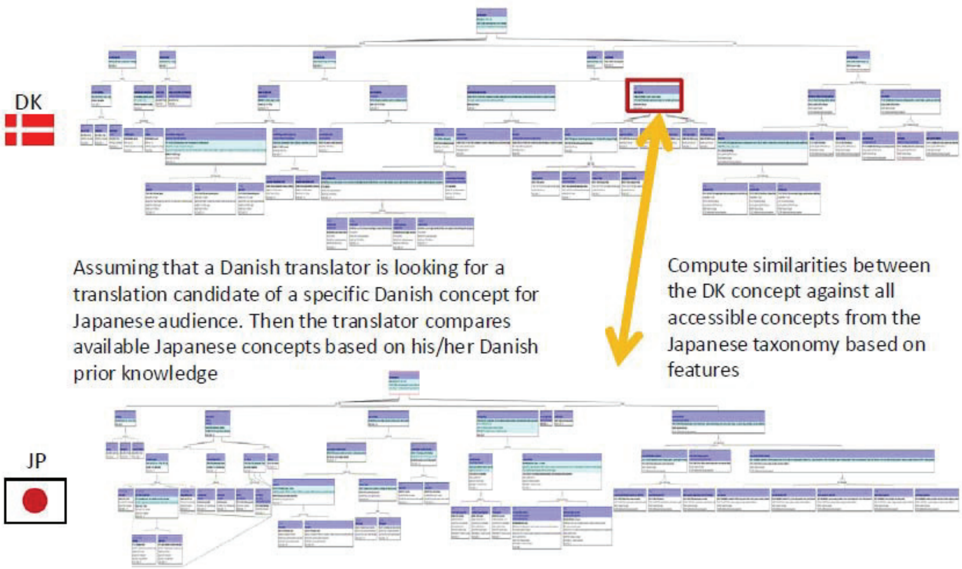


Figure 3-4: Image of ontology mapping: the SL-oriented translation

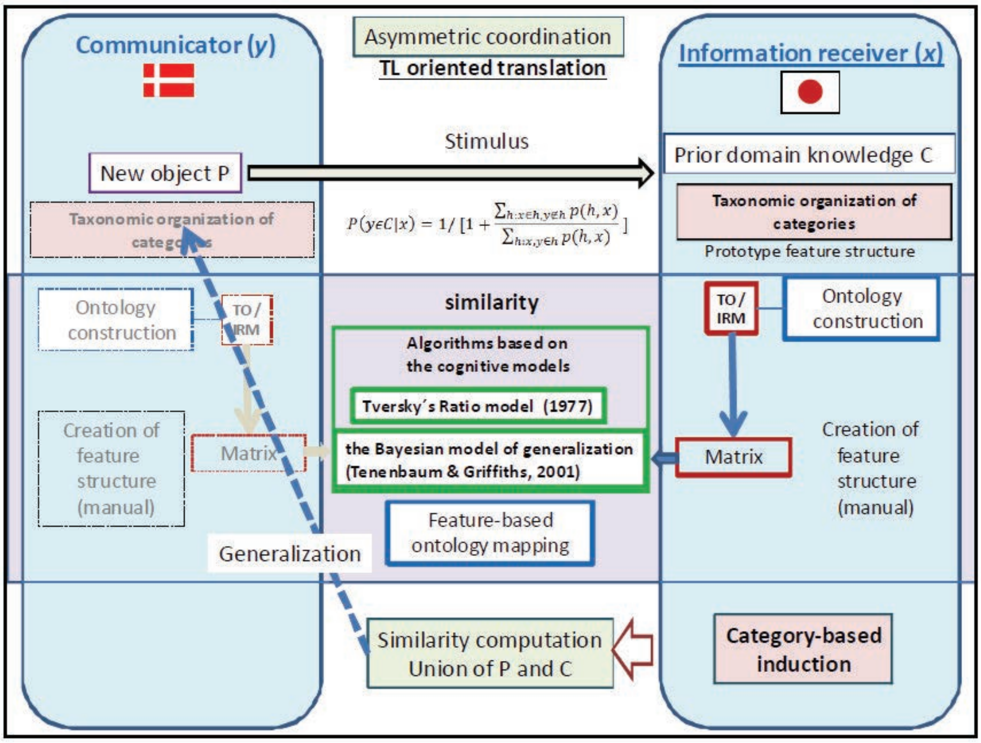


Figure 3-5: Four elements from the Revised RTC framework integrated into the COM framework (TL-oriented)

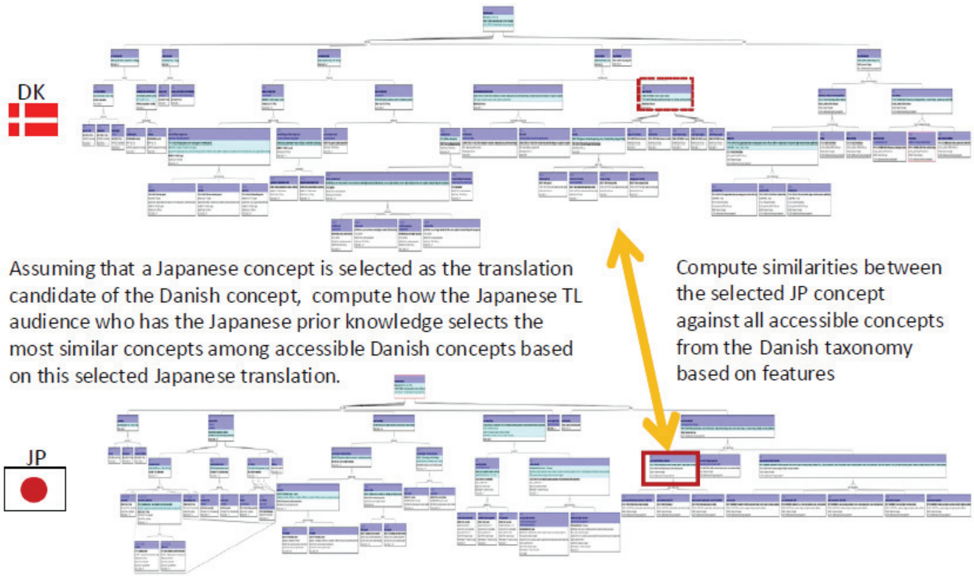


Figure 3-6: Image of ontology mapping: the TL-oriented translation

Chapter 4.

Methodologies

Figures 3-3 and 3-5 in the previous chapter illustrated the idea of the COM framework as an integration of selected elements from the Relevance Theory of Communication and from the Knowledge Approach. As was outlined, the four key elements are: **a)** the asymmetric coordination; **b)** the contextual effects *Q* generated by the union of a new object *P* and prior knowledge *C*; **c)** the taxonomic organization of categories; and **d)** the category-based induction.

To fulfill the requirements for the elements a) and b), methods for representing the element c) a taxonomic organization of categories and methods for performing the element d) the category-based inductions, should be identified. As a starting point, the Terminological Ontology (TO) method (Madsen et al., 2004) whose uniqueness is feature specifications and subdivision criteria which enables one to construct a rigidly structured concept system is selected as candidate method. Based on this representation of the domain-knowledge, two types of algorithms for implementing the element d), i.e. performing category-based induction, are selected. Accordingly, the asymmetric feature-based similarity algorithm (Tversky, 1977) described in (Murphy, 2004) and the approach referred to as the Bayesian Model of Generalization (BMG) (Tenenbaum & Griffiths, 2001) that is a powerfully extended version of Tversky's original model are considered as ideal candidate algorithms. For the purpose of comparison, these two types of algorithms are applied to datasets obtained from the TOs. In addition, for increasing the objectivity of the comparative analyses, these algorithms are directly applied to the standardized datasets created by a third party, the UNESCO Institute for Statistics (UIS). The comparison of these algorithms for performing category-based induction enables me to answer my first sub-question **a)** what algorithms are suitable for aligning TOs and mapping Culturally-specific Concepts

(CSCs) ? The workflow of the first comparative analyses is depicted in Figure 4-1.

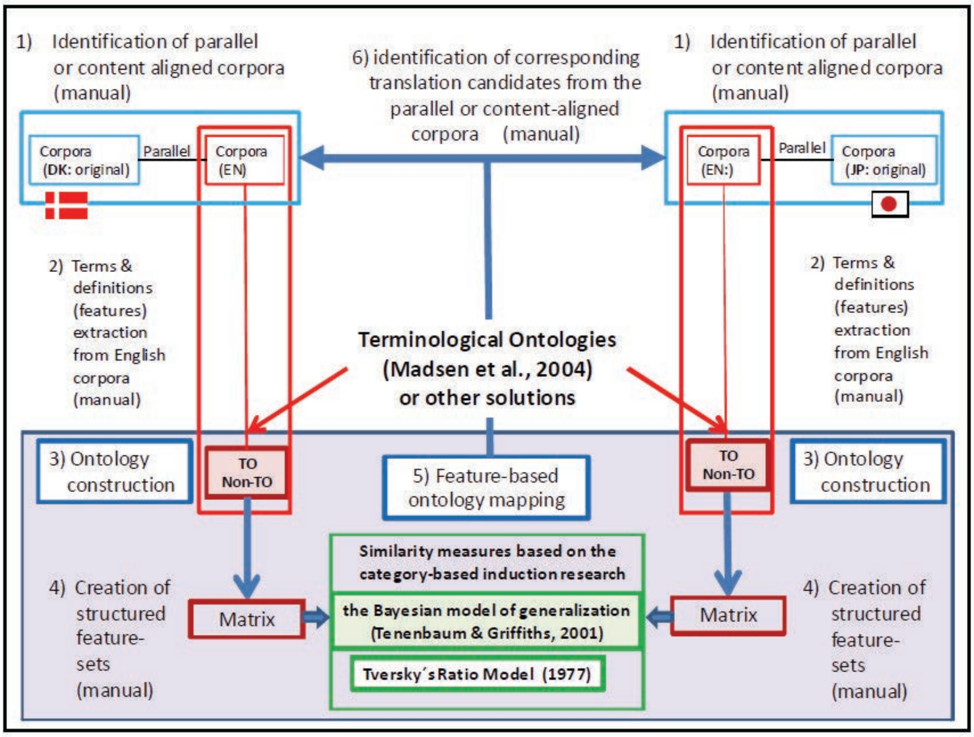


Figure 4-1: Workflow for Step 1 in the COM framework

(Step 1: comparison of the mapping algorithms)

The aforementioned comparative analyses of the mapping algorithms identify that the BMG (Tenenbaum & Griffiths, 2001) is the most optimal algorithm in the following chapter. Based on this finding, as a second step, whether the TO method is suitable for representing domain knowledge for the purpose of cross-cultural communications is further investigated. Accordingly, a novel unsupervised machine learning method, the so-called Infinite Relational Model (IRM) (Kemp et al., 2006), is selected as a non-TO method for representing domain knowledge for further comparative analyses. The IRM is a flexible model that can co-cluster multiple variables simultaneously. Hence the IRM can be applied to datasets in several different ways. In order to identify an effective strategy of applying the IRM for the purpose of the CSC mapping, a preliminary comparative study is first implemented. The study consists of three sub-experiments to test three strategies: **1)** applying the IRM directly to two CSC-feature matrices, respectively representing the educational domain knowledge in Japan and Denmark for first categorizing them into categorical classes that are to be subsequently compared and aligned (Figure 4-2); **2)** applying the IRM directly to a matrix where the two CSC-feature matrices respectively representing the Danish- and Japanese educational domain knowledge are merged (Figure 4-3); and **3)** applying the BMG to directly compute similarity relations between CSCs in the two cultures at hand, thereafter to apply the IRM for clustering CSCs in the respective cultures into categorical classes (Figure 4-4). The results of this comparative study indicate that the third strategy for combining the BMG and the IRM (the BMG+IRM approach) is the most optimal strategy. Accordingly, this BMG+IRM approach is compared against the approach of applying the TO combined with the BMG (the TO+BMG approach), in order to answer my second sub-question: **b)** within the defined framework of the COM, what are the advantages or disadvantages of using the TO compared with other approaches for systematically representing and organizing conceptual features as culturally-dependent prior-knowledge? The workflow of the second comparative analyses is illustrated in Figure 4-5.

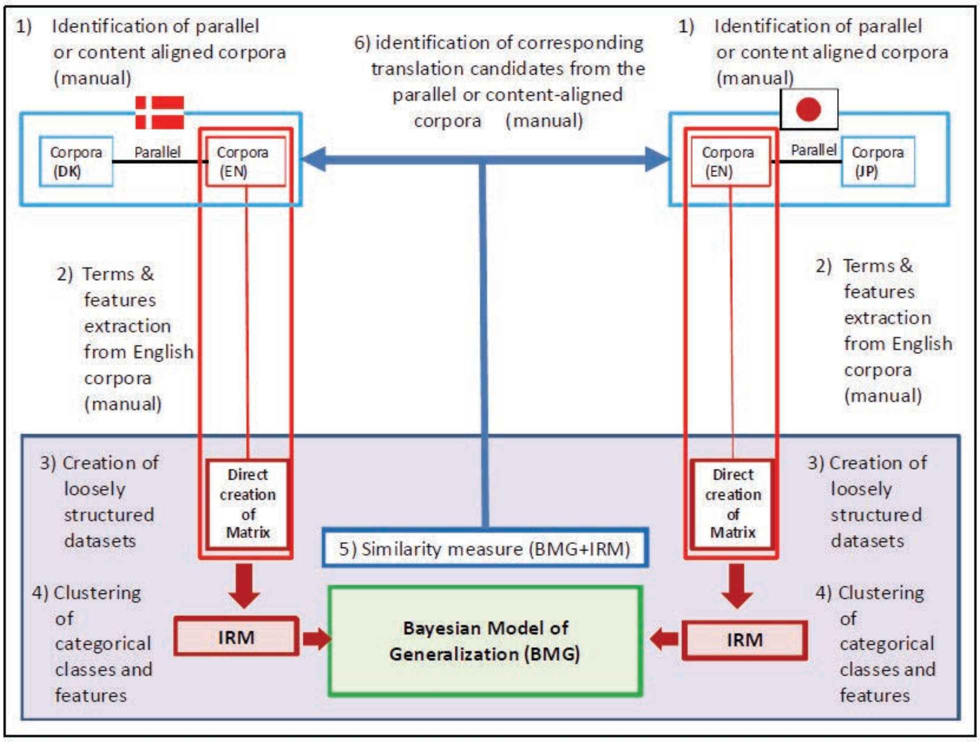


Figure 4-2: Step 2 Application of the IRM – Strategy 1

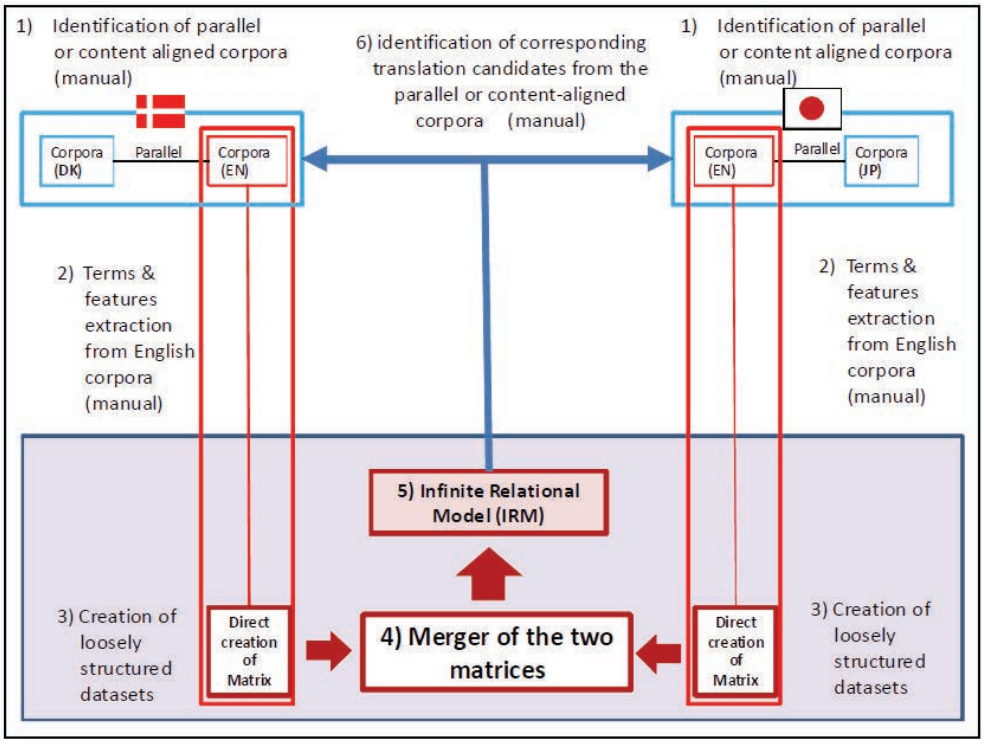


Figure 4-3: Step 2 Application of the IRM – Strategy 2

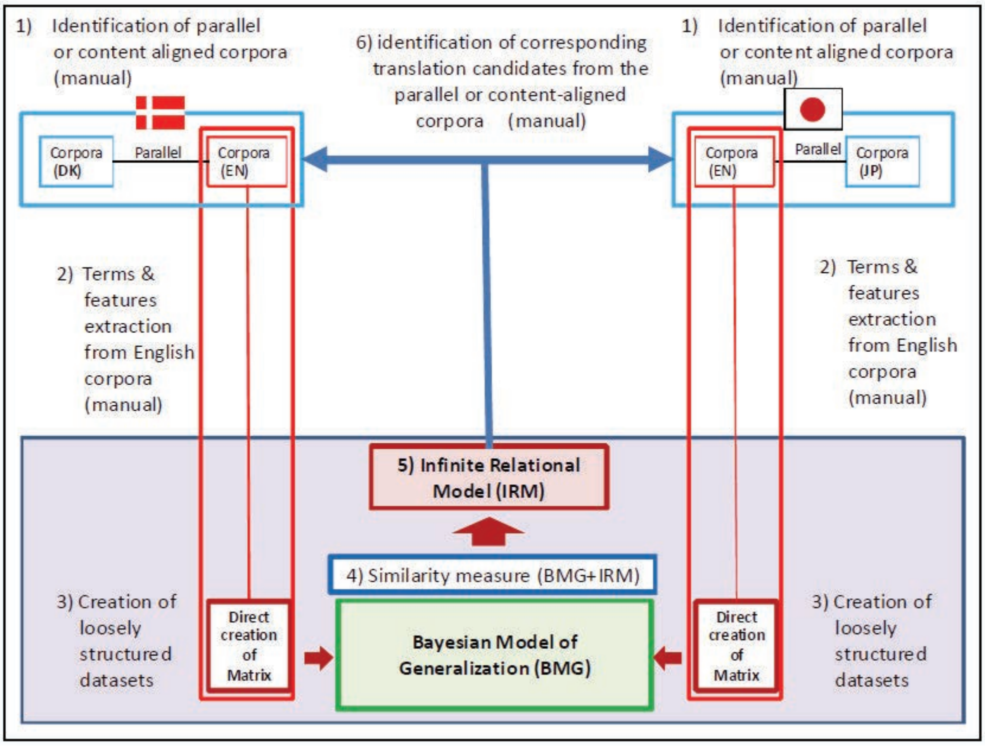


Figure 4-4: Step 2 Application of the IRM – Strategy 3

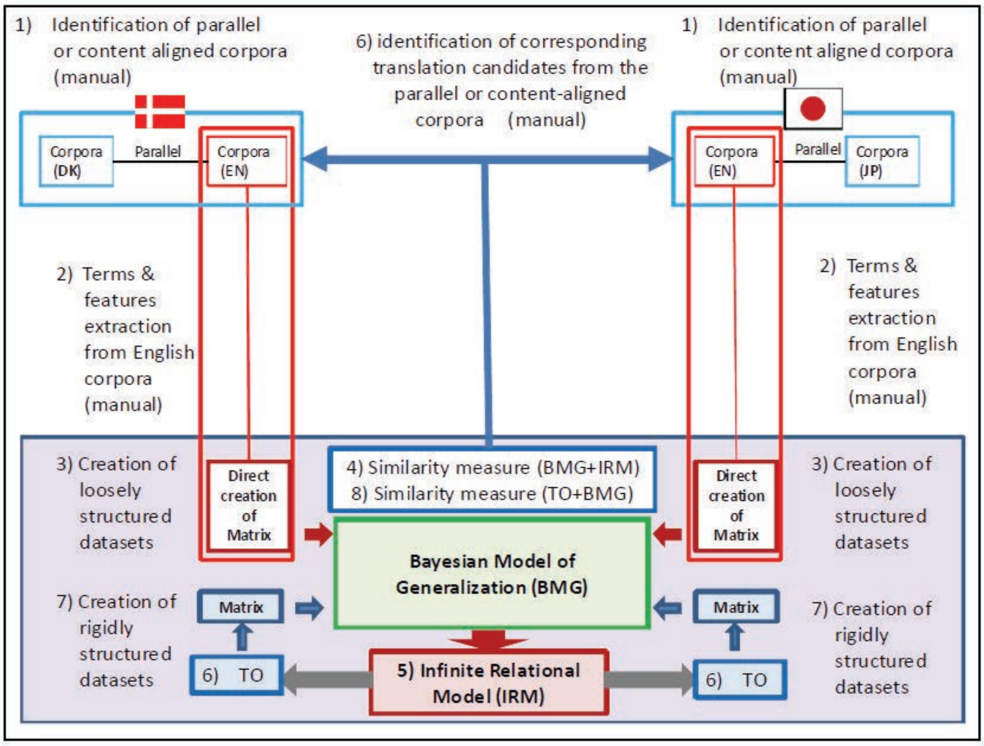


Figure 4-5: Workflow for Step 2 in the COM framework

(Step 2: comparison of the knowledge representation method)

In the following sections, the methods employed for generating datasets as well as algorithms employed for the respective comparative analyses are reviewed.

1. Step 1: Comparative analyses of algorithms for mapping TOs (the TO+Tversky algorithm vs. the TO+BMG algorithm)

1.1. Identification of corpora and manual knowledge extraction (Module 1-2 in Figure 4-1)

The main focus of the COM framework is on domain-specific knowledge that is culturally-rooted in a specific country, e.g. the educational system, the social system, the legal system, or traditional events etc. Strictly speaking, cultural prior knowledge possessed by an individual must vary from person to person. For convenience, the COM framework assumes that the average population in a specific country has some general knowledge that is culturally-rooted as so-called domain knowledge e.g. related to a country-specific educational system. In this thesis, the educational systems in Denmark, Japan and Germany are selected as culturally-specific domain knowledge possessed by the average population in each of these respective countries. The motivation is based on the assumption that if people in a specific country is born and grown up with particular cultural roots, they must inherently possess general domain knowledge about e.g. the educational system in their particular society. A pre-condition assumed here, is that populations in each of these countries do not have substantial knowledge about the educational systems in the other considered countries.

Based on this pre-condition, the country-specific domain knowledge of the educational system in each of these respective countries has been identified from documents officially published from authorities in these respective countries. All these text documents describing the educational systems in the three chosen countries, DK, GE and JP, are officially published in English by reliable authorities of each of the countries, respectively. Thus, all English translated terms and expressions in their original languages can be considered as official terms. It means that it is feasible to assume that one can identify terminological expressions in each original language based on these documents. This enables us to eventually identify translation equivalences linking between, in this case, Danish, German and Japanese. However, in this thesis, I only focus on English documents that describe language-dependent concepts (CSCs) in the three cultures. By identifying a country-specific, domain-specific corpus officially written in English, it becomes feasible to manually extract English expressions of concepts and their definitions. These English expressions and their definitional features that are identified in their definitions can be used for constructing a hierarchical structure of categories.

For the first comparative analyses, documents downloaded from the Eurydice web-site published by the Education, Audiovisual and Culture Executive Agency under the EU commission have been used as text corpora for the Danish and German educational systems. The constitution of the respective documents is as follows:

	Germany	**Denmark**
Tokens in text	165,854	89,192
Types (distinct words)	6,466	4,388
Type/token ratio (TTR)	3.90	4.92

A document structure describing the educational systems in these two countries is harmonized based on the standardized format following UNESCO's ISCED (International Standard Classification of Education). Therefore, it is assumed here to be simpler to represent structured knowledge that can effectively be aligned with each other. However, as Figure 4-3 shows, the conceptual structures of the respective educational systems are quite different from each other – in this case UK, DK and GE are considered in parallel. Thus, it is in fact a challenge to map concepts existing in Denmark and Germany as is clearly evident from the illustration.

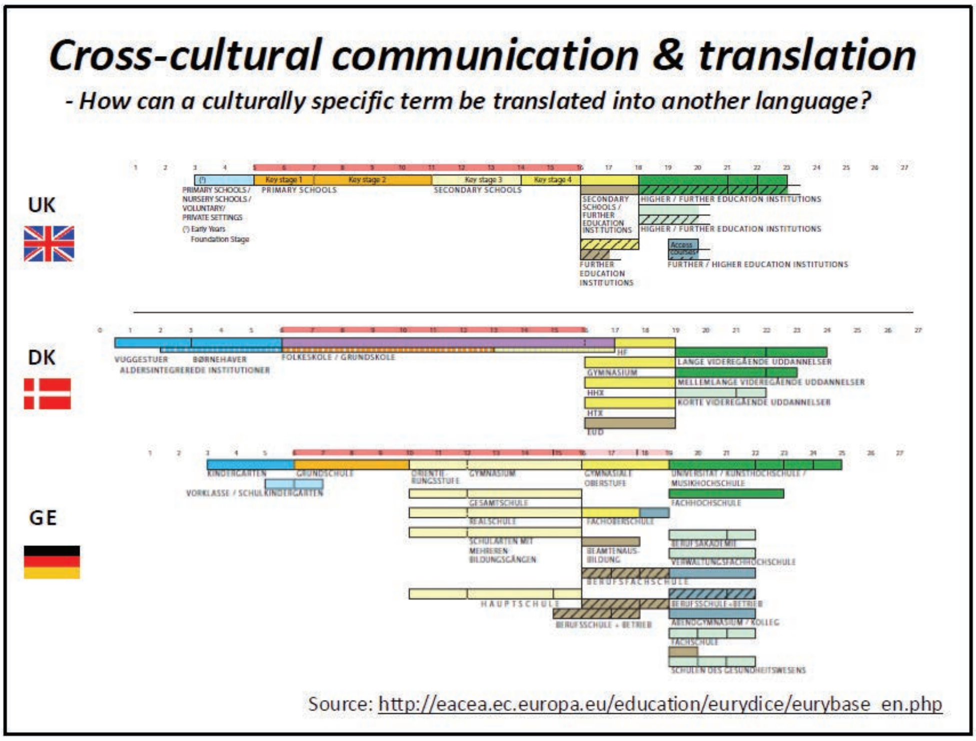


Figure 4-6: ISCED categorization (UK/DK/GE)

To summarize, the aforementioned data preparation method corresponds to the COM framework modules 1) and 2) proposed in Chapter 1 in the following way: **1)** identification of domain specific content-aligned (or parallel) corpora consisting of SL-English and English-TL language combinations; and **2)** terms and features extractions from content-aligned English corpora which describe domain specific terms and their definitions in English for the respective domain-knowledge in the SL- and TL cultures. In order to achieve module **3)** from Chapter 1 (see also: Figure 4-1), the method for constructing a structured set of CSCs for the respective domain-knowledge in the SL- and TL countries based on the information extracted in module **2)** is outlined hereafter.

1.2. Knowledge representation: Terminological Ontology (Module 3 in Figure 4-1)

Before opening a larger discussion on the principles of TO, I will first try to summarize the historical development of the theory of terminology. This is mainly because the traditional theory of terminology is in fact far from the idea of the cognitive aspects of the COM framework.

The theory of terminology was first introduced by (Wüster, 1968). Wüster's original objectives of terminology were *to eliminate ambiguity from technical languages by means of standardisation of terminology in order to make them efficient tools of communication* (Cabré, 2003). Kageura (2002: 17) summarizes that *the traditional theory of terminology addresses the relation between concepts and terms, starting from concepts and focusing on the present state of the conceptual structure and its representation.* These strict principles are also reflected in the ISO standards on terminological principles, c.f. ISO 704 (2000). Contrary to this traditional approach, Temmerman (2000) introduced so-called socio-cognitive aspects based on the importance of a communication-oriented and discourse-centred view (Cabré 1999; 2000; Temmerman, 2000). Temmerman (2000) more concretely suggests that it is necessary to distinguish between *concepts* and *categories.* She argues that many categories are fuzzy and cannot be absolutely classified by logical and ontological means because categories evolve, terms change in meaning, and general human understanding constantly develops.

In parallel to these chronological trends of terminology, a different approach – the knowledge-based approach - has been steadily evolving since the late 90es originating in Denmark (Madsen, Thomsen and Vikner, 1999). The backbone of the knowledge-based approach is centered on terminological concept modeling resulting in the TO approach. In this approach, information about concepts is identified from text corpora or in discussions with specialists in a particular subject field. The knowledge-based approach responds to the issues of "fuzziness of categories" originally taken up by Temmerman (2000) in the following way: Assuming that there is no distinction between concepts and categories, some difference in views on concepts may lead to the creation of different versions of a TO. A TO is developed within a knowledge sharing community, then dynamically updated and validated. If it is necessary to share knowledge with other communities, TOs developed in different communities should be compared, aligned and merged as needed. Based on this view, terminological ontologists argue that concepts are defined in a language dependent context, and therefore TOs are inherently language dependent.

The principles of the TO have been developed in the research and development project called CAOS - Computer-Aided Ontology Structuring – where the aim has been to develop a computer system designed to enable semi-automatic construction of ontologies (Madsen, Thomsen and Vikner, 2004; 2005). The uniqueness of the TO is its feature specifications and subdivision criteria (Madsen et al., 2004a; 2004b). The use of feature specifications is subject to principles and constraints described in detail in (Madsen et al., 2004a). Most importantly, a concept automatically inherits all feature specifications of its superordinate concepts. According to Madsen et al. (2004a), *this principle models the principle of traditional terminology that 'the intension of the subordinate concept includes the intension of the superordinate concept'* (ISO 704: 5.4.2.2; cf. also Madsen 1999: 21). Secondly, subdivision criteria that have been used for many years in terminology work are strictly implemented in the TO by introducing dimensions and dimension specifications (Madsen et al., 2004a; 2004b). This enables the CAOS prototype to perform consistency checking which helps in constructing TOs. A dimension of a concept is an attribute occurring in a non-inherited feature specification of one or more of its subordinate concepts. Values of the dimension allow a distinction among sub-concepts of the concept in question. For example, a dimension of the concept "academic degree" is [LENGTH OF EDUCATION] whose values are [2-3 year | minimum 4 years]. These dimension values distinguish the sub-concepts: "junior college" and "university" in Figure 4-7 below. The dimension can only occur in feature specifications on sister concepts and a given value can only appear on one of these sister concepts. This second principle implies the third principle, that is, a concept must be distinguished from each of its nearest superordinate concepts by at least one feature specification. In the TO, a concept dimension and its feature values are registered as (DIMENSION : [value1, value2, ...]). In the case of Fig. 4-7, *dimension specification* of a concept "academic degree" is represented as (LENGTH OF EDUCATION : [2-3 year | minimum 4 years]. This dimension specification subdivides the concept "academic degree" into two sub concepts "junior college" and "university" which respectively possess primary features, [LENGTH OF EDUCATION : 2-3 year] and [LENGTH OF EDUCATION : minimum 4 years]. The features that subdivide these two concepts are called *primary feature specifications* which are differentiated from other feature specifications that are inherited from superordinate concepts. It is also allowed to define one or more dimension specification of a concept, e.g. an example shown as the concept "degree" in Fig. 4-7. In this way, a concept must be distinguished from each of its nearest superordinate concepts as well as from each of its sister concepts by at least one feature specification (Madsen et al., 2004a; 2004b).

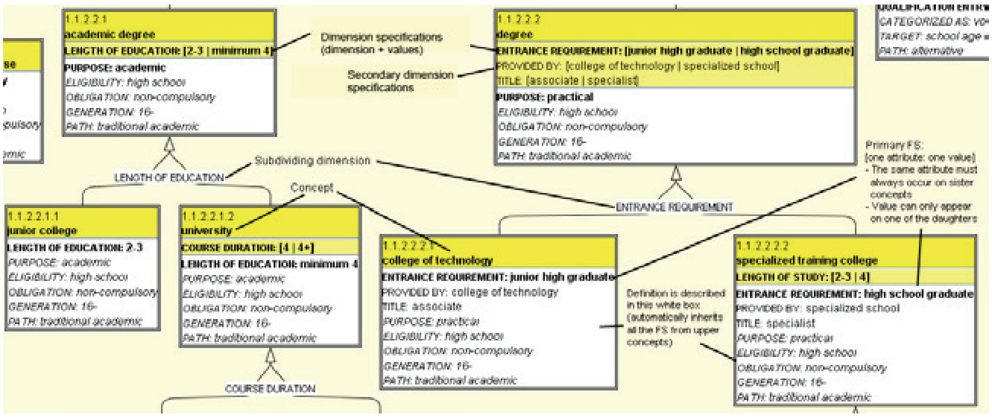


Figure 4-7: Terminological Ontology example

These principles are fairly intuitive and reasonably consistent with the hierarchical structure of categories described in the previous chapter (Murphy, 2004). On the other hand, the principles of the TO also defines more strict rules derived from the traditional view of terminology that aims at proper standardization. For example, a principle of uniqueness of dimension defines that *a given dimension may only occur on one concept in an ontology*. Madsen et al. (2004a) argues that *uniqueness of dimensions contributes to create coherence and simplicity in the ontological structure, because concepts that are characterized by means of a certain common dimension must appear as descendants of the same superordinate concept.* In the same way, Madsen et al. (2004) also defines the uniqueness of primary feature specifications as *a given primary feature specification can only appear on one of the daughters*. The argument is that these uniqueness principles *make it possible to a certain extent to carry out automatic placing of concepts* *into an ontology.* Another point is that the TO principles allows poly-hierarchy structures so that one concept may be related to two or more superordinate concepts. This principle is inconsistent with the principle of the taxonomic organization of categories described in (Murphy, 2004).

Although some of the principles are consistent with the principles of human taxonomic organization of categories described in (Murphy, 2004), the question arises whether these strict artificial rules are applicable to the COM framework when the framework aims at simulating cognitive processes of human category-based inductions. In a way, this issue is quite similar to Temmerman's (2000) question about "fuzziness of categories" arguing that categories cannot be absolutely classified by logical and ontological means. In order to analyze this specific issue about the strictness of principles vs. the fuzziness of category formation, the second comparative analysis is conducted in this thesis by applying a neutral unsupervised learning method called the Infinite Relational Model (IRM) proposed by a group of cognitive scientists, Kemp, Tenenbaum, Griffiths, Yamada and Ueda as described in the next section. This will hopefully enable me to answer the following previously raised key challenge in a concrete way: within the defined framework of cross-cultural concepts mapping, what are the advantages or disadvantages of using TOs compared with other approaches for systematically representing and organizing conceptual features as culturally-dependent prior-knowledge? Before explaining the IRM algorithm, the algorithms for mapping TOs: i.e. Tversky's Ratio Model (Tversky, 1977) and the Bayesian Model of Generalization (Tenenbaum & Griffiths, 2001) are reviewed in the following sections.

1.3. Datasets creation (Module 4 in Figure 4-1)

From the TOs constructed in the previous section, structured feature sets that represent definitions of each concept are enlisted in a country-specific matrix. Feature values extracted from respective country-specific English corpora might express the same feature in different ways. These inconsistent expressions of a feature are manually aligned. For example, Tables 1 and 2 are part of the feature structures used in the empirical studies described in the next chapter (Chapter 5, section 4).

Table 4-1: Terms and features extracted from a TO (German)

ID	Term	Feature-values
G2	preschool education	{ISCED97, children & young, ISCED0}
G5	kindergärten	{ISCED97, children & young, ISCED0, child welfare, 3-6y.o.}
G7	schulkindergärten & vorklassen	{ISCED97, children & young, ISCED0, preparation}
G10	primary education	{ISCED97, children & young, ISCED1}
G11	primary school	{ISCED97, children & young, ISCED1, <6-10y.o.<}
G13	secondary education	{ ISCED97, children & young, ISCED2+3}
G14	lower secondary level	{ISCED97, children & young, ISCED2+3, <10-16y.o.<}
G15	school offering one single course	{ISCED97, children & young, ISCED2+3, <10-16y.o.< , single}
G16	hauptschule	{ISCED97, children & young, ISCED2+3, <10-16y.o.< , single , general basic, 5-9^{th} grade}
G18	gymnasium	{ISCED97, children & young, ISCED2+3, <10-16y.o.< , single, intensified, 5-12/13^{th} grade}
G19	schools offering several courses	{ISCED97, children & young, ISCED2+3, <10-16y.o.< , several}

Table 4-2: Terms and features extracted from a TO (Denmark)

ID	Term	Feature-values
D2	pre primary	{ISCED97, children & young, ISCED0}
D4	kindergarten	{ISCED97, children & young, ISCED0, 3-6y.o.}
D6	single structure	{ISCED97, children & young, ISCED1+2}
D7	alternative structure	{ISCED97, children & young, ISCED1+2, alternative}
D9	efterskole or youth school	{ISCED97, children & young, ISCED1+2, alternative, compulsory, <14-18y.o.<}
D10	efterskole	{ISCED97, children & young, ISCED1+2, alternative, compulsory, <14-18y.o.<, boarding school, approved by state}
D11	youth school	{ISCED97, children & young, ISCED1+2, alternative, compulsory, <14-18y.o.<, day-to-day, public municipal council}
D14	municipal school	{ISCED97, children & young, ISCED1+2, formal teaching, municipality}
D16	0-9th form	{ISCED97, children & young, ISCED1+2, compulsory}
D17	0th form	{ISCED97, children & young, ISCED1+2, compulsory, preparation}
D18	1-9th form	{ISCED97, children & young, ISCED1+2, compulsory, general basic}
D19	10th form	{ISCED97, children & young, ISCED1+2, optional}

For applying the plurality of cognitive models explained in the subsequent sections, the following systematic operation has been performed in advance:

1. All feature values existing in two country-specific ontologies to be aligned, are respectively registered in a country-specific matrix.
2. If feature values in the two matrices to be aligned are completely overlapping (e.g. “ISCED0-pre-primary” in DK and “ISCED0-pre-primary” in GE in Tables 1 and 2), the feature columns in question should be merged into one column.
3. If a feature is possessed by a concept, the numeric value should be “1”, otherwise “0” in the matrices.
4. If a feature value in one matrix is completely included in a feature value in the other matrix (e.g. “ISCED1+2” in DK and “ISCED1” in GE), a concept possessing the feature that includes the other feature (e.g. Danish “ISCED1+2”) should have numeric value “1” in both feature columns (e.g. “ISCED1+2” in DK and “ISCED1” in GE). It means that a concept possessing a feature value that is included in the other feature (e.g. German “ISCED1”) should have numeric value “1” only in the feature column in question.
5. If feature values in the Danish and German matrices are partly overlapping (e.g. “ISCED1+2” in DK and “ISCED2+3” in GE), a dummy column referring to the exact overlapping feature value (e.g. “ISCED2” for both DK and GE) is created. In this example, a Danish concept possessing a feature “ISCED 1+2” should have numeric value “1” in both “ISCED 1+2” and “ISCED2” columns, but not in the “ISCED2+3” column.

In this way, the country-specific matrices representing the Danish, and German educational systems are respectively generated from the TOs.

1.4. Mapping algorithm: Tversky’s Ratio Model (Module 5 in Figure 4-1)

The concept of similarity has a long history within the area of psychological theory. Tversky (1977) states that *similarity serves as an organizing principle by which individuals classify objects, form concepts, and make generalizations.* Tversky´s view of similarity is distinguished from the traditional theoretical analysis c.f. (Shepard, 1987) on two points: **1)** while the theoretical analysis of similarity relations has been dominated by the continuous metric space models, Tversky argues that *the assessment of similarity between objects may be better described as a comparison of features rather than as the computation of metric distance between points;* and **2)** although *similarity has been viewed by both philosophers and psychologists as a prime example of a symmetric relation*, the asymmetric similarity relation has been demonstrated in (Tversky, 1977) based on several empirical evidences.

Based on these two points, Tversky proposed a classic set-theoretic model of similarity, later coined Tversky's Ratio Model, as described by the following equation:

$$sim\ (y,x) = \frac{f(Y \cap X)}{f(Y \cap X) +\ \alpha * f(Y - X) + \beta * f(X - Y)} \qquad (1)$$

Here, X and Y are the feature sets of object x and object y, respectively. f denotes a measure over the feature sets. In this thesis, the function f is defined as additive in the series of empirical studies described in the next chapter. ($Y \cap X$) represents the sets of features present in both X and Y, (Y-X) represents the sets of features present in Y but not in X, and (X-Y) represents the sets of features present in X but not in Y. Since the similarity score in equation (1) is normalized, the obtained score lies between 0 and 1. α and β are free parameters representing an asymmetric relationship between X and Y. Assignment of these parameters severely influences the similarity measurements. When defining $\alpha = \beta = 1$, $sim\ (y,x) = f(Y \cap X)\ /f(Y \cup X)$ corresponds to the well-known algorithm, Jaccard's coefficient measure (Jaccard, 1901). When defining $\alpha = 1$ and $\beta = 0$, $\mathrm{sim}\ (\mathrm{y,x}) = \mathrm{f(Y \cap X)}\ /\mathrm{f(Y)}$ corresponds to what is found in e.g. (Bush & Mosteller, 1951). The uniqueness of Tversky´s view, is this asymmetric similarity that has been originally demonstrated in (Tversky, 1977) based on several empirical evidences. His argument is that similarity does not necessarily form a symmetric relation. When $\alpha = \beta$, this similarity formula assesses the degree to which object x and y are similar to each other. It means that $\mathrm{sim}\ (\mathrm{y,x}) = \mathrm{sim}\ (\mathrm{x,y})$. On the other hand, if the parameters α and β differ, this symmetric relation does not hold. Tversky (1977) explains that *if sim(y,x) is interpreted as the degree to which y is similar to x, then y is the subject of the comparison and x is the referent. Hence the features of the subject are weighted more heavily than the features of the referent (i.e., α > β). Consequently, similarity* *is reduced more by the distinctive feature of the subject than by the distinctive features of the referent.*

Tversky (1977) argues that his model *offers explications of similarity, prototypicality, and family resemblance* discussed in the previous chapter (Murphy, 2004; see also Rosch & Mervis, 1975). Based on his idea of typicality and asymmetric relations between a subject and a referent, as well as on the Relevance Theory of Communication that inherits the asymmetric co-ordination between communicator and audience on the choice of code and context, translation should provide the set of assumptions that are adequately relevant to the audience. And the stimulus (that is translation) produced by the translator should be such that it avoids gratuitous inferential processing effort on the audience´s part. Considering that *similarity serves as an organizing principle by which individuals classify objects, form concepts, and make generalizations* (Tversky, 1977), the most similar concept to an SL concept, that is identified in the audience´s taxonomic organization of categories through feature matching, could be the set of assumptions which are adequately relevant to the audience. Based on this hypothesis, Tversky's Ratio Model assigning different combinations of α and β parameters is applied to datasets obtained based on the methods described in the previous section

1.5. Mapping algorithm: Bayesian Model of Generalization (Module 5 in Figure 4-1)

Tenenbaum and Griffiths (2001) argue that Tversky´s Ratio Model is remarkably similar to the Bayesian Model of Generalization (BMG), that is rooted in Shepard´s theory of the generalization problem. Hence, it is possible to unify these two classically opposing approaches to similarity and generalization. In (Tenenbaum and Griffiths, 2001), three crucial questions of learning, after Chomsky 1986, are addressed in order to explain the BMG: ***1)*** *what constitutes the learner´s knowledge about the consequential region;* ***2)*** *how the learner uses that knowledge to decide how to generalize; and* ***3)*** *how the learner can acquire that knowledge from the example encountered*. For instance, one example x of some consequence R is given. It is assumed that x can be represented as a point in a continuous metric psychological space and R corresponds to some region (referred to as *consequential region*) of this space. A task of the learner is to infer the probability that a newly encountered object y will fall within R given the observation of the example x. This conditional probability can be expressed as: $P(y\epsilon R|x)$. In order to compute the conditional probability, Tenenbaum and Griffiths (2001) first answer the question: the learner´s knowledge about the consequential region is represented as a probability distribution $p(h|x)$ over an a priori-specified hypothesis space H of possible consequential regions $h\epsilon H$. Prior to observing *x*, this distribution for the prior probability is $p(h)$, then becomes the posterior probability $p(h|x)$, after observing *x*. According to (Tenenbaum and Griffiths, 2001), the learner uses this knowledge for generalization by summing the probabilities $p(h|x)$ of all hypothesized consequential regions that contain *y* as follows

$$P(y\epsilon R|x) = \sum_{h:y\in h} p(h|x) \tag{2}$$

Tenenbaum and Griffiths (2001) further describe how a rational learner arrives at $p(h|x)$ from $p(h)$ after the generalization, through the use of Bayes´ rule as follows

$$P(h|x) = \frac{p(x|h)p(h)}{p(x)} \tag{3}$$

Tenenbaum & Grifiths (2001) argue that equation (2) can be similar to Tversky's Ratio Model. In order to demonstrate this, they have reformulated Tversky´s Ratio Model to the following formula:

$$sim(y,x) = 1/\left[1 + \frac{\alpha \cdot f(Y - X) + \beta \cdot f(X - Y)}{f(Y \cap X)}\right] \tag{4}$$

As defined in the previous section, *X* and *Y* are the feature sets representing *x* and *y,* respectively. f denotes a measure over the feature sets, here considered as an additive function. (*Y*⋂*X*) represents the sets of features present in both *Y* and *X*, (*Y*-*X*) represents the sets of features present in *Y* but not in *X*, and (*X*-*Y*) represents the sets of features present in *X* but not in *Y.* α and β are free parameters representing an asymmetric relationship between *Y* and *X*. A key point in this thesis is to clarify which variable is defined as concept in an SL- or a TL culture. According to Tversky (1977), *if sim(y,x) is interpreted as the degree to which y is similar to x, then y is the subject to the comparison and x is the referent.* Considering that *similarity serves as an organiz-*

ing principle by which individuals classify objects, form concepts, and make generalizations (Tversky, 1977)*,* the most similar concept to an SL concept which is correctly identified in a TL culture of an audience through a feature mapping could, from a communicator's point of view, be the set of assumptions which are adequately relevant to the audience. Hence, the referent *x* should be defined as an SL concept and the subject *y,* that is to be compared with *x,* should be defined as a TL concept according to Tversky's Ratio Model. This definition should be applied to all algorithms derived from Tversky's Ratio Model.

Keeping this definition in mind, Tenenbaum and Griffiths (2001) have reformulated equation (2) in a mathematically equivalent form as follows:

$$P\ (y \in R|x) = \ 1/[1 + \frac{\sum_{h:x\in h, y\notin h} p(h,x)}{\sum_{h:x,y\in h} p(h,x)}] \tag{5}$$

A key point emphasized by Tenenbaum & Griffiths (2001) here is that *the bottom sum ranges over all hypotheses that include both x and y, while the top sum ranges over only those hypotheses that include x but not y. If we identify each feature k in Tversky's framework with a hypothesized subset h, where an object belongs to h if and only if it possesses feature k, and if we make the standard assumption that the measure f is additive, then the Bayesian model as expressed in equation (5) corresponds formally to Tversky's Ratio Model (4) with asymmetric parameters α=0, β=1.*

It means that if the free parameters in equation (4) is set as *α=0, β=1,* this algorithm is formally corresponding to equation (5) of the BMG which compute *the conditional probability that y falls under R (Consequential Region) given the observation of the example x* (Tenenbaum & Griffiths, 2001). The consequential region *R,* in this thesis, indicates the categorical region where a subject *y* belongs. In equation (5), a hypothesized subset *h* is defined as the region where a concept belongs to *h*, if and only if, it possesses feature *k* (Tenenbaum & Griffiths, 2001). Hence, in the COM framework, *y* has been implicitly considered as a newly encountered object existing in the TL ontology that should be aligned to the referent ontology of the SL. It means that by exchanging assignment of variables *x* and *y*, the algorithm defined in equation (5) also computes the probabilities that the audience in a TL culture generalizes an SL concept from a stimulus presented by a communicator.

Another key point of the BMG is that $P(h, x) = P(x|h)P(h)$ in equation (5) represents the weight assigned to the consequential subset h in terms of the example x. This can be achieved by specifically assigning the weight $P(h, x)$ based on the "strong sampling scheme" defined in (Tenenbaum & Griffiths, 2001) as follows:

$$P(x\,|h) = \begin{cases} 1/|h| & if\ x \in h \\ 0 & otherwise \end{cases} \tag{6}$$

Here, $|h|$ indicates the size of the region h (Tenenbaum & Griffiths, 2001). According to Tenenbaum & Griffiths (2001), *what likelihood function P(x | h) is determined by how we think the process that generated the example x relates to the true consequential region for R.* For example, Shepard's Universal Law of Generalization (1986), Heit's (1998) Bayesian analysis of inductive reasoning, and the standard machine learning literature (Haussler et al., 1994; Mitchell, 1997) argue that *the example x and consequential region R are sampled independently, and x just happens to land inside C* (Tenenbaum & Griffiths, 2001). Thus, the likelihood is defined in a binary fashion in the following way:

$$P(x\,|h) = \begin{cases} 1 & if\ x \in h \\ 0 & otherwise \end{cases} \tag{7}$$

Opposed to this, Tenenbaum (1999) argues that *under many conditions, it is more natural to treat x as a random positive example of R, which involves the stronger assumption that x was explicitly sampled from R.* This argument leads to the "strong sampling scheme" defined by equation (6). In the COM framework, the number of objects possessing the k^{th} feature in the referent ontology is considered as the size of the region h. For example, if the feature "compulsory education" is possessed by three objects in Table 2, the weight is assigned as 1/3. This strong sampling can intuitively be illustrated in a situation where the feature "objects that have four legs" is given to us as an example. We immediately imagine that this object must be something related to an animal or possibly a piece of furniture. We unconsciously limit the hypothetical region within a narrower region in order to achieve a more effective generalization. Finally, Tenenbaum & Griffiths (2001) explain that the prior $P(h)$ is not constrained in their analysis so that *it can accommodate arbitrary flexibility across contexts*. Hence, in this thesis, we set $P(h) = 1$. However, strictly speaking, $P(h)$ could be computed based on probabilities that a hypothesized subset h is identified among all possible hypothesized subsets h that are assumed from a dataset in question. It will be demonstrated, that this strong sampling scheme applied for the BMG works effectively compared with the widely used Jaccard coefficient similarity measure, as well as Tversky's Ratio Model in the COM framework by the fourth and fifth empirical studies described in the next chapter (Chapter 5, Sections 4 and 5).

1.6. Non-TO: Controlled datasets for an objective comparison (Module 4 in Figure 4-1)

In the aforementioned comparative analysis, the TOs are semi-automatically developed based on the principles and rules of the TO and the datasets are systematically prepared in the form of matrices. However, the subjectivity of the analysis is one of the critical points in this empirical setting of the comparative analysis to be conducted in Chapter5, Section 4. In order to increase the objectivity of the comparative analysis, datasets prepared by a third party should be employed. Accordingly, datasets prepared by the UNESCO Institute for Statistics (UIS) who collected data from UNESCO Member States on an individual basis is employed for comparing the performance of Tversky's Ratio Model and the Bayesian Model of Generalization (BMG).

According to UIS, the purpose of collecting data from UNESCO Member States is to *map the Member States' national education systems according to the International Classification of Education (ISCED)*. UIS aims for *Member States to report their data in an internationally comparative framework.* These datasets from all over the world are downloadable from UIS' web-site[6]. Here, Japanese and Danish datasets will be used for the analysis in Chapter 5, Section 5. Each dataset consists of educational terms defined by several pre-defined feature dimensions such as the ISCED level, programme destination and orientation, starting age, cumulative duration of education, and entrance requirements. Most feature dimension values are pre-defined, i.e. for the programme destination dimension, values are

[6] http://www.uis.unesco.org/education/ISCEDmappings/Pages/default.aspx

pre-defined as [general | pre-vocational | vocational] as shown in Table 4-3.

Table 4-3: Example of original datasets obtained from UIS: feature structure of selected concepts.

Concept ID	ISCED level	Programme destination (A/B/C)	Programme orientation (G/P/V)	Theoretical cumulative duration at ISCED 5	Position in the national degree / qualification structure (intermediate, first, second, etc...)	Position in the tertiary education structure (Bachelor-Master-PhD)	Minimum entrance requirement (ISCED level or other)	Theoretical starting age	Theoretical duration of the programme	Theoretical cumulative years of education at the end of the programme	Does the programme have a work based element? (Y/N)	Programme specifically designed for adults (Y/N)	Programme specifically designed for part-time attendance (Y/N)
						Danish concepts							
D1	0		G					2-5 years	4 years		No	No	No
D2	0		G					5-6 years	1 year		No	No	No
D3	1		G					6-7 years	6 years	6 years	No	No	No
D4	2	A	G				1	12-13 years	3-4 years	9-10 years	No	No	No
D7	3	C	V				2A	16-30 years	3-5 years	14 years	Yes	No	No
D19	5	B		Short	1st		3A, 3C	18-50 years	0,5-4 years	13-15 years	No	Yes	Yes
D20	5	B		Short	1st		3A, 3C	20-30 years	2-3 years	14 years	No	No	No
D21	5	A		Medium	1st	Bachelor	3A	18-50 years	2-4 year	13-15 years	No	Yes	Yes
D22	5	A		Medium	1st	Bachelor	3A	20-30 years	3-5 years	16 years	Yes	No	No
D23	5	A		Medium	1st	Bachelor	3A	20-30 years	3 years	15-16 years	No	No	No
						Japanese Concepts							
J35	5	B		Short	Intermediate		3 ABC	18	2-3	14-15	No	No	No
J36	5	B		Short Medium	Intermediate		5B	20	1+	15+	No	No	No
J37	5	B		Short	Intermediate		3	18	2-3	14-15	No	No	Yes
J38	5	B		Short	Intermediate		3	18	2	14	No	No	No
J40	5	B					3	18	1+	13+	No	No	No
J41	5	A		Medium	1st	Bachelor	3	18	4	16	No	No	No
J42	5	A		Long	1st	Bachelor	3	18	6	18	No	No	No
J44	5	A		Long	Intermediate		5A 1st,M	22	1+	17+	No	No	No

One of the challenges of using these datasets is how to map the numeric feature values of dimensions such as "theoretical starting age" and "cumulative duration of the programme." For example, in the Danish educational system, the starting age of upper secondary school is defined as "16-17 years old" and its cumulative years of education is "12-13 years". On the other hand, the Japanese educational system is a so called "single-track system" meaning that the starting age of upper secondary school is exactly defined as "15 years old" and its cumulative years of education is "12 years". To handle this difficulty in an objective and systematic manner, the same principles of generating matrices defined in Section 4.1.3 have been applied for the columns highlighted in Table 4-3.

In order to objectively perform the qualitative comparative analysis of mapping algorithms, simpler datasets that do not contain the numeric feature dimensions/values (highlighted in Table 4-3) have been prepared as control data. It means that these simpler datasets only contain the standardized feature dimensions/values defined by UIS. The results of the comparative analysis employing these controlled datasets are presented in (Chapter 5, Section 5).

2. Step 2: Comparative analyses of knowledge representation methods (the TO+BMG approach vs. the BMG+IRM approach)

The comparative analyses of the mapping algorithms in Step 1 identify that the Bayesian Model of Generalization (BMG) (Tenenbaum & Griffiths, 2001) is the most optimal algorithm in the next chapter (Chapter 5, Sections 4 and 5). Based on this finding, as a second step, whether the TO method is suitable for representing domain knowledge for the purpose of cross-cultural communication is further investigated. Accordingly, a non-TO method for representing domain knowledge, i.e. the Infinite Relational Model (IRM) (Kemp et al., 2006), is selected as candidate for further comparative analyses. The IRM is a flexible model that can co-cluster multiple variables simultaneously. Hence the IRM can be applied to datasets in several different ways. In order to identify an effective strategy of applying the IRM for the purpose of culturally-specific concept (CSC) mapping, a preliminary comparative study is first implemented. The study consists of three sub-experiments to test three strategies: **1)** applying the IRM directly to two CSC-feature matrices, respectively representing the educational domain knowledge in Japan and Denmark, for first categorizing them into categorical classes that are to be subsequently compared and aligned; **2)** applying the IRM directly to a matrix where the two CSC-feature matrices respectively representing the Danish- and Japanese educational domain knowledge are merged; and **3)** applying the BMG to directly compute similarity relations between CSCs in the two cultures at hand, thereafter to apply the IRM for clustering CSCs in the respective cultures into categorical classes. The results of this comparative study shown in Chapter 5, Section 6, indicate that the third strategy for combining the BMG and the IRM (the BMG+IRM approach) is the most optimal strategy.

Based on these findings, the second comparative analysis of comparing the two methods of representing knowledge: the TO and the non-TO approaches are implemented. In this analysis, for comparing and assessing advantages and disadvantages of the TO approach under controlled settings, the same datasets consisting of a list of CSCs and features that are manually extracted from text corpora are used as data source. While the TOs are constructed from this data source for the TO+BMG approach, the BMG is directly applied to these datasets for the BMG+IRM approach (Chapter 5, Section 7). The principles of the data preparation as well as the IRM algorithm, is explained below.

2.1 Data preparation (Module 3 in Figure 4-2)

In the second comparative analysis, two datasets, respectively representing the Danish educational domain knowledge and the Japanese educational domain knowledge, are created. The datasets consist of educational terms and their definitional features that are manually extracted from text corpora. The Japanese corpora used for this experiment are: **1)** "Outline of the Japanese School System" published by the Center for Research on International Cooperation in Educational Development (CRICED), University of Tsukuba, Japan; and **2)** "Higher Education in Japan" published by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT). The Danish documents are downloaded from the Euridice web-site published by the Education, Audiovisual and Culture Executive Agency under the EU commission. These corpora are written in English and hence it is feasible to identify original expressions of educational terms in the respective languages from existing parallel- or content aligned corpora. This enables one to eventually achieve translation between Japanese CSCs and Danish CSCs through the English term mapping.

The CSCs and their definitions, all written in English, are manually extracted from the text corpora, e.g. the Danish CSC "municipal school (DA: *folkeskole*)" and its definition "*a comprehensive school covering both primary and lower secondary education, i.e. one year of pre-school class, the first (grade 1 to 6) and second (grade 7-9/10) stage basic education, or in other words it caters for the 6-16/17-year-*

olds". From this definition, a feature list consisting of "comprehensive school" "primary and lower secondary education" "basic education" "targeted for 6-16/17 years old" is created. This definition also implies that "municipal school" is categorized into three sub-CSCs "preschool class", "first stage" and "second stage", respectively having their features "one year preschool education" "1-6 grades" and "7-9/10 grades". These sub-CSCs are supposed to inherit features defined in the superordinate CSC, in this case "municipal school". In this way, 59 Danish CSCs and 54 Japanese CSCs and their features all written in English are listed up. In addition, some features are manually standardized, e.g. a feature "continuing education for adults" in Denmark is standardized with a feature "opportunities for life-long learning" in Japan. Finally, in order to handle the problem of numeric feature values which the system does not inherently handle, the same principles for creating the matrices defined in Section 4.1.3 is applied.

Accordingly, in total 229 features are registered in the two matrices, respectively representing the Danish- and Japanese educational systems. In each matrix, if a feature is possessed by a CSC, the numeric value "1" appears, otherwise "0" is assigned. In these matrices, the Danish- and Japanese CSCs are respectively denoted as D_i and J_j and feature IDs are assigned as f_k. Both the Danish- and Japanese CSCs and their features are alphabetically registered. These matrices are used for first comparing three strategies of applying the IRM for mapping CSCs (Chapter 5, Section 6) and secondly, for comparing the two approaches: the TO+BMG approach and the BMG+IRM approach (Chapter 5, Section 7).

2.2 Infinite Relational Model (Module 4 in Figure 4-2)

According to Kemp, et al. (2006), a key feature of the IRM is to *automatically choose an appropriate number of clusters using a prior that favors small numbers of clusters, but has access to a countably infinite collection of clusters*. In (Kemp et al., 2006), the observed data are considered as *m* relations involving *n* types. For the experimental strategies **1)**, **2)** and **3)** in Chapter 5, Section 6, the simplest model: dealing with two types with a single two-place relation R: $T_1 \times T_2 \rightarrow \{0, 1\}$ is applied. More specifically, in strategies **1)** and **2)** T_1 corresponds to either Danish and/or Japanese CSCs and T_2 corresponds to features, while in strategy **3)**, T_1 and T_2, respectively, corresponds to Danish CSCs and Japanese CSCs.

The principle of generating clusters in the IRM, according to (Kemp et al., 2006), is based on a distribution over partitions induced by a so-called Chinese Restaurant Process (CRP) (Pitman, 2002). The CRP starts a partition process with a single cluster containing a single object. The i^{th} object has possibilities to belong to either of the following:

- A new cluster with probability: γ / (i-1+ γ)
- An existing cluster with probability: n_a / (i-1+ γ)

Here, n_a is the number of objects already assigned to cluster a, and γ is a parameter (Kemp et al., 2006). The CRP continues until all the objects belong to clusters. Hence, the distribution over clusters for object i conditioned on the cluster assignments of objects 1, …, i-1 is defined as (Kemp et al., 2006):

$$P(z_i = a | z_1, \dots, z_{i-1}) = \begin{cases} \gamma/(i-1+\gamma) & a \text{ is a new cluster} \\ n_a/(i-1+\gamma) & n_a > 0 \end{cases} \quad (8)$$

(Kemp et al., 2006) explains that *the distribution on z induced by the CRP is exchangeable: the order in which objects are assigned to clusters can be permuted without changing the probability of the resulting partition. P(z) can therefore be computed by choosing an arbitrary ordering and multiplying conditional probabilities. Since new objects can always be assigned to new clusters, the IRM effectively has access to a countably infinite collection of clusters.*

From the clusters generated by the CRP, relations are generated based on the following generative model:

- As described above, for the cluster assignment of objects
  $z \mid \gamma \sim$ CRP (γ)
- For link probabilities between clusters
  η (a, b) | $\beta \sim$ Beta (β, β)
- For links between objects
  R (i, j) | z,η ~ Bernoulli (η (z_i, z_j))

In the above generative model, we set parameters β=1, and γ=log ($J^{(i)}$) where $J^{(i)}$ is the number of CSCs in the i^{th} mode.

In here, relationships are assumed to be conditionally independent given cluster assignments (Kemp et al., 2006). The eventual purpose of the generative model is to identify a cluster z that maximizes $P(z|R)$. Based on the generative model defined above, relations from clusters are generated by:

$$P(R \mid z, \eta) = \prod_{ab} (\eta_{ab})^{m^{+}_{ab}} (1 - \eta_{ab})^{m^{-}_{ab}} \tag{9}$$

where m^{+}_{ab} refers to the total number of links between categorical classes a and b; and m^{-}_{ab} refers to the total number of non-links between categorical classes a and b. The conjugate prior η_{ab} is in the aforementioned generative model defined as:
$\eta(a, b) \mid \beta \sim \text{Beta}(\beta, \beta)$. Accordingly, the conjugate prior η_{ab} is integrated out by the following:

$$P(R \mid z) = \prod_{a,b \in \mathrm{N}} \frac{Beta(m^{+}_{ab} + \beta, m^{-}_{ab} + \beta)}{Beta\ (\beta, \beta)} \tag{10}$$

From formulae (3) and (5), the IRM identifies z that maximizes: $P(z|R) \propto P(R|z)\,P(z)$. According to (Kemp et al., 2006), the expected value of η_{ab} given z, is:

$$\frac{m^{+}_{ab} + \beta}{m^{-}_{ab} + m^{+}_{ab} + 2\beta} \tag{11}$$

The mathematical procedure for the inference is further described in (Mørup et al., 2010)[7]. The solutions displayed in the empirical studies in Chapter 5, Sections 6 and 7, are based on the sample with highest likelihood.

[7] The application of the IRM to my research problem has been encouraged by Ass. Prof. Morten Mørup, Section for Cognitive Systems, DTU Informatics at the Technical University of Denmark. The mathematical and technical instructions as well as the IRM tool have been provided by Ass. Prof. Morten Mørup. Thus my contribution for the empirical studies in Chapter 5, Section 6 and 7, is the entire design of the empirical study, data preparation and analyses of the output data obtained.

Chapter 5.

Empirical studies

This chapter consists of seven published papers, each of which represents the gradual progress of my PhD-research based on the COM framework whose methodologies are defined in the previous chapter. I first summarize an overview of these papers in Table 5-1, in contrast to the methodological workflow described in the previous chapter.

Table 5-1: Overview of empirical studies in contrast to the methodologies

	Paper title	**Description**	**Methodology**
1	*Application of classical psychological theory to terminological ontology alignment*	As a preliminary study, whether Tversky's model is applicable to data-sets obtained from the TO method is investigated based on manually computed data on a smaller scale, in order to identify the potential for simulating cognitive theories/models explaining real-life cross-cultural communication scenarios	Preliminary study
2	*Alignment of remote cultures*	Whether the alignment of two language-dependent TOs is a potential method for optimizing the relevance required in cross-cultural communications is further investigated on a slightly larger scale, by assessing the precision/recall score comparing the results against human mapping	Preliminary study
3	*Asymmetric similarity and cross-cultural communication process*	The assignment of different α and β parameters to Tversky's model is defined based on the BMG theory which demonstrates the mathematical equivalence between Tversky's model and the BMG. Three different ways of assigning parameters in Tversky's model are investigated by applying these to datasets obtained from TOs.	Preliminary study

4	*Feature-based Ontology Mapping from an Information Receivers' Viewpoint*	Two types of cognitive models: Tversky´s model and the BMG are applied to datasets created based on the TO approach	Step 1. Modules 1-5: TO
5	*Cross-cultural Concept Mapping of Standardized Datasets*	Two types of cognitive models: Tversky´s model and the BMG are applied to datasets directly obtainable from a third party (the UNESCO Institute for Statistics)	Step 1. Modules 1-5: non-TO
6	*Application of the Infinite Relational Model combined with the Bayesian Model of Generalization for Effective Cross-Cultural Knowledge Transfer*	A preliminary comparative study of investigating an optimal strategy of applying the IRM for the purpose of CSC mapping	Step 2. Modules 1-5: BMG+IRM
7	*Flexible- or Strict Taxonomic Organization? - Impact on culturally-specific knowledge transfer*	A comparison of the two methods of representing knowledge: TO (the TO+BMG approach) and the non-TO (the BMG+IRM approach) and an assessment of advantages and disadvantages of the TO under controlled settings	Step 2. Comparison between Modules 1-5: BMG+IRM vs. Modules 1-8: TO+BMG

1. Application of classical psychological theory to terminological ontology alignment [8]

Preliminary study:

As a preliminary study, this paper investigates whether Tversky's model (Tversky, 1977) is applicable to data-sets obtained from the TO method (Madsen et al., 2004) based on manually computed data on a smaller scale, in order to identify the potential for simulating cognitive theories/models explaining real-life cross-cultural communication scenarios.

[8] In Proceedings of Natural Language Processing & Cognitive Sciences 2011, August 20-21, Copenhagen Business School (peer-reviewed full paper)

Application of Classical Psychological Theory to Terminological Ontology Alignment

Fumiko Kano Glückstad

Copenhagen Business School
Dept. of International Language Studies and Computational Linguistics
Dalgas Have 15, DK-2000, Frederiksberg, Denmark
fkg.isv@cbs.dk

Abstract. Terminological Ontology [1] is a method used for knowledge sharing and domain-specific translation practices, and could potentially be suitable to apply and for simulating the cognitive theories/models explaining real-world inter-cultural communication scenarios. In this paper, we investigate - as a preliminary study - whether Tversky´s contrast model [2] is applicable to data-sets obtained from the Terminological Ontology method. The eventual purpose of this study is to propose an approach for identifying potential translation candidates by optimizing relevance between concepts in two remote languages such as European and Asian languages.

Keywords: Terminology, Ontology, Translation, Inter-Cultural Communication, Set-theory, Similarity, Generalization, Relevance

1 Introduction

The role of ontology in a multilingual setting is considered an emerging challenge to Semantic Web development. As a consequence, there are several major ongoing projects, such as the MONNET project on Multilingual Ontologies for Networked Knowledge [3] and the KYOTO project on Knowledge-Yielding Ontologies for Transition-Based Organization [4]. Though both projects deal with translation of terms from a Source Language (SL) to a Target Language (TL), they focus on linking lexical data through an interoperable common ontology rather than on optimizing relevance between concepts that are potentially measurable based on diverse models derived from cognitive theory.

Terminological Ontology (TO) is a domain-specific ontology used for knowledge sharing [1], which normally is applied in terminology work, cf. for example [5]. The unique characteristics of TO that differentiate it from other types of ontologies are feature specifications and subdivision criteria [6]. A feature specification consists of a feature dimension and its value. Hence, a representation of a whole concept is a feature structure, i.e. a set of feature specifications corresponding to the unique set of characteristics that constitutes that particular concept [1][6]. Terminological ontologists argue that concepts are defined in a language-dependent context, and therefore TO is language-dependent. TO is developed within a knowledge sharing community, then dynamically updated and validated. If it is necessary to share knowledge with other communities, TOs developed in different communities should be compared, aligned and merged upon necessity. While the aforementioned two mainstream projects, MONNET and KYOTO, both deal with complex ontologies involving huge data-sets, TO usually handles smaller amounts of concepts.

Considering this, a point that should be emphasized in this work is that TO could potentially be a suitable tool to apply and for simulating cognitive theories/models explaining a real-world inter-cultural communication scenario. Thus, in the next chapter a problem existing in the real-world is explained by use of the Relevance Theory of Communication [7]. Next, the concept of similarity, which has a long history in psychological theory, is reviewed in Chapter 3. We consider how the theory proposed by Tversky [2] is applied in the context of the Relevance Theory of Communication. In Chapter 4, the empirical analysis is performed to assess the potential of applying the models based on [2] to TOs. Chapter 5 discusses findings and future work followed by conclusions in Chapter 6.

2 Real-World Problem

Imagine a situation where a non-native English speaking European and an Asian are debating in English about the issue of academic degree systems in their respective cultures. While a German might be explaining about the Doctor of Science (Habilitation) degree (the highest achievable academic degree in Germany after obtaining a Doctor of Philosophy degree), a Japanese might be having the highest possible academic degree in Japan in his mind which is a Doctor of Philosophy degree (also frequently referred to as Doctor of Science in Japan). This imagined conversation shows a typical scenery revealing a deep inherent misconception between the two communicating parties since each of them have their own conceptual - and correct - understanding of the highest obtainable academic degree in their respective cultures.

This example may further create problems for a translator who is going to translate academic titles into the language of the other party. When a translator translates the term of the German Doctor of Science Degree into Japanese, the first condition he/she has to fulfill is that his/her translation should convey the same meaning as the original German meaning. Gutt [8] explains that this requires *the receptors to familiarize themselves with the context envisaged for the original text.* Now the question is, when a Japanese receptor is not familiar with the German language and its academic culture, how should this particular German academic title then be translated into Japanese?

The proposal [8] of applying the Relevance Theory of Communication [7] is a key to address this issue. This theory focuses on how people share thoughts with one another and views communication as principally an inferential process. It means that the essential task of the communicator is to produce a *stimulus* from which an audience can infer what set of thoughts or assumptions the communicator intends to convey [8]. Hence the second condition the translator has to fulfill is that his/her translation should explicitly provide a set of assumptions that are adequately relevant to the audience. The issue here is how the translator should create the *stimulus* (that is *translation*) optimally relevant to the audience. Assuming that both German and Japanese have their respective conceptual structures of the academic system rooted in their own cultures, translation candidates that have optimally relevant relationships identified from these two conceptual systems could avoid the gratuitous inferential processing effort on the audience´s part.

The optimization of the relevance between two concepts could be well explained by the cognitive theory, Tversky´s Set-theoretic model of similarity [2]. Thus the next Chapter reviews Tversky´s model and considers how this model could be used in the context of optimizing the relevance of communication.

3 Tversky´s Set-theoretic Model of Similarity

The concept of similarity has a long history within the area of psychological theory. Tversky´s view of similarity [2] is distinguished from the traditional theoretical analysis (c.f. [9]) on two points: 1) while the theoretical analysis of similarity relations has been dominated by the continuous metric space models, [2] argues that *the assessment of similarity between objects may be better described as a comparison of features rather than as the computation of metric distance between points;* and 2) although *similarity has been viewed by both philosophers and psychologists as a prime example of a symmetric relation*, the asymmetric similarity relation has been demonstrated in [2] based on several empirical evidences.

Based on these two points, [2] proposed a classic feature-set model of similarity as follows:

$$sim(a,b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha \cdot f(A-B) + \beta \cdot f(B-A)} \qquad \textbf{(1)}$$

Here, A and B are the feature sets of object *a* and object *b*. f denotes a measure over the feature sets. (A∩B) represents the sets of features present in both A and B, (A-B) represents the sets of features present in A but not in B, and (B-A) represents the sets of features present in B but not in A. α and β are free parameters representing an asymmetric relationship between A and B. Since the similarity score in this equation is normalized, the obtained score lies between 0 and 1.

An interesting point is that the application of Tversky´s model requires *a limited list of relevant features and the representation of an object as a collection of features that is viewed as a product of a prior process of extraction and compilation* [2]. In fact, the principle of TO in a way follows rigid rules of categorization. This can systematically extract the collection of features based on the subdividing dimensions. Therefore, the hypothesis is that Tversky´s model is applicable to data-sets extracted from the terminological ontologies. Another important point in the context of the Relevance Theory of Communication is that translation should provide the set of assumptions that are adequately relevant to the audience, and the stimulus (that is translation) produced by the translator is such that it avoids gratuitous inferential processing effort on the audience´s part. Considering that *similarity serves as an organizing principle by which individuals classify objects, form concepts, and make generalizations* [2], the most similar concept to a SL concept, which is identified in the audience´s culture through the feature matching, could be the set of assumptions which are adequately relevant to the audience. Thus, the second hypothesis is that the optimization of the relevance required in an inter-cultural communication can be achieved by aligning the ontological graphs (conceptual hierarchies) and feature specifications which constitute concepts in the two language-dependent terminological ontologies. In order to assess these hypotheses, terminological ontologies are developed from corpora describing real-world concepts in the two remote cultures. The similarity score of the selected concepts are computed by applying Tversky´s model [2] based on the collection of features extracted from these ontologies. This is dealt with in the next chapter.

4 Feature Matching Based on Tversky´s Model

4.1 Corpora

Texts describing the Japanese educational system have been identified from the "Multilingual Living Information[1]" site provided by the Council of Local Authorities for International Relations and from a pamphlet entitled "Higher Education in Japan[2]" published by the Japanese Ministry of Education, Culture, Sports, Science and Technology. For the Danish educational system, documents that are downloaded from the Eurydice web-site[3] published by the Education, Audiovisual and Culture Executive Agency under the EU commission have been used as text corpus. All these documents are officially published in English by reliable authorities of each country. Thus, all English translated terms and expressions in their original languages are considered as official terms. It means that it is feasible to identify terminological expressions in an original language from documents published by the respective authorities. This enables one to eventually identify translation equivalences linking between, in this case, Danish and Japanese. In this study, only the English documents describing language-dependent concepts in the two cultures are used as text corpora.

The Eurydice publishes documents describing the educational systems in a majority of the EU member countries both in English and in their native languages. It means that the same method can, in principle, be applied to other language combinations.

4.2 Ontology Construction

The terms and their definitions describing the educational systems in each country are manually identified from the respective English documents. Based on these terms and their definitions, terminological ontologies representing the educational system in each

1 http://www.clair.or.jp/tagengorev/en/j/index.html

2 http://www.mext.go.jp/english/koutou/detail/1287370.htm

3 http://eacea.ec.europa.eu/education/eurydice/index_en.php

of the two countries are developed using a Computer Aided Ontology Structuring prototype (CAOS) that is based on the TO principles defined in [1]. As described in Chapter 1, the uniqueness of TO is feature specifications and subdivision criteria [6]. A feature specification is presented as attribute-value pair - for example in Figure 1, [ENTRANCE REQUIREMENT: high school graduate]. Thus, a representation of a whole concept is a feature structure, i.e. a set of feature specifications corresponding to the unique set of characteristics that constitutes that particular concept [1]. In Figure 1, each box that represents a particular concept is divided into three layers: 1) top layer, lexical representation (term), 2) middle layer, dimension specifications, and 3) bottom layer, feature structure (set of feature specifications).

The use of feature specifications is subject to principles and constraints described in detail in [1]. Most importantly, a concept automatically inherits all feature specifications of its superordinate concepts. Secondly, polyhierarchy is allowed so that one concept may be related to two or more superordinate concepts. On the other hand, subdivision criteria that have been used for many years in terminology work are strictly implemented in TO by introducing dimensions and dimension specifications [1][6]. This enables the CAOS prototype to perform consistency checking which helps in constructing ontologies [1]. A dimension of a concept is an attribute occurring in a non-inherited feature specification of one or more of its subordinate concepts [1][6]. Values of the dimension allow a distinction among sub-concepts of the concept in question. In Figure1, the concept "academic degree" has the dimension [LENGTH OF EDUCATION] whose values are [2-3 years | minimum 4 years]. These dimension values distinguish the sub-concepts: "junior college degree" and "university degree". This clarification makes it much easier to identify subdivision criteria and differentiating characteristics [6]. The same feature attribute can only occur on sister concepts and a given value can only appear on one of these sister concepts. In this way a concept must be distinguished from each of its nearest superordinate concepts as well as from each of its sister concepts by at least one feature specification [1][6].

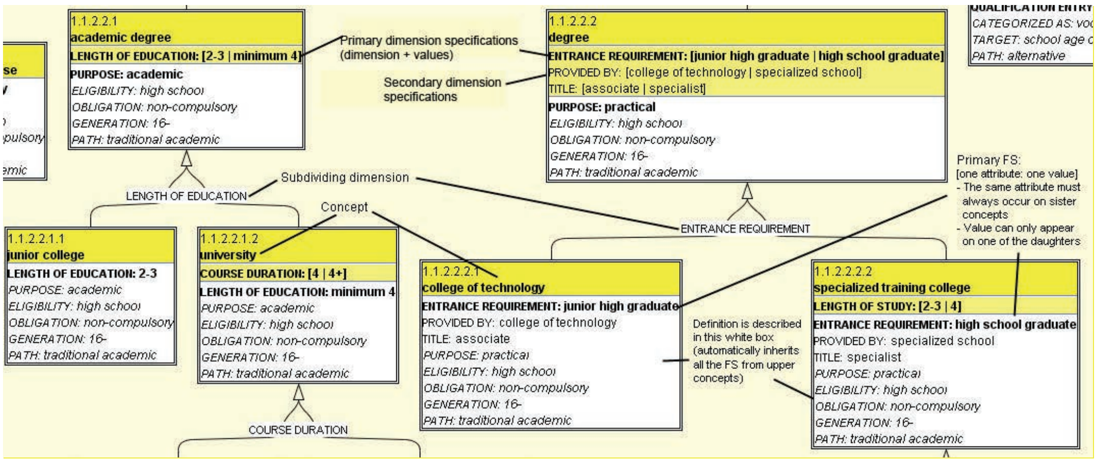


Fig. 1 Example of the Terminological Ontology.

By using the CAOS prototype that performs the consistency checking of the TO principles, the two educational ontologies are developed based on the terms and definitions manually extracted from the corpora.

4.3 Feature Matching Based on Tversky´s Contrast Model

The basic techniques generally used in the ontology matching are string-based (lexical) matching, graph-based (structural) matching, and feature-based matching [10]. Accordingly, string-based matching is manually performed as the first step. The Japanese and Danish educational system ontologies, respectively, contain 42 and 65 concepts consisting of terms (lexical representations) and their feature specifications. Among these, only two terms are completely matched. It indicates that graph-based matching based on the lexically matched nodes is not sufficient in this case. Therefore, feature-based matching is manually implemented in a top-down manner as the second step. The third highest dimension in the Japanese ontology and the highest dimension in the Danish ontology are "generation". Hence, the two ontologies are categorized into the following three blocks based on feature values of this generation dimension: "0-6 years old"; "6-15 years old"; and "16 years old and above."

Table 1. List of terms and feature sets.

Japanese concepts	
a:high school	A= {formal education, 16 years +, non-compulsory, lower secondary graduate}
b:general course	B= {formal education, 16 years +, non-compulsory, lower secondary graduate, general}
c:specialized training course	C= {formal education, 16 years +, non-compulsory, lower secondary graduate, specialized}
d:technical course	D= {formal education, 16 years +, non-compulsory, lower secondary graduate, specialized, technical}
e:business course	E= {formal education, 16 years +, non-compulsory, lower secondary graduate, specialized, business}
f: higher education	F= {formal education, 16 years +, non-compulsory, secondary graduate}
Danish concepts	
g:upper secondary education	G= {16 years +, lower secondary graduate}
h:general upper secondary education	H= {16 years +, lower secondary graduate, access to higher education}
i:gymnasium	I= {funded by state, 16 years +, lower secondary graduate, access to higher education}
j:business college	J= {self governing, 16 years +, lower secondary graduate, access to higher education}
k:HHX program	K= {self governing, 16 years +, lower secondary graduate, access to higher education, business}
l:HTX program	I= {self governing, 16 years +, lower secondary graduate, access to higher education, technical}
m:vocational or technical education	M={16 years +, lower secondary graduate, access to labor market}
n: tertiary education	N={secondary graduate, project and research}

In this study, the focus is on the block having feature value "16 years old and above." From this block, some of the sub-concepts and their feature values listed under the Japanese term, "non-compulsory education", and the Danish terms, "upper secondary education" and "tertiary education" are manually selected in Table 1 (due to the paper space, redundant data – e.g. concepts having similar constitution of feature sets – has intentionally been omitted). In order to apply Tversky´s model, synonymous feature expressions identified from the country specific corpora are approximately standardized by hand in Table 1. One thing to notice in Table 1 is that if a concept is categorized into several sub-concepts based on a dimension, an extra feature specification is added to each of them according to the principles of TO. Hence it is possible to observe the hierarchical structure from the feature values listed in Table 1.

Now the question is how to assign the asymmetric parameters in accordance to the translation direction. In [2], the direction of asymmetry is determined by the relative salience of the stimuli, in other words, the variant is more similar to the prototype than vice versa. Thus, *if sim(a,b) is interpreted as the degree to which a is similar to b, then a is the subject of the comparison and b is the referent. Hence the features of the subject are weighted more heavily than the features of the referent.* When considering a translation scenario, translators´ task is to identify a concept in audiences´ conceptual structure that is optimally relevant to the concept in the SL. It means that the stimulus selected by a translator should to the maximum extent be similar to a concept in the SL concept. Therefore, the features of a stimulus should be weighted more heavily than the ones of an SL concept in accordance to [2]. Hence, the asymmetric parameters are manually set as α=0.7 and β=0.3 in this empirical study. The result is shown in Table 2 and 3. In Table 2, the Danish concepts (*g-n*) are set as subject of the comparison and the Japanese (*a-f*) as referent. Opposite to this, the Japanese concepts (*a-f*) are set as subject of the comparison and the Danish (*g-n*) as referent in Table 3.

Table 2. Tversky´s similarity score: *a-f* (JP) as referent (SL), *g-n* (DK) as stimulus (TL).

SL	*g*	*h*	*i*	*j*	*k*	*l*	*m*	*n*
a	**.77**	.61	.50	.50	.43	.43	.61	.00
b	**.69**	.56	.47	.47	.4	.4	.56	.00
c	**.69**	.56	.47	.47	.4	.4	.56	.00
d	**.63**	.51	.43	.43	.38	.57	.51	.00
e	**.63**	.51	.43	.43	.57	.38	.51	.00
f	**.38**	.30	.25	.25	.21	.21	.30	**.38**

Table 3. Tversky´s similarity score: *g-n* (DK) as referent (SL), *a-f* (JP) as stimulus (TL).

SL	*a*	*b*	*c*	*d*	*e*	*f*
g	**.58**	.49	.49	.41	.41	.29
h	**.54**	.45	.45	.39	.39	.27
i	**.50**	.43	.43	.37	.37	.25
j	**.50**	.43	.43	.37	.37	.25
k	.47	.4	.4	.35	**.53**	.23
l	.47	.4	.4	**.53**	.35	.23
m	**.54**	.45	.45	.39	.39	.27
n	.00	.00	.00	.00	.00	**.29**

Note1: Bold font is the highest score in each row

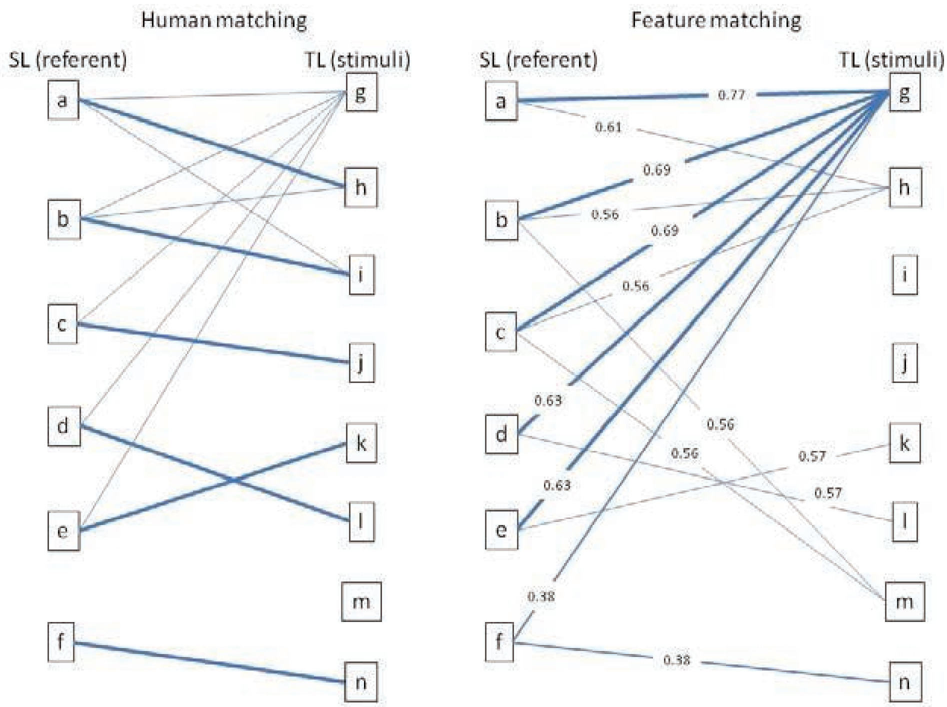


Fig. 2 Comparison with human matching: Danish terms as stimuli

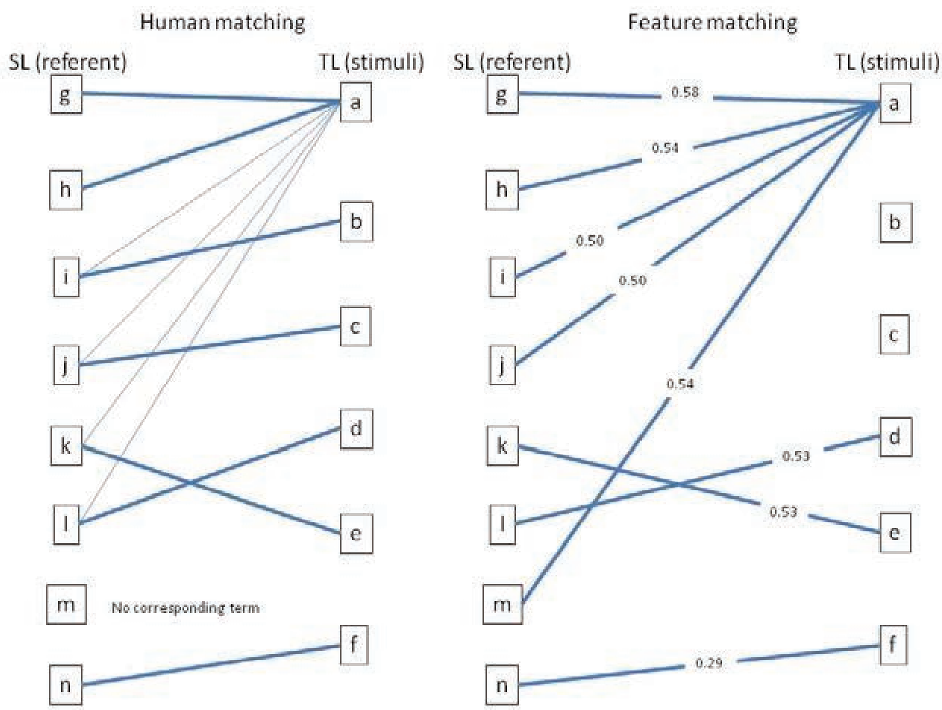


Fig. 3 Comparison with human matching: Japanese terms as stimuli

In Table 2 and 3, the scores with bold fonts are the highest scores in each row. From these tables, it can be interpreted that a concept with the highest score has the most optimal relevance to an SL concept.

The first notable point is that the application of the asymmetric parameters resulted in the asymmetric bidirectional relationships in most of the links between Danish and Japanese concepts. To be more precise, it is less optimal to use the Japanese term as stimulus, when conveying the original meaning of the Danish concept (e.g. the asymmetric score for term *c* as stimulus and *g* as referent: 0.49). On the other hand, it is more optimal to use the Danish term as stimulus, when conveying the original meaning of the Japanese concept (the asymmetric score for term *g* as stimulus and *c* as referent: 0.69).

Another point is that the majority of the identified optimal stimuli in Table 2 and 3 were the most general terms located at the highest hierarchy in the data-sets. From this viewpoint, it is difficult to assess whether the application of Tversky´s model to the terminological ontologies is successful. Hence the feature matching results are compared with the human matching results in Figure 2 and 3. In the human matching charts (the left side of the figures), the bold line indicates the ideal stimulus (optimal translation candidate) and the slim line indicates the acceptable stimuli (reasonably acceptable translation candidates) for each SL term. In the feature matching charts (the right side of the figures), the bold line indicates the optimal stimulus having the highest score for each SL term. The slim line indicates the stimuli that are not the highest score for a SL term, but having scores over 0.55.

5 Discussion

As described in Chapter 4, among all the concepts between the two ontologies, only two terms were completely matched in the string-based matching. This indicates that the English educational terminology used in respective knowledge sharing communities is immensely dissimilar, even though the educational concepts existing in the two countries are relatively similar. From this observation, it can be elaborated that it is very complicated to link concepts in two remote languages. This is because language resources having direct links between two remote languages are usually very limited, and therefore a pivot translation via English is often required both for dictionary-based human translations and for statistically-based machine translations. This also emphasizes the necessity for carefully analyzing how the meanings of a concept in one culture can be conveyed to a person in another culture through English as lingua franca.

From this viewpoint, the empirical analysis in Chapter 4 showed modest progress. The results listed in Table 2 and 3 as well as Figure 2 and 3 indicate that Tversky´s contrast model [2] is to a certain extent applicable to data-sets extracted from terminological ontologies. The application of the asymmetric parameters showed an interesting indication that it is less optimal to use the Japanese term, e.g. “c: high school-specialized course” as stimulus for a Japanese audience, when conveying the original meaning of the Danish concept e.g. “g: upper secondary education” (the asymmetric score is 0.49). On the other hand, it is more optimal to use the Danish term e.g. "g: upper secondary education" as stimulus for a Danish audience, when conveying the original meaning of the Japanese concept e.g. "c: high school-specialized course" (the asymmetric score is 0.69). Even though concepts in two cultures are mapped to each other, it is not necessarily true that a translational equivalence holds in a bidirectional way, if the two concepts are not 100% identical. This result can be explained as follows: when considering equation (1) of Tversky´s model, it is obvious that if features for the parameter α (features that are in the feature set A but not in set B: (A-B)) increases, the similarity score severely decreases. The reason for the aforementioned result (the use of Japanese term as stimulus is less optimal) is that, when we categorize the two ontologies into three blocks, we select the "generation" dimension that is the third highest dimension in the Japanese educational system. Therefore, the feature values of the first- and second highest dimensions have been inherited to all the Japanese concepts listed in Table 1. This indicates that the asymmetric parameters of Tversky´s model, to a certain degree, reflect the hierarchical structure hidden behind the feature structures of the terminological ontologies.

A final point is that, in most cases, the identified optimal stimuli based on the similarity score in Table 2 and 3 is the most general term located at the highest hierarchy in the data-sets. According to the principle of TO, when a concept is subdivided into several sub-concepts based on a dimension, an extra feature is added to each sub-concept. Hence it is often the case that concepts having more features are more specific sub-concepts. It means that the lower a concept is located in the ontology, the more features the concept inherits from superordinate concepts. If dimensions and their values at the lower part of the two ontologies are not consistent, all the inherited features are simply acting as noise in the data-sets. The positive interpretation could be that Tversky´s model is applicable to identify corresponding pairs with less noise, in other words, pairs that optimally share common features with less noise. The negative interpretation could be that Tversky´s model has limitations in identifying corresponding pairs at the optimally specific level. Considering communication in the real world, it is not incorrect to say that the relevance required in the communication is achieved in this way, since people can usually achieve a mutual understanding much easier at a reasonably general level than at a very specific level. However, as Figure 2 and Figure 3 illustrate, the optimal translation candidates selected by human are the optimally specific concepts. Hence one of the challenges is to identify the reasonably specific terms from noisy data-sets. In order to achieve this, additional investigations (e.g. implementing the feature matching for the all concepts in the two ontologies) are required.

Another future challenge is to further investigate and compare this empirical study with data-sets obtained from terminological practices as well as from the translation practices in the real world. The data-sets used in this study are English documents published by the EU commission for the Danish educational system. The EU commission has used English terminology that is standardized based on the International Standard Classification of Education (ISCED) defined by the United Nations Educational, Scientific and Cultural Organization. Therefore, the documents describing the educational system for the majority of EU member countries are based on the standardized classification and English terminology. Hence, it may be much easier to align the educational system ontologies constructed from documents published by the EU member countries. On the other hand, the Japanese ontology does not conform to the same standardized classification and terminology. By applying Tversky´s contrast model to the different data-set combinations, it may be possible to investigate the behavior of data in different scenarios. Extending from this viewpoint, the further development of the CAOS prototype for automating the knowledge extraction, ontology construction and update described in [11] could efficiently be synchronized in order to link language-specific concepts existing in different countries. To be more concrete, it might be effective to automatically extract knowledge from the domain-specific corpora based on pre-defined feature dimensions derived from a standardized classification and the terminology e.g. the ISCED classification.

Finally, Tversky´s contrast model has been extended by several researchers in different disciplines. Especially in the area of cognitive science, Tenebaum and Griffiths [12] proposed a framework that subsumes Tversky´s model of similarity by recasting Shepard´s universal law of generalization [10] in a more general Bayesian framework. Frank et al. [13] further extended this framework in [12] in order to model informative communication based on [7]. Hence, it is an obvious future challenge to apply these extended cognitive models to the aforementioned different combinations of data-sets.

6 Conclusion

In this paper, the applicability of Tversky´s contrast model derived from the cognitive theory [2] to data-sets extracted based on the Terminological Ontology method is investigated. The study indicates that the application of [2] to [1] could, to a certain extent, enable one to analyze not only the degree of relevance between concepts in two cultures, but also the degree of asymmetric relationship between the concepts. By extending [2] to e.g. [12][13], it may be feasible to investigate further how meanings of a concept in one culture can be effectively conveyed to a another culture through English as lingua franca. However, further investigations using data-sets obtained from terminological practices as well as from translation practices are needed in order to clarify the limitations pointed out in this study.

Acknowledgments. I would like to express my gratitude to my supervisors, Hanne Erdman Thomsen and Daniel Hardt for their valuable guidance and support on my Ph.D. project. I would also like to thank my colleagues in the psycholinguistic reading group at Copenhagen Business School, who provided me with a selection of highly relevant papers and constant inspiration in the area of intercultural communication and translation theories.

References

1. Madsen, B.N., Thomsen, H.E., Vikner, C.: Principles of a System for Terminological Concept Modelling. In: Proc. The 4th International Conference on Language Resources and Evaluation, ELRA, 15-19 (2004)
2. Tversky, A.: Features of Similarity. In: Psychological Review 84, 327-352 (1977)
3. Declerck, T., Krieger, H.U., Thomas, S.M., Buitelaar, P., O´Riain, S., Wunner, T., Maguet, G., McCrae, J., Spohr, D., Montiel-Ponsoda, E.: Ontology-based Multilingual Access to Financial Reports for Sharing Business Knowledge Across Europe. In: Rooz, J., Ivanyos, J. Eds.: Internal Financial Control Assessment Applying Multilingual Ontology Framework, Budapest: HVG Press, 67-76 (2010)
4. Vossen, P., Agirre, E., Calzolari, N., Fellbaum, C., Hsieh, S., Huang, C., Isahara, H., Kanzaki, K., Marchetti, A., Monachini, M., Neri, F., Raffaelli, R., Rigau, G., Tescon, M.: KYOTO: A System for Mining, Structuring and Distributing Knowledge Across Languages and Cultures. In: Proc. The 6th International Conference on Language Resources and Evaluation, Morocco, 1462-1469 (2008)
5. ISO 704: Terminology Work – Principles and Methods. Genève: ISO (2000)
6. Madsen, B.N., Thomsen, H.E., Vikner, C.: Comparison of Principles Applying to Domain Specific Versus General Ontologies. In: Oltramari, A., Paggio, P., Gangemi, A., Pazienza, M.T., Calzolari, N., Pedersen, B.S., Simov, K. (eds.): OntoLex 2004: Ontologies and Lexical Ressources in Distributed Environments. ELRA, 90-95 (2004)
7. Sperber D., Wilson, D.: Relevance: Communication and Cognition. Blackwell, Oxford (1986)
8. Guut, E.A.: A Theoretical Account of Translation – without a translation theory. In: Target: International Journal of Translation Studies, 2/2, 135-164 (1990)
9. Shepard, R.N.: Toward a Universal Law of Generalization for Psychological Science. In: Science, New Series, Vol. 237, No. 4820, 1317-1323 (1987)
10. Euzenat J., Shvaiko, P.: Ontology Mapping. Springer-Verlag, Berlin Heidelberg (2007)
11. Madsen B.N., Thomsen H.E., Halskov J., Lassen T.: Automatic Ontology Construction for a National Term Bank. In: Proc. Terminology and Knowledge Engineering Conference 2010, Dublin, 502-533 (2010)
12. Tenenbaum, J.B., Griffiths, T.L.: Generalization, Similarity, and Bayesian Inference. In: Behavioral and Brain Sciences 24, 629-640 (2001)
13. Frank, M. C., Goodman, N. D., Lai, P., Tenenbaum, J. B.: Informative Communication in Word Production and Word Learning. In: Proc. The 31st Annual Meeting of the Cognitive Science Society (2009)

2. Alignment of remote cultures[9]

Preliminary study:

This paper further investigates on a slightly larger scale than the previous, whether the alignment of two language-dependent TOs is a potential method for optimizing the relevance required in cross-cultural communications, by assessing the precision/recall score comparing the results against human mapping.

[9] In Proceedings of Culture and Computing 2011, October, Kyoto, Japan. (peer-reviewed full paper: nominated as one of four finalists for the Best Paper Award)

Alignment of Remote Cultures

In contrast to the Relevance Theory of Communication

Fumiko Kano Glückstad
Dept. of International Language Studies and Computational Linguistics
Copenhagen Business School, CBS
Frederiksberg, Denmark
e-mail: fkg.isv@cbs.dk

Abstract— **As the role of ontology in a multilingual setting becomes important to Semantic Web development, it becomes necessary to understand and model how an original conceptual meaning of a Source Language word is conveyed into a Target Language translation. Terminological ontology [1] is a tool used for knowledge sharing and domain-specific translation, and could potentially be suitable for simulating the cognitive models explaining real-world inter-cultural communication scenarios. In this paper, a framework referred to as the Relevance Theory of Communication [2] is contrasted to an empirical study applying Tversky´s contrast model [3] to data-sets obtained from the terminological ontology. The results indicate that the alignment of two language-dependent terminological ontologies is a potential method for optimizing the relevance required in inter-cultural communication, in other words, for identifying corresponding concepts existing in two remote cultures.**

Keywords - Terminology, Ontology, Inter-cultural communication, Set-theory, Similarity, Relevance, Translation

I. Introduction

The role of ontology in a multilingual setting is an emerging challenge for Semantic Web development. As a consequence, there are several major ongoing projects, such as MONNET on Multilingual Ontologies for Networked Knowledge project [4] and KYOTO on Knowledge-Yielding Ontologies for Transition-Based Organization [5]. Though both projects deal with translation of terms from a Source Language (SL) to a Target Language (TL), they focus on linking lexical data through an interoperable common ontology rather than on optimizing relevance between concepts that are potentially measurable based on diverse models derived from the cognitive theory.

Another major project, the Language Grid [6] is an infrastructure that combines language resources such as bilingual dictionaries and language processing tools. Its technology is based on an ontology called Language Service Ontology and Semantic Web Service Technology. These two technologies enable one to combine bilingual dictionaries, e.g. Japanese-English and English-Danish dictionaries, and natural language processing tools from all over the world. Thus translations between languages, e.g. Japanese-Danish, can be realized through the transitive translation using English as a pivot language. However, this approach focuses on combining the existing language resources rather than the cognitive process required in human translation. A question is how well such a pivot translation can convey an original conceptual meaning of an SL word into a TL translation.

Terminological Ontology (TO) is a domain-specific ontology used for knowledge sharing [1], which normally is applied in terminology work, cf. for example [7]. The unique points of TO that differentiate it from other types of ontologies are feature specifications and subdivision criteria [8]. A feature specification consists of a feature dimension and its value. Thus, a representation of a whole concept is a feature structure, i.e. a set of feature specifications corresponding to the unique set of characteristics that constitutes that particular concept [1][8]. Terminological ontologists argue that concepts are defined in a language dependent context, and therefore TO is language dependent. TO is developed within a knowledge sharing community, then dynamically updated and validated. If it is necessary to share knowledge with other communities, TOs developed in different communities should be compared, aligned and merged as needed. While the aforementioned two mainstream projects, MONNET and KYOTO, both deal with complex ontologies involving huge data-sets, TO usually handles smaller amounts of concepts.

Consequently, a point that should be emphasized in this work is that TO could potentially be a suitable method to apply and for simulating cognitive models explaining a real-world inter-cultural communication scenario. Thus, in this paper, a framework referred to as the Relevance Theory of Communication [2] is contrasted to an empirical study applying a classical psychological model [3] to data-sets obtained from TO [1]. The eventual purpose of this study is to propose an approach for identifying potential translation candidates by optimizing relevance between concepts in two remote languages such as a set of European and Asian languages. The expectation is that such translation candidates should provide better cognitive understandings of SL concept-meanings among a TL audience. However, in this preliminary study, the scope is limited to align ontologies constructed from English texts describing the educational systems in two remote cultures.

The rest of the paper is structured as follows: the Relevance Theory of Communication is reviewed in the next section by use of a real-world example. Section 3 describes a classical psychological model applied to this preliminary study. In Section 4, the empirical analysis is performed to assess the potential of applying the models based on [3] to the terminological ontologies. Section 5 discusses findings and future work followed by conclusions in Section 6.

II. THE RELEVANCE THEORY OF COMMUNICATION

Imagine a situation where a non-native English speaking European and an Asian are debating in English about the issue of academic degree systems in their respective countries. While a Danish might be explaining about the Doctor of Science degree (the highest achievable academic doctor degree in Denmark after obtaining a Doctor of Philosophy that is considered a lower doctorate degree), a Japanese might be having the highest possible academic degree in Japan in his mind which is a Doctor of Philosophy degree (also frequently referred to as Doctor of Science in Japan). This imagined conversation shows a typical scenario revealing a deep inherent misconception between the two communicating parties since each of them have their own conceptual - and correct - understanding of the highest obtainable academic degree in their respective cultures.

This example may further create problems for a translator who is going to translate academic titles into the language of the other party. When a translator translates the term for the Danish Doctor of Science Degree into Japanese, the first condition he/she has to fulfill is that his/her translation should convey the same meaning as the original Danish meaning. Gutt [9] explains that this requires *the receptors to familiarize themselves with the context envisaged for the original text.* Now the question is, when a Japanese receptor is not familiar with the Danish language and its academic culture, how should this particular Danish academic title be translated into Japanese?

The proposal [9] of applying the Relevance Theory of Communication [2] might be a key to address this issue. This theory focuses on how people share thoughts with one another and views communication as principally an inferential process. It means that the essential task of the communicator is to produce a *stimulus* from which the audience can infer what set of thoughts or assumptions the communicator intends to convey [9]. Hence the second condition the translator has to fulfill is that his/her translation should explicitly provide a set of assumptions that are adequately relevant to the audience. The set of assumptions the translator provides should therefore be a stimulus that avoids unnecessary inferential processing effort required for the audience. The issue here is how the translator should create such *stimulus* (that is translation) optimally relevant to the audience. Assuming that both Danish and Japanese have their respective conceptual structures of the academic system rooted in their own culture, translation candidates that have optimally relevant relationships identified from these two conceptual systems could avoid the gratuitous inferential processing effort on the audience´s part.

The optimization of the relevance between two concepts could be well explained by the cognitive model, Tversky´s contrast model [3]. Thus the next section reviews Tversky´s model and considers how his model could be used in the context of optimizing the relevance of communication.

III. TVERSKY´S CONTRAST MODEL

The concept of similarity has a long history within the area of cognitive science. Tversky´s view of similarity [3] is distinguished from the traditional theoretical analysis (c.f. [10]) on two points: 1) while the theoretical analysis of similarity relations has been dominated by the continuous metric space models, [3] argues that *the assessment of similarity between objects may be better described as a comparison of features rather than as the computation of metric distance between points;* and 2) although *similarity has been viewed by both philosophers and psychologists as a prime example of a symmetric relation*, the asymmetric similarity relation has been demonstrated in [3] based on several empirical evidences.

Based on these two points, [3] proposed a classic feature-set model of similarity as follows:

$$sim(a,b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha \cdot f(A-B) + \beta \cdot f(B-A)} \qquad (1)$$

Here, A and B are the feature sets of object *a* and object *b*. *f* denotes a measure over the feature sets. ($A \cap B$) represents the sets of features present in both A and B, (A-B) represents the sets of features present in A but not in B and (B-A) represents the sets of features present in B but not in A. α and β are free parameters representing an asymmetric relationship between A and B. Since the similarity score in this equation is normalized, the obtained score lies between 0 and 1.

An interesting point is that the application of Tversky´s model requires *a limited list of relevant features and the representation of an object as a collection of features that is viewed as a product of a prior process of extraction and compilation* [3]. In fact, the principle of TO in a way follows rigid rules of categorization. This can systematically extract the collection of features based on subdividing dimensions. Therefore, the hypothesis is that Tversky´s model could be applicable to data-sets extracted from the terminological ontologies. Another important point in the context of the Relevance Theory of Communication is that translation should provide the set of assumptions that are adequately relevant to the audience. Considering that *similarity serves as an organizing principle by which individuals classify objects, form concepts, and make generalizations* [3], the most similar concept to an SL concept, which is identified in the audience´s culture through the feature matching, could be the set of assumptions which are adequately relevant to the audience. Thus, the second hypothesis is that the optimization of the relevance required in an inter-cultural communication can be achieved by aligning the feature specifications which constitute concepts in the two language-dependent terminological ontologies. In order to assess the first hypothesis, terminological ontologies are developed from corpora describing real-world concepts in the two remote cultures. The similarity score of the selected concepts are computed by applying Tversky´s model [3] based on the collection of features extracted from these ontologies. This is dealt with in the next section.

IV. Ontology Alignment

A. Corpora

Texts describing the Japanese educational system have been identified from the "Multilingual Living Information" site provided by the Council of Local Authorities for International Relations and from a pamphlet entitled "Higher Education in Japan" published by the Japanese Ministry of Education, Culture, Sports, Science and Technology. For the Danish educational system, documents that are downloaded from the Eurydice web-site published by the Education, Audiovisual and Culture Executive Agency under the EU commission have been used as text corpus. All these documents are officially published in English by reliable authorities of each country. Thus, all English translated terms and expressions in their original languages are considered as official terms. It means that it is feasible to identify terminological expressions in an original language from documents published by the respective authorities. This enables one to eventually identify translation equivalences linking between, in this case, Danish and Japanese. In this study, only the English documents describing about language-dependent concepts in the two cultures are used as text corpora.

B. Ontology construction

The terms and their definitions describing the educational systems in each country are manually identified from the respective English corpora. Based on these terms and their definitions, terminological ontologies representing the educational system in each of the two countries are developed using the Computer Aided Ontology Structuring prototype (CAOS) that is based on TO principles defined in [1]. As described in Section 1, the uniqueness of TO is feature specifications and subdivision criteria [8]. A feature specification is presented as attribute-value pair. Thus, a representation of a whole concept is a feature structure, i.e. a set of feature specifications corresponding to the unique set of characteristics that constitutes that particular concept [1].

The use of feature specifications is subject to principles and constraints described in detail in [1]. Most importantly, a concept automatically inherits all feature specifications of its superordinate concepts. Secondly, polyhierarchy is allowed so that one concept may be related to two or more superordinate concepts. On the other hand, subdivision criteria that have been used for many years in terminology work are strictly implemented in TO by introducing dimensions and dimension specifications [1][8]. This enables the CAOS prototype to perform consistency checking which helps in constructing ontologies [1]. A dimension of a concept is an attribute occurring in a non-inherited feature specification of one or more of its subordinate concepts [1][8]. Values of the dimension allow a distinction among sub-concepts of the concept in question. For example, a dimension of a concept "higher education" is [TITLE] whose values are [associate | bachelor]. These dimension values distinguish the sub-concepts: "junior college" and "university (undergraduate)". The dimension can only occur on sister concepts and a given value can only appear on one of these sister concepts. In this way a concept must be distinguished from each of its nearest superordinate concepts as well as from each of its sister concepts by at least one feature specification [1][8].

By using the CAOS prototype that performs the consistency checking of TO principles, the two educational ontologies are developed based on the terms and definitions manually extracted from the corpora.

C. Application of Tversky´s contrast model

Based on the TO principles, the Japanese and Danish educational ontologies respectively containing 40 and 54 concepts are developed. All concepts consist of features systematically extracted from the terminological ontology. For example, the Japanese term, "specialized training college – specialist course", consists of features {formal education, require upper secondary exam, profession oriented education, 2-3 years program duration}. On the contrary, the Danish term, "short cycle education", consists of features {require upper secondary exam, project & research based, profession oriented education, administered by Ministry of Education, 1-3 years program duration, vocational college, earn credit to medium or longer cycle education}. These features existing in a Danish concept and a Japanese concept are used for computing a similarity score based on Tversky´s model (equation 1). In order to apply this model, synonymous expressions of features extracted from the country specific corpora are approximately standardized by hand. Accordingly, in case of the aforementioned examples, $(A \cap B)$ is 2, since the features {require upper secondary exam, profession oriented education} exist in both the Danish and Japanese terms. As for the asymmetric similarity, it is defined in [3] that *if sim(a,b) is interpreted as the degree to which a is similar to b, then a is the subject of the comparison and b is the referent. Hence the features of the subject are weighted more heavily than the features of the referent*. When considering a translation scenario, translators´ task is to identify a concept in audiences´ conceptual structure that is optimally relevant to the concept in the SL. It means that the stimulus selected by a translator should to the maximum extent be similar to a concept in the SL concept. Therefore, the features of a stimulus should be weighted more heavily than the ones of an SL concept in accordance to [3]. Hence, the asymmetric parameters are manually set as α=0.7 and β=0.3 in this empirical study. It means that the similarity score becomes smaller, when (A-B) is larger than (B-A) in the equation (1). In case of the aforementioned example of "specialized training college – specialist course" and "short cycle education", (A-B) is 5 and (B-A) is 2 when the Danish term is set as variant. On the other hand, (A-B) is 2 and (B-A) is 5, when the Japanese term is set as variant. Accordingly, the asymmetric similarity scores are, 0.328 and 0.408, respectively. In this way, the asymmetric similarity scores are computed in all possible combinations between the Japanese and the Danish concepts. Examples of the mapped concepts based on the similarity scores are illustrated in Figure 1 and 2. The concepts listed in Figure 1 and 2 are automatically selected based on the similarity scores. It means that Tversky´s model roughly categorizes clusters that are more or less similar to each other. In these figures, the bold lines are the highest scores equal to or over 0.5 obtained for each SL concept, and the slim lines are the highest score lower than 0.5, or the second highest scores equal or higher than 0.5. The bold dotted lines are ideal corresponding concepts and the slim dotted lines are acceptable corresponding concepts defined by human.

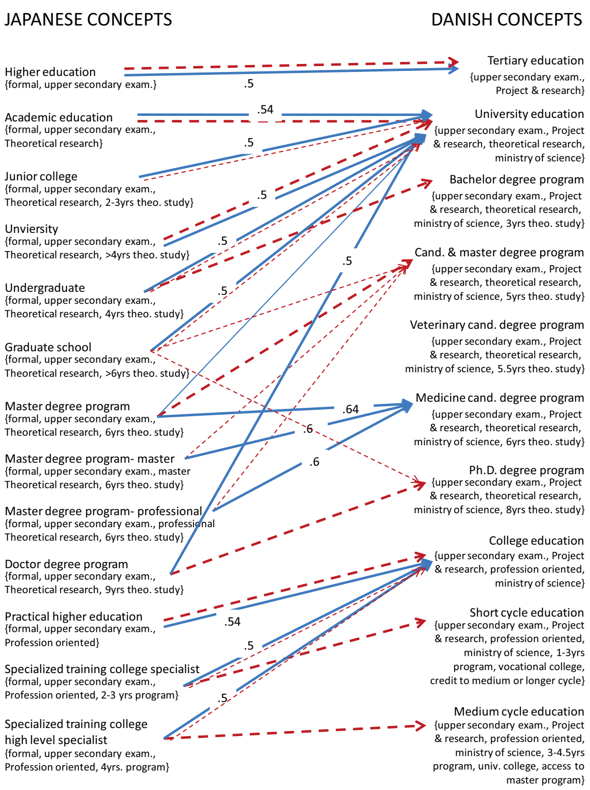


Figure 1. Tversky´s similarity score (Danish as variant)

Based on the aforementioned criteria, precision and recall scores are computed based on the following equations.

$$precision = \frac{\{corresponding\ pairs\} \cap \{mapped\ pairs\}}{\{mapped\ pairs\}} \quad (2)$$

$$recall = \frac{\{corresponding\ pairs\} \cap \{mapped\ pairs\}}{\{corresponding\ pairs\}} \quad (3)$$

The precision and recall scores are calculated in two ways. For the first scores (referred to as "strict"), only the ideal corresponding pairs (bold dotted lines) are considered as corresponding concepts and only the highest scores (bold lines) are considered as most similar concepts. For the second scores (referred to as "relaxed"), the acceptable corresponding pairs (slim dotted lines) are included in addition to the ideal corresponding pairs and the highest score lower than 0.5, or the second highest scores equal or higher than 0.5 (slim lines) are included in addition to the most similar concepts. The results are summarized in Table 1.

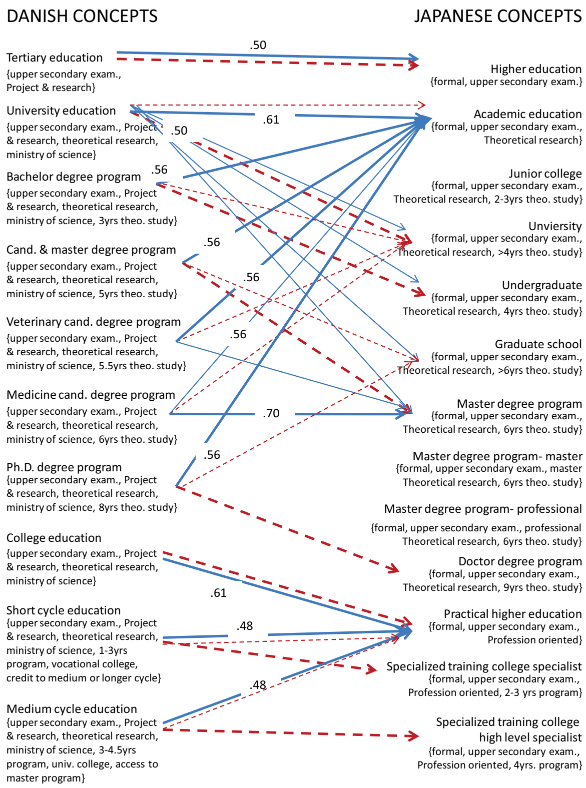


Figure 2. Tversky´s similarity score (Japanese as variant)

TABLE I. PRECISION AND RECALL

	Srict alignment				Relaxed alignment			
	Corresponding pairs	Mapped pairs	Precision	Recall	Corresponding pairs	Mapped pairs	Precision	Recall
DA as variant	26	36	27.8 %	38.5 %	72	77	44.2 %	47.2 %
	10				34			
JA as variant	26	32	28.1 %	34.6 %	53	62	25.8 %	30.2 %
	9				16			

V. DISCUSSION

A. Findings

In this empirical study, the terminological ontologies are constructed from English texts describing the educational systems in two remote cultures. As Figure 1 and 2 illustrate, the English expressions of concepts in the two ontologies are in most cases not identical. This indicates that the English educational terminology used in respective knowledge sharing communities is immensely dissimilar, even though the educational concepts existing in the two countries are relatively similar. From this fact, it can be elaborated that it is a very complicated task to link concepts in two remote languages. This is because language resources having direct links between two remote languages are usually very limited, and therefore it often requires a pivot translation using English both for dictionary-based human translations and for statistically-based machine translations. This also emphasizes the necessity for carefully analyzing how a concept in one culture can be conveyed to a person in another culture through English as lingua franca.

In this study, the feature-based asymmetric similarity has been computed for all combinations of concepts in two cultures in both directions. Accordingly, concepts in the two ontologies were approximately clustered into 5 groups. Figure 1 and 2 show one of the clusters automatically extracted from this process. These figures indicate that concepts referring to education for people who have completed the secondary education are adequately extracted from the two ontologies. This can be explained in the following way: The principles of Terminological Ontology - e.g. inheritance of feature specifications, uniqueness of dimensions, and uniqueness of primary feature specifications - enable features to reflect the structure of ontologies. For example, a feature value, "upper secondary exam" in Figure 1 and 2 has been specified only once when defining the concept "higher education" and "tertiary education" respectively in the Japanese and the Danish ontologies. Since all of the sub-concepts under these two terms inherit this feature value, the concepts referring to education for people who completed the upper secondary exam have systematically been extracted and grouped as a cluster.

The computation of the asymmetric similarity scores for all combinations of concepts within each cluster resulted in identifying a concept in a target ontology that has the highest optimal relevance to a source concept. Figure 1 and 2 show that, in most cases, the identified optimal stimuli based on the asymmetric score was the most general term located at the highest node within the cluster. According to the principle of TO, when a concept is subdivided into several sub-concepts based on a dimension, an extra feature is added to each sub-concept. Hence it is always the case that concepts having more features are more specific sub-concepts. It means that the lower a concept is located in the ontology, the more features the concept inherits from superordinate concepts. If dimensions and their values at the lower part of the two ontologies are not consistent, all the inherited features are simply noise in the data-sets. Therefore, the identified optimal stimuli were in most cases more general and abstract terms rather than specific terms. The positive interpretation of this phenomenon could be that Tversky´s model is applicable to identify corresponding pairs with less noise, in other words, pairs that optimally share common features with less noise. The negative interpretation could be that Tversky´s model has limitations in identifying corresponding pairs at the optimally specific level. Considering communication in the real world, it is not incorrect to say that the relevance required in the communication is achieved in this way, since people can often achieve mutual understandings much easier at a reasonably general level than at a very specific level. However, there is a room for further investigating this phenomenon in the future.

Another notable point is that, the asymmetric similarity measure based on common- and distinctive features seems to be a very useful approach for the translation and inter-cultural communication. The results shown in Table 1 indicate that the precision and recall scores for Danish as variant are generally higher than Japanese as variant. The reason could be that the Japanese ontology has extra layers of the hierarchy compared to the Danish one. This made it difficult to identify a Japanese concept as stimulus. Even though concepts in two cultures are mapped to each other, it is not necessarily true that a translational equivalence holds both ways, if the two concepts are not 100% identical. Furthermore, in culturally dependent domains, some concepts simply do not exist in another culture. In such situations, reasonably abstract concepts identified as a corresponding pair in the target ontology could be considered as the optimal translation. This cannot be achieved by symmetric similarity measures.

Finally, a major problem has also been identified in this empirical study. In a cluster consisting of concepts referring to the Danish continuing education and Japanese alternative education, some mismatches have been identified. This is because the Japanese education system has been classified into two categories, "formal education" and "alternative education" at the highest level of the ontological hierarchy, while the Danish education system has been classified based on the "age and entrance qualification" dimension at the highest level. Under the Danish continuing education concept, there are two types of education, "formal education" and "non-formal education". Since the feature value "formal education" in the two ontologies was considered as a common feature, all the Japanese formal educations that are from the primary education to the higher education and the Danish formal educations under the Danish continuing education targeted for adults were considered as corresponding concepts. This strongly influenced the precision and recall scores in Table 1. This example indicates that the selection of feature dimensions and values upon the development of ontology is a very sensitive issue.

B. Future challenges

One of the challenges pointed out from the aforementioned findings is to identify reasonably specific corresponding concepts from noisy data-sets. One approach is to investigate each situation where features are considered as either noisy or useful data. It is pointed out in [11] that Tversky´s contrast model allows any kind of features and any feature weights whatsoever. However, to improve the mapping quality, some semantic questions such as what qualifies as a feature and what determines the feature weights should be investigated. Considering the aforementioned example, even though the feature value “formal education” is considered as a common feature, if other features in the Danish and Japanese concepts consist of features that hold a conflicting relationship, e.g. one has a feature “6 years old or above” and the other “adult”, the impact of common feature should be smaller. How to weight each feature may be a key to improving the mapping quality. The Bayesian model proposed in [11] explains some aspects of the origins and the dynamics of feature weights. Another question raised in [11] is that Tversky´s formal theory does not explain why and how the two subsets of distinctive features (A-B) and (B-A) should be given different weights. In fact, the parameters used in this empirical study have been randomly selected. This issue is dealt with in the Bayesian model as well. Hence, one way to approach this is to extend Tversky´s model based on the idea proposed in [11] and apply it to the same data-sets.

Another way of approaching the challenge is to investigate how feature dimensions and values are selected during the ontology construction phase. Considering the technical aspects of the ontology matching, handling of the ontology construction may not be the scope for the improvement of matching results. However, for the purpose of translation and inter-cultural communication of domain specific terms that are constrained by a cultural boundary (such as education system, social system, legal system, financial system), the standardization is considered as an important process for sharing knowledge across country borderlines. In fact, the documents describing the educational system for the majority of EU member countries are based on the standardized classification and terms on the Eurydice web-site, since the EU commission has employed English terminology that is standardized based on the International Standard Classification of Education (ISCED) defined by the United Nations Educational, Scientific and Cultural Organization. Therefore, it may be much easier to align the educational system ontologies within the EU member countries. Further investigation and comparison of this empirical study with data-sets obtained from standardization practices as well as from the translation practices in the real world would contribute to the optimization of the relevance required in an inter-cultural communication scenario.

VI. Conclusions

In this paper, a framework referred to as the Relevance Theory of Communication proposed by [2] is contrasted to an empirical study applying Tversky´s contrast model [3] to data-sets obtained from the Terminological Ontology method [1]. The results demonstrate that the application of Tversky´s contrast model to the data-sets extracted from the terminological ontologies is to a certain degree an effective approach. However, further investigations are needed to assess whether the optimization of the relevance required in an inter-cultural communication can be achieved by aligning the feature specifications which constitute concepts in the two language-dependent terminological ontologies.

Acknowledgment

I would like to express my gratitude to my supervisors, Hanne Erdman Thomsen and Daniel Hardt for their valuable guidance and support on my Ph.D. project.

References

[1] B.N. Madsen, H.E. Thomsen, and C. Vikner, “Principles of a system for terminological concept modelling,” Proc. The 4th International Conference on Language Resources and Evaluation, ELRA, 2004, pp. 15-19

[2] D. Sperber and D. Wilson, “Relevance: Communication and Cognition”, Blackwell, Oxford, 1986

[3] A., Tversky, “Features of similarity,” Psychological Review 84, 1977, pp. 327-352

[4] T. Declerck, H.U. Krieger, S.M. Thomas, P. Buitelaar, S. O´Riain, T. Wunner, G. Maguet, J. McCrae, D. Spohr, and E. Montiel-Ponsoda, “Ontology-based multilingual access to financial reports for sharing business knowledge across Europe,” Internal Financial Control Assessment Applying Multilingual Ontology Framework, J. Rooz, J. Ivanyos, Eds. Budapest: HVG Press, 2010, pp. 67-76

[5] P. Vossen, E. Agirre, N. Calzolari, C. Fellbaum, S. Hsieh, C. Huang, H. Isahara, K. Kanzaki, A. Marchetti, M. Monachini, F. Neri, R. Raffaelli, G. Rigau, and M. Tescon, “KYOTO: A system for mining, structuring and distributing knowledge across languages and cultures,” Proc. The 6th International Conference on Language Resources and Evaluation, Morocco, 2008, pp. 1462-1469

[6] T. Ishida, “Language Grid: An infrastructure for intercultural collaboration,” Proc. IEEE/IPSJ Symposium on Applications and the Internet (SAINT 06), 2006, pp. 96-100

[7] ISO 704, “Terminology work — Principles and methods,” Genève: ISO, 2000

[8] B.N. Madsen, H.E. Thomsen, C. Vikner, “Comparison of principles applying to domain specific versus general ontologies,” OntoLex 2004: Ontologies and Lexical Ressources in Distributed Environments, A. Oltramari, P. Paggio, A. Gangemi, M.T. Pazienza, N. Calzolari, B.S. Pedersen, K. Simov, Eds. ELRA, 2004, pp. 90-95

[9] E.A. Guut, “A theoretical account of translation – without a translation theory,” Target: International Journal of Translation Studies, 2/2, 1990, pp.135-164

[10] R.N. Shepard, “Toward a universal law of generalization for psychological science,” Science, New Series, Vol. 237, No. 4820, 1987, pp. 1317-1323

[11] J.B. Tenenbaum, and T.L. Griffiths, “Generalization, similarity, and bayesian inference,” Behavioral and Brain Sciences 24, 2001, pp. 629-640

3. Asymmetric similarity and cross-cultural communication process[10]

Preliminary study:

In this paper, the assignment of different α and β parameters to Tversky's model (Tversky, 1977) is defined based on the BMG theory (Tenenbaum & Griffiths, 2001) which demonstrates the mathematical equivalence between Tversky's model and the BMG. Three different ways of assigning parameters in Tversky's model are investigated by applying (α = 1, β = 0; α = 0.5, β = 0.5; α = 0, β = 1) to datasets obtained from TOs.

[10] In Proceedings of Terminology & Artificial Intelligence 2011, November, Paris, France. (Peer-reviewed full paper: Selected as winner of the Best Paper Award by student authors)

Asymmetric Similarity and Cross-Cultural Communication Process

Fumiko Kano Glückstad
Dept. of International Language Studies and Computational Linguistics
Copenhagen Business School
Dalgas Have 15, DK-2000 Frederiksberg, Denmark
fkg.isv@cbs.dk

Abstract

Terminological Ontology (Madsen et al., 2004a) is a method used for knowledge sharing and domain-specific translation practices, and could potentially be highly suitable for simulating the cognitive models explaining real-life cross-cultural communication scenarios. In this work, we apply cognitive models (Tversky, 1977; Tenenbaum & Griffiths, 2001) to data-sets obtained from the Terminological Ontology method. A key discussion is how the asymmetric parameters set in the models influence the identification of equivalent relationships between language specific concepts existing in two different cultures.

1 Introduction

The role of ontology in a multilingual setting is considered an emerging challenge to Semantic Web development. As a consequence, there are several major ongoing projects, such as MONNET on Multilingual Ontologies for Networked Knowledge project (Declerck et al., 2010) and KYOTO on Knowledge-Yielding Ontologies for Transition-Based Organization (Vossen et al., 2008). Though both of these projects deal with translation of terms from a Source Language (SL) to a Target Language (TL), they focus on linking lexical data through an interoperable common ontology (Cimiano et al., 2010) rather than on optimizing relevance between concepts that are potentially measurable based on diverse models derived from cognitive theory.

Terminological Ontology (TO) is a domain-specific ontology used for knowledge sharing (Madsen et al. 2004a), which normally is applied to terminology work, cf. for example (ISO, 2000). The unique points of TO that differentiate it from other types of ontologies are feature specifications and subdivision criteria (Madsen et al. 2004b). A feature specification consists of a feature dimension and its value. Thus, the representation of a whole concept is a feature structure, i.e. a set of feature specifications corresponding to the unique set of characteristics that constitutes that particular concept (Madsen et al. 2004a; 2004b). Terminological ontologists argue that concepts are defined in a language dependent context, and therefore TO is language dependent. TO is developed within a knowledge sharing community, then dynamically updated and validated. If it is necessary to share knowledge with other communities, TOs developed in different communities should be compared, aligned and merged as needed. While the aforementioned two mainstream projects, MONNET and KYOTO, both deal with complex ontologies involving huge data-sets, TO usually handles smaller amounts of concepts.

Considering this, one of the points that should be emphasized in this work is that TO could potentially be a suitable tool for simulating cognitive models explaining a real-life cross-cultural communication scenario. That is, how the original meanings of a cultural-specific concept in one culture are best conveyed to a person in another culture. In this paper, Tversky´s contrast model is applied to data-sets obtained from terminological ontologies. A main focus here is the asymmetric similarity proposed in Tversky´s contrast model. In order to discuss this key issue, we first review relevant cognitive models: Tversky´s contrast model (Tversky, 1977) and the Bayesian model of generalization (Tenenbaum and Griffiths, 2001) in the next section. In Section 3, an experiment is performed to assess how the asymmetric parameters set in Tversky´s contrast model influence the identification of equivalent relationships between language specific concepts existing in two different cultures. Section 4 discusses findings and future work followed by conclusions in Section 5.

2 Cognitive models

Within the area of cognitive science, there have traditionally been two opposing views of similarity and generalization: set-theoretic matching models (c.f. Tversky, 1977) and continuous metric space models (c.f. Shepard, 1987). The uniqueness of Tversky´s view is the asymmetric similarity that has been demonstrated in (Tversky 1977) based on several empirical evidences. Tenenbaum and Griffiths (2001) argue that Tversky´s model is remarkably similar to the Bayesian model of generalization that is rooted in Shepard´s theory of the generalization problem. Hence, it is possible to unify these two classically opposing approaches to similarity and generalization.

In (Tenenbaum and Griffiths, 2001), three crucial questions of learning (after Chomsky 1986) are addressed in order to explain the Bayesian model of generalization: *1) what constitutes the learner´s knowledge about the consequential region; 2) how the learner uses that knowledge to decide how to generalize; and 3) how the learner can acquire that knowledge from the example encountered*. For instance, one example *x* of some consequence C is given. It is assumed that x can be represented as a point in a continuous metric psychological space and C corresponds to some region (referred to as *consequential region*) of this space. A task of the learner is to infer the probability that a newly encountered object *y* will fall under C given the observation of the example *x*. This conditional probability can be expressed as: $P(y\epsilon C|x)$. In order to compute the conditional probability, Tenenbaum and Griffiths (2001) first answer the question: the learner´s knowledge about the consequential region that is represented as a probability distribution $p(h|x)$ over an a priori-specified hypothesis space $\mathcal{H}$ of possible consequential regions $h\epsilon\mathcal{H}$. Prior to observing *x*, this distribution is the prior probability $p(h)$, then becomes the posterior probability $p(h|x)$ after observing *x*. According to Tenenbaum and Griffiths (2001), the learner uses this knowledge for generalization by summing the probabilities $p(h|x)$ of all hypothesized consequential regions that contain *y* as follows:

$$P(y\epsilon C|x) = \sum_{h:y\in h} p(h|x) \quad (1)$$

Tenenbaum and Griffiths (2001) further describe how a rational learner arrives at $p(h|x)$ from $p(h)$ after the generalization, through the use of Bayes´ rule as follows:

$$P(h|x) = \frac{p(x|h)p(h)}{p(x)} \quad (2)$$

They further compare this Bayesian model of generalization to Tversky´s contrast model:

$$sim(y,x) = 1/\left[1 + \frac{\alpha \cdot f(Y-X) + \beta \cdot f(X-Y)}{f(Y\cap X)}\right] \quad (3)$$

Here, X and Y are the feature sets representing *x* and *y*, respectively. *f* denotes a measure over the feature sets. (Y⋂X) represents the sets of features present in both Y and X, (Y-X) represents the sets of features present in Y but not in X, and (X-Y) represents the sets of features present in X but not in B. α and β are free parameters representing an asymmetric relationship between Y and X. For comparing this formula with equation (1), Tenenbaum and Griffiths (2001) express equation (1) in a mathematically equivalent form as follows:

$$P(y\epsilon C|x) = 1/\left[1 + \frac{\sum_{h:x\in h, y\notin h} p(h,x)}{\sum_{h:x,y\in h} p(h,x)}\right] \quad (4)$$

Here, $p(h,x) = p(x|h)p(h)$. According to (Tenenbaum and Griffiths, 2001), *the bottom sum ranges over all hypothesis that include both x and y, while the top sum ranges over only those hypotheses that include x but not y. If we identify each feature k in Tversky´s framework with a hypothesized subset h, where an object belongs to h if and only if it possesses feature k, and if we make the standard assumption that the measure f is additive, then the Bayesian model as expressed in equation (4) corresponds formally to*

Tversky´s contrast model with asymmetric parameter α=0, β=1.

Now, what we will pursue in this work is to apply Tversky´s contrast model to data-sets obtained from two culturally dependent terminological ontologies, in other words, perform the terminological ontology matching based on Tversky´s contrast model. Rooted in Tversky´s contrast model and the Bayesian model of generalization, three combinations of asymmetric parameters are set to compute the similarities proposed by Tversky (1977): 1) α=0, β=1 which corresponds to the Bayesian model of generalization (referred to as the Bayesian framework); 2) α=0.5, β=0.5 which represents the symmetric similarity; and 3) α=1, β=0 based on Tversky´s definition (referred to as Tversky´s default model): *if sim(y,x) is interpreted as the degree to which y is similar to x, then y is the subject of the comparison and x is the referent. Hence the features of the subject are weighted more heavily than the features of the referent.*

3 Experiment

3.1 Corpora

Documents downloaded from the Eurydice website published by the Education, Audiovisual and Culture Executive Agency under the EU commission have been used as text corpora. The documents describe the educational systems in a majority of the EU member countries both in English and in their native languages based on the ISCED classification[1]. It means that it is in principle possible to develop an ontology of any of these EU member countries. In this study, documents describing the Danish- and German educational systems are used. All corpora are officially published in English by reliable authorities of each country, respectively. Thus, all English translated terms and expressions in their original languages can be considered as official terms. It means that it is feasible to identify terminological expressions in an original language from these documents. This enables one to eventually identify translation equivalences linking between, in this case, Danish and German. However, in this study, we will only focus on English documents that describe language-dependent concepts in two cultures.

3.2 Ontology construction

The terms and their definitions describing the educational systems in each country are manually extracted from the respective English corpora. In order to extract features that can be matched at a later step, definitions containing features defined in the standardized ISCED classification are intentionally extracted. Based on these terms and their definitions, ontologies representing the educational system in each of the two countries are developed using a Computer Aided Ontology Structuring prototype (CAOS) that is based on the Terminological Ontology principles defined in (Madsen et al., 2004a).

As described in Section 1, the uniqueness of TO is feature specifications and subdivision criteria (Madsen et al., 2004a; 2004b). The use of feature specifications is subject to principles and constraints described in detail in (Madsen et al., 2004a). Most importantly, a concept automatically inherits all feature specifications of its superordinate concepts. Secondly, polyhierarchy is allowed so that one concept may be related to two or more superordinate concepts. On the other hand, subdivision criteria that have been used for many years in terminology work are strictly implemented in TO by introducing dimensions and dimension specifications (Madsen et al., 2004a; 2004b). This enables the CAOS prototype to perform consistency checking which helps in constructing ontologies. A dimension of a concept is an attribute occurring in a non-inherited feature specification of one or more of its subordinate concepts. Values of the dimension allow a distinction among sub-concepts of the concept in question. For example, a dimension of a concept "pre primary education" is [AGE] whose values are [0-3 | 3-6]. These dimension values distinguish the sub-concepts: "nursery" and "kindergarten". The dimension can only occur on sister concepts and a given value can only appear on one of these sister concepts. In this way, a concept must be distinguished from each of its nearest superordinate concepts as well as from each of its sister concepts by at least one feature specification (Madsen et al., 2004a; 2004b).

By using the CAOS prototype that performs the consistency checking of TO principles, the two educational ontologies are developed based

[1] http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm

on the terms and definitions manually extracted from the corpora

3.3 Computation of asymmetric similarities

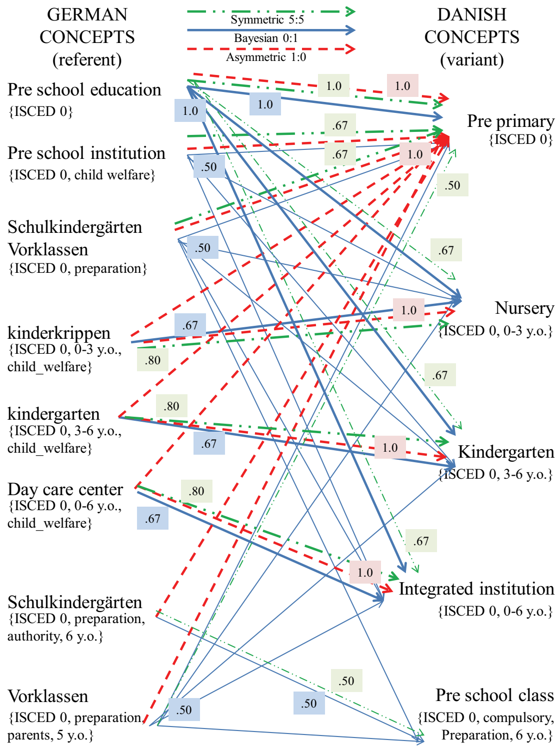


Figure 1. Similarity scores (German as referent)

Based on the TO construction method, the Danish and German educational ontologies respectively containing 51 and 57 concepts are developed. Figures 1 and 2 show that a concept consists of a terminological expression (e.g. “kindergarten”) and a set of feature values (e.g. {ISCED 0, 3-6 y.o., child_welfare}). Here, one of the concepts in either the Danish or the German ontology is considered as object *x*, while one of the concepts in the other ontology is considered as object *y* in terms of the cognitive models described in Section 2. All the concepts, that are referred to as object *x* or *y*, consist of feature specifications systematically extracted from the TO construction.

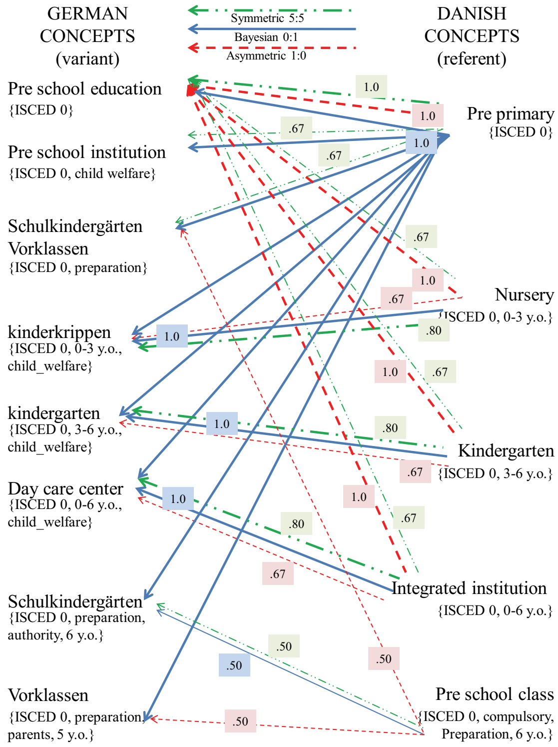


Figure 2. Similarity scores (Danish as referent)

For computation of the three different types of similarities described in Section 2, the extracted feature values of each concept in the two ontologies are considered as feature sets X or Y in equation (3). According to Tversky (1977), *if sim(y,x) is interpreted as the degree to which y is similar to x, then y is the subject of the comparison and x is the referent.* The issue here is which party should be referent, when considering the cross-cultural communication scenario where a person in one culture (communicator) is conveying original meanings of a cultural-specific concept in his/her culture to a person in the other culture (audience). Sperber and Wilson (1986) view the cross-cultural communication as principally an inferential process. It means that the essential task of the communicator is to produce a *stimulus* from which the audience can infer what set of thoughts or assumptions the communicator intends to convey. Hence, the results of the ontology matching should identify the *stimulus* that is optimally relevant to the audience while conveying the original meanings of a source concept. The communicator´s task is then to identify a concept in the audiences´ conceptual structure that is optimally relevant to the source concept. It means that the stimulus selected by a communicator should to the maximum extent be similar to the source concept. Therefore, the features of a stimulus should be weighted more heavily than the ones of a source concept in Tversky´s default model. In other words, a stimulus should be considered as variant y, and a source concept as referent x in equation (3). Hence the asymmetric parameter for y (*stimulus*) is weighted higher in Tversky´s default model described in Section 2.

Based on this definition, the three types of similarities are computed in all possible combinations between the German and the Danish concepts in both directions. Examples of the aligned concepts based on the similarity scores are illustrated in Figures 1 and 2. Figure 1 illustrates a cross-cultural communication scenario where a German conveys original meaning of source concepts to a Danish audience, while Figure 2 illustrates the opposite pattern. The concepts listed in these figures are automatically selected based on the similarity computation. It means that Tversky´s contrast model roughly categorizes clusters that are more or less similar to each other. In these figures, the bold lines are the highest scores equal to or over 0.55 obtained for each source concept, and the slim lines indicate the highest scores lower than 0.55, or the second highest scores equal or higher than 0.55. The straight lines in the figures represent results based on the Bayesian framework, the double dotted lines are based on the symmetric similarity, and the single dotted lines are based on Tversky´s default model.

4 Discussions

4.1 Findings

In this experiment, the three types of similarities have been computed for all combinations of concepts in two cultures in both directions. Accordingly, concepts in the two ontologies were approximately clustered into 5 groups. Figures 1 and 2 show one of the clusters automatically extracted from this process. These figures indicate that concepts referring to education targeted for children under school age are adequately extracted from the two ontologies. This can be explained in the following way: The principles of Terminological Ontology - e.g. inheritance of feature specifications, uniqueness of dimensions, and uniqueness of primary feature specifications - enable features to reflect the structure of ontologies.

For example, a feature value, "ISCED 0" in Figures 1 and 2 has been specified only once when defining the concepts, "preschool education" and "pre primary" respectively in the German and the Danish ontologies. Since all of the subconcepts under these two terms inherit this feature value, the concepts categorized as "ISCED 0 level" have systematically been extracted and grouped as a cluster. One critical point in this experiment is that the ISCED classification has been used as the highest feature dimension. In fact, the corpora based on the ISCED classification have intentionally been selected in this study, in order to extract standardized terminologies. However, in these corpora, the German educational system consists of "preschool education (level 0)", "primary education (level 1)", "secondary education (level 2+3)", "tertiary education (level 5)" and "continuing education for adults", whereas the Danish educational system consists of "pre primary education (level 0)", "single structure education (level 1+2)", "upper secondary education (level 3)", "tertiary education (level 5)" and "continuous education for adults". Since only 100% identical features are considered as common features between two concepts and common features other than the ISCED level do not exist in each concept at level 1 and 2 in both ontologies, no concepts representing the primary and lower secondary education have been extracted as matched pairs in this study. This problem could be solved by defining a matching algorithm that can detect such partly overlapping features.

Another important issue that has been a main focus of this study is that the pattern of the bold straight lines (the Bayesian framework) in Figure 1 is almost identical to the bold single dotted lines (Tversky´s default model) in Figure 2. In the same way, the bold single dotted lines in Figure 1 are identical to the bold straight lines in Figure 2. The only difference is the direction of arrows. While the arrows rooted from several source concepts points to a single general term in Tversky´s default model, the arrows rooted from a single general term points to several specific terms in the Bayesian framework. This phenomenon can be explained in the following way.

In Tversky´s default model, the computation of the similarity scores for all combinations of concepts within each cluster results in identifying a concept in a target ontology that has the highest optimal relevance to a source concept. Figures 1 and 2 show that, in most cases, the identified optimal stimuli based on Tversky´s default model is the most general term located at the highest node within the cluster. According to the principle of TO, when a concept is subdivided into several subconcepts based on a dimension, an extra feature is added to each subconcept. Hence it is always the case that concepts having more features are more specific subconcepts. It means that the lower a concept is located in the ontology, the more features the concept inherits from superordinate concepts. If dimensions and their values at the lower part of the two ontologies are not consistent, all the inherited features are distinctive features that can be considered as simply noise in the data-sets. Therefore, the identified optimal stimuli were in most cases more general and abstract terms rather than specific terms. A question is whether these identified general stimuli are really the most optimal stimuli that convey original meanings of source concepts to the audience in the other culture. This could be answered by the results obtained from parameters based on the Bayesian framework.

Equation (4) – the Bayesian model of generalization expresses the conditional probability that a newly encountered object *y* will fall under C (consequential region) given the observation of the example *x* (Tenenbaum and Griffiths, 2001). Contrasting this equation to this experiment, example *x* is considered as a source concept in the referent ontology, a hypothesized subset *h* as the region where a concept belongs to *h* if and only if it possesses feature *k,* and *y* as a newly encountered concept in the target ontology that should be aligned to the referent ontology. Based on the Bayesian framework, the asymmetric parameter in equation (3) is set as $\alpha=0$, $\beta=1$. It means that, when the number of distinctive features possessed by a source concept increases, the probability that a newly encountered concept *y* in the target ontology falls under the consequential region decreases. The results based on the Bayesian framework indicate the probability that the audience in the target culture generalizes a source concept from a stimulus presented by a communicator. Therefore, the probability that the audience generalize from a provided stimulus is higher for more general and abstract source concepts than it is for specific source concepts. This explains the phenomenon identified in Figures 1 and 2.

When considering the cross-lingual terminology management, the main purpose for terminologists is to precisely clarify the original meaning of source terms to a target audience. From this viewpoint, the aforementioned results look unsatisfactory. However, if an identified stimuli in the target culture contains features that are irrelevant to a source concept, this will create misconceptions among the audience. Hence, to select an optimally general term is the most secure approach. The terminologist´s job is then to clarify the distinctive features that differentiate the particular source concept from this identified general term. Although further improvements and developments are required, the approach presented in this paper is potentially effective for cross-lingual terminology management and reflects the human generalization process involved in the cross-cultural communication, especially for the culturally-dependent domains.

4.2 Future challenges

Tenenbaum and Griffiths (2001) points out some weaknesses of Tversky´s contrast model. To be more precise, Tversky´s contrast model *allows any kind of features and any feature weights whatsoever*. However, to improve the matching quality, some semantic questions such as what qualifies as a feature and what determines the feature weights should be investigated. For example, the problem of the partly overlapping features that have not been identified as common features in this study could be weighted lower compared to other features. On the other hand, some primary features could be weighted higher. How to weight each feature may thus be a key to improving the matching performance. Another question raised in (Tenenbaum and Griffiths, 2001) is that Tversky´s formal theory does not explain why and how the two subsets of distinctive features (Y-X) and (X-Y) should be given different weights. In this experiment, the simplest binary parameter described in the Bayesian model has been applied. However, Tenenbaum and Griffiths´ extended Bayesian model (2001) proposes a method called "strong sampling" that treats *x* as a random positive example of C, which involves the stronger assumption that *x* is explicitly sampled from C. This approach makes one to set more fine-grained asymmetric parameters based on the size of the hypothetical region *h*. Frank et al. (2009) has further extended their Bayesian model and proposed the informative communication framework based on the Relevance Theory of Communication (Sperber & Wilson, 1986). Hence, it is an obvious future challenge to apply these extended cognitive models to the aforementioned different combination of datasets.

5 Conclusions

In this paper, cross-cultural ontology matching has been performed based on the contrastive analysis of the Bayesian model of generalization and Tversky´s contrast model (Tenenbaum and Griffiths, 2001). The results show that in most cases the identified optimal stimuli based on Tversky´s default model is the most general term located at the highest node within the automatically extracted cluster. On the other hand, the results based on the Bayesian framework indicate the probability that the audience in the target culture can generalize a source concept from a stimulus presented by a communicator. The probability that the audience generalize from a provided stimulus is higher for more general and abstract terms than it is for specific terms. Both identified results are in most cases corresponding mutually. Although further improvements and developments are required, the approach presented in this work is potentially effective for cross-lingual terminology management that reflects the human generalization process involved in cross-cultural communication especially in culturally-dependent domains.

Acknowledgments

I would like to express my gratitude to my supervisors, Hanne Erdman Thomsen and Daniel Hardt for their valuable guidance and support on my Ph.D. project.

References

Cimiano, P., et al. 2010. A Note on Ontology Localization. In Journal of Applied Ontology Vol. 5, No. 2, IOS Press, 127-137.

Chomsky, 1986. Language and Problems of Knowledge. Managua lectures.

Declerck, T., et al. 2010. Ontology-based multilingual access to financial reports for sharing business knowledge across Europe. In Rooz, J., Ivanyos, J. (Eds.) Internal Financial Control Assessment Applying Multilingual Ontology Framework, Budapest: HVG Press, 67-76.

Frank, M. C., Goodman, N. D., Lai, P., Tenenbaum, J. B. 2009. Informative communication in word production and word learning. In Proc. The 31st Annual Meeting of the Cognitive Science Society.

ISO 704:2000 Terminology Work – Principles and Methods. Genève, International Standards Organisation.

Madsen, B.N., Thomsen, H.E., Vikner, C., 2004a. Principles of a system for terminological concept modelling. In Proc. The 4th International Conference on Language Resources and Evaluation, ELRA, 15-19.

Madsen, B.N., Thomsen, H.E., Vikner, C., 2004b. Comparison of principles applying to domain specific versus general ontologies. In Oltramari A., et al. (Eds.) OntoLex 2004: Ontologies and Lexical Ressources in Distributed Environments. ELRA, 90-95

Shepard, R.N., 1987. Toward a universal law of generalization for psychological science. In Science, New Series, Vol. 237, No. 4820, 1317-1323.

Sperber D., Wilson, D., 1986. Relevance: Communication and Cognition. Blackwell, Oxford.

Tenenbaum, J.B., Griffiths, T.L., 2001. Generalization, Similarity, and Bayesian Inference. In Behavioral and Brain Sciences 24, 629-640

Tversky, A., 1977. Features of Similarity. In: Psychological Review 84, 327-352.

Vossen, P., et al. 2008. KYOTO: A system for mining, structuring and distributing knowledge across languages and cultures. In Proc. The 6th International Conference on Language Resources and Evaluation, Morocco, 1462-1469.

4. Feature-based Ontology Mapping from an Information Receivers' Viewpoint[11]

In this paper, the two types of cognitive models: Tversky´s model (further divided into three different measures according to the different assignment patterns of parameters) and the BMG (assigning feature weights based on the strong sampling scheme), are applied to datasets created based on the TO approach. This corresponds to Modules 1-5, when the "TO" is employed in Module 3, in the following figure.

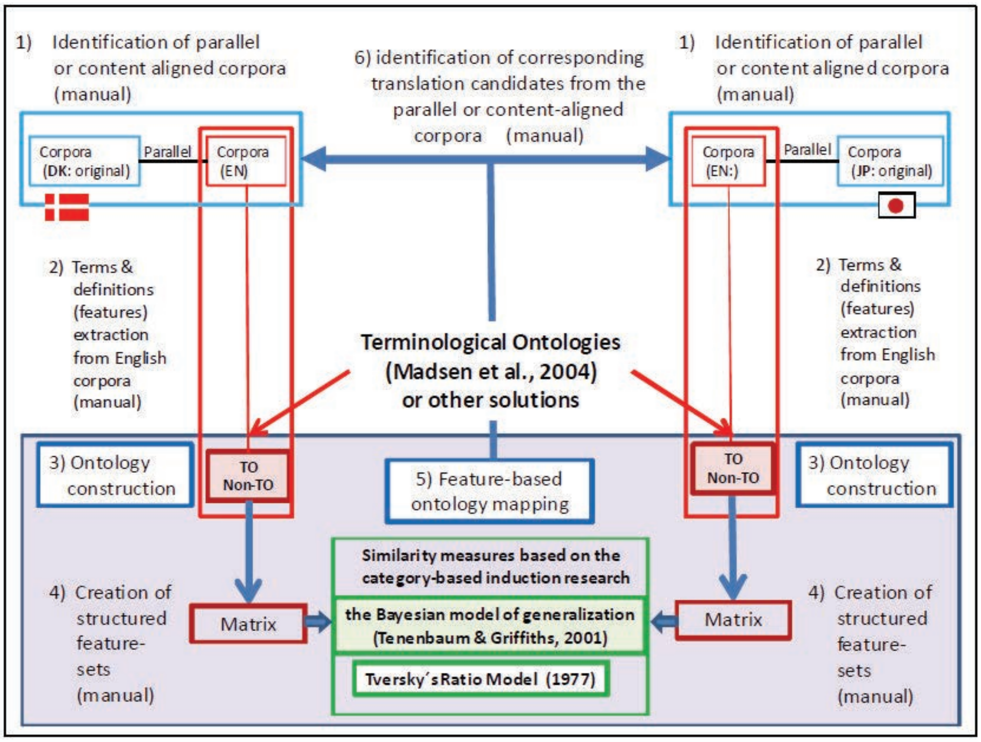


Figure 5-1: Workflow for Step 1 in the COM framework

(Step 1: comparison of the mapping algorithms)

[11] Peer-reviewed full paper: 9th International Workshop on Natural Language Processing and Cognitive Science - NLPCS 2012, Wroclaw, Poland, 28 June 2012

Feature-based Ontology Mapping from an Information Receivers' Viewpoint

Fumiko Kano Glückstad[1] and Morten Mørup[2]

[1] Copenhagen Business School, Dept. of International Business Communication
Dalgas Have 15, DK-2000 Frederiksberg, Denmark
fkg.ibc@cbs.dk
[2] Technical University of Denmark, DTU Informatics, Section for Cognitive Systems
Richard Pedersens Plads, DK-2800 Kgs Lyngby, Denmark
mm@imm.dtu.dk

Abstract. This paper compares four algorithms for computing feature-based similarities between concepts respectively possessing a distinctive set of features. The eventual purpose of comparing these feature-based similarity algorithms is to identify a candidate term in a Target Language (TL) that can optimally convey the original meaning of a culturally-specific Source Language (SL) concept to a TL audience by aligning two culturally-dependent domain-specific ontologies. The results indicate that the Bayesian Model of Generalization [1] performs best, not only for identifying candidate translation terms, but also for computing probabilities that an information receiver successfully infers the meaning of an SL concept from a given TL translation.

1. Introduction

Accelerated by the recent internet revolution with its fast-paced globalization, cross-cultural communication, e.g. between an Asian and a European, becomes inherently challenging due to the lack of sufficient linguistic resources directly bridging remote languages. This challenge is not only caused by the lack of linguistic resources, but also by differences in human perception of similar concepts existing in diverse socio-cultural communities. The MONNET on Multilingual Ontologies for Networked Knowledge project [2] and the KYOTO project on Knowledge-Yielding Ontologies for Transition-based Organization [3] are some typical major projects that deal with such multilingual issues based on ontological methodologies. The approaches taken in these major research projects are thoroughly analyzed in [4] based on three dimensions: international (standardized) vs. culturally-influenced domains; functional (conceptual) vs. documental (lexical) localization; and finally interoperable vs. independent ontology. The work presented here challenges this multilingual issue by mapping independent ontologies from a culturally-influenced domain in a functional manner. The work is part of an overall framework for investigating how background knowledge possessed by a Source Language (SL) communicator and a Target Language (TL) reader should be represented and linked in light of various cognitive processes involved in cross-cultural communication. Background knowledge is considered as the average domain knowledge possessed by average citizens in a specific socio-cultural community and assumed to be represented as domain ontology. We employ a knowledge representation method known as Terminological Ontology (TO) [5] by constructing two culturally-dependent TOs respectively representing the Danish- and the German educational systems. A specific purpose is to identify the most optimal algorithm of mapping culturally-influenced domain knowledge existing in two cultures using taxonomically organized hierarchical feature-structures obtained from these TOs. A candidate algorithm is the so-called Bayesian Model of Generalization (BMG) [1], a novel cognitive model that considers the hierarchical feature-structure as prior knowledge of an SL communicator or a TL audience, depending on the assignment of variables to be explained in next section. More specifically, the BMG computes asymmetric (uni-directional) similarities based on feature values either from an SL communicator- or a TL audience's viewpoint by considering the prior knowledge as cultural bias. The asymmetric coordination in communication is also well illustrated in the Relevance Theory of Communication [6]. Accordingly, the BMG is compared against Tversky's set-theoretic model [7] that has previously been tested in [8].

In Section 2, the similarity measures applied here are further explained in detail. Section 3 describes an experiment applying four different feature-based similarity measures to data-sets obtained from the TOs, respectively representing concepts in the educational systems in Denmark and Germany. Section 4 discusses the analysis of the results followed by conclusions in Section 5.

2. Feature-based similarity algorithms

The first three similarity measures are derived from Tversky´s ratio model. This model is defined as follows [7]:

$$sim\,(y,x) = 1/[\,1 + \frac{\alpha * f(Y-X) + \beta * f(X-Y)}{f(Y \cap X)}] \qquad (1)$$

In equation (1), *X* and *Y*, respectively, represent feature sets of objects *x* and *y,* and *f* is considered as additive function. ($Y \cap X$) represents common features present in both *Y* and *X*, (*Y*-*X*) denotes distinctive features existing in *Y* but not in *X*, and (*X*-*Y*) denotes distinctive features in *X* but not in *Y*. α and β are free parameters which enables one to compute an asymmetric similarity relationship between object *x* and *y*. Accordingly, three combinations of parameter values are assigned in the previous study [8]: A) α=1 and β=1: which corresponds to the Jaccard Similarity Coefficient [9] that represents a symmetric similarity relationship between object *x* and *y*; B) α=1 and β=0: which only computes distinctive features present in *Y*, not in *X*; and C) α=0 and β=1: which only computes distinctive features present in *X*, not in *Y.*

As briefly stated above, a key point is to clarify which variable is defined as a concept in an SL- or a TL culture. According to Tversky [7], *if sim(y,x) is interpreted as the degree to which y is similar to x, then y is the subject to the comparison and x is the referent.* This definition should be applied to all three parameter settings defined in here. Keeping this definition in mind, an additional key point is that Tenenbaum & Griffiths [1] demonstrate that Tversky's model C) is formally corresponding to the following equation which forms the basis of the BMG explained below. Equation (2) computes *the conditional probability that y falls under C (Consequential region) given the observation of the example x* [1]. The consequential region *C* in our work indicates the categorical region where a subject *y* belongs.

$$P\,(y \in C|x) = 1/[1 + \frac{\sum_{h:x\in h, y\notin h} p(h,x)}{\sum_{h:x,y\in h} p(h,x)}] \tag{2}$$

In equation (2), a hypothesized subset *h* is defined as the region where a concept belongs to *h*, if and only if, it possesses feature *k* [1]. Thus, $P(h,\ x) = P(x|h)P(h)$ in equation (2) represents the weight assigned to the consequential subset *h* in terms of the example *x*. Accordingly, the BMG - algorithm D) - is considered as a model where the weight $P(h,\ x)$ is - based on the strong sampling scheme defined in [1] - specifically assigned to Tversky's model C). The weight is defined as follows [1]:

$$P(x\,|h) = \begin{cases} 1/|h| & if\ x \in h \\ 0 & otherwise \end{cases} \tag{3}$$

Here, $|h|$ indicates the size of the region *h* [1]. In our work, the number of objects possessing the k^{th} feature in the referent ontology is considered as the size of the region *h*. [1] explains that the prior $P(h)$ is not constrained in their analysis so that *it can accommodate arbitrary flexibility across contexts*. Hence in this work, we set $P(h) = 1$. In the following experiments the BMG is compared against the three parameter settings defined for Tversky's Ratio model.

3. Experiment

3.1 Data preparation

Data source: The data-sets used in [8] have been used as original data sources. They are based on document corpora obtainable from the Eurydice web-site published by the Education, Audiovisual and Culture Executive Agency under the EU commission. These documents describe the German- and Danish educational systems both in English and in the original languages based on the ISCED classification. Hence, it is feasible to identify terminological expressions in the original language from these documents and eventually identify translation equivalences linking between German and Danish. Hence, language-dependent terms and their definitions describing the educational systems in the two cultures have manually been extracted from the respective English corpora for developing TOs. The reason that these non-remote European languages are employed in this work is that these documents are written in accordance with the standardized template defined by the Eurydice, which may better provide for a well controlled experiment for assessing the similarity measures.

One of the key principles for developing the TOs is that a concept automatically inherits all feature specifications of its super-ordinate concepts [5]. A dimension of a concept is an attribute occurring in a non-inherited feature specification of one or more of its sub-ordinate concepts. The values of the dimension allow a distinction among sub-concepts of the concept in question. For example, a dimension of the concept "pre-primary education" is [AGE] whose values are [0-3 | 3-6]. These dimension values distinguish the sub-concepts: "nursery" and "kindergarten". The dimension can only occur on sister concepts and a given value can only appear on one of these sister concepts. In this way, a concept must be distinguished from each of its nearest super-ordinate concepts as well as from each of its sister concepts by at least one feature specification [5]. These principles enable us to generate well-structured feature sets that are assumed to be useful for the feature-based similarity computations. Tables 1 and 2 show examples of the expressed feature structures.

Table 1. Example of German Data Source (Terms and Feature sets)

ID	Term	Feature-values
G2	preschool education	{ISCED97, children & young, ISCED0}
G5	kindergärten	{ISCED97, children & young, ISCED0, child welfare, 3-6y.o.}
G7	schulkindergärten & vorklassen	{ISCED97, children & young, ISCED0, preparation}
G10	primary education	{ISCED97, children & young, ISCED1}
G11	primary school	{ISCED97, children & young, ISCED1, <6-10y.o.<}
G13	secondary education	{ ISCED97, children & young, ISCED2+3}
G14	lower secondary level	{ISCED97, children & young, ISCED2+3, <10-16y.o.<}
G15	school offering one single course	{ISCED97, children & young, ISCED2+3, <10-16y.o.< , single}
G16	hauptschule	{ISCED97, children & young, ISCED2+3, <10-16y.o.< , single , general basic, 5-9th grade}
G18	gymnasium	{ISCED97, children & young, ISCED2+3, <10-16y.o.< , single, intensified, 5-12/13th grade}
G19	schools offering several courses	{ISCED97, children & young, ISCED2+3, <10-16y.o.< , several}

Table 2. Example of Danish Data Source (Terms and Feature sets)

ID	Term	Feature-values
D2	pre primary	{ISCED97, children & young, ISCED0}
D4	kindergarten	{ISCED97, children & young, ISCED0, 3-6y.o.}
D6	single structure	{ISCED97, children & young, ISCED1+2}
D7	alternative structure	{ISCED97, children & young, ISCED1+2, alternative}
D8	home tuition	{ ISCED97, children & young, ISCED1+2, alternative, compulsory, 6-16y.o}
D9	efterskole or youth school	{ISCED97, children & young, ISCED1+2, alternative, compulsory, <14-18y.o.<}
D10	efterskole	{ISCED97, children & young, ISCED1+2, alternative, compulsory, <14-18y.o.<, boarding school, approved by state}
D11	youth school	{ISCED97, children & young, ISCED1+2, alternative, compulsory, <14-18y.o.<, day-to-day, public municipal council}
D14	municipal school	{ISCED97, children & young, ISCED1+2, formal teaching, municipality}
D16	0-9th form	{ISCED97, children & young, ISCED1+2, compulsory}
D17	0th form	{ISCED97, children & young, ISCED1+2, compulsory, preparation}
D18	1-9th form	{ISCED97, children & young, ISCED1+2, compulsory, general basic}
D19	10th form	{ISCED97, children & young, ISCED1+2, optional}

Creation of feature-term matrices: In order to compute similarities, matrices referring to the German- and Danish educational systems which, respectively, consist of 58 and 52 terms are manually generated from the feature sets. Feature value columns are defined in the following way:

1. All feature values existing in the Danish and German data sources are registered in both matrices.

2. If feature values in the Danish and German matrices are completely overlapping (e.g. “ISCED0-pre-primary” in DK and “ISCED0-pre-primary” in GE), the feature columns in question should be merged into one column.
3. If a feature is possessed by a term, the numeric value should be “1”, otherwise “0” in the matrices.
4. If a feature value in one matrix is completely included in a feature value in the other matrix (e.g. “ISCED1+2” in DK and “ISCED1” in GE), a term possessing the feature that includes the other feature (e.g. Danish “ISCED1+2”) should have numeric value “1” in both feature columns (e.g. “ISCED1+2” in DK and “ISCED1” in GE). It means that a term possessing a feature value that is included in the other feature (e.g. German “ISCED1”) should have numeric value “1” only in the feature column in question.
5. If feature values in the Danish and German matrices are partly overlapping (e.g. “ISCED1+2” in DK and “ISCED2+3” in GE), a dummy column referring to the exact overlapping feature value (e.g. “ISCED2” for both DK and GE) is created. In this example, a Danish term possessing a feature “ISCED 1+2” should have numeric value “1” in both “ISCED 1+2” and “ISCED2” columns, but not in the “ISCED2+3” column.

In this way, we create the German matrix consisting of 58 terms x 117 feature values and the Danish matrix consisting of 52 terms x 117 feature values.

3.2 Similarity computation

The basic idea of similarity computation here is to identify a translation candidate from concepts existing in a TL culture. Assuming that SL communicators and TL information receivers have general conceptualization of culturally-dependent domains - in this case the educational system in each country - all combinations of similarities between TL- and SL terms are computed. When computing similarities based on the three settings of Tversky’s model and the BMG described in Section 2, the variables: “terms subject to comparison” and “referent terms” are consistently defined across the four feature-based similarity algorithms: **A)** Tversky: α=1 and β=0 (Jaccard); **B)** Tversky: α=1 and β=0; **C)** Tversky: α=0 and β=1; and **D)** the BMG.

3.3 Results

Figure 1 shows the most typical patterns of similarity scores obtained from the aforementioned four algorithms from top to bottom: A), B), C), and D). Figure 1-a indicates that algorithms A) and B) are relatively identical, showing that rather general German terms, such as G2 and G46, score higher similarities. On the other hand, algorithms C) and D) show that the terms from G2 to G9, all of which are within the category of preschool education in Germany, score the highest. Especially, the BMG clearly identify the series of German preschool educations, all of which are targeted for children under the school age categorized as ISCED0. Since - in the simplified formulae of the BMG - the sum of distinctive features possessed by *referent* (variable: *x*) but not *subject to comparison* (variable: *y*) and common features possessed by both *x* and *y* become denominator, the eventual score results in the value "1". From a communicator's viewpoint, it is reasonable to consider that, based on prior knowledge of Danish SL concepts, all the German TL terms that possess feature ISCED0 targeted for children under the school age are categorized as objects belonging to D2.

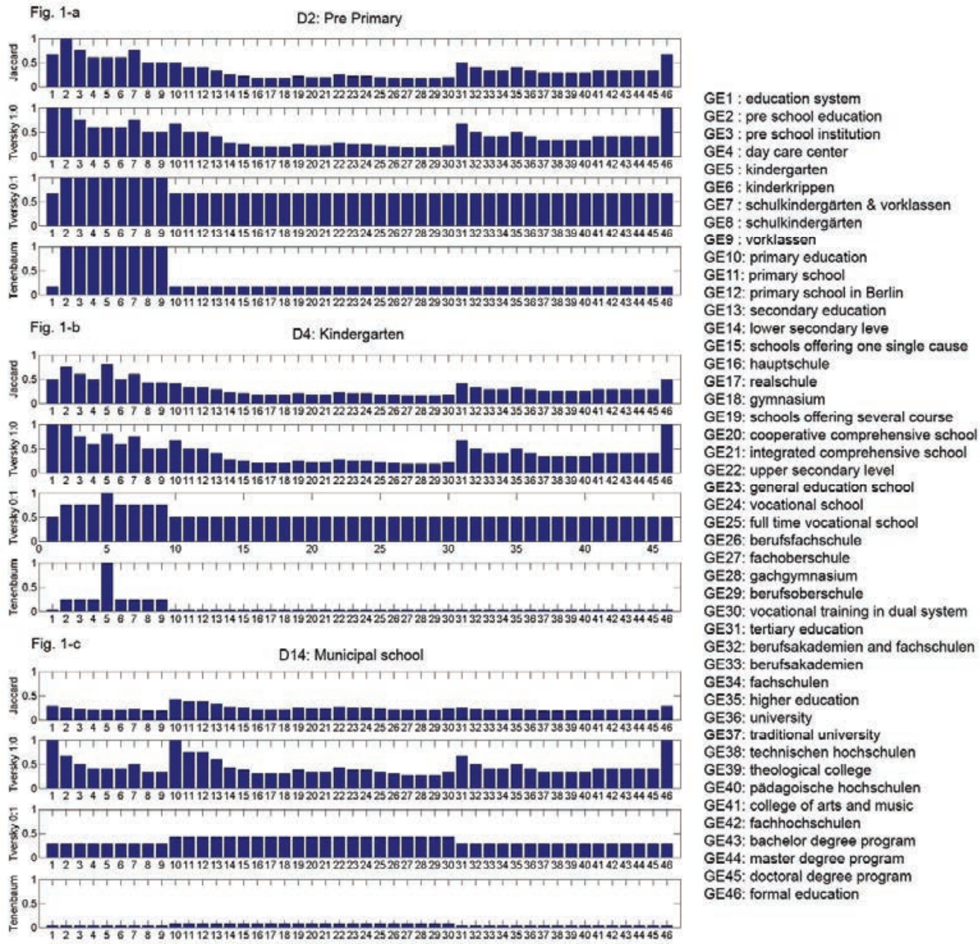


Fig. 1. Similarity scores: German as variant, Danish as referent

In Figure 1-b, algorithm B) identifies general German terms such as G1, G2 and G46 as the most similar terms to D4: kindergarten. On the other hand, all other algorithms indicate that term G5: "kindergärten" has the highest similarity in terms of D4. Especially, the BMG clearly points out this implication selecting G5 as the most similar concept to D4, because the *size principle* weight the feature value, "3-6 years old", which are possessed only by D4 and G5, heavier than other features.

Figure 1-c for the algorithms C) and D) show that the series of terms referring to the German primary- and secondary education have slightly higher similarity scores than other terms. However, the scores in the BMG are particularly low. When inspecting the feature values of D14, it becomes clear that D14 contains two distinctive features ("formal teaching" and "municipality") that are not possessed by

any German terms. In addition, the fully- and partly overlapping common features are possessed by many terms in both German and Danish, which result in assigning lower feature weights due to the *size principle* of the BMG [1]. It should be noticed that, when contrasting the feature set of D14 with the definition from the text corpus: *a comprehensive school covering both primary and lower secondary education, i.e. one year of pre-school class, the first (grade 1 to 6) and second (grade 7-9/10) stage basic education, or in other words it caters for the 6-16/17-year-olds,* it turns out that no decisive features (age, grade etc.) that describe D14 are included in the feature set. Hence, the result in Figure 1-c could potentially be significantly improved if the quality of data source is reconsidered.

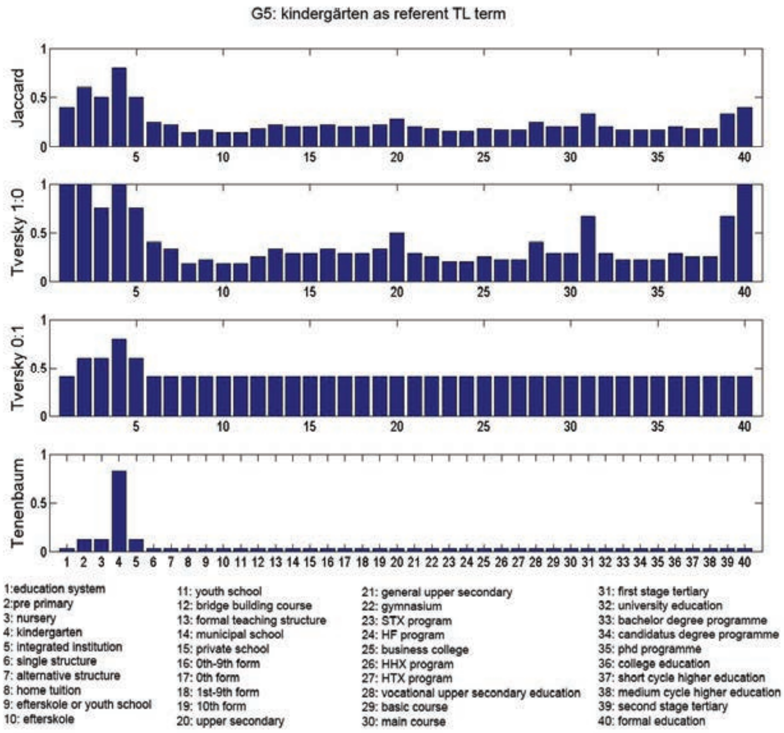


Fig. 2. Generalization probabilities: Danish as variant, German as referent

As described in Section 2, the BMG can, by exchanging assignment of variables *x* and *y*, also compute probabilities that a TL audience generalizes a source concept from a stimulus presented by an SL communicator. Hence, in Figure 2, Danish SL concepts are defined as subjects to comparison and German TL terms as referent. This computes probabilities, from a German TL reader´s viewpoint, that he/she possibly infers the meanings of Danish SL concepts based on his/her prior knowledge of the German educational concepts when a German TL term is given as translation. Although Figure 2 shows that all four algorithms scored the highest for D4, it demonstrates that, due to the assigned feature weights, the BMG clearly indicates that a German TL audience will, from the given TL stimulus G5, likely infer D4. Another noteworthy point is that similarity relations between D4 and G5 are not symmetrical, e.g. the BMG result in Figure 2 is 82.3%, while it is 100% in Figure 1- b.

4. Discussions

By inspecting similarity scores of all combinations between Danish and German concepts, the results obtained from the BMG seem to reasonably identify optimally specific translation candidates if the structured feature sets are properly prepared.

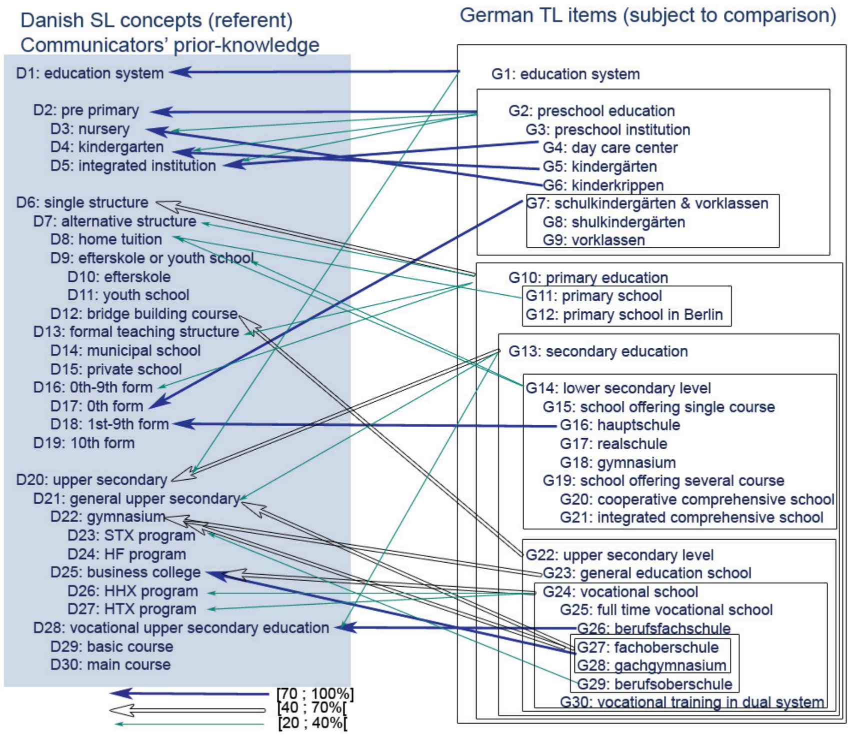


Fig. 3. Ontology mapping overview: from a Danish communicator's viewpoint

For further analyzing the performance of the BMG, Figures 3 outlines corresponding relationships between the Danish SL concepts and the German TL terms from a Danish communicator's viewpoint. The corresponding relationships are depicted by the three patterns: **1)** solid thick arrows, when the probability scores are 70% or higher; **2)** transparent thick arrows, when the probability scores are 40% or higher and below 70 %; and **3)** thin arrows, when the probability scores are 20% or higher and below 40%.

Figure 3 also indicates that a communicator who has prior-knowledge of the Danish educational system (gray filled square box) observes each German TL concept as translation candidate and assess whether each German TL concept falls under the class of each Danish SL concept. A more concrete and imaginable picture would be that a communicator whose mother tongue is Danish seeks for a translation candidate in his/her non-native language (German). For example, in a situation where a Danish communicator looks for a German translation candidate for a concept D2, all of the German terms within the relevant transparent square box, from G2 to G9, respectively falls under the class D2 with the probability range [70 ; 100].

On the other hand, Figure 4 illustrates that a German TL information receiver possibly generalizes the meanings of Danish SL concepts from a given German TL translation as stimulus based on his/her prior knowledge of the German educational concepts (gray filled square box). For instance, if he/she observes a German stimulus, G3, he/she will likely infer some of the Danish source concepts within the relevant transparent square box, from D2 to D5, with the probability range [40 ; 70[that is lower than the case of the German stimulus, G2 with the probability range [70 ; 100].

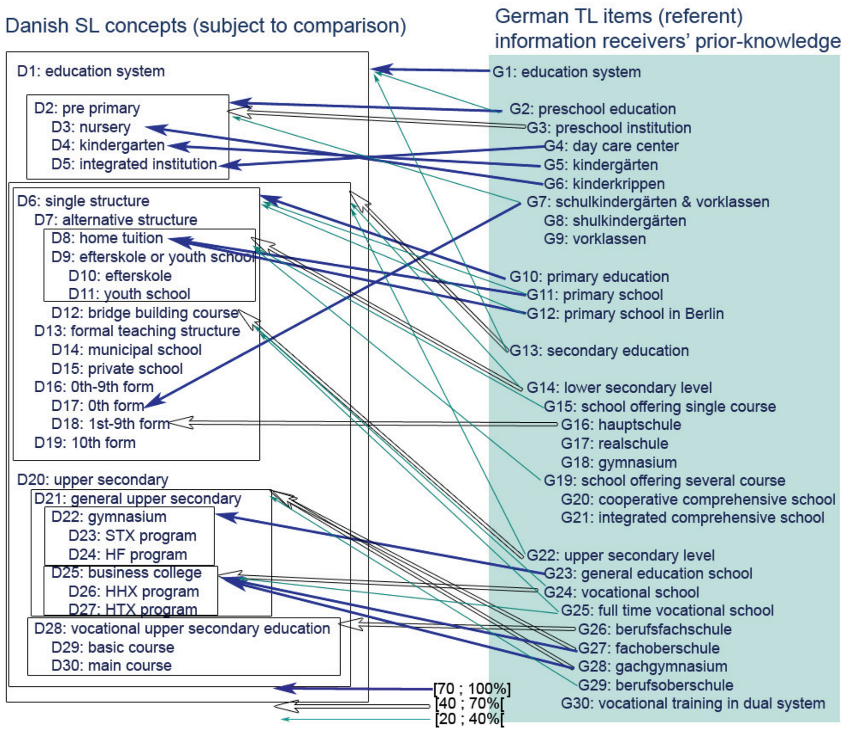


Fig. 4. Ontology mapping overview: from a German audience's viewpoint

Although the BMG [1] can be quite useful as algorithm for linking multilingual culturally-specific concepts existing for two cultures, there are still some unsatisfactory results that have been identified in this study. For example, in both Figures 3 and 4, the German concept, G11, has relations with D8. According to our intuitive assessment based on the basic domain-knowledge, G11 should rather be relevant to some of the concepts among D13-D18. When inspecting the feature sets of G11 and D8/ D13-D18, it becomes obvious that, while G11 contains a feature "[10-16 y.o.]", D13-D18 which refers to the Danish formal primary education for children 6-16 years old does not contain features referring to age range. Instead, D8 contains the important definitional feature about the age. This problem has been caused, not by the BMG, but by the particularly strict principles for constructing TOs which may risk causing the elimination of important features. This issue indicates that, if some decisive features are lacking or some irrelevant features are included, the results obtained from the BMG can immediately be affected. Hence, a future attempt would be to investigate how to generate appropriate feature sets, that is, a more flexible taxonomic organization of feature structures based on terms and definitions identified in domain-specific parallel corpora. This may improve the mapping of culturally-specific concepts applying the BMG. Another key point is that the analysis performed here is a rather subjective assessment. Hence, for future undertakings, it is necessary to identify an appropriate method based on assessments using human subjects.

5. Conclusions

In this work, the Bayesian Model of Generalization [1] and Tversky's set-theoretic model [7] have been applied to data-sets consisting of culturally-specific concepts and of features extracted from data sources based on Terminological Ontologies [5]. The results indicate that, if input data-sets consisting of culturally-specific concepts and of feature-values in two cultures are properly prepared, the BMG [1] can be uniquely used not only for identifying a TL translation candidate, but also for estimating probabilities of how a TL information receiver generalizes an SL concept from a given TL translation. To successfully promote the next step for an overall framework, a human based assessment of concept mappings as well as an improvement of the method to create highly appropriate feature sets, will be required.

References

1. Tenenbaum, J.B., Griffiths, T.L., 2001. Generalization, Similarity, and Bayesian Inference. In: Behavioral and Brain Sciences 24, 629-640.
2. Declerck, T., et al. 2010. Ontology-based multilingual access to financial reports for sharing business knowledge across Europe. In: Rooz, J., Ivanyos, J. (Eds.) Internal Financial Control Assessment Applying Multilingual Ontology Framework, Budapest: HVG Press, 67-76.
3. Vossen, P., et al. 2008. KYOTO: A system for mining, structuring and distributing knowledge across languages and cultures. In: Proc. The 6th International Conference on Language Resources and Evaluation, Morocco, 1462-1469.
4. Cimiano, P., et al. 2010. A Note on Ontology Localization. In: Journal of Applied Ontology Vol. 5, No. 2, IOS Press, 127-137.
5. Madsen, B.N., Thomsen, H.E., Vikner, C., 2004a. Principles of a system for terminological concept modelling. In: Proc. The 4th International Conference on Language Resources and Evaluation, ELRA, 15-19.
6. Tversky, A., 1977. Features of Similarity. In: Psychological Review 84, 327-352.
7. Sperber, D., Wilson, D., 1986. Relevance: Communication and Cognition. Oxford: Blackwell
8. Glückstad, F.K. 2011. Asymmetric Similarity and Cross-Cultural Communication Process. In: 9th International Conference on Terminology and Artificial Intelligence: Proceedings of the Conference. 8-10 November 2011, Paris, France. Paris : Institut National des Langues et Civilisations Orientales , 2011. p. 59-65.
9. Jaccard, P. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37: 547–579.

5. Cross-cultural Concept Mapping of Standardized Datasets[12]

In this paper, the two types of cognitive models: Tversky´s model (further divided into three different measures according to the different assignment patterns of parameters) and the BMG (assigning feature weights based on the strong sampling scheme) are applied to datasets directly obtainable from a third party (the UNESCO Institute for Statistics). This corresponds to Modules 1-5, when the “Non-TO” is employed in Module 3, in the following figure.

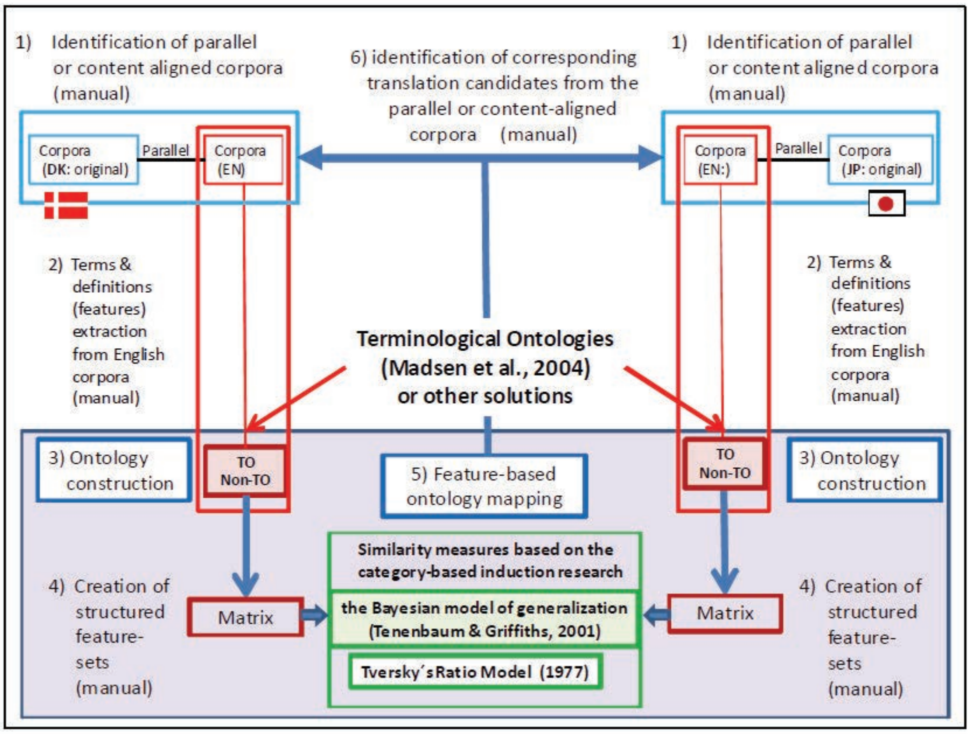


Figure 5-1’: Workflow for Step 1 in the COM framework (Step 1: comparison of the mapping algorithms)

[12] Peer-reviewed full paper: Workshop of Challenges to Knowledge Representation in Multilingual Contexts, Terminology and Knowledge Engineering Conference 2012, Madrid, Spain, June 19-22, 2012

Cross-Cultural Concept Mapping of Standardized Datasets

Fumiko Kano Glückstad

Copenhagen Business School, Dept. of International Business Communication
Dalgas Have 15, DK-2000 Frederiksberg, Denmark
{fkg.ibc , fkg.isv}@cbs.dk

Keywords: feature-based similarity, cross-cultural communication, multilinguality, Bayesian model of generalization, categorization

Abstract. This work compares four feature-based similarity measures derived from cognitive sciences. The purpose of the comparative analysis is to verify the potentially most effective model that can be applied for mapping independent ontologies in a culturally influenced domain [1]. Here, datasets based on standardized pre-defined feature dimensions and values, which are obtainable from the UNESCO Institute for Statistics (UIS) have been used for the comparative analysis of the similarity measures.

1 Introduction

The recent internet revolution and its globalization impact has brought about new possibilities for people located at opposite sides of the globe to real-time dynamically communicate with each other. Although we most often use English as a common communication code, misunderstandings are almost unavoidable in such cross-cultural communications. This implies that multilinguality is a highly increasing demand that can correctly link concepts existing in diverse socio-cultural communities. This work challenges these multilingual issues based on the following pragmatic- and cognitive theories: the Relevance Theory of Communication [2] and the Knowledge Effects involved in category-based inductions [3]. A key point in these models is that a symmetric choice of code and context is not plausible in a cross-cultural communication scenario because the two communicating parties are unlikely to share an identical cognitive environment [2]. If e.g. a new object existing in a Source Language (SL) culture is introduced to a person in a Target Language (TL) culture, the TL reader will compare this new object with something he/she knows in advance (prior knowledge). This implies that feature-based asymmetric similarity measures play a key role for the communicating human cognitive mind.

In the ontology research domain, ref [1] compares several multilingual ontology frameworks such as the KYOTO project [4] and the MONNET project [5] based on a number of dimensions used in categorizing different types of ontology localization projects [6]. These dimensions are: *International (standardized) vs. culturally influ-*

adfa, p. 1, 2011.

enced domains; functional vs. documental localization; and interoperable vs. independent ontology. In this paper, potentially applicable asymmetric similarity measures that can be used for mapping *independent ontologies* in a *culturally influenced domain* are compared based on qualitative analyses. To increase the objectivity of the comparative analysis of the four different feature-based similarity measures, datasets based on standardized pre-defined feature dimensions and values, which are obtainable from the UNESCO Institute for Statistics (UIS) have been employed.

In the following, Section 2 describes the experimental settings of this work followed by a summary of results in Section 3, and summarizing with concluding remarks in Section 4.

2 Experimental Settings

2.1 Datasets

Datasets used in this experiment has been obtained from UIS who collected data from UNESCO Member States on an individual basis. The purpose of collecting data, according to UIS is to *map the Member States' national education systems according to the International Classification of Education (ISCED)*. UIS aims for *Member States to report their data in an internationally comparative framework.* These datasets from all over the world are downloadable from UIS' web-site[1]. Here, Japanese and Danish datasets have been used for the analysis. Each dataset consists of educational terms defined by several pre-defined feature dimensions such as ISCED level, programme destination and orientation, starting age, cumulative duration of education, and entrance requirements. Most feature dimension values are pre-defined, i.e. for the programme destination dimension, values are pre-defined as [general | pre-vocational | vocational].

One of the challenges of using these datasets is how to map the numeric feature values of dimensions such as "starting age" and "cumulative duration of education." For example, in the Danish educational system, the starting age of upper secondary school is defined as "16-17 years old" and its cumulative years of education is "12-13 years". On the other hand, the Japanese educational system is a so called "single-track system" meaning that the starting age of upper secondary school is exactly defined as "15 years old" and its cumulative years of education is "12 years". To handle this difficulty in an objective and systematic manner, the following procedure has been implemented: **1)** If a feature value in one country is completely included in a feature value in the other country (e.g. a feature "6-12 y.o." in Japan is completely included in a feature "6-17 y.o." in Denmark), a term possessing the feature that includes the other feature (a term possessing "6-17 y.o." should also possess "6-12 y.o."), and **2)** If two features from the respective countries are partly overlapping (e.g. "13-15 y.o." in Japan and "14-17 y.o." in Denmark), a dummy feature referring to the exact overlapping range (i.e. "14-15 y.o.") is created. In this example, a Japanese term that possesses "13-15 y.o." should also possess the dummy feature "14-15 y.o." In the same

[1] http://www.uis.unesco.org/education/ISCEDmappings/Pages/default.aspx

way, a Danish term that possesses “14-17 y.o.” should also possess the dummy feature “14-15 y.o.”.

In order to objectively assess feature-based similarity measures, simpler datasets that do not contain these ambiguous feature dimensions/values have been prepared as control data. It means that these simpler datasets only contain the standardized feature dimensions/values defined by UIS. Based on these, similarity scores are computed by applying the four feature-based similarity measures described in the following.

2.2 Similarity computation

In this work, the first three similarity algorithms defined below based on Tversky′s Ratio Model are considered as baseline algorithms [7]:

$$sim\ (y,x) = 1/[\ 1 + \frac{\alpha * f(Y-X) + \beta * f(X-Y)}{f(Y \cap X)}] \tag{1}$$

Equation (1) computes the degree to which object *y* is similar to *x*, when objects *x* and *y*, respectively, consist of feature sets *X* and *Y*. In here, object *x* is considered referent and object *y* as subject of comparison according to the definitions of [7]. In equation (1) f is considered as additive function and α and β as free parameters. ($Y \cap X$) represents common features present in both *Y* and *X*, (*Y-X*) denotes distinctive features existing in *Y* but not in *X*, and (*X-Y*) in *X* but not in *Y*. In [9], three algorithms were defined based on different parameter settings: **i)** α=1 and β=1: which corresponds to the Jaccard Similarity Coefficient representing a symmetric similarity relationship between objects *x* and *y*; **ii)** α=1 and β=0: which only computes distinctive features present in *Y*, not in *X*; and **iii)** α=0 and β=1: which only computes distinctive features present in *X*, not in *Y*.

Here, a referent object *x* should be defined as an SL concept and a subject object *y* that is to be compared with *x* should be defined as a TL concept according to [7]. This definition should be applied to all three algorithms defined above. Keeping this definition in mind, an additional key point is that Tenenbaum & Griffiths [8] argue that the third algorithm is formally corresponding to the following equation (2) of the Bayesian Model of Generalization (BMG), which computes *the conditional probability that y falls under C (Consequential region) given the observation of the example x* [8]. Here, the consequential region *C* indicates the categorical region to where a subject *y* belongs.

$$P\ (y \in C|x) = \ 1/[1 + \frac{\sum_{h:x \in h, y \notin h} p(h,x)}{\sum_{h:x,y \in h} p(h,x)}] \tag{2}$$

In equation (2), a hypothesized subset *h* is defined as the region where a concept belongs to *h*, if and only if, it possesses feature *k* [8]. It means that *y* is considered as a newly encountered object existing in the TL ontology that should be aligned to the referent ontology of the SL according to Tversky’s definition [7].

P(h, x) = *P(x|h)P(h)* above represents the weight assigned to the consequential subset *h* in terms of the example *x*. Therefore, as the fourth similarity algorithm, the weight *P*(*h*, *x*) is specifically assigned to the third algorithm based on the strong sampling scheme defined in [8] as follows:

$$P(x\,|h) = \begin{cases} 1/|h| & if\ x \in h \\ 0 & otherwise \end{cases} \tag{3}$$

Here, |*h*| indicates the size of the region *h* [8]. In our work, the number of objects possessing the k^{th} feature in the referent ontology is considered as the size of the region *h*. [8] explains that the prior *P*(*h*) is not constrained in their analysis so that it can accommodate arbitrary flexibility across contexts. Hence in this work, *P*(*h*) = 1.

3 Results and Data Analysis

Concept ID	ISCED level	Programme destination (A/B/C)	Programme orientation (G/P/V)	Theoretical cumulative duration at ISCED 5	Position in the national degree / qualification structure (intermediate, first, second, etc...)	Position in the tertiary education structure (Bachelor-Master-PhD)	Minimum entrance requirement (ISCED level or other)	Theoretical starting age	Theoretical duration of the programme	Theoretical cumulative years of education at the end of the programme	Does the programme have a work based element? (Y/N)	Programme specifically designed for adults (Y/N)	Programme specifically designed for part-time attendance (Y/N)
Danish concepts													
D1	0		G					2-5 years	4 years		No	No	No
D2	0		G					5-6 years	1 year		No	No	No
D3	1		G					6-7 years	6 years	6 years	No	No	No
D4	2	A	G				1	12-13 years	3-4 years	9-10 years	No	No	No
D7	3	C	V				2A	16-30 years	3-5 years	14 years	Yes	No	No
D19	5	B		Short	1st		3A, 3C	18-50 years	0,5-4 years	13-15 years	No	Yes	Yes
D20	5	B		Short	1st		3A, 3C	20-30 years	2-3 years	14 years	No	No	No
D21	5	A		Medium	1st	Bache-lor	3A	18-50 years	2-4 year	13-15 years	No	Yes	Yes
D22	5	A		Medium	1st	Bache-lor	3A	20-30 years	3-5 years	16 years	Yes	No	No
D23	5	A		Medium	1st	Bache-lor	3A	20-30 years	3 years	15-16 years	No	No	No
Japanese Concepts													
J35	5	B		Short	Inter-mediate		3 ABC	18	2-3	14-15	No	No	No
J36	5	B		Short Medium	Inter-mediate		5B	20	1+	15+	No	No	No
J37	5	B		Short	Inter-mediate		3	18	2-3	14-15	No	No	Yes
J38	5	B		Short	Inter-mediate		3	18	2	14	No	No	No
J40	5	B					3	18	1+	13+	No	No	No
J41	5	A		Medium	1st	Bache-lor	3	18	4	16	No	No	No
J42	5	A		Long	1st	Bache-lor	3	18	6	18	No	No	No
J44	5	A		Long	Inter-mediate		5A 1st,M	22	1+	17+	No	No	No

Table 1. Example of original datasets obtained from UIS: feature structure of selected concepts. The shadowed columns are feature values that are considered only for *a* graphs in Figures 1-2

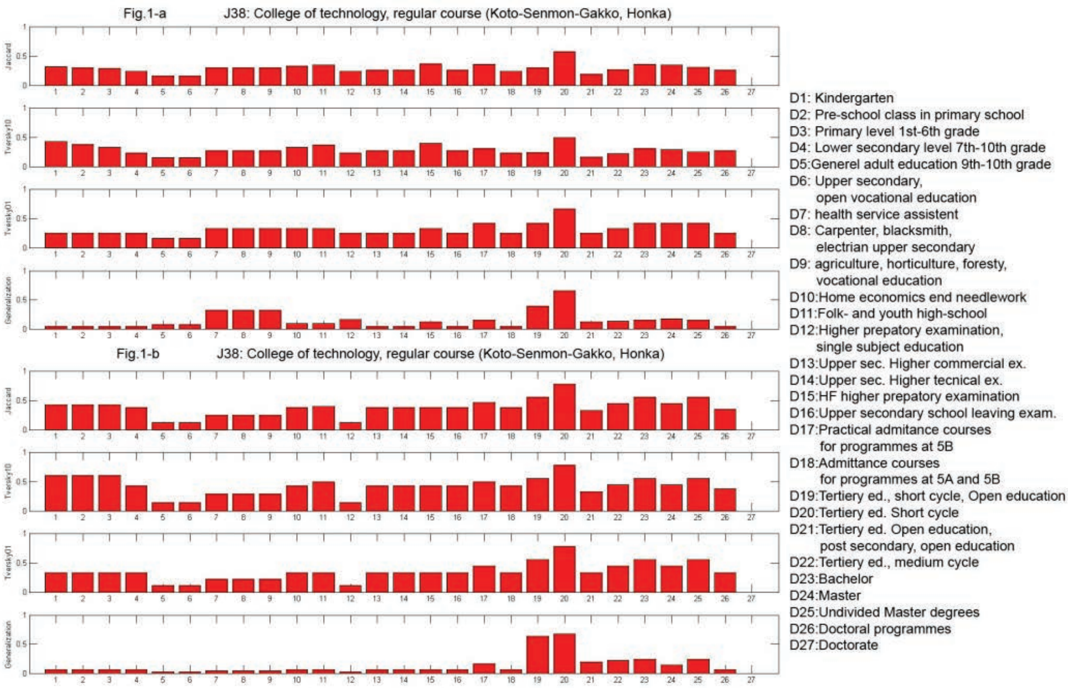


Fig. 1. Similarity scores: J38: college of technology as referent

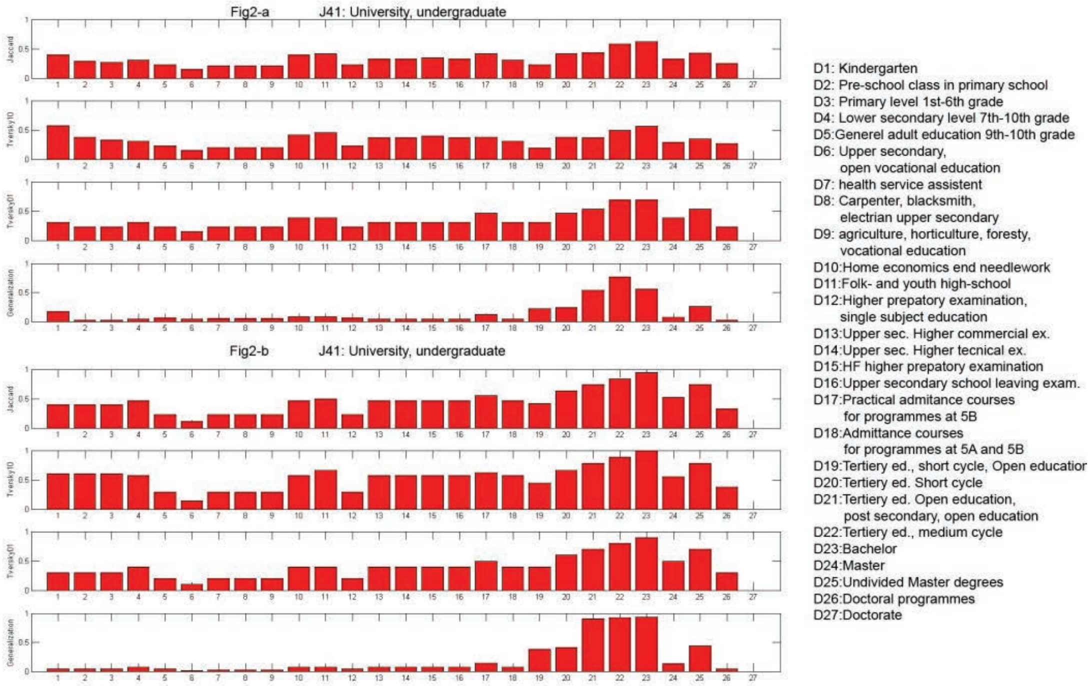


Fig. 2. Similarity scores: J41: University, undergraduate as referent

Although the datasets obtained from UIS have been developed for the purpose of statistical comparative analysis and mapping of the educational concepts among the Member States, no definite mapping pairs are proposed in a concrete form. This implies that the judgment of mapping depends on human evaluators in the respective

Member States countries. Consequently, the evaluation of results in this work focuses on a qualitative analysis, e.g. what kind of feature structures affect the results of similarity computation, instead of a quantitative analysis, e.g. recall-precision measures.

Figures 1 and 2, respectively, show similarity scores of the Japanese concepts, "J38: college of technology, regular course (高等専門学校本科: *Koto-Senmon-Gakko, Honka*)" and "J41: university, undergraduate (大学学部: *Daigaku Gakubu*)" against all accessible Danish concepts listed in the UIS dataset. The Japanese concept J38 is, from the authors' own subjective point of view, an "atypical" concept compared to the more universally used concepts such as J41. While the upper part of the figures marked as *a* are the similarity scores computed based on feature dimensions/values including the numeric feature values described in Section 2, the lower part of the figures marked as *b* are computed without these feature dimensions/values.

The first thing to be noticed between the *a* and *b* graphs in general, is that the higher the number of feature values that are possessed by two concepts in question, i.e. in case of the *a* graphs, the lower the similarity scores. In particular, the first to third similarity scores in the *a* graphs show rather flat and ambiguous results. This is because the way the datasets have been created for mapping the feature values of dimensions such as "starting age" and "cumulative duration of education" simply increases the number of features. Among these, distinctive features will act as noise in the similarity computation, and hence the similarity scores decrease. In contrast to the first three similarity measures, the size principle in the fourth algorithm (BMG) effectively identifies specific concepts that are more similar than others, in all figures. For example for both J38, "D19: Tertiary, short cycle, open education" and "D20: Tertiary, short cycle education"; and for J41, "D21: Tertiary, post secondary open education", "D22: Tertiary, medium cycle education", and "D23: Bachelor" are respectively identified as the most similar concepts. On the other hand, the first to third similarity measures indicate that the aforementioned Danish concepts are only slightly more similar than the others. In addition, other Danish concepts referring to the pre-primary to lower secondary educations, i.e. D1-D4 are also considered slightly more similar than the others. Finally, the fourth similarity measure in Figure 1-*a* also identify that the Danish concepts referring to the vocational upper secondary educations, i.e. D7-9 are more similar than the others.

The results shown in Figures 1-2 indicate that the fourth similarity measure (BMG) seems to be the most effective algorithm. However, to conclude on this observation, it is necessary to investigate how the feature structures of each concept reflect the similarity computation. Table 1 shows the feature structures of selected concepts that are affected in the similarity results shown in Figures 1-2. Table 1 explains why the Danish concepts referring to the pre-primary to lower secondary educations, i.e. D1-D4 score higher with the first to third algorithms. There are two reasons for this. The first reason that apply especially for the first and second algorithms is that these algorithms consider distinctive features possessed by Danish concepts (y: subject to comparison), while the third and fourth algorithms consider ones possessed only by Japanese concepts (x: referent). Hence all feature values listed in the "programme orientation" column possessed by the Danish concepts strongly affect the similarity scores. The second reason is that the first to third algorithms equally consider all features that are shared between two concepts in question based on additive functions. It means that for example all feature values with "no" that are matched between the two concepts are counted as "1". On the other hand, the BMG consider a feature value that is shared by many concepts as less important, which reduces similarity scores of all less relevant concepts such as pre-primary and primary education concepts. Another point is that the BMG detects that "J38: college of technology" is relatively similar to the Danish concepts referring to the vocational upper secondary educations, i.e. D7-9 in Figure 1-*a*. This is in fact true since the Japanese college of technology is a higher educational institution that is targeted for students who have graduated from lower secondary school and wish to acquire vocational skills based on 5 years education which consists of 3 years of upper secondary education and 2 years of vocationally oriented post-secondary education. The relevance between J38 and D7-9 has been effectively detected by balanced effects of feature values, i.e. feature value "14" of "cumulative duration" affects as decisive feature and other less important features reduce similarity scores of other irrelevant Danish upper secondary concepts.

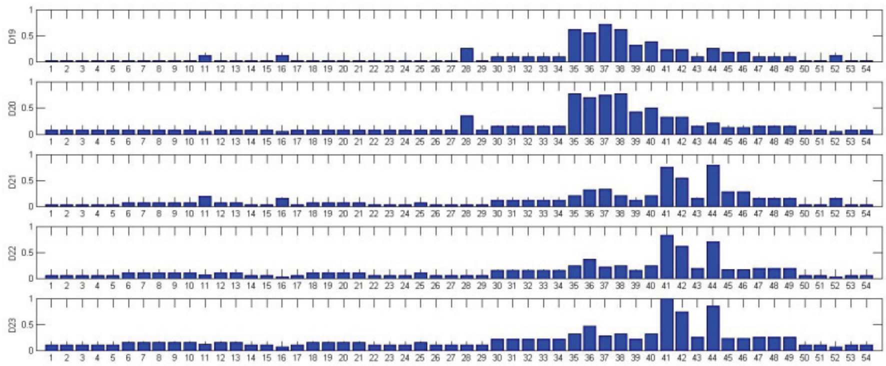


Fig. 3**.** Similarity scores: D19; D20; D21; D22; D23 as referent

Finally, equation (2) of the BMG theoretically explains that the model computes probabilities that a new object y falls under a hypothesized categorical region C provided that example x (prior knowledge) is observed. It means that by replacing variables x and y, it is possible to compute similarities from the Danish side, i.e. how a person who has prior knowledge of the Danish educational system selects the most similar Japanese concept as a feedback function. The results in Figure 3 show that Japanese concepts J35-37 referring to short cycle higher education provided at junior colleges and "J39: college of technology", are identified as the most similar concepts for Danish concepts "D19: Tertiary short cycle open education" and "D20: Tertiary short cycle education". In the same way, J41, J42 and J44, all of which are the Japanese bachelor degree programmes are detected as the most similar concepts for the Danish concepts "D21: Tertiary post-secondary open education", "D22: Tertiary medium cycle education" and "D23: Bachelor". These results demonstrate that, in these standardized datasets, uni-directional similarity relations from both the Japanese- and the Danish sides, are effectively computed. The feedback function of computing similarities from a Japanese or a Danish evaluator's viewpoint may be useful for detecting asymmetric similarity relations, when mapping *independent ontologies* in a *culturally influenced domain* [1]. The theoretical argument of applying asymmetric similarity measures considering human prior knowledge is further discussed from a cognitive- and pragmatic point of view in [10].

4 Conclusions

In this work, four feature-based similarity measures are applied to the standardized datasets consisting of pre-defined feature dimensions/values developed by the UIS. The purpose of this comparison is to verify the similarity measures based on the objectively developed datasets. The results demonstrate that the BMG provides for the most effective cognitive model for identifying the most similar corresponding concepts existing for a targeted socio-cultural community.

Reference

1. Cimiano P., Montiel-Ponsoda E., Buitelaar P., Espinoza M., Gómez-Pérez A.: A Note on Ontology Localization. In: Journal of Applied Ontology Vol. 5, No. 2, IOS Press, (2010) 127-137.
2. Sperber, D., Wilson, D.: Relevance: Communication and Cognition. Blackwell, Oxford (1986)
3. Murphy,G. L.: The Big Book of Concepts. The MIT Press Cambridge, Massachusetts (2004)
4. Vossen P., Agirre E., Calzolari N., Fellbaum C., Hsieh S., Huang C.R., Isahara H., Kanzaki K., Marchetti A., Monachini M., Neri F., Raffaelli R., Rigau G., Tescon M., VanGent J.: KYOTO: A system for mining, structuring and distributing knowledge across languages and cultures. In: Proc. The 6th International Conference on Language Resources and Evaluation, Morocco, (2008) 1462-1469.
5. Declerck T., Krieger H.U., Thomas S.M., Buitelaar P., O'Riain S., Wunner T., Maguet G., McCrae J., Spohr D., Montiel-Ponsoda.E.: Ontology-based multilingual access to financial reports for sharing business knowledge across Europe. In: Rooz, J., Ivanyos, J. (Eds.) Internal Financial Control Assessment Applying Multilingual Ontology Framework, Budapest: HVG Press (2010) 67-76.
6. Espinoza, M., Montiel-Pansoda, E., Gómez-Pérez, A.: Ontology localization. In: Proc. The fifth International Conference on Knowledge Capture KCAP 09, ACM Press (2009) 33-40.
7. Tversky, A.: Features of similarity. Psychological Review, Vol., 84(4; 4), (1977) 327-352.
8. Tenenbaum, J. B., Griffiths, T. L.: Generalization, Similarity, and Bayesian Inference. Behavioral and Brain Sciences, Vol.24(4; 4) (2001) 629-640.
9. Glückstad, F.K.: Asymmetric Similarity and Cross-Cultural Communication Process. In: 9th International Conference on Terminology and Artificial Intelligence: Proceedings of the Conference. 8-10 November 2011, Paris, France. Paris : Institut National des Langues et Civilisations Orientales (2011) 59-65.
10. Glückstad F.K.: Bridging Remote Cultures: Influence of cultural prior-knowledge in cross cultural communication, In: Proc. The 26th Annual Conference of the Japanese Society for Artificial Intelligence: the Alan Turing Year Special Session on AI Research That Can Change The World (IOS-2), Yamaguchi, Japan (2012)

6. Application of the Infinite Relational Model combined with the Bayesian Model of Generalization for effective cross-cultural knowledge[13]

This paper investigates an optimal strategy of applying the IRM (Kemp et al., 2006) for the purpose of CSC mapping. More specifically, three strategies are tested: **1)** applying the IRM directly to two CSC-feature matrices, respectively representing the educational domain knowledge in Japan and Denmark for first categorizing them into categorical classes that are to be subsequently compared and aligned; **2)** applying the IRM directly to a matrix where the two CSC-feature matrices respectively representing the Danish- and Japanese educational domain knowledge are merged; and **3)** applying the BMG to directly compute similarity relations between CSCs in the two cultures at hand, thereafter to apply the IRM for clustering CSCs in the respective cultures into categorical classes. The strategies 1-3 respectively correspond to the following figures: Figs. 5-2, 5-3 and 5-4.

[13] In: International Organized Session – Application Oriented Principles of Machine Learning and Data Mining, The 26th Annual Conference of the Japanese Society for Artificial Intelligence, 12-15 June, 2012, Yamaguchi, Japan

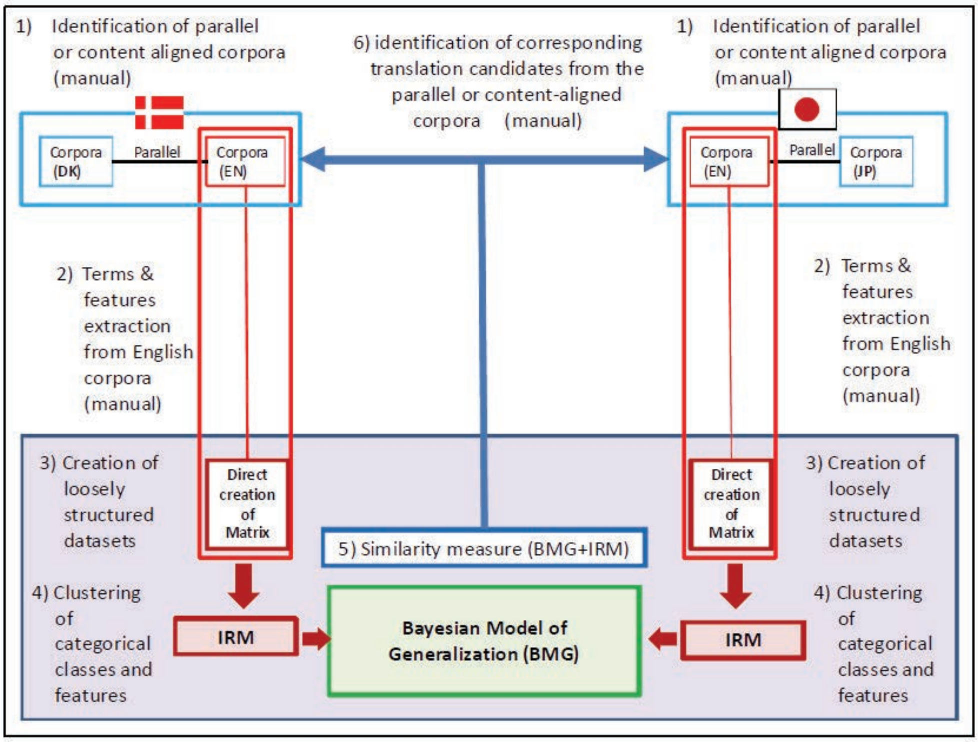


Figure 5-2: Step 2 Application of the IRM – Strategy 1

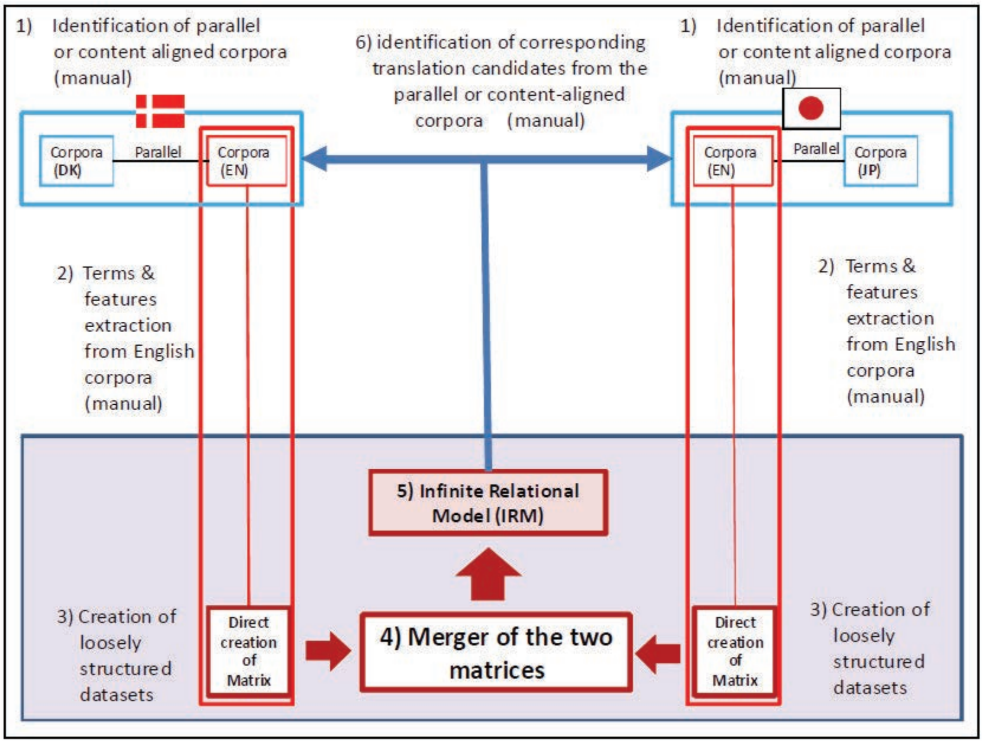


Figure 5-3: Step 2 Application of the IRM – Strategy 2

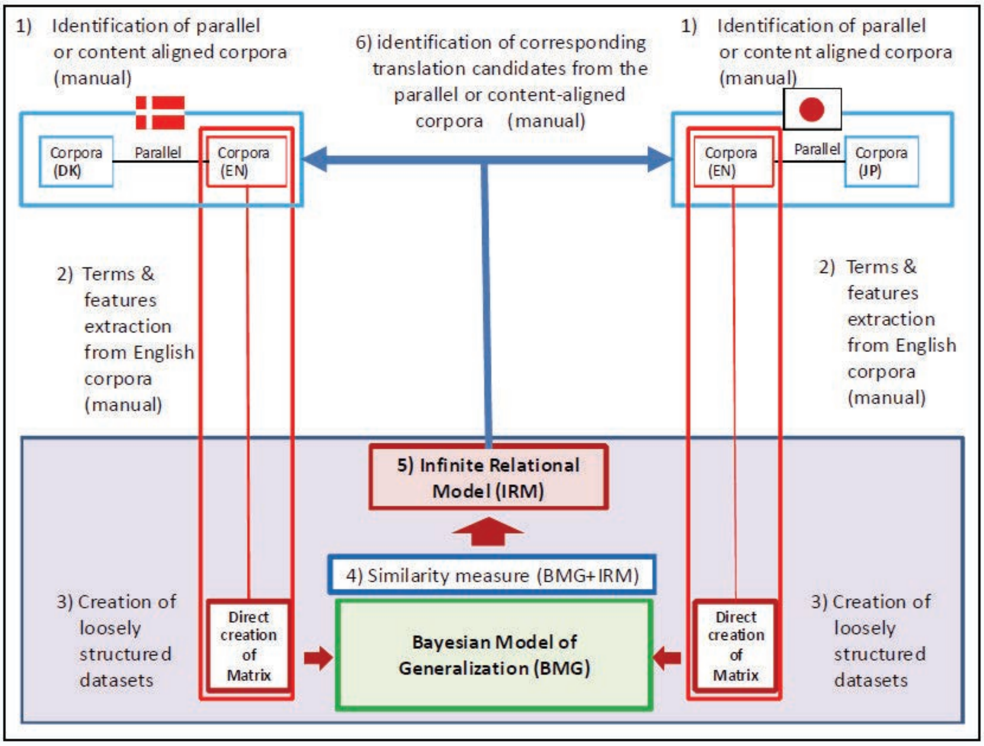


Figure 5-4: Step 2 Application of the IRM – Strategy 3

1K1-IOS-1a-4

Application of the Infinite Relational Model combined with the Bayesian Model of Generalization for Effective Cross-Cultural Knowledge Transfer

Fumiko Kano Glückstad [*1]

Morten Mørup [*2]

[*1] Dept. of International Business Communication
Copenhagen Business School (CBS)

[*2] Section for Cognitive Systems
DTU Informatics
Technical University of Denmark

This paper investigates how the Infinite Relational Model (IRM) [Kemp 2006], a novel unsupervised machine learning method, is effectively applied to loosely-structured datasets consisting of concepts and features for the purpose of mapping Culturally Specific Concepts (CSCs) in a multi-cultural context. The aim of this investigation is two-fold: **i)** to identify an effective strategy of applying the IRM for the purpose of CSC-mapping; and **ii)** to investigate possibilities of applying the IRM for efficiently constructing feature-based ontologies that are multi-culturally interoperable. Accordingly, three strategies are tested in our experiments: **1)** applying the IRM directly to two CSC-feature matrices, respectively representing the educational domain knowledge in Japan and Denmark for first categorizing them into categorical classes that are to be subsequently compared and aligned; **2)** applying the IRM directly to a matrix where the two CSC-feature matrices respectively representing the Danish- and Japanese educational domain knowledge are merged; and **3)** applying the Bayesian Model of Generalization (BMG) [Tenenbaum 2001] to directly compute similarity relations between CSCs in the two cultures at hand, thereafter to apply the IRM for clustering CSCs in the respective cultures into categorical classes. The results indicate that the third strategy seems to be the most effective approach for not only clustering CSCs into more specific and appropriate categorical classes but also for capturing complex relationships between each categorical classes existing in the two cultures.

1. Introduction

The recent internet revolution with its fast-paced globalization has brought about new possibilities for people located thousands of miles apart to real-time communicate with each other. Although we mostly use English as a common communication code, misunderstandings are almost unavoidable in contemporary cross-cultural communications. This implies that multilinguality is inherently challenged to effectively support human perception of concepts existing in diverse socio-cultural communities. Within the ontology research domain, there are several large-scale frameworks that link multi-cultural information in a complex manner. [Cimiano, 2011] compares these multilingual ontology frameworks such as the KYOTO project [Vossen, 2008] and the MONNET project [Declerck, 2010] based on a number of dimensions used in categorizing different types of ontology localization projects [Espinoza, 2009]. These dimensions are: *international (standardized) vs. culturally influenced domains; functional vs. documental localization; and interoperable vs. independent ontology* [Cimiano, 2011].

In [Glückstad, 2012-a], [Glückstad, 2012-b], potentially applicable feature-based similarity measures that can be used for mapping *independent ontologies* in a *culturally influenced domain* in a *functional* manner are compared based on qualitative analyses. These analyses identified that the Bayesian Model of Generalization (BMG) [Tenenbaum 2001] is the most intuitive and effective measure of mapping CSCs existing in two cultures. The application of the BMG requires highly appropriate datasets consisting of CSCs and their definitional features. In [Glückstad, 2012-a], an empirical study was performed with datasets obtained from a semi-automatic feature-based ontology construction method known as Terminological Ontology (TO) proposed by [Madsen 2004]. The results from that study indicated that particularly strict rules for constructing TOs may risk causing the elimination of important features. It means that the original TO-approach may require a more flexible taxonomic organization of feature structures.

Inspired by the previous works, we here investigate how the Infinite Relational Model (IRM) [Kemp 2006], a novel unsupervised machine learning method, is combined with the BMG for efficiently mapping CSCs and is applied for constructing more flexible feature structures of taxonomies. Both the IRM and the BMG applied in this work are originally proposed by cognitive scientists [Kemp, 2006] and [Tenenbaum, 2001]. While [Kemp, 2006] emphasizes that the IRM considers the semantic knowledge problem from the viewpoint of: *how representations of semantic knowledge are acquired*, [Tenenbaum, 2001] addresses three crucial questions of learning raised by [Chomsky, 1986]: **1)** *What constitutes the learner's knowledge?*; **2)** *How does the learner use that knowledge to decide how to generalize;* and **3)** *How can the learner acquire that knowledge from the example encountered?* Thus, the combination of the IRM and the BMG is potentially an interesting attempt from a philosophical point of view. However, in this paper we focus on how the IRM and the BMG are efficiently combined from a practical point of view, and the philosophical and pragmatic discussions are dealt with in a subsequent research report [Glückstad, 2012-c].

Relevant works that apply the BMG to a practical problem of mapping ontologies do not exist to our knowledge. However,

Contact: Fumiko Kano Glückstad, CBS, DK-2000 Frederiksberg, Denmark, E-mail: {fkg.ibc | fkg.isv}@cbs.dk

Tversky's set-theoretic model [Tversky, 1977] on which the BMG is based, is widely known in the area of ontology matching such as shown in [Huang, 2010] and [de Souza, 2004]. The IRM has been applied to diverse research domains among others in the area of neuroimaging where functional groups and their interactions are extracted by the IRM [Mørup, 2010] and in the area of collaborative filtering and topic modeling [Xu, 2006], [Hansen, 2011].

In the next section, the experimental settings and strategies employed in this work are explained in detail, followed by brief reviews of the BMG and the IRM in Section 3. Section 4 analyzes the results obtained from the three experimental strategies. In Section 5, we discuss some critical issues and future perspectives, followed by conclusions in Section 6.

2. Experimental settings

2.1 Data source

In this work, we first create two datasets, respectively representing the Danish educational domain knowledge and the Japanese educational domain knowledge. The datasets consists of educational terms and their definitional features that are manually extracted from text corpora. The Japanese corpora used for this experiment are: **1)** "Outline of the Japanese School System" published by the Center for Research on International Cooperation in Educational Development (CRICED), University of Tsukuba, Japan; and **2)** "Higher Education in Japan" published by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT). The Danish documents are downloaded from the Euridice web-site published by the Education, Audiovisual and Culture Executive Agency under the EU commission. These corpora are written in English and hence it is feasible to identify original expressions of educational terms in the respective languages from existing parallel- or content aligned corpora. This enables us to eventually achieve translation between Japanese CSCs and Danish CSCs through the English term mapping.

The CSCs and their definitions, all written in English, are manually extracted from the text corpora, e.g. the Danish CSC "municipal school (DA: *folkeskole*)" and its definition "*a comprehensive school covering both primary and lower secondary education, i.e. one year of pre-school class, the first (grade 1 to 6) and second (grade 7-9/10) stage basic education, or in other words it caters for the 6-16/17-year-olds*". From this definition, we create a feature list consisting of "comprehensive school" "primary and lower secondary education" "basic education" "targeted for 6-16/17 years old". This definition also implies that "municipal school" is categorized into three sub-CSCs "preschool class", "first stage" and "second stage", respectively having their features "one year preschool education" "1-6 grades" and "7-9/10 grades". These sub-CSCs are supposed to inherit features defined in the subordinate CSC, in this case "municipal school". In this way, 59 Danish CSCs and 54 Japanese CSCs and their features all written in English are listed up. In addition, some features are manually standardized, e.g. a feature "continuing education for adults" in Denmark is standardized with a feature "opportunities for life-long learning" in Japan. Finally, the following operations are manually implemented: **a)** If a feature value in one country is completely included in a feature value in the other country (e.g. the feature "6-12 y.o." in Japan is completely included in the feature "6-17 y.o." in Denmark), a CSC possessing the feature that includes the other feature (a CSC possessing "6-17 y.o." should also possess "6-12 y.o."; and **b)** If two features from the respective countries are partly overlapping (e.g. "13-15 y.o." in Japan and "14-17 y.o." in Denmark are partly overlapping), a dummy feature referring to the exact overlapping range (e.g. "14-15 y.o.") is created. In this example, a Japanese CSC possessing "13-15 y.o." should also possess the dummy feature "14-15 y.o.". In the same way, a Danish CSC possessing "14-17 y.o." should also possess the dummy feature "14-15 y.o.".

Accordingly, in total 229 features are registered in the two matrices, respectively representing the Danish- and Japanese educational systems. In each matrix, if a feature is possessed by a CSC, the numeric value "1" appears, otherwise "0" is assigned. In these matrices, the Danish- and Japanese CSCs are respectively denoted as D_i and J_j and feature IDs are assigned as f_k. Both the Danish- and Japanese CSCs and their features are alphabetically registered. Hence, it requires systematic taxonomic organization of CSCs for achieving effective CSC-mapping between the two cultures.

2.2 Experimental strategies

[Kemp 2006] emphasizes that the IRM considers the semantic knowledge problem from the viewpoint of: *how representations of semantic knowledge are acquired*, instead of starting from the viewpoint of how these systems can be represented. This has inspired us to investigate how semantic knowledge is acquired from the limited domain text corpora and how this can effectively be represented for the purpose of cross-cultural knowledge transfer. Accordingly, three strategies are tested in the experiments: **1)** applying the IRM directly to the respective CSC-feature matrices for first categorizing them into categorical classes that are to be subsequently compared and aligned; **2)** applying the IRM directly to a matrix where the two CSC-feature matrices, respectively representing the Danish- and Japanese educational domain knowledge are merged; and **3)** applying the BMG to directly compute similarity relations between CSCs in the two cultures, and thereafter to apply the IRM for clustering CSCs in the respective cultures into categorical classes. This implies that, in case of strategy **1)** and **2)**, the IRM clusters CSCs based on CSC-feature links, whereas, in case of strategy **3)**, the clustering is based on CSC-CSC links between Danish and Japanese CSCs which are identified by the BMG. These three strategies are compared in Section 4 and 5.

3. Methods

3.1 The Bayesian Model of Generalization (BMG)

The BMG [Tenenbaum 2001] is a cognitive model, which uniquely unifies the two following classically opposing approaches to similarity and generalization: Tversky's *set theoretic model of similarity* [Tversky 1977] and Shepard´s *continuous metric space model of similarity and generalization* [Shepard 1987]. The philosophical background of the BMG described in [Tenenbaum 2001] is reviewed in [Glückstad, 2012-

c]. A key point in the BMG is to compute *the conditional probability that y falls under C (Consequential Region) given the observation of the example x* based on the following formula:

$$P\ (y \in C|x) = 1/[1 + \frac{\sum_{h:x \in h, y \notin h} p(h,x)}{\sum_{h:x,y \in h} p(h,x)}] \qquad (1)$$

The consequential region C in our work indicates the categorical region where a new object y belongs. In equation (1), a hypothesized subset h is defined as the region where a concept belongs to h, if and only if, it possesses feature k [Tenenbaum 2001]. It means, for our work, that y is considered as a newly encountered object existing in a Source Culture (SC) domain when y is introduced to a Target Culture (TC) audience and the TC audience compares this new object y with an observed data x which is part of his/her prior knowledge (referent dataset).

Another unique point of the BMG is that $P(h, x) = P(x|h)P(h)$ in equation (1), represents the weight assigned to the consequential subset h in terms of the example x. This can be achieved by specifically assigning the weight $P(h, x)$ based on the strong sampling scheme defined in [Tenenbaum 2001] as follows:

$$P(x\,|h) = \begin{cases} 1/|h| & if\ x \in h \\ 0 & otherwise \end{cases} \qquad (2)$$

Here, $|h|$ indicates the size of the region h [Tenenbaum 2001]. In our work, the number of objects possessing the k^{th} feature in the referent dataset is considered as the size of the region h. [Tenenbaum 2001] explains that the prior $P(h)$ is not constrained in their analysis so that it can accommodate arbitrary flexibility across contexts. Hence in this work, $P(h) = 1$.

3.2 The Infinite Relational Model (IRM)

According to [Kemp, 2006], a key feature of the IRM is to *automatically choose an appropriate number of clusters using a prior that favors small numbers of clusters, but has access to a countably infinite collection of clusters*. In [Kemp 2006], the observed data are considered as m relations involving n types. For the experimental strategies **1)**, **2)** and **3)** in our work, we apply the simplest model: dealing with two types with a single two-place relation R: T_1 x $T_2 \rightarrow \{0, 1\}$. More specifically, in strategies **1)** and **2)** T_1 corresponds to either Danish and/or Japanese CSCs and T_2 corresponds to definitional features, while in strategy **3)**, T_1 and T_2 respectively corresponds to Danish CSCs and Japanese CSCs.

The principle of generating clusters in the IRM, according to [Kemp, 2006], is based on a distribution over partitions induced by a so-called Chinese Restaurant Process (CRP) [Pitman, 2002]. The CRP starts a partition process with a single cluster containing a single object. The i^{th} object has possibilities to belong to either of the following:

- A new cluster with probability: $\gamma / (i\text{-}1+ \gamma)$
- An existing cluster with probability: $n_a / (i\text{-}1+ \gamma)$

Here, n_a is the number of objects already assigned to cluster a, and γ is a parameter [Kemp, 2006]. The CRP continues until all the objects belong to clusters. Hence, the distribution over clusters for object i conditioned on the cluster assignments of objects 1, …, i-1 is defined as [Kemp, 2006]:

$$P(z_i = a|z_1, \dots, z_{i-1}) = \begin{cases} \gamma/(i-1+\gamma) & a \text{ is a new cluster} \\ n_a/(i-1+\gamma) & n_a > 0 \end{cases} \qquad (3)$$

[Kemp, 2006] explains that *the distribution on z induced by the CRP is exchangeable: the order in which objects are assigned to clusters can be permuted without changing the probability of the resulting partition. P(z) can therefore be computed by choosing an arbitrary ordering and multiplying conditional probabilities. Since new objects can always be assigned to new clusters, the IRM effectively has access to a countably infinite collection of clusters.*

From the clusters generated by the CRP, relations are generated based on the following generative model:

- As described above, for the cluster assignment of objects
  $z \mid \gamma \sim$ CRP (γ)
- For link probabilities between clusters
  $\eta\ (a, b) \mid \beta \sim$ Beta (β, β)
- For links between objects
  $R\ (i, j) \mid z,\eta \sim$ Bernoulli ($\eta\ (z_i, z_j)$)

In the above generative model, we set parameters β=1, and γ=log (Ji) where Ji is the number of concepts in each mode.

In here, relationships are assumed to be conditionally independent given cluster assignments [Kemp, 2006]. The eventual purpose of the generative model is to identify a cluster z that maximizes $P(z|R)$. Based on the generative model defined above, relations from clusters are generated by:

$$P(R\,|z,\eta) = \prod_{ab} (\eta_{ab})^{m_{ab}^{+}} (1-\eta_{ab})^{m_{ab}^{-}} \qquad (4)$$

where m^{+}_{ab} refers to the total number of links between categorical classes a and b; and m^{-}_{ab} refers to the total number of non-links between categorical classes a and b. The conjugate prior η_{ab} is in the aforementioned generative model defined as: $\eta(a, b) \mid \beta \sim$ Beta(β, β). Accordingly, the conjugate prior η_{ab} is integrated out in the following:

$$P(R\,|z) = \prod_{a,b \in N} \frac{Beta(m_{ab}^{+}+\beta, m_{ab}^{-}+\beta)}{Beta\ (\beta,\beta)} \qquad (5)$$

From formulae (3) and (5), the IRM identifies z that maximizes: $P\ (z|R) \propto P\ (R|z)\ P(z)$. According to [Kemp, 2006], the expected value of η_{ab} given z is:

$$\frac{m_{ab}^{+}+\beta}{m_{ab}^{-}+m_{ab}^{+}+2\beta} \qquad (6)$$

The procedure for the inference is further described in [Mørup 2010]. The solutions displayed in the following are based on the sample with highest likelihood.

4. Results

Based on the BMG and the IRM reviewed above, empirical studies are performed based on the experimental strategies described in Section 2.

4.1 Experimental strategy 1

The first experimental strategy is in a way the most natural approach to judge how an ontology is learned from data consisting of CSCs and features that respectively represent specific domain knowledge existing in two cultures. Thus the IRM is directly applied to the CSC-feature matrices, respectively created from the aforementioned English corpora describing the Danish- and the Japanese educational systems. Accordingly, 59 Danish CSCs and 229 features are simultaneously clustered into 5 and 10 categorical classes in Figure 1-a. In the same way, 54 Japanese CSCs and 229 features are respectively clustered into 6 and 11 categorical classes in Figure 1-b. In both Figure 1-a and 1-b, the unsorted graph shows the relations between the CSCs and their features. It means that each dot represents a relation that a CSC possesses a specific feature in the matrix. The upper right graphs in the figures show the graph sorted according to extracted assignments of clusters computed by the IRM algorithm. The bottom-left graphs show the distribution of CSCs over the extracted categorical classes, and the bottom-right graphs show the distribution of features over the extracted clusters. The bottom-center graphs correspond to the graphs sorted according to extracted assignments of clusters, which indicates the density of relationships between a Danish (or a Japanese) categorical class and a feature cluster. In Figure 2, all members (i.e. specific CSCs) for each categorical class are respectively listed for the Danish and Japanese educational domain knowledge.

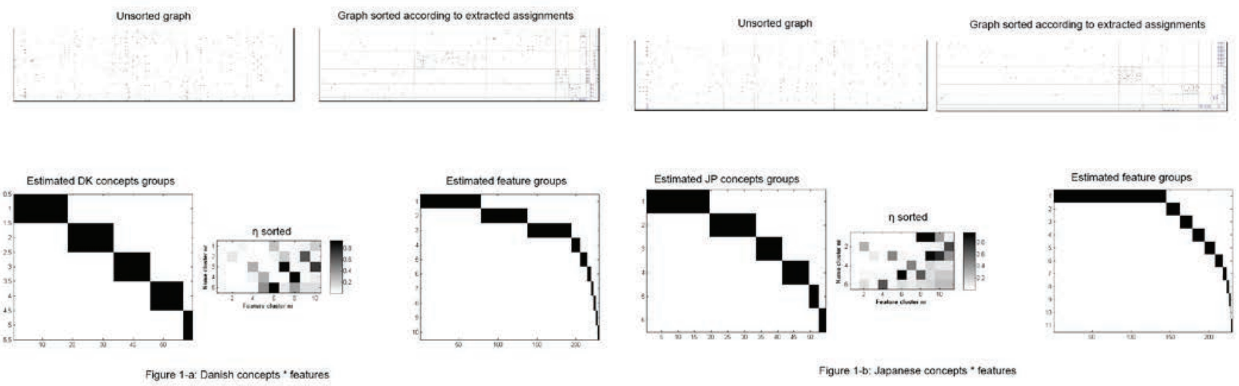


Figure 1: IRM clustering based on the Strategy 1 (Left: Danish CSCs * features / Right: Japanese CSCs * features)

Cluster 1 (Pre primary + adult education)
'D13' continuing education and training for adults
'D14' continuing vocational education and training
'D15' day-care facilities
'D22' further education for adults
'D23' general adult education
'D26' higher preparatory single subjects courses
'D28' higher education for adults
'D32' integrated institution
'D33' Kindergarten (børnehave)
'D38' Nursery/Créche/vuggestue
'D39' open education
'D41' preparatory adult education programme for adults
'D42' pre-primary education
'D43' preschool class
'D49' technical college
'D5' basic adult education
'D55' Vocational college
'D8' business college

Cluster 2 (Primary + alternative primary + Doctoral)
'D1' 10th form
'D12' Continuation school (efterskole)
'D17' doctoral degree
'D18' Dual system
'D19' first stage
'D20' Folk High School
'D21' Full time system
'D30' home tuition
'D36' municipal school/Folkeskole
'D37' non-formal adult education
'D44' private school
'D46' school-leaving examinations
'D47' second stage
'D48' primary and lower secondary education
'D59' youth school

Cluster 3 (Upper secondary)
'D24' general upper secondary education
'D25' gymnasium/STX
'D27' HHX (højere handels eksamen)
'D29' higher preparatory exam
'D31' HTX (højere teknisk eksamen)
'D34' main course
'D54' upper secondary education
'D56' vocational education and training
'D57' vocational upper secondary education
'D58' youth education programmes
'D6' basic course
'D7' Basic vocational education and training

Cluster 4 (Tertiary)
'D10' Centres for Higher Education
'D11' college sector
'D2' Academies of professional higher education
'D3' Academy profession degree programmes
'D40' PhD programme
'D45' professional bachelor programme
'D50' tertiary education
'D51' University Bachelor degree programme
'D52' University college
'D53' University education
'D9' Candidates degree programme

Cluster 5 (open education)
'D16' Diploma programmes
'D35' master programme
'D4' Advanced adult program

Cluster 1 (Upper secondary)
'J1' Advanced course (Senkoka)
'J13' Fisheries course
'J14' Full time course
'J16' General course (Honka)
'J17' Graded course
'J19' Home-economics course
'J2' Agriculture course
'J31' Nursing course
'J32' Ordinary education course
'J33' Part time course
'J44' Public upper secondary school
'J47' Specialized course (Bekka)
'J48' Specialized education course
'J50' Technology course
'J54' Upper secondary school
'J6' Commerce course
'J7' Comprehensive education course
'J8' Correspondence course
'J9' Credit course

Cluster 2 (Tertiary + Primary)
'J11' Doctoral degree
'J12' Elementary school
'J18' Graduate school
'J20' Junior college
'J26' Master's degree
'J28' National elementary school
'J29' National university
'J36' Private elementary school
'J38' Private university
'J39' Professional graduate school
'J40' Public elementary school
'J43' Public university
'J51' Undergraduate department
'J52' University

Cluster 3 (alternative post compulsory)
'J15' General course
'J27' Miscellaneous schools
'J3' colleges of technology
'J4' colleges of technology economy, IT management course
'J46' Specialized course
'J49' Specialized training college
'J5' colleges of technology – industrial course
'J53' Upper secondary course

Cluster 4 (Pre Primary + post compulsory)
'J10' Day care center
'J21' kindergarten
'J22' Kindergarten 1 year course
'J23' Kindergarten 2 years course
'J24' Kindergarten 3 years course
'J30' nursery school
'J34' post-compulsory educational institution
'J35' Pre-school education

Cluster 5 (Secondary)
'J37' Private six-year secondary school
'J42' Public six-year secondary school
'J45' six-year secondary school/ comprehensive secondary school

Cluster 6 (Lower secondary)
'J25' lower secondary school
'J41' Public lower secondary school

Figure 2: CSC members that constitute each categorical class based on Strategy 1 (Left: Danish / Right: Japanese)

For convenience, each categorical class has been named in Figure 2 within a parenthesis based on members that constitute the specific categorical class in question. As Figure 2 shows, some categorical classes (e.g. Danish classes 3, 4 and 5; and Japanese classes 1, 3, 5, and 6) are successfully formed only with CSCs that are related to the respective categorical classes such as "upper secondary", "open education", "secondary", and "lower secondary". However, the rest of the categorical classes are partly formed with CSCs that represent different categorical classes. For example, the Danish categorical class 1 consists of CSCs that are supposed to belong to "pre primary" and "adult education" and Japanese categorical class 2 consists of CSCs that are supposed to belong to "tertiary" and "primary". When observing Figure 1-a, the successful Danish categorical class 3 "upper secondary" has a very dense relationship with feature cluster 7 consisting of "16-18 years old" and "post compulsory education" and with feature cluster 10 consisting of "upper secondary education" and "vocational perspectives". In the same way; the Danish categorical class 5 representing degree programs targeted for adults has a dense relationship with feature cluster 6 consisting of features: "opportunities for lifelong learning", "part time", "possibilities for combining education and work", "occupational function", and "open education". Figure 1-b shows another notable point that the Japanese categorical classes 1 and 3 both have a dense relationship with feature cluster 9 consisting of "non-compulsory educational school" and "post-compulsory education". However, the Japanese categorical class 1 - "upper secondary" - has also a strong relationship with feature cluster 8 consisting of "16-18 years old". Also the Japanese categorical class 3 - "alternative post compulsory" - has another relationship with feature cluster 10 consisting of "education + practical training". This indirectly indicates that the Japanese categorical classes 1 and 3 both belong to a super-ordinate category (although it does not exist in the dataset) referring to "post-compulsory education". This kind of information could be useful for representing knowledge in a taxonomical structure, e.g. for constructing Terminological Ontologies [Madsen 2004].

The results of experimental strategy 1 indicate that if few decisive features exist for representing a categorical class, the IRM effectively sorts CSCs that relate to these decisive features. However, when relationships between categorical classes and feature clusters are weak, there is a tendency that CSCs that belongs to different categorical classes start to be mixed into one class.

4.2 Experimental strategy 2

The second strategy is to apply the IRM directly to the matrix where the two CSC-feature matrices respectively representing the Danish- and Japanese educational domain knowledge are merged. The aim of this second strategy is to assess whether the IRM can directly be used for mapping CSCs existing in two cultures. Accordingly, in total 113 CSCs (59 Danish and 54 Japanese) and 229 features are respectively clustered into 8 categorical classes and 16 feature clusters as shown in Figure 3. Figure 4 lists 8 categorical classes and their cluster members.

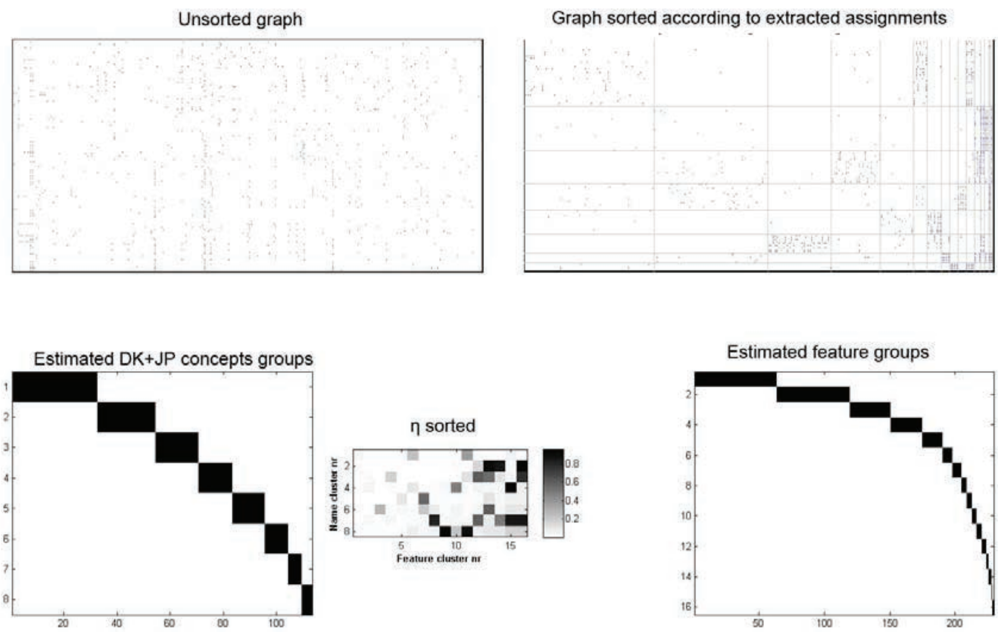


Figure 3: IRM clustering based on Strategy 2 (Danish + Japanese CSCs * features)

The results in Figure 4 show that 3 out of 8 categorical classes are mono-cultural categories. More specifically, class 4 solely consists of Danish CSCs referring to the Danish adult education; and the classes 7 and 8 solely consist of Japanese CSCs respectively referring to the Japanese alternative post-compulsory education and graduate school. It means that the purpose of mapping CSCs existing in the two cultures has not been achieved in terms of these three categorical classes. On the other hand, one interesting finding is that the Danish mono-cultural categorical class 4 and the Japanese mono-cultural categorical class 7 share feature cluster 15 consisting of "opportunities for lifelong learning" according to the η-sorted graph in Figure 3. Although concepts existing in different cultures have not been grouped in the same class, the IRM enables us to identify a relationship indicating which categorical class share which feature cluster in the η-sorted graph. This type of information is highly useful for constructing feature-based ontologies as well as for CSC-mappings. Finally, the rest of the five bi-cultural categorical classes result in rather ambiguous categories. For example, categorical class 2 that is the group of upper secondary CSCs, consists of the most abstract Danish upper secondary CSCs and all concrete Japanese upper secondary CSCs. On the contrary, categorical class 3 that is also the group of upper secondary CSCs representing concrete Danish upper secondary education for both general and vocational purposes are grouped together with the Japanese college of technologies (JP: *Koto Senmon Gakko*) which provides 5 years of practical and vocational education consisting of 3 years upper secondary and 2 years post-secondary education. Accordingly, the mapping results obtained from the second strategy seems not to be optimal.

Cluster 1 (pre primary + tertiary)
'D10' Centres for Higher Education
'D11' college sector
'D13' continuing education and training for adults
'D15' day-care facilities
'D17' Doctoral degree
'D2' Academies of professional higher education
'D3' Academy profession degree programmes
'D32' integrated institution
'D33' Kindergarten (børnehave)
'D38' Nursery/Créche/vuggestue
'D40' PhD programme
'D42' pre-primary education
'D43' preschool class/0th form/børnehave klasse
'D45' professional bachelor programme
'D50' tertiary education
'D51' University Bachelor degree programme
'D52' University college
'D53' University education
'D9' Candidates degree programme
'J10' Day care center
'J20' Junior college
'J21' kindergarten
'J22' Kindergarten 1 year course
'J23' Kindergarten 2 years course
'J24' Kindergarten 3 years course
'J29' National university
'J30' nursery school
'J35' Pre-school education
'J38' Private university
'J43' Public university
'J51' Undergraduate department
'J52' University

Cluster 4 (adult education)
'D14' continuing vocational education and training
'D16' Diploma programmes/Diplomuddannelse
'D20' Folk High School
'D22' further education for adults
'D23' general adult education
'D26' higher preparatory single subjects courses
'D28' higher education for adults
'D35' master programme
'D37' non-formal adult education
'D39' open education
'D4' Advanced adult program
'D41' preparatory adult education programme for adults
'D5' basic adult education

Cluster 2 (upper secondary)
'D54' upper secondary education
'D58' youth education programmes
'J1' Advanced course (Senkoka)
'J13' Fisheries course
'J14' Full time course
'J16' General course (Honka)
'J17' Graded course
'J19' Home-economics course
'J2' Agriculture course
'J31' Nursing course
'J32' Ordinary education course
'J33' Part time course
'J34' post-compulsory educational institution
'J44' Public upper secondary school
'J47' Specialized course (Bekka)
'J48' Specialized education course
'J50' Technology course
'J54' Upper secondary school
'J6' Commerce course
'J7' Comprehensive education course
'J8' Correspondence course
'J9' Credit course

Cluster 3 (general- and vocational upper secondary + college of technology)
'D24' general upper secondary education
'D25' gymnasium/STX
'D27' HHX (højere handels eksamen)
'D29' higher preparatory exam
'D31' HTX (højere teknisk eksamen)
'D34' main course
'D49' technical college
'D55' Vocational college
'D56' vocational education and training
'D57' vocational upper secondary education
'D6' basic course
'D7' Basic vocational education and training
'D8' business college
'J3' colleges of technology
'J4' colleges of technology – economy, IT management course
'J5' colleges of technology – industrial course

Cluster 5 (primary + lower secondary)
'D1' 10th form/bridge building
'D19' first stage
'D30' home tuition
'D36' municipal school/Folkeskole
'D44' private school
'D46' school-leaving examinations
'D47' second stage
'D48' primary and lower secondary education
J12 elementary school
'J28' National elementary school
'J36' Private elementary school
'J40' Public elementary school

Cluster 6 (alternative lower secondary + secondary school)
'D12' Continuation school (efterskole)
'D18' Dual system
'D21' Full time system
'D59' youth school
'J25' lower secondary school
'J37' Private six-year secondary school
'J41' Public lower secondary school
'J42' Public six-year secondary school
'J45' six-year secondary school/ comprehensive secondary school

Cluster 7 (alternative post compulsory)
'J15' General course
'J27' Miscellaneous schools
'J46' Specialized course
'J49' Specialized training college
'J53' Upper secondary course

Cluster 8 (graduate school)
'J11' Doctoral degree
'J18' Graduate school
'J26' Master's degree
'J39' Professional graduate school

Figure 4: CSC members that constitute each categorical class based on Strategy 2 (Danish + Japanese concepts)

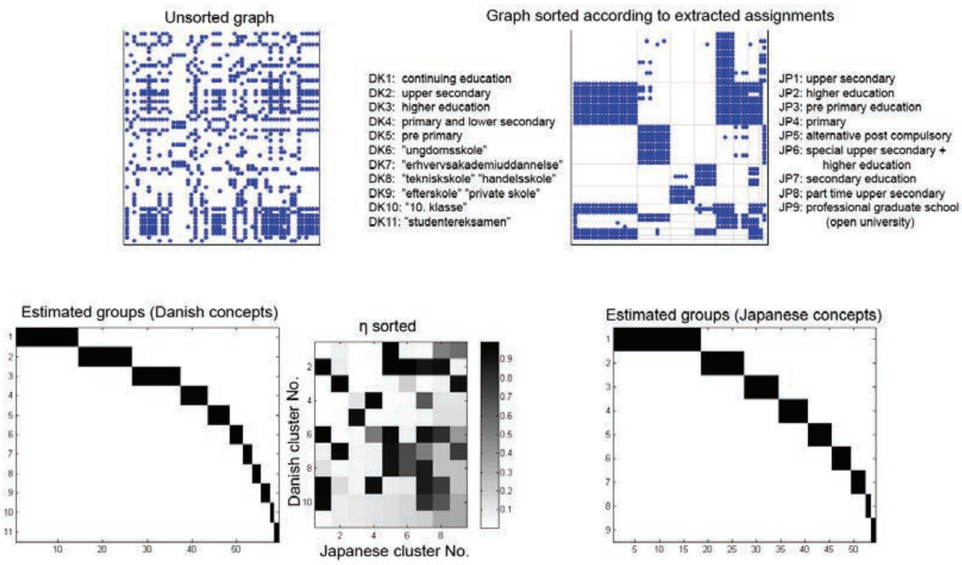


Figure 5: IRM clustering based on the Strategy 3 (BGM + IRM)

Cluster 1 (continuing education)
D13:continuing education and training for adults
D14:continuing vocational education and training
D16:Diploma programmes
D20:Folk High School
D22:further education for adults
D23:general adult education
D26:higher preparatory single subjects courses
D28:higher education for adults
D35:master programme
D37:non-formal adult education
D39:open education
D4:Advanced adult program
D41:preparatory adult education programme for adults
D5:basic adult education

Cluster 2 (upper secondary)
D24:general upper secondary education
D25:gymnasium/STX
D27:HHX (højere handels eksamen)
D29:higher preparatory exam/HF
D31:HTX (højere teknisk eksamen)
D34:main course
D54:upper secondary education
D56:vocational education and training (I-VET)
D57:vocational upper secondary education
D58:youth education programmes
D6:basic course
D7:Basic vocational education and training

Cluster 9 (alternative lower secondary)
D12:Continuation school (efterskole)
D44:private school

Cluster 10 (bridge building course)
D1:10th form/bridge building

Cluster 11 (lower secondary graduate)
D46:school-leaving examinations

Cluster 3 (higher education)
D10:Centres for Higher Education
D11:college sector
D17:doctoral degree
D2:Academies of professional higher education
D40:PhD programme
D45:professional bachelor programme
D50:tertiary education
D51:University Bachelor degree programme
D52:University college
D53:University education
D9:Candidates degree programme

Cluster 4 (primary and lower secondary)
D19:first stage
D30:home tuition
D36:municipal school/Folkeskole
D43:preschool class/0th form/børnehave klasse
D47:second stage
D48:primary and lower secondary education

Cluster 5 (pre primary)
D15:day-care facilities
D32:integrated institution
D33:Kindergarten (børnehave)
D38:Nursery/Créche/vuggestue
D42:pre-primary education

Cluster 6 (youth school)
D18:Dual system
D21:Full time system
D59:youth school

Cluster 7 (vocational academy)
D3:Academy profession degree programmes
D55:Vocational college

Cluster 8 (vocational college)
D49:technical college
D8:business college

Cluster 1 (upper secondary)
J1:Advanced course (Senkoka)
J13:Fisheries course
J14:Full time course
J16:General course (Honka)
J17:Graded course
J19 :Home-economics course
J2 :Agriculture course
J31 :Nursing course
J32 :Ordinary education course
J44 :Public upper secondary school
J47 :Specialized course (Bekka)
J48 :Specialized education course
J50 :Technology course
J54 :Upper secondary school
J6 :Commerce course
J7 :Comprehensive education course
J8 :Correspondence course
J9 :Credit course

Cluster 5 (alternative post compulsory)
J15 :General course
J27 :Miscellaneous schools
J46 :Specialized course
J49 :Specialized training college
J53 :Upper secondary course

Cluster 6 (special upper secondary + higher education)
J3 :colleges of technology
J34: post-compulsory educational institution
J4 :colleges of technology – economy, IT management course
J5 :colleges of technology – industrial course

Cluster 7 (secondary)
J37 :Private six-year secondary school
J42 :Public six-year secondary school
J45 :six-year secondary school

Cluster 2 (higher education)
J11 :Doctoral degree
J18 :Graduate school
J20 :Junior college
J26 :Masters degree
J29 :National university
J38 :Private unviersity
J43 :Public university
J51 :Undergraduate department
J52 :university

Cluster 3 (pre primary)
J10 :Day care center
J21 :kindergarten
J22 :Kindergarten 1 year course
J23 :Kindergarten 2 years course
J24 :Kindergarten 3 years course
J30 :nursery school
J35 :Pre-school education

Cluster 4 (primary)
J12 :elementary school
J25 :lower secondary school
J28 :National elementary school
J36 :Private elementary school
J40 :Public elementary school
J41 :Public lower secondary school

Cluster 8 (part time upper secondary)
J33 :Part time course

Cluster 9 (professional graduate school)
J39 :Professional graduate school

Figure 6: CSC members that constitute each categorical class based on Strategy 3 (Left: Danish / Right: Japanese)

4.3 Experimental strategy 3

The third strategy is to apply the BMG to directly compute similarity relations between CSCs existing in the two cultures, and thereafter to apply the IRM in order to cluster CSCs in the respective countries into categorical classes. This enables us not only to observe the inter-relations of categorical classes existing in the two cultures, but also to instantly scrutinize more specific similarity relations between each category member (i.e. CSCs) existing in the two cultures. Accordingly, 59 Danish CSCs and 54 Japanese CSCs are simultaneously clustered into 11 and 9 categorical classes, respectively shown in Figure 5. In this figure, the unsorted graph shows existing links between the Danish CSCs and the Japanese CSCs identified by the computation of the BMG. It means that each dot represents a link established between a Danish CSC and a Japanese CSC, when they share at least one common feature. The upper right graph in Figure 5 shows the graph sorted according to extracted assignments of categorical classes computed by the IRM algorithm. The bottom left- and right graphs show the distribution of concepts over the extracted categorical classes, respectively for the Danish- and Japanese CSCs. The bottom center graph corresponds to the graph sorted according to extracted assignments of categorical classes, which indicates the density of relationships between a Danish categorical class and a Japanese categorical class.

The results in Figure 6 show that both the Danish- and the Japanese CSCs are clustered into a more fine-grained level compared with the results obtained from the first- and second experimental strategies. Almost all members in each categorical class in Figure 6 are grouped together with other members that are supposed to belong to the same categorical class. For example, some CSCs such as the Japanese "J3: college of technology (JP: *koto-senmon-gakko*)" and the Danish "D36: municipal school (DK: *folkeskole*)" are CSCs that are difficult to be categorized in a multi-cultural context. While, in the first- and second experimental strategies, these CSCs have been included in a more ambiguous larger categorical class where CSCs that are supposed to belong to different categorical classes have been grouped together, J3 and D36 are respectively grouped into a more specific and independent categorical class, i.e. the Japanese categorical class 6 and the Danish categorical class 4, in this third strategy. One of the noteworthy points in the third strategy is that, when observing the η-sorted graph in Figure 5, it is possible to study more complex relationships of categorical classes in a cross-cultural context. For instance, the Japanese categorical class 6 where "J3: college of technology" belongs, has a strong relationship with the Danish categorical class 2 "upper secondary" class, but also has a little weaker relationship with both the Danish categorical classes 7 and 8, which respectively represent "vocational academy" and "vocational college" categories providing a 2 years post-secondary degree in Denmark. The observation of the η-sorted graph in Figure 5 further provide a clear picture of how each country-specific categorical class is related to categorical classes existing in another country in a very complex and comprehensible manner. This kind of overview of how categorical classes in different cultures are inter-related is highly useful and valuable not only for mapping CSCs but also for constructing ontologies in a multi-cultural context.

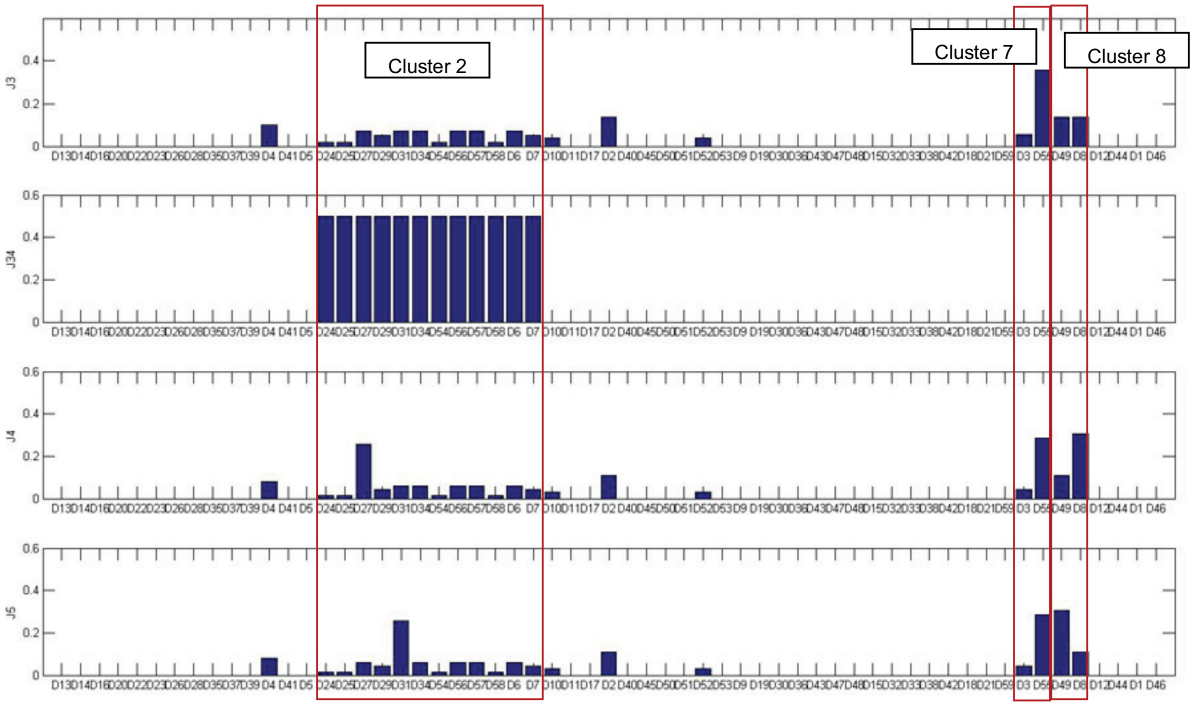


Figure 7: BMG similarity relations between the category members of Japanese categorical class 6 and all the Danish concepts

5. Discussions and future perspectives

The experimental results presented in this work indicate that the third strategy seems to be the most effective approach for clustering CSCs into more specific and appropriate categorical classes. In addition, this strategy enables us to capture complex relationships existing between each categorical class in the two cultures. However, the drawback of the third strategy is that it is not able to assess how each categorical class is related with features. On the other hand, the first- and second strategies enable us to analyze how features and each categorical class are related to each other. The shortcoming of these strategies is that the clustered categorical classes are rather ambiguous and some categorical classes are mixed with members that are supposed to belong to other categorical classes.

Another advantage of the third strategy is that the direct application of the BMG enables us to analyze further specific similarity relations between category members of the respective categorical classes existing in the two cultures. Figure 7 illustrates how the category members of the Japanese categorical class 6 in Figure 6 are related with the category members of the Danish categorical classes 2, 7 and 8. As discussed in the previous section, the η-sorted graph in Figure 5 shows that the Japanese categorical class 6 has the strongest relationship with the Danish categorical class 2 and slightly weaker relationship with the Danish categorical classes 7 and 8. Figure 7 explains these relationships between the classes by showing that all the category members of the Danish categorical class 2 share at least one feature with all the category members of the Japanese categorical class 6, while only 75% of the category members of the Danish categorical classes 7 and 8, respectively, share at least one feature with 75% of the category members of the Japanese categorical class 6. On the other hand, when observing individual relationships between category members between the Japanese- and the Danish categorical classes, similarity relationships are not necessarily strong in most of the combinations. Here, the selection of feature-based similarity measures plays in to the considerations.

In this work, we have selected the BMG as the most suitable feature-based similarity measure. However, for implementing the IRM based on the third experimental strategy, it is possible to apply other feature-based similarity measures, such as the Jaccard similarity coefficient [Tan, 2005], [Jaccard, 1901] and Tversky´s set-theoretic model [Tversky, 1977], which compute similarities based on common features and distinctive features possessed by the two CSCs in question. Comparative qualitative analyses of applying different feature-based similarities to CSC-mapping are further discussed in [Glückstad, 2012-a] and [Glückstad, 2012-b], and our arguments of applying the BMG from both cognitive- and pragmatic point of views are discussed in details in another session of this conference [Glückstad, 2012-c]. Thus, in this work, we focus on how to interpret the results obtained from the BMG shown in Figure 7. As explained in Section 2, equation (1) computes the conditional probability that a new observed object y falls under a consequential region C given the learner's prior knowledge that have already been observed as x. It means that in relation to Figure 7, a scenario could be that a Japanese person who has prior knowledge of the Japanese educational system is learning the Danish educational system by comparing similarities of individual Danish educational CSCs. Thus, all the knowledge about Japanese CSCs are considered as x and the individual Danish CSC as y in equations (1) and (2) in Section 2. In other words, it can be interpreted that all definitional features possessed by a Japanese CSC are considered as prior knowledge of the Japanese learner and can act as noise (or cultural bias) if a feature is not possessed by a Danish concept taken in comparison. In addition, the uniqueness of the BMG is to reflect the importance of features for the similarity computation by considering features that are possessed by many concepts as less important and features that are possessed by fewer concepts as more important and decisive. In case of “J5: college of technology – industrial course (JP: *Koto-senmon-gakko, sangyo-koosu*)”, Danish CSCs - “D31: HTX (Danish upper education that is specialized in technical and natural science)”, “D55: vocational college (Danish educational institution that offer vocationally-oriented upper- and 2 year post-secondary education)”, and “D8: technical college (Danish vocational college specialized in technical and natural science)” – they are identified as the most similar concepts. In the respective cases, one or few decisive features such as “specialized in technical and natural science”, “offering 2 year post-secondary degree” strongly influence the similarity computation and differentiate the confidence levels of similarity from other concepts. Another interesting pattern of applying the BMG as shown in Figure 7 is related to “J34: post-compulsory educational institution (JP: *Gimu-kyoiku-go-no-kyoiku-kikan*)”. The J34-CSC is a very abstract concept only possessing two features. It means that no other distinctive features influence as noise in the similarity computation. Accordingly, the result indicates that the Japanese CSC, J34, likely covers all category members of the Danish categorical class 2 as the most similar concepts identified in the Danish culture. In this way, the combination of the BMG and the IRM could possibly be used for first capturing the abstract relationship between categorical classes and next for further analyze the individual similarity relationship between CSCs in order to achieve more fine-grained CSC-mappings.

A critical point in this work is that the datasets have been created in a way where the authors systematically but manually extracted CSCs and definitional features from the applied text corpora. Although the procedure has been systematized, it is hence not possible to perfectly eliminate human subjectivities. Accordingly, one of our future challenges is to further investigate what types of datasets are suitable for CSC-mapping applying the BMG+IRM solution. One possibility would be to apply this solution to a more complex dataset obtained from other major multilingual ontology projects such as the Monnet project [Declerck, 2010]. Another possibility is to apply the BMG+IRM solution to datasets consisting of CSCs and features that are automatically extracted from text corpora [Lassen, 2012]. This may eventually lead to not only an automated CSC-mapping but also to an automatic ontology learning applying the IRM.

Another aim of this work is, as briefly mentioned in the previous paragraph, to investigate possibilities of applying the IRM for efficiently constructing feature-based ontologies that are multi-culturally interoperable. From this viewpoint, the method of constructing Terminological Ontologies (TOs) in [Madsen, 2004] that are in accordance with [ISO 2000] could be highly suitable for the IRM application. The uniqueness of the TO method is its feature specifications and subdivision criteria [Madsen 2004], [Madsen 2005]. While the use of feature specifications is subject to principles and constraints, the TO-approach allows for so-called poly-hierarchy structures. It means that one CSC may be related to two or more super-ordinate CSCs. Accordingly, the IRM could potentially be used as an effective pre-processing step of constructing TOs that are interoperable in a multi-cultural context. Mainly for two reasons: **1)** the IRM may indicate which features influence the formation of categorical classes as the results from the first- and second experimental strategies in this work have shown; and **2)** the IRM based on the third experimental strategy may cluster CSCs into more specific and appropriate categorical classes that may capture complex relationships between each categorical classes existing in the two cultures. Hence, we need a solution to achieve these strategies at one time. Here, it is important to remind that the design we have chosen for this work is the simplest design of the IRM dealing with two types with a single two-place relation R: T_1 x $T_2 \rightarrow \{0, 1\}$, and as described in [Kemp, 2006], the IRM design can be a more complex model clustering three relations simultaneously. Accordingly, our obvious future challenge would be to investigate what kind of complex model is applicable for constructing ontologies in collaboration with the automatic TO construction project [Madsen, 2010].

6. Conclusions

In this work, we investigated the application of the IRM [Kemp 2006] to the loosely-structured datasets consisting of CSCs and features representing two cultures for the purpose of mapping CSCs in a multi-cultural context. The results from the three applied experimental strategies indicate that the combination of the BMG and the IRM seems to be the most effective approach for not only clustering CSCs into more specific and appropriate categorical classes but also for capturing complex relationships between each categorical classes existing in the two cultures. In addition, the direct application of the BMG to the datasets enables us to effectively analyze further specific similarity relations between category members existing in the two cultures. However, in order to conclude on this, it is necessary to investigate the performance of the BMG+IRM solution with different types of datasets that are purely objectively generated. Another important finding is the potential application of the IRM for the automatic construction of feature-based ontologies among others, TOs [Madsen 2004]. The results obtained from the first- and second experimental strategies indicate that this may potentially be achieved by designing a more complex IRM. Although further research is required, the application of the IRM to the multi-cultural ontologies seems to provide a diverse potential in this research domain.

Acknowledgement

We would like to express our thanks to the DanTermBank project members, among others, Bodil Nistrup Madsen, Hanne Erdman Thomsen and Tine Lassen at the Copenhagen Business School for valuable and relevant discussions related to our work.

References

[Chomsky 1986] Chomsky, N., Language and Problems of Knowledge: The Managua lectures. Cambridge, MA: MIT Press, 1986

[Cimiano 2010] Cimiano P., Montiel-Ponsoda E., Buitelaar P., Espinoza M., Gómez-Pérez A. A Note on Ontology Localization., Journal of Applied Ontology Vol. 5, No. 2, IOS Press, 2010.

[Declerck 2010] Declerck T., Krieger H.U., Thomas S.M., Buitelaar P., O'Riain S., Wunner T., Maguet G., McCrae J., Spohr D., Montiel-Ponsoda.E. Ontology-based Multilingual Access to Financial Reports for Sharing Business Knowledge Across Europe., In Internal Financial Control Assessment Applying Multilingual Ontology Framework, J. Rooz, J. Ivanyos, Eds. Budapest: HVG Press, 2010

[de Souza 2004] de Souza K.X.S., Davis J., Aligning Ontologies and Evaluating Concept Similarities, In On the Move to Meaningful Internet Systems 2004: Lecture Notes in Computer Science, Volume 3291, Springer, 2004.

[Espinoza 2009] Espinoza, M., Montiel-Ponsoda, E., Gómez-Pérez, A. Ontology Localization. In Proc. the 5th International Conference on Knowledge Capture (KCAP) New York, NJ, USA, 2009

[Glückstad 2012-a] Glückstad F.K., Mørup M., Feature-based Ontology Mapping from an Information Receivers' Viewpoint, (under review)

[Glückstad 2012-b] Glückstad F.K., Cross-Cultural Concept Mappings of Standardized Datasets, (under review)

[Glückstad 2012-c] Glückstad F.K., Bridging Remote Cultures: Influence of Cultural Prior-Knowledge in Cross-Cultural Communication, In Proc. The 26th Annual Conference of the Japanese Society for Artificial Intelligence, Japan, June 2012.

[Hansen 2011] Hansen T.J., Mørup M., Hansen L. K., Non-parametric Co-clustering of Large Scale Sparse Bipartite Networks on the GPU, In IEEE International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, 2011.

[Huang 2010] Huang H.H., Kuo Y.H., Cross-lingual Document Representation and Semantic Similarity Measure: A fuzzy set and rough set approach, In IEEE Transaction on Fuzzy Systems, Volume 18:6, IEEE, 2010.

[ISO 2000] ISO 704. Terminology Work — principles and methods, ISO, 2000

[Jaccard 1901] Jaccard, P., Étude comparative de la eistribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37: 1901

[Kemp 2006] Kemp, C., Tenenbaum, J.B., Griffiths, T.L. Yamada, T., Ueda, N., Learning Systems of Concepts with an Infinite Relational Model., In Proc. the 21st National AAAI Conference on 1:381-388, 2006.

[Lassen 2012] Lassen, T., A Corpus Compilation and Processing Prototype for Terminology Work., In Proc. the 10th International Conference on Terminology and Knowledge Engineering, Spain, June 2012.

[Madsen 2004] Madsen, B.N., Thomsen, H.E. and Vikner, C., Principles of a System for Terminological Concept Modelling. In Proc. the 4th International Conference on Language Resources and Evaluation. ELRA, 2004.

[Madsen 2005] Madsen, B.N., Thomsen, H.E. and Vikner, C., Multidimensionality in Terminological Concept Modelling. In Proc. of the 7th International Conference on Terminology and Knowledge Engineering, Copenhagen, 2005

[Madsen 2010] Madsen N.B., Thomsen H.T., Halskov J., Lassen T., Automatic Ontology Construction for a National Term Bank, In Proc. Terminology and Knowledge Engineering Conference 2010, Dublin, 2010.

[Mørup 2010] Mørup M., Madsen K.H., Dogonowski A.M., Siebner H., Hansen L.K., Infinite Relational Modeling of Functional Connectivity in Resting State fMRI., In Neural Information Processing Systems, 2010.

[Pitman, 2002] Pitman J., Combinatorial Stochastic Processes. Notes for Saint Flour Summer School, 2002.

[Shepard 1987] Shepard R.N., Towards a Universal Law of Generalization for Psychological Science. Science, 237, 1987.

[Tan 2005] Tan P.N., Steinbach M., Kumar V., Introduction to Data Mining, Pearson Education, Inc. 2006.

[Tenenbaum 2001] Tenenbaum, J. B., & Griffiths, T. L., Generalization, Similarity, and Bayesian Inference. Behavioral and Brain Sciences, Vol.24(4; 4), 2001

[Tversky 1977] Tversky, A., Features of Similarity. Psychological Review, Vol., 84(4; 4), 1977

[Vossen 2008] Vossen P., Agirre E., Calzolari N., Fellbaum C., Hsieh S., Huang C.R., Isahara H., Kanzaki K., Marchetti A., Monachini M., Neri F., Raffaelli R., Rigau G., Tescon M., VanGent J. KYOTO: A system for mining, structuring and distributing knowledge across languages and cultures. In Proc. the 6th International Conference on Language Resources and Evaluation, Morocco, 2008.

[Xu 2006] Xu Z., Tresp V., Yu K., Kriegel H.P., Infinite Hidden Relational Models, In Proc. 22nd Conf. on Uncertainty in Artificial Intelligence (UAI'06) Cambridge, MA, 2006

7. Flexible- or strict taxonomic organization? – Impact on culturally-specific knowledge transfer[14]

This paper compares the two methods of representing knowledge: "TO" (the TO+BMG approach) and the "non-TO" (the BMG+IRM approach) and an assessment of advantages and disadvantages of the TO under the controlled settings. The empirical study conducted in this paper, respectively, corresponds to Modules 1-8 and Modules 1-5 for the TO+BMG approach and the BMG+IRM approach.

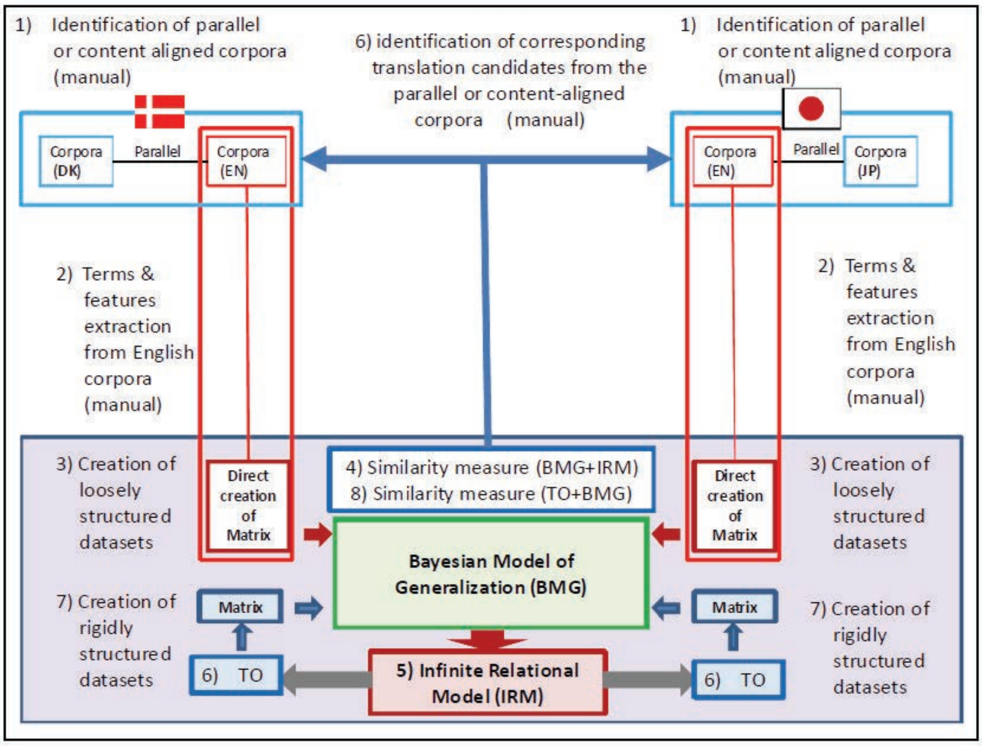


Figure 5-5: Workflow for Step 2 in the COM framework

(Step 2: comparison of the knowledge representation method)

---

[14] Peer-reviewed extended abstract accepted as full-paper, In: Terminology and Knowledge Engineering Conference 2012, Madrid, Spain, 20-21 June 2012

Flexible or Strict Taxonomic Organization ?

Impact on culturally-specific knowledge transfer

Fumiko Kano Glückstad[1] and Morten Mørup[2]

[1] Copenhagen Business School, Dept. of International Business Communication
Dalgas Have 15, DK-2000 Frederiksberg, Denmark
{fkg.ibc , fkg.isv}@cbs.dk

[2] Technical University of Denmark, DTU Informatics, Section for Cognitive Systems
Richard Pedersens Plads, DK-2800 Kgs Lyngby, Denmark
mm@imm.dtu.dk

Abstract. This work compares methods for constructing feature-based ontologies that are supposed to be used for culturally-specific knowledge transfer. The methods to be compared are the Terminological Ontology (TO) [1], a method of constructing ontology based on strict principles and rules, and the Infinite Relational Model (IRM) [2], a novel unsupervised machine learning method that learns multi-dimensional relations among concepts and features from loosely structured datasets. These methods are combined with a novel cognitive model, the Bayesian Model of Generalization (BMG) [3] that maps culturally-specific concepts existing in two cultures in an effective and intuitive manner.

1 Introduction

The fast-paced globalization driven by the current internet revolution has brought about new possibilities for people scattered around the planet to dynamically communicate with each other anytime and anywhere. It has become a familiar situation where e.g. an Asian- and a European are communicating at online channels via a transit of their second language - English - as the most common communication "code". However, even though the majority of us are using the most common language – English – while communicating internationally, misunderstandings are commonly occurring with all its derived challenges that are almost unavoidable in cross-cultural communications. When it comes to translations between an Asian and a European language, the problems are, without doubt, magnified tremendously. There are at least two major reasons for such cross-cultural misunderstandings: 1) often, it is impossible to identify a semantically 100% equivalent culturally-specific concept (CSC) for the other culture at hand; and 2) prior-knowledge of CSCs that is deeply rooted in one's own cultural background such as traditions, social systems etc. inherently influences any inter-cultural communication scenario. For example, if an Asian and a European are discussing about the educational system in general, their prior knowledge in this communication is deeply influenced by their respective prior knowledge about how educational systems are functioning in their respective coun-

adfa, p. 1, 2011.

tries. Hence, the question might immediately arise on how we can possibly improve the transfer of the original conceptual meanings of Source Language (SL) concepts to a Target Language (TL) audience, when there are no 100% equivalent concepts existing in the two cultures in consideration?

One of the major multilingual ontology frameworks, the MONNET project on Multilingual Ontologies for Networked Knowledge[1], addresses this problem by defining "multilingual ontology localization" based on different dimensions: *international (standardized) vs. culturally influenced domains; functional vs. documental localization; and interoperable vs. independent ontology* [4]. Within the MONNET framework, Declerck et al. [5] employ a standardized domain-specific ontology (i.e. XBRL taxonomies) linked with linguistic information from the domain lexicon and Aguado-de-Cea et al. [6] addresses the multilingual problems in culturally influenced domains by focusing on interoperability among geospatial resources in different natural languages.

Our work here addresses these issues based on pragmatic and cognitive theories: the Relevance Theory of Communication [7] and the Knowledge Effects involved in category-based inductions [8] discussed in [9]. The main idea in these theories is that a symmetric choice of code and context is unrealistic in cross-cultural communication because the two communicating parties are unlikely to share identical cognitive environments [7]. If a new object existing in a Source Language (SL) culture is introduced to an audience in a Target Language (TL) culture, the audience compares the new object with something it knows in advance. This implies that feature-based asymmetric similarity measures play an important role when bridging semantic knowledge existing in two cultures. In [10] and [11], potentially applicable feature-based similarity measures that could be used for mapping *independent ontologies* in a *culturally influenced domain* are compared based on qualitative analyses. These analyses identify that the Bayesian Model of Generalization (BMG), a novel cognitive model proposed by [3] is the most intuitive and effective measure of mapping CSCs in two cultures.

Accordingly, our standpoint here is to identify what kind of taxonomical datasets is suitable for the application of the BMG. The application of the BMG requires highly appropriate datasets consisting of CSCs and their features. In [10], the BMG was applied to datasets obtained from a semi-automatic feature-based ontology construction method known as Terminological Ontology (TO) by Madsen et al. [1]. The results from [10] indicate that particularly strict rules for constructing TOs may risk causing the elimination of important features. It means that the original TO-approach may require a more flexible taxonomic organization of feature structures.

Inspired by the previous works, we investigate in [12] how the Infinite Relational Model (IRM) [2], a novel unsupervised machine learning method, can be applied to loosely-structured datasets consisting of CSCs and features that are manually extracted from text corpora. In [12], three strategies have been tested in the experiments: **1)** applying the IRM directly to two CSC-feature matrices respectively representing the educational domain knowledge in Japan and Denmark for first categorizing them into

---

[1] http://www.monnet-project.eu

categorical classes that are to be afterwards compared and aligned; **2)** applying the IRM directly to a matrix where the two CSC-feature matrices respectively representing the Danish- and Japanese educational domain knowledge are merged; and **3)** applying the BMG [3] to directly compute similarity relations between CSCs in the two cultures, thereafter applying the IRM for clustering CSCs in the respective cultures into categorical classes. The results from the three experimental strategies indicate that the third strategy - the combination of the BMG and IRM seems to be the most effective approach for not only clustering CSCs into more specific and appropriate categorical classes but also for capturing complex relationships between each of the categorical classes existing in the two cultures. In addition, the direct application of the BMG to the datasets enables one to effectively analyze further specific similarity relations between category members existing in the two cultures. These results imply that loosely-structured datasets are sufficient for mapping CSCs in a multilingual context.

Interestingly, Kemp et al. [2] points out that researchers who start with the viewpoint of - *how the semantic knowledge should be represented as a system of relations* - often devise complex representational schemes. On the contrary, Kemp et al. [2] challenges this issue from the viewpoint of - *how representations of semantic knowledge are acquired*. This viewpoint of [2] and the results obtained from [12] motivate us to further analyze what are the advantages and disadvantages of representing semantic knowledge in accordance with rather strict principles of constructing ontologies for the purpose of cross-cultural knowledge transfer. Accordingly, in this work, CSC mapping results obtained from the application of the BMG to datasets obtained from TOs (the TO+BMG approach) is compared against the direct application of the BMG to datasets obtained from text corpora, of which results are subsequently clustered by the IRM application (the BMG+IRM approach).

In the next section, we first review our previous work applying the BMG+IRM approach in [12], followed by the experimental settings of applying the TO+BMG approach in Section 3. Section 4 analyzes results obtained from the experiment followed by discussion and future perspective in Section 5 and conclusions in Section 6.

2 The BMG+IRM approach

2.1 Datasets creation

The data source used for the BMG+IRM approach in [12] consists of educational terms and their definitional features that are manually extracted from text corpora respectively describing the Japanese and the Danish educational system. The Japanese corpora are published by Center for Research on International Cooperation in Educational Development, University of Tsukuba, Japan and the Japanese Ministry of Education, Culture, Sports, Science and Technology. The Danish corpora are published by the Education, Audiovisual and Culture Executive Agency under the EU commission. These corpora are all written in English and it is feasible to identify original expressions of educational terms in the respective languages from existing parallel- or content aligned corpora. This enables us to eventually achieve translation between Japanese CSCs and Danish CSCs through the English term mapping.

CSCs and their definitions all written in English are manually extracted from the text corpora, e.g. the Danish CSC "municipal school (DA: *folkeskole*)" and its definition "*a comprehensive school covering both primary and lower secondary education, i.e. one year of pre-school class, the first (grade 1 to 6) and second (grade 7-9/10) stage basic education, or in other words it caters for the 6-16/17-year-olds*". From this definition, we create a feature list consisting of "comprehensive school" "primary and lower secondary education" "basic education" "targeted for 6-16/17 years olds". This definition also implies that "municipal school" is categorized into three sub-CSCs "preschool class", "first stage" and "second stage" respectively having their features "one year preschool education" "1-6 grades" "7-9/10 grades". These sub-CSCs are supposed to inherit features defined in the subordinate CSC, in this case "municipal school". In this way, 59 Danish CSCs and 54 Japanese CSCs and their features all written in English are listed. In addition, some features are manually aligned, e.g. the feature "continuing education for adults" in Denmark is aligned with the feature "opportunities for life-long learning" in Japan. Finally, in order to handle feature values which the system does not inherently handle, the following operations are manually implemented: **a)** If a feature value in one country is completely included in a feature value in the other country (e.g. the feature "6-12 y.o." in Japan is completely included in the feature "6-17 y.o." in Denmark), a CSC possessing the feature that includes the other feature (a CSC possessing "6-17 y.o." should also possess "6-12 y.o."; and **b)** If two features from the respective countries are partly overlapping (e.g. "13-15 y.o." in Japan and "14-17 y.o." in Denmark are partly overlapping), a artificial feature referring to the exact overlapping range (e.g. "14-15 y.o.") is created. In this example, a Japanese CSC that possesses "13-15 y.o." should also possess the artificial feature "14-15 y.o.". In the same way, the Danish CSC that possesses "14-17 y.o." should also possess the artificial feature "14-15 y.o.".

Accordingly, in total 229 features are registered in the two matrices respectively representing the Danish and Japanese educational systems. In each matrix, if a feature is possessed by a CSC, the numeric value "1", otherwise "0" is assigned. In these matrices, the Danish and Japanese CSCs are respectively denoted as D_i and J_j and feature IDs are assigned as f_k. Both the Danish- and Japanese CSCs and their features are alphabetically registered. To these datasets, the BMG is first applied, and thereafter the IRM is applied to the obtained BMG results for simultaneously clustering respective Danish- and Japanese CSCs based on CSC-CSC links which are identified by the BMG.

2.2 Bayesian Model of Generalization (BMG)

The BMG [3] is a cognitive model, which uniquely unifies two classically opposing approaches to similarity and generalization: Tversky's set theoretic model of similarity [13] and Shepard´s continuous metric space model of similarity and generalization [14]. The philosophical background of BMG is reviewed in [9]. A key point in BMG is to compute the conditional probability that y falls under C (Consequential Region) given the observation of the example x based on the following formula:

$$P\,(y \in C|x) = 1/[1 + \frac{\sum_{h:x\in h, y\notin h} p(h,x)}{\sum_{h:x,y\in h} p(h,x)}] \tag{1}$$

The consequential region C in our work indicates the categorical region where a new object y belongs. In equation (1), a hypothesized subset h is defined as the region where a concept belongs to h, if and only if, it possesses feature k [3]. It means that, in our work, y is considered as a newly encountered object existing in a SL domain when y is introduced to a TL audience and the TL audience compares this new object y with an observed data x which is part of his/her prior knowledge (referent dataset).

Another unique point in the BMG is that $P(h, x) = P(x|h)P(h)$ in equation (1) represents the weight assigned to the consequential subset h in terms of the example x. This can be achieved by specifically assigning the weight $P(h, x)$ based on the strong sampling scheme defined in [3] as follows:

$$P(x\,|h) = \begin{cases} 1/|h| & if\ x \in h \\ 0 & otherwise \end{cases} \tag{2}$$

Here, $|h|$ indicates the size of the region h [3]. In our work, the number of objects possessing the k^{th} feature in the referent dataset is considered as the size of the region h. [3] explains that the prior $P(h)$ is not constrained in their analysis so that it can accommodate arbitrary flexibility across contexts. Hence in this work, $P(h) = 1$.

2.3 Infinite Relational Model (IRM)

According to [2], a key feature of the IRM is to *automatically choose an appropriate number of clusters using a prior that favors small numbers of clusters, but has access to a countably infinite collection of clusters*. In [2], the observed data are considered as m relations involving n types. The experiments performed in [12] apply the simplest model: dealing with two types with a single two-place relation R: T_1 x $T_2 \rightarrow \{0, 1\}$. More specifically, in the BMG+IRM approach shown below, T_1 and T_2 respectively corresponds to Danish CSCs and Japanese CSCs.

The principle of generating clusters in IRM, according to [2], is based on a distribution over partitions induced by a so-called Chinese Restaurant Process (CRP) [15]. The CRP starts a partition process with a single cluster containing a single object. The i^{th} object has possibilities to belong to either of the two options in equation (3). Hence, the distribution over clusters for object i, conditioned on the cluster assignments of objects 1, …, i-1 is defined as:

$$P(z_i\ = a|z_1, \dots, z_{i-1}) = \begin{cases} \gamma/(i-1+\gamma) & a \text{ is a new cluster} \\ n_a/(i-1+\gamma) & n_a > 0 \end{cases} \tag{3}$$

Here, n_a is the number of objects already assigned to cluster a, and γ is a parameter [2]. The CRP continues until all the objects belong to clusters. [2] explains that *the distribution on z induced by the CRP is exchangeable: the order in which objects are* *assigned to clusters can be permuted without changing the probability of the resulting partition. P(z) can therefore be computed by choosing an arbitrary ordering and multiplying conditional probabilities. Since new objects can always be assigned to new clusters, the IRM effectively has access to a countably infinite collection of clusters.*

From the clusters generated by the CRP, relations are generated based on the following generative model:

- For the cluster assignment of objects
  $z \mid \gamma \sim \mathrm{CRP}(\gamma)$
- For link probability between clusters
  $\eta(a, b) \mid \beta \sim \mathrm{Beta}(\beta, \beta)$
- For links between objects
  $R(i, j) \mid z,\eta \sim \mathrm{Bernoulli}(\eta(z_i, z_j))$

In the above generative model, we set parameters β=1, and $\gamma=\log(J^{(i)})$ where $J^{(i)}$ is the number of concepts in the i'th mode. In here, relationships are assumed to be conditionally independent given cluster assignments [2]. The eventual purpose of the generative model is to identify a cluster z that maximizes $P(z \mid R)$. Based on the generative model defined above, relations from clusters are generated by:

$$P(R \mid z, \eta) = \prod_{ab} (\eta_{ab})^{m^{+}_{ab}} (1 - \eta_{ab})^{m^{-}_{ab}} \tag{4}$$

where m^{+}_{ab} refers to the total number of links between categorical classes a and b; and m^{-}_{ab} refers to the total number of non-links between categorical classes a and b. From the conjugate prior η_{ab} defined above, η_{ab} is integrated out in the following:

$$P(R \mid z) = \prod_{a,b\in \mathrm{N}} \frac{Beta(m^{+}_{ab}+\beta, m^{-}_{ab}+\beta)}{Beta\ (\beta,\beta)} \tag{5}$$

From formulae (3) and (5), the IRM identifies the z that maximizes:
$P(z|R) \propto P(R|z)\, P(z)$. According to [2], the expected value of η_{ab} given z is:

$$\frac{m^{+}_{ab}+\beta}{m^{-}_{ab}+m^{+}_{ab}+2\beta} \tag{6}$$

The procedure for the inference is further described in [16]. The solutions displayed in the following are based on the sample with highest likelihood.

2.4 Brief reviews of the CSC mapping

As reported in [12], 59 Danish CSCs and 54 Japanese CSCs are simultaneously clustered into 11 and 9 categorical classes by applying first the BMG and subsequently the IRM. In Figure 1, the unsorted graph shows existing links between the Danish CSCs and the Japanese CSCs identified by the computation of the BMG. It means that each dot represents a link established between a Danish CSC and a Japanese CSC, when they share at least one common feature. The upper right graph in Figure 1 shows the graph sorted according to extracted assignments of categorical classes computed by the IRM algorithm. The bottom left- and right graphs show the distribution of concepts over the extracted categorical classes, respectively for the Danish- and Japanese CSCs. The bottom center graph corresponds to the graph sorted according to extracted assignments of categorical classes, which indicates the density of relationships between a Danish- and a Japanese categorical class.

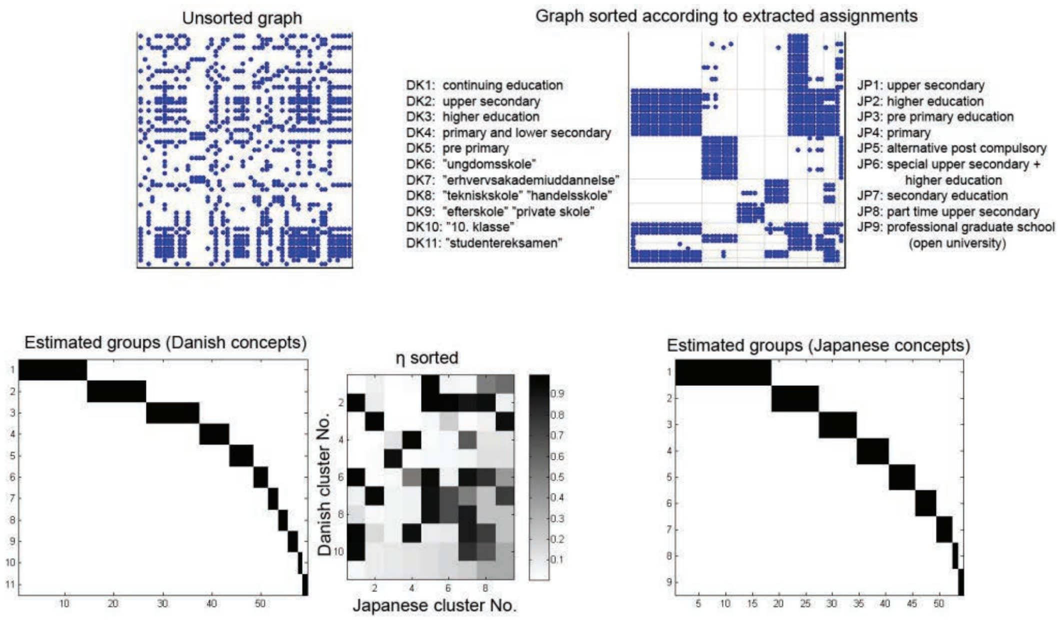


Fig. 1. IRM clustering based on the BMG+IRM approach (adapted from ref. [12])

The results in Figure 2 show cluster members that constitute each categorical class. Almost all members in each categorical class in Figure 2 are grouped together with other members that are supposed to belong to the same categorical class. For example, some CSCs such as the Japanese "J3: college of technology (JP: koto-senmon-gakko)" and the Danish "D36: municipal school (DK: folkeskole)" are CSCs that are difficult to categorize in a multi-cultural context. Compared with the other two experimental strategies conducted in [12], these CSCs (J3 and D36) are respectively grouped into a more specific and independent categorical class, i.e. the Japanese categorical class 6 and the Danish categorical class 4. One of the noteworthy points is that, when observing the η-sorted graph in Figure 1, it is possible to study more complex relationships of categorical classes in a cross-cultural context. For instance, the Japanese categorical class 6 where J3 belongs, has a strong relationship with the Danish categorical class 2 "upper secondary" class, but also has a little weaker relationship with both classes 7 and 8, which respectively represent "vocational academy" and "vocational college" categories providing a 2 years post-secondary degree in Denmark. The observation of the η-sorted graph in Figure 1 further provides a clear picture of how each country-specific categorical class is related to categorical classes existing in another country in a very complex and comprehensible manner. This kind of overview of how categorical classes in different cultures are inter-related is highly useful and valuable not only for mapping CSCs but also for constructing ontologies in a multi-cultural context.

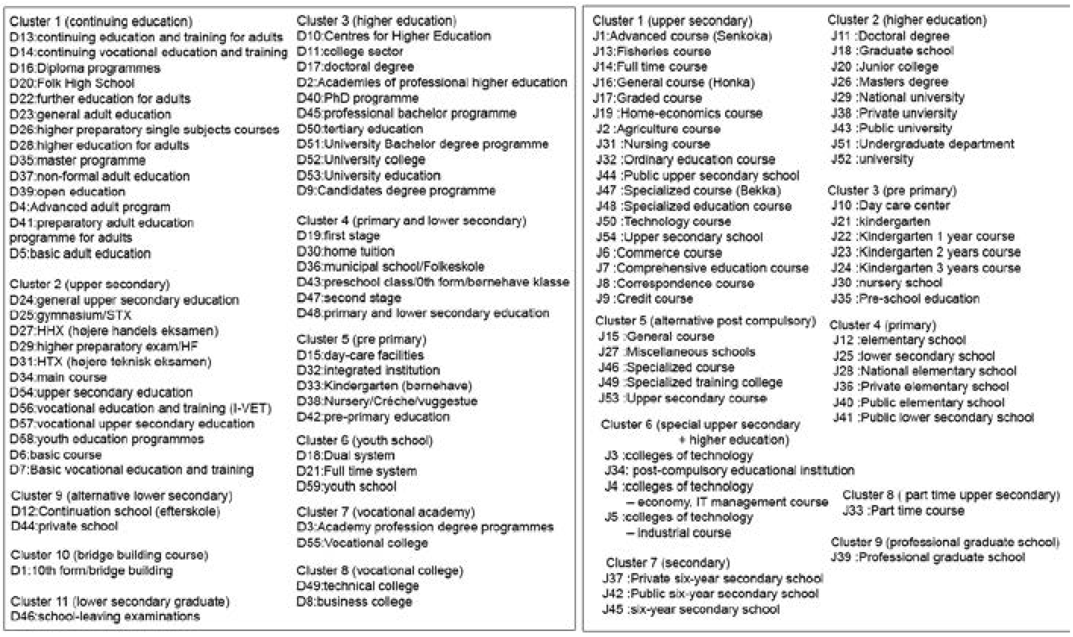


Fig. 2. CSC members that constitute each categorical class [Left: Danish / Right: Japanese] (adapted from ref. [12])

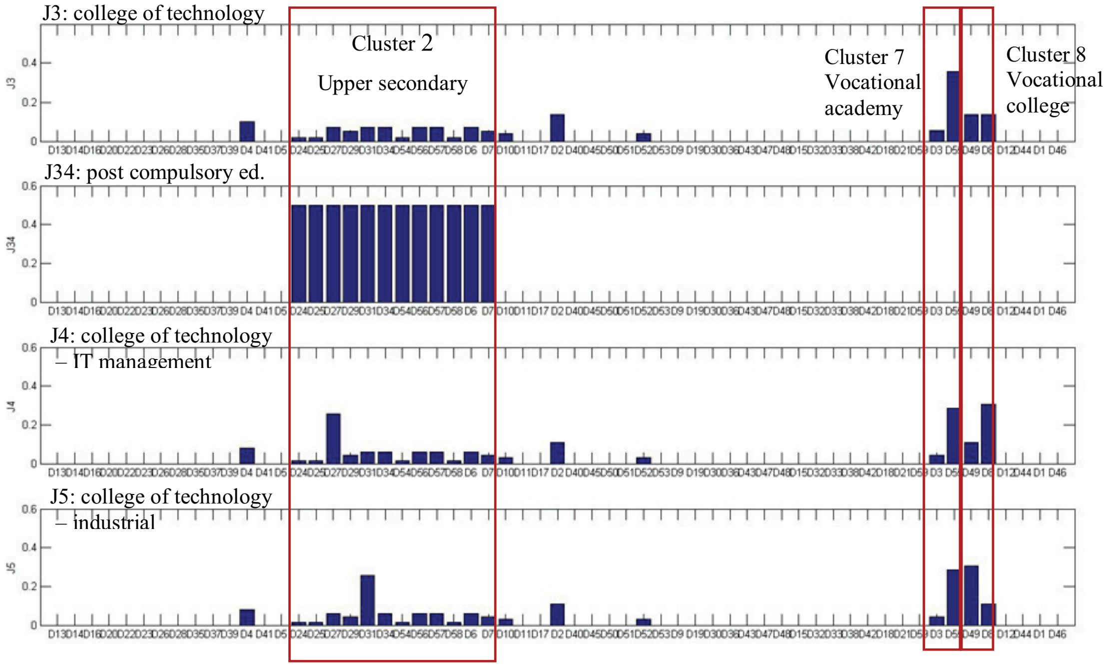


Fig. 3. BMG similarity relations between the category members of Japanese categorical class 6 and all the Danish concepts (adapted from ref. [12])

Another advantage of the BMG+IRM approach is that the direct application of the BMG enables us to analyze further specific similarity relations between category members of the respective categorical classes existing in the two cultures. Figure 3 illustrates how the category members of the Japanese categorical class 6 are related with the category members of the Danish categorical classes 2, 7 and 8. As discussed in the above, the η-sorted graph in Figure 1 shows that the Japanese categorical class 6 has the strongest relationship with the Danish categorical class 2 and slightly weaker relationship with the Danish categorical classes 7 and 8. Figure 3 explains these relationships between the classes by showing that all the category members of the Danish class 2 share at least one feature with all the category members of the Japanese class 6, while only 75% of the category members of the Danish classes 7 and 8, respectively, share at least one feature with 75% of the category members of the Japanese class 6. On the other hand, when observing individual relationships of category members between the Japanese- and the Danish classes, similarity relationships are not necessarily strong in most of the combinations. Here, the selection of feature-based similarity measures plays in to the considerations.

As explained in Section 2.2, equation (1) computes the conditional probability that a new observed object y falls under a consequential region C given the learner's prior knowledge that have already been observed as x. It means that, in relation to Figure 3, a scenario could be that a Japanese who has prior knowledge of the Japanese educational system is learning the Danish educational system by comparing similarities of individual Danish educational CSCs. Thus, all the knowledge about Japanese CSCs is considered as x and the individual Danish CSC as y in equations (1) and (2). In other words, it can be interpreted that all definitional features possessed by a Japanese CSC are considered as prior knowledge of the Japanese learner and can act as noise (or cultural bias) if a feature is not possessed by a Danish concept taken in comparison. In addition, the uniqueness of the BMG is to reflect the importance of features for the similarity computation by considering features that are possessed by many concepts as less important and features that are possessed by fewer concepts as more important and decisive. In case of "J5: college of technology – industrial course", three Danish CSCs, "D31: HTX (upper education specialized in technical and natural science)", "D55: vocational college (educational institution offering vocationally-oriented upper- and 2 year post-secondary education)", and "D8: technical college (vocational college specialized in technical and natural science)", are identified as the most similar concepts. In the respective cases, one or a few decisive features such as "specialized in technical and natural science", "offering 2 year post-secondary degree" strongly influence the similarity computation and differentiate the confidence levels of similarity from other concepts. Another interesting pattern indicated in Figure 3 is that a very abstract concept "J34: post-compulsory educational institution" likely covers all category members of the Danish categorical class 2 as the most similar concepts identified in the Danish culture. This is because J34 possesses only two features, which implies that no other distinctive features influence as noise in the similarity computation. In this way, the combination of the BMG and the IRM can possibly be used for first capturing the abstract relationship between categorical classes and next for further analyzing the individual similarity relationship between CSCs in order to achieve more fine-grained CSC-mappings.

3 The TO+BMG approach

3.1 Data sources

As described in Section 1, the main purpose of this work is to compare the TO+BMG approach against the aforementioned BMG+IRM approach. In order to identify weaknesses and strengths of the TO+BMG approach, a strict experimental setting has been defined. It means that TOs should be constructed from exactly the same data - consisting of 59-Danish/54-Japanese CSCs and 229 features - employed in the BMG+IRM approach. This is indeed not a typical procedure for terminological ontologists, because the TO usually employs information about concepts identified from text corpora or in discussions with specialists in a particular subject field. Consequently, it turns out to be an extremely difficult task for human to overview relations among concepts and feature values from the alphabetically listed data. Accordingly, information about the categorical classes identified by the BMG+IRM approach and their relations to features identified from the categorically sorted data in [12] is used as a hint of constructing TOs in accordance with the rules and principles explained in the following.

3.2 Terminological Ontology (TO)

The principles of Terminological Ontologies have been developed in the research and development project called CAOS - Computer-Aided Ontology Structuring - whose aim has been to develop a computer system designed to enable semi-automatic construction of ontologies [1]. The uniqueness of TO is given by its feature specifications and subdivision criteria [1]. A feature specification is presented as attribute-value pair - for example as shown in Figure 4, [FIELDS: technology]. Thus, a representation of a whole concept is a feature structure, i.e. a set of feature specifications corresponding to the unique set of characteristics that constitutes that particular concept [1]. In Figure 4, each box that represents a particular concept is divided into three layers: **1)** top layer, lexical representation (term), **2)** middle layer, dimension specifications, and **3)** bottom layer, feature structure (set of feature specifications).

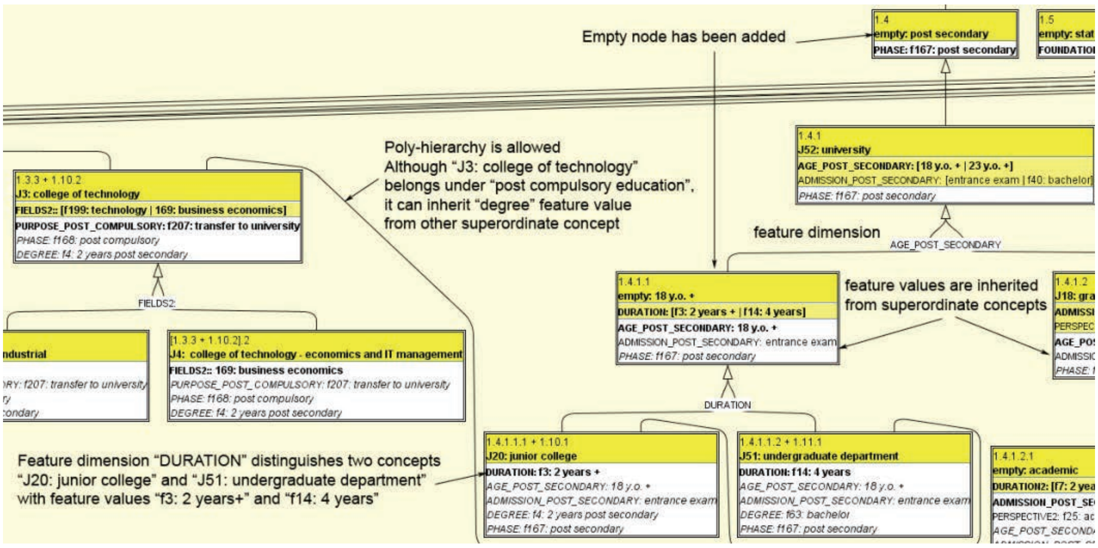


Fig. 4. Terminological Ontology

The use of feature specifications is subject to principles and constraints described in detail in [1]. Most importantly, a concept automatically inherits all feature specifications of its superordinate concepts. This approach is fairly intuitive and reasonably consistent with the hierarchical structure of categories that are generally discussed by cognitive scientists [1]. However, the principles of TO also defines more strict rules derived from the traditional view of terminology that aims at proper standardization. For example, the principle of uniqueness of dimension defines that *a given dimension may only occur on one concept in an ontology*. [1] argues that *uniqueness of dimensions contributes to create coherence and simplicity in the ontological structure, because concepts that are characterized by means of a certain common dimension must appear as descendants of the same superordinate concept.* In the same way, [1] also defines the uniqueness of primary feature specifications as *a given primary feature specification can only appear on one of the daughters*. The argument is that these uniqueness principles *make it possible to a certain extent to carry out automatic placing of concepts into an ontology.* Another important principle is that the TO approach allows polyhierarchy structures so that one concept may be related to two or more superordinate concepts.

Following the aforementioned principles, TOs representing the Danish- and the Japanese educational systems have been constructed. It means that, although some feature specifications are considered as very important in several categorical classes, the TO principles allow us to use them only once in the ontology. For example, features defining the academic degree system such as "f40: Bachelor", "f140: Master" are very common and decisive features that could be useful for mapping CSCs existing in different cultures. However, although there are two types of Bachelor or Master degree programmes provided as "university education" under "tertiary education" and as "open education" under "continuing education for adults" in the Danish educational system, the TO principles allow us to use this only once in the ontology. As a solution, terminological ontologists argue that this problem is solved by creating empty concepts (nodes) respectively possessing attribute-value pairs "degree: Bachelor" and "degree: Master" at a higher level of the ontology, from which the two "Bachelor" concepts belonging to different superordinate concepts can inherit "degree: Bachelor" in the polyhierarchical structure. Another important issue is as described above that the TO principles do not allow us to define the same dimension in several places in an ontology and to have conflicting attribute-value pairs in a hierarchy. It means that the number of features that are listed in an ontology by these strict TO principles is rather limited, whereas the BMG+IRM approach enables us to enlist all features identified in the text corpora without any constraints. Due to these strict TO principles, the number of CSCs for the Danish- and Japanese educational system, respectively, decrease to 47 and 50. Instead, 22 and 14 empty nodes have been added to the respective datasets. The number of features is also reduced from 229 to 99 based on the TO principles.

4 Results

Figure 5 shows the respective results of the BMG similarity computation obtained from the TO+BMG approach in the left column and from the BMG+IRM approach in the right column. Each graph can be interpreted in the same way as Figure 3 has been interpreted. It means that in relation to Figures 5, a scenario could be that a Japanese who possesses prior knowledge of the Japanese educational system is learning the Danish educational system by comparing similarities of individual Danish educational CSCs. The effects of feature weights also work in the same way as indicated by Figure 3. Thus features that are possessed by many Japanese CSCs are considered as less important and features that are possessed by fewer Japanese CSCs are considered as more important and decisive.

The first thing to notice in Figure 5 is that, since equation (1) of the BMG computes only one feature "f75" (compulsory) possessed by the Japanese "empty-compulsory" CSC, the "empty-compulsory" CSC in the TO column identifies six Danish CSCs with the similarity score 1.0. Secondly, the TO+BMG approach identifies "D19: first stage (primary)" as the most similar CSC to the Japanese "J12: elementary school" CSC, while the BMG+IRM approach identifies the four Danish CSCs: "D19: first stage (primary)", "D36: private elementary school", "D44: private school", "D48: single structure education". This indicates that the TO+BMG approach precisely identifies the Danish CSC which represents the exactly same target range of age, while the BMG+IRM approach identifies all the Danish CSCs which provide the primary education to children in the age range: 7-12 years old. When inspecting the feature structure of these CSCs in Table 1, it turns out that the TO+BMG approach computes similarities with Danish CSCs that possess *either* the feature "7-12" (years old) *or* "f75" (compulsory). Since f75 is possessed by many more Japanese CSCs, the feature weight is reduced based on equation (2) of the BMG. On the other hand, since the feature "7-12" is possessed by fewer Japanese CSCs, this feature is weighted higher in the computation. Accordingly D19 is identified as the most similar CSC, despite the fact that D19 shares only one feature "7-12" with J12. On the other hand, the BMG+IRM approach considers any of the features:

“7-12”, “f65” (basic education) and “f75” (compulsory) that are possessed by Danish CSCs. Hence, the Danish CSCs that possess *all* three features are identified as the most similar CSCs to J12. A key point that should be noticed in Figure 5 is that the Danish CSCs in the TO do not simultaneously possess the two features “7-12” and “f75”. This is due to the strict principle of TO that *a given dimension may only occur on one concept in an ontology* [1]. The third example is that the TO+BMG approach identifies “D12: continuing school” which is targeted for 14-17 years old children, and “D44: private school” as the most similar concepts to “J36: private elementary school”. In this case, the TO+BMG approach considers the Danish CSCs that possess the feature “f190” (private) together with “f75” (compulsory) as much more relevant CSCs than D19 which possesses the feature “7-12”. A problem here is that the attribute of the value “f190” (private) overshadows the age dimension of “7-12”. Thus the meaning of “elementary” in “J36: private elementary school” is not appropriately reflected in the similarity computation. On the contrary, the BMG+IRM approach allows features that describe a specific CSC to be registered unlimitedly. Accordingly, D44 which possess both features “7-12” and “f190” together with “f75” are considered more similar than D12 which possess only “f190” and “f75”.

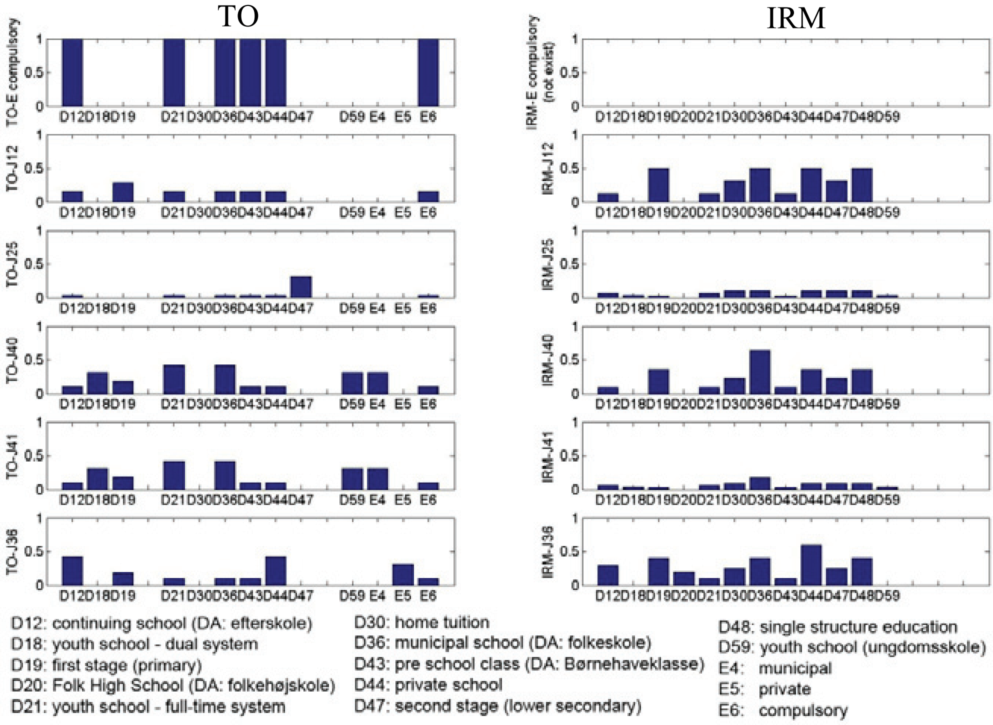


Fig. 5. Comparison of results obtained from the TO+BMG and from the BMG+IRM J12: elementary school, J25: lower secondary school, J40: public elementary school, J41: public lower secondary school, J36: private elementary school

Table 1. List of feature values

	TO	IRM		TO	IRM
E	F75	N/A	D12	14-17,16-17, f75, f174, f190	14-15, 14-17, 16-17, f18, f54, f75, f173, f174, f186, f190
J12	7-12, f17, f36, f75	7-12, f17, f36, f65, f75, f82	D19	7-12, f174	7-12, f65, f75, f122, f174
J25	13-15, f10, f75, f108	13-15, 14-15, f10, f24, f36, f59, f60, f75, f84, f85, f108, f134, f187, f189	D36	f75, f143, f174	13-15, 6-16, 7-12, f65, f75, f117, f143, f174
J40	7-12, f17, f36, f75, f143	7-12, f17, f36, f65, f75, f82, f143	D44	f75, f174, f190	13-15, 16-17, 6-17, 7-12, f65, f75, f174, f184, f190
J36	7-12, f17, f36, f75, f190	7-12, f17, f36, f65, f75, f82, f190	D48	f174	13-15, 6-16, 7-12, f65, f75

5 Discussions

The results presented in Section 4 indicate that the TO+BMG approach, due to its strict principles, inherits a disadvantage of delimiting the registration of features that are potentially important for the multi-cultural CSC mapping. Terminological ontologists usually claim that the strict principles are an advantage for *making it possible to a certain extent to carry out automatic placing of concepts into an ontology* [1]. It may truly be applicable when the TO is used for sharing knowledge in a specific knowledge community. In addition, it should be emphasized that one of the advantages of TO is the clarification of domain knowledge in a specific knowledge community and the ability to compare the *hierarchical structures* across different communities. For example, the TO approach enables us to generate a new concept in one community which exists in the other community. Such concept generation is not feasible based on the similarity computation approach, i.e. the BMG+IRM approach. However, it should be noticed that, as stated in Section 1, our standpoint in this work has been to identify what kind of taxonomic datasets is suitable for the application of the BMG whose purpose is to deal with the problem of cross-cultural communication that are not symmetrically coordinated [7]. The CSC mapping in a multi-cultural context is not the same as a mono-cultural knowledge sharing. A dimension which is considered important in one culture may not be equally important in the other culture. By eliminating some feature attributes and their values due to the strict rules, there is a substantial risk of eliminating important nuances hidden behind a concept. In this respect, the BMG which is a cognitive model of generalization is designed to compute such fuzzyness of the human mind by reflecting prior knowledge of a learner who compares a new object with something he/she knows in advance.

As [2] points out, the IRM challenges the problem of knowledge representation from the viewpoint of - *how representations of semantic knowledge are acquired.*

This viewpoint implies that, instead of directly challenging *how the semantic knowledge should be represented as a system of relations,* it may be an idea to integrate the ontology learning model, i.e. the IRM, originating from the cognitive sciences as a pre-processing to the ontology construction. This has been empirically indicated in this work where information about the categorical classes identified by the BMG+IRM approach and their relations to features identified from the categorically sorted data in [12] was indeed very useful for constructing TOs. In addition, as shown in Figure 1, the IRM can provide an image of how the Danish- and the Japanese categorical classes are inter-related. This kind of information could be highly useful if we are going to develop culturally-specific ontologies for the respective cultures which are inter-operable in a multi-cultural context. This implies that, for example, the features that influence the formation of categorical classes in both cultures could be prioritized as necessary feature dimensions when constructing TOs. This may prevent eliminating important features that could be used for computing similarities based on the BMG. An attractive aspect of the IRM is that the model can be a more complex clustering of three or more relations simultaneously. Hence, this could in principle be applied for multi-cultural modeling as well. These results imply that the integration of all methods, i.e. the BMG+IRM+TO approach, could be our future potential that enables us not only to map CSCs by respecting nuances of each concept existing in each respective culture, but also to construct TOs that are cross-culturally interoperable as well as mono-culturally clarified.

6 Conclusions

In this paper, two approaches of mapping CSCs are compared. The results show that the TO+BMG approach, due to its strict principles, has a disadvantage of delimiting dimensions that are potentially important for the CSC mapping. On the contrary, the BMG combined with the IRM, both of which originate from the cognitive sciences, is designed to compute the fuzzyness of human mind by reflecting prior knowledge of a learner who compares a new object with something he/she knows in advance. By considering the advantage of the TO that is the clarification of domain knowledge in a mono-cultural knowledge community, the BMG+IRM+TO approach may be an optimal solution, which may enables us not only to map CSCs by respecting nuances of each concept in the respective cultures, but also to construct TOs that are multi-culturally interoperable. This remains as a grand future challenge.

Acknowledgement: We would like to express our thanks to the DanTermBank project members, among others, Bodil Nistrup Madsen, Hanne Erdman Thomsen and Tine Lassen at the Copenhagen Business School for valuable and relevant discussions related to our work.

Reference

1. Madsen, B.N., Thomsen, H.E. and Vikner, C.: Principles of a system for terminological concept modelling. In: Proc. the 4th International Conference on Language Resources and Evaluation. ELRA, (2004).
2. Kemp, C., Tenenbaum, J.B., Griffiths, T.L. Yamada, T., Ueda, N.: Learning systems of concepts with an infinite relational model. In: Proc. the 21st National Conference on Artificial Intelligence. (2006)
3. Tenenbaum, J. B., & Griffiths, T. L.: Generalization, Similarity, and Bayesian inference. In: Behavioral and Brain Sciences, Vol.24(4; 4), (2001)
4. Cimiano P., Montiel-Ponsoda E., Buitelaar P., Espinoza M., Gómez-Pérez A.: A Note on Ontology Localization. In: Journal of Applied Ontology Vol. 5, No. 2, IOS Press, (2010) 127-137.
5. Declerck T., Krieger H.U., Thomas S.M., Buitelaar P., O'Riain S., Wunner T., Maguet G., McCrae J., Spohr D., Montiel-Ponsoda.E.: Ontology-based multilingual access to financial reports for sharing business knowledge across Europe. In: Rooz, J., Ivanyos, J. (Eds.) Internal Financial Control Assessment Applying Multilingual Ontology Framework, Budapest: HVG Press (2010) 67-76.
6. Aguado-de-Cea G., Gómez-Pérez A., Montiel-Ponsoda E., Vilches-Blázquez L.M.: Rivière or Fleuve? Modelling Multilinguality in the Hydrographical Domain. In: Proc. Workshop on the Multilingual Semantic Web at the 19th International World Wide Web Conference, USA (2010)
7. Sperber, D., Wilson, D.: Relevance: Communication and Cognition. Blackwell, Oxford (1986)
8. Murphy,G. L.: The Big Book of Concepts. The MIT Press Cambridge, Massachusetts (2004)
9. Glückstad F.K.: Bridging Remote Cultures: Influence of cultural prior-knowledge in cross cultural communication, In: Proc. The 26th Annual Conference of the Japanese Society for Artificial Intelligence, Japan, (2012)
10. Glückstad F.K., Mørup M.: Feature-based ontology mapping from an information receivers' viewpoint, (under review)
11. Glückstad F.K.: Cross-Cultural Concept Mappings of Standardized Datasets, (under review)
12. Glückstad F.K., Mørup M.: Application of the Infinite Relational Model combined with the Bayesian Model of Generalization for effective cross-cultural knowledge transfer. In: Proc. The 26th Annual Conference of the Japanese Society for Artificial Intelligence, Japan, (2012)
13. Tversky, A.: Features of similarity. In: Psychological Review, Vol., 84(4; 4), (1977) 327-352.
14. Shepard R.N.: Towards a universal law of generalization for psychological science. In: Science, 237 (1987).
15. Pitman J.: Combinatorial stochastic processes. Notes for Saint Flour Summer School, (2002)
16. Mørup M., Madsen K.H., Dogonowski A.M., Siebner H., Hansen L.K.: Infinite Relational Modeling of functional connectivity in resting state fMRI. In: Neural Information Processing Systems (2010)

Chapter 6.

Discussions

1. Summary of the empirical results

The raised key research question in this thesis has been: **How should the background knowledge possessed by an SL communicator and a TL receiver be represented and linked in light of various cognitive processes involved in cross cultural communication?**

As previously explained, my hypothesis started by the assumption that the TO method (Madsen et al., 2004) could potentially be used for representing prior-knowledge of the culturally-specific domain knowledge, which is required for the inference process in a cross-cultural human communication scenario. I investigated this research issue from the view of two dimensions: **a)** what algorithms are suitable for mapping TOs? and **b)** within the defined framework of COM, what are the advantages or disadvantages of using the TO approach compared with other approaches for systematically representing and organizing conceptual features as culturally-dependent prior-knowledge?

In order to answer the first sub-question posed in a), I considered several cognitive models originating from the categorization and concept learning research as potential ontology mapping algorithms, which reflect the inference process involved in cross-cultural communication. I implemented several preliminary studies of mapping two culturally-dependent TOs. The empirical studies started from a very primitive and manual feature-mapping approach, gradually applying a rather basic cognitive model, Tversky's Ratio model (Tversky, 1977) by assigning three different patterns of parameter settings, and eventually identifying a further advanced algorithm that has been well-recognized as a novel cognitive model, the Bayesian Model of Generalization (BMG) (Tenenbaum & Griffiths, 2001) as the best possible algorithm for the COM framework. The fourth and fifth empirical studies in Chapter 5 indicated that the BMG with the "strong sampling scheme" sensitively reflects the feature structure of concepts in comparison, and both intuitively and effectively identified the most similar concepts by reflecting on the learner's prior knowledge.

For addressing the second sub-question posed in b), I implemented two types of comparative analyses by applying the Infinite Relational Model (IRM) (Kemp et al., 2006), a novel unsupervised machine learning method, to loosely structured datasets consisting of CSCs and features that are manually extracted from text corpora. In the sixth empirical study in Chapter 5, three strategies of applying the IRM have first been tested: **1)** applying the IRM directly to two CSC-feature matrices, respectively representing the educational domain knowledge in Japan and Denmark for first categorizing them into categorical classes that were to be subsequently compared and aligned; **2)** applying the IRM directly to a matrix where the two CSC-feature matrices, respectively representing the Danish- and Japanese educational domain knowledge were merged; and **3)** applying the BMG to directly compute similarity relations between CSCs in the two cultures, thereafter applying the IRM for clustering CSCs in the respective cultures into categorical classes (the BMG+IRM approach). The results from the three experimental strategies indicated that the third strategy - the BMG+IRM approach - seemed to be the most effective approach for not only clustering CSCs into more specific and appropriate categorical classes but also for capturing complex relationships between each of the categorical classes existing in the two cultures. In addition, the direct application of the BMG to the datasets enabled me to effectively analyze further specific similarity relations between category members existing in the two cultures. These results indicated that loosely-structured datasets seemed to be sufficient for mapping CSCs in a multi-cultural context. The seventh empirical study further analyzed what were the advantages and disadvantages of representing semantic knowledge in accordance with rather strict principles of constructing ontologies for the purpose of cross-cultural knowledge transfer. More specifically, CSC mapping results obtained from the application of the BMG to datasets obtained from TOs (the TO+BMG approach) was compared against the direct application of the BMG to datasets obtained from text corpora, of which results were subsequently clustered by the IRM application (the BMG+IRM approach). The results indicated that the TO+BMG approach, due to its strict principles, has a disadvantage of delimiting the dimensions that may potentially be important for the CSC mapping. Although some feature specifications are considered as very important in several categorical classes, the TO principles allow us to use them only once in the ontology. For example, features defining the academic degree system such as "bachelor" and "master" are very common and decisive features that could be useful for mapping CSCs existing in different cultures. However, although there are several types of Bachelor- or Master degree programmes provided as e.g. "university education" under "tertiary education" and as "open education" under "continuing education for adults" in the Danish educational system, the TO principles allow us to use this only once in the ontology. Terminological ontologists argue that this problem is solved by creating empty concepts (nodes) respectively possessing attribute-value pairs "degree: Bachelor" and "degree: Master" at a higher level of the ontology, from which the two "Bachelor" concepts belonging to different superordinate concepts can inherit "degree: Bachelor" in a poly-hierarchical structure. In addition, it should be emphasized that one of the advantages of TO is the clarification of domain knowledge in a specific knowledge community and the ability to compare the *hierarchical structures* across different communities. The aforementioned principle of poly-hierarchical structures enables us to generate a new concept in one community which exists in the other community. Such concept generation is not feasible based on the similarity computation approach, i.e. the BMG+IRM approach. On the contrary, the BMG combined with the IRM, both of which originate from the cognitive sciences, was designed to compute the fuzziness of the human mind by reflecting prior knowledge of a learner who compares a new object with something he/she knows in advance.

In the next section, I discuss how the aforementioned results can be interpreted in contrast to the Relevance Theory of Communication (Sperber & Wilson, 1986) and the Knowledge Effects (Murphy, 2004) which are naturally fostering the ground theories that support the COM framework.

2. Contrast to the Relevance Theory of Communication and the Knowledge Effects

What I aimed at demonstrating, by the series of empirical studies, is the simulation of asymmetric relations in cross-cultural communication between a communicator and a receiver. This has been achieved under the framework of COM that can be best described as a merger of the Relevance Theory of Communication (RTC) (Sperber & Wilson, 1986) and the so-called Knowledge Effects (Murphy, 2004).

The first point emphasized in the COM framework, based on the RTC and the Knowledge Effects, is that the symmetric choice of code and context is unfeasible to apply for a cross-cultural communication scenario, because the communicating parties are for obvious reasons unlikely to share identical cognitive environments. Due to the use of different languages, different conceptualization of society, different domain knowledge rooted in their respective cultures etc., the asymmetric coordination of code and context between the cross-cultural communicators comprise a significantly more realistic view (Sperber & Wilson, 1986).

Secondly, in the RTC, Sperber & Wilson (1986) raise two important dimensions in terms of *relevance*. In short, those are: **i)** how to assess the degree of relevance; and **ii)** how the contexts are chosen. In a way, these dimensions, respectively, correspond naturally to my posed sub research questions in this thesis: **a)** what algorithms are suitable for mapping TOs? ; and **b)** within the defined framework of COM, what are the advantages or disadvantages of using TO compared to other approaches for systematically representing and organizing conceptual features as culturally-dependent prior-knowledge?

In terms of question **i)** above, Sperber & Wilson (1986) argue that there are two factors to be taken into account for assessing the degree of relevance. The first factor is obviously the contextual effects which is new knowledge generated based on the union of new information *P* and background information *C*. From Murphy's viewpoint, our background knowledge is structured in our memories based on our past experiences that correspond to the background information *C*. When a new object *P* is observed, we compare and categorize it in relation to concepts stored in our memory as prior knowledge *C*, and outline a new form of concept *Q*. Hence, according to Sperber & Wilson (1986), a new object *P* is relevant in a context *C* to the extent that its contextual effects in this context C are large; and a new object P is relevant in a context C to the extent that the effort required to process it in this context C is small. For addressing the aforementioned sub-question **a)**, I applied, in the series of empirical studies in Chapter 5, Tversky's Ratio model rooted in the Prototype Approach (Rosch & Mervis, 1975) and the BMG that supports the Knowledge Approach (Murphy, 2004). The results obtained from these studies indicated that the BMG with the strong sampling scheme sensitively reflects the feature structure of concepts under comparison, and both intuitively and effectively identifies the most similar concepts by reflecting on the learner's prior knowledge. Indeed, the interpretation of the results is highly inspired by the way Tenenbaum & Griffiths originally formulated the BMG and their subsequent analysis of the BMG against Tversky's model (Tenenbaum & Griffiths, 2001):

1) *Interpreting this formal correspondence between our Bayesian model of generalization and Tversky's set-theoretic models of similarity* [equation (5) and (4) in Chapter 4 in this thesis] *is compared by the fact that in general the relation between similarity and generalization is not well understood. A number of authors have proposed that similarity is the more primitive cognitive process and forms (part of) the basis for our capacity to generalize inductively (Quine, 1969; Rips, 1975; Osherson et al., 1990). But from the stand point of reverse-engineering the mind and explaining why human similarity of generalization computation take the form that they do, a satisfying theory of similarity is more likely depending upon a theory of generalization than vice versa. The problem of generalization can be stated objectively and given a principled rational analysis, while the question of how similar two objects are is notoriously slippery and underdetermined (Goodman, 1972). We expect that, depending on the context of judgment, the similarity of y to x may involve the probability of generalizing from x to y, or from y to x, or some combination of those two. It may also depend on other factors altogether. Qualifications aside, interesting consequences nonetheless follow just from the hypothesis that similarity somehow depends on generalization, without specifying the exact nature of the dependence.*

2) *Most fundamentally, our Bayesian analysis provides a rational basis for the qualitative form of set-theoretic models of similarity. For instance, it explains why similarity should in principle depend on both the common and the distinctive features of objects. Tversky (1977) asserted as an axiom that similarity is a function of both common and distinctive features, and he presented some empirical evidence consistent with that assumption, but he did not attempt to explain why it should hold in general. ... Our rational analysis* [equation (5) in Chapter 4 in this thesis]*, in contrast, explains why both kinds of features should matter in general, under the assumption that similarity depends on generalization. The more hypothesized consequential subsets that contain both x and y (common features of x and y), relative to the number that contain x (distinctive features of x), the higher the probability that a subset known to contain x will also contain y.*

3) *The hypothesis that similarity depends in part on generalization explains why similarity may in principle be an asymmetric relationship, that is, why the similarity of x to y may differ from the similarity of y to x. Tversky (1977) presented compelling demonstrations of such asymmetries and showed that they could be modelled in his set-theoretic framework if the two subsets of distinctive features X-Y and Y-X have different measures under f and are given different weights in* [equations (1) and (4) in Chapter 4 in this thesis]. *But Tversky's formal theory does not explain why those two subsets should be given different weights; it merely allows this as one possibility. In contrast, the probability of generalizing from x to y is intrinsically an asymmetric function, depending upon the distinctive features of x but not those of y. To the extent that similarity depends on generalization, it inherits this intrinsic asymmetry.*

Up to now, the issue of similarity and generalization has ambiguously been dealt with in both the revised RTC and the COM frameworks proposed in this thesis. After considering Tenenbaum & Griffiths' view of similarity and generalization, a cross-cultural communication scenario can be drawn in the revised RTC framework and illustrated in a more concrete way. Figure 6-1 shows how a Danish audience, who is told by a Japanese communicator that "A Japanese omelet will be served today", will generalize the meaning of the original Japanese source concept "*Okonomi-yaki*" from the English stimulus "*Japanese omelet*". Each Dane has no idea how "*Okonomi-yaki*" looks like and he/she only knows that the dish is referred to as "*Japanese omelet*" in English (as pivot). His/her conceptualization of the cuisine is obviously rooted in the Danish culture. Hence, when he/she hears the word "*omelet*", he/she first induces from category labels on how "*omelet*" looks like, i.e. "made of egg", "fried on a pan", "optional ingredients can be chosen" naturally rooted in his/her prototype image of "*omelet*", as illustrated in Figure 6-1. This mental process is emphasized by Murphy (2004) by stating that *category-based inductions are extremely common in communication.* Consequently, what a Danish receiver generalizes from the word "*omelet*" may very well end up in a mental picture that better corresponds to "*Japanese ome-rice*" which is completely different from what the Japanese communicator originally intended to convey - i.e. the *true* meaning of "*Okonomi-yaki*". Hence, the probability of generalizing from *"omelet"* to either *"Okonomi-yaki"* or *"Japanese ome-rice"* can rationally be computed, from the viewpoint of Tenenbaum & Griffiths (2001), in the framework of the revised RTC which inherits the *asymmetric coordination of communication*. Of course, in a real communication scenario, a Danish audience does not know what the Japanese communicator intends to convey. Although the following "gedanken experiment" is rather hypothetical, let us assume that we, from a third party's viewpoint (*aka* "helicopter perspective"), already know what the Japanese communicator intends to convey in terms of the *true* meaning of "*Okonomi-yaki*" and let us assume we also have access to both the Danish audience's and the Japanese communicator's taxonomic organizations of food categories. Thus, we can see the whole feature structure of all Japanese food categories in the taxonomy. Then, what we are computing with the BMG (Tenenbaum & Griffiths, 2001) is the probability of generalizing a Japanese food category from a given English stimulus "*omelet*". If the Danish audience generalizes to other food categories with much higher probability than "*Okonomi-yaki*", this communication may likely end up in a huge misunderstanding. That is exactly the scenario I have tried to illustrate in Figure 6-1. Then, what we really need in order to help the communicating parties here, is to identify a Danish food category that inherently comprise a higher probability of generalizing to the original Japanese "*Okonomi-yaki*". In other words, and based on Tenenbaum & Griffiths constructions, we should somehow "reverse-engineer" which Danish context *C* that can eventually generate an optimal contextual effect for introducing the Danish audience to the new information *P*, i.e. "*Okonomi-yaki*". The label representing such an optimal context *C* ought to be the most effective translation that can convey the original meaning of "*Okonomi-yaki*" for the Danish audience. We can even add a supplemental description on how the new information *P* "*Okonomi-yaki*" and the most optimal context *C*, differ by explaining distinctive features between them.

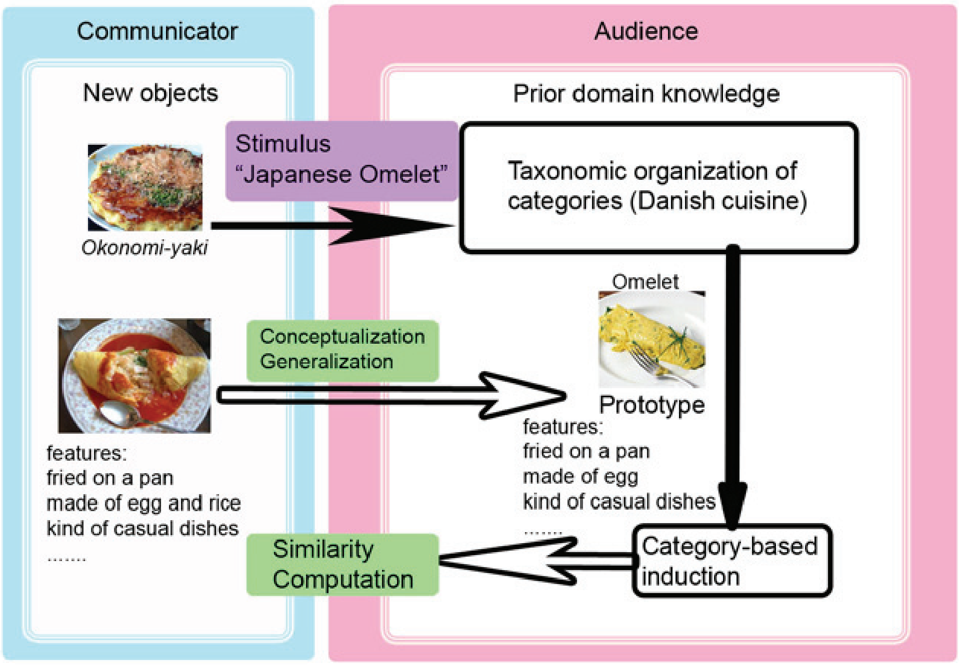


Figure 6-1: Example of asymmetric relations in cross-cultural communication between a communicator and an audience

Another key point of the BMG is that the aforementioned Knowledge Effects is nicely explained in this generalization process. As Murphy (2004) points out, the relation between a new concept and the prior knowledge is bi-directional: i.e. the learning process of a new concept can be highly influenced by the prior knowledge, while the novel information of this newly introduced concept may also influence the prior knowledge. The Knowledge Approach considers concepts as part of our general knowledge of the world. It means that concepts should be consistent with whatever else we know (Keil 1989; Murphy and Medin 1985). In order to maintain such consistency, people use their prior knowledge to reason about a new object and to decide on what category it is, or whether to learn a new category. In such situations, subjects use their prior knowledge when learning categories. Therefore, *people are primarily learning how features are related to the different categories* (Murphy & Wisnienski, 1989). For the computation of the BMG, prior knowledge has been considered as taxonomically organized feature lists. In the empirical studies of simulating the COM framework, either a communicator or an audience's prior knowledge has been considered as feature-sets that are created by TOs or are directly extracted from text corpora. Subsequently, the feature weights have been assigned based on the *size principle* defined by Tenenbaum & Griffiths (2001) for the prior knowledge of the parties in question. The comparative analysis in the fourth and fifth empirical studies of this thesis demonstrated that if feature-sets are appropriately prepared, the size principle effectively works e.g. as can be seen by the example illustrated in Figure 6-2.

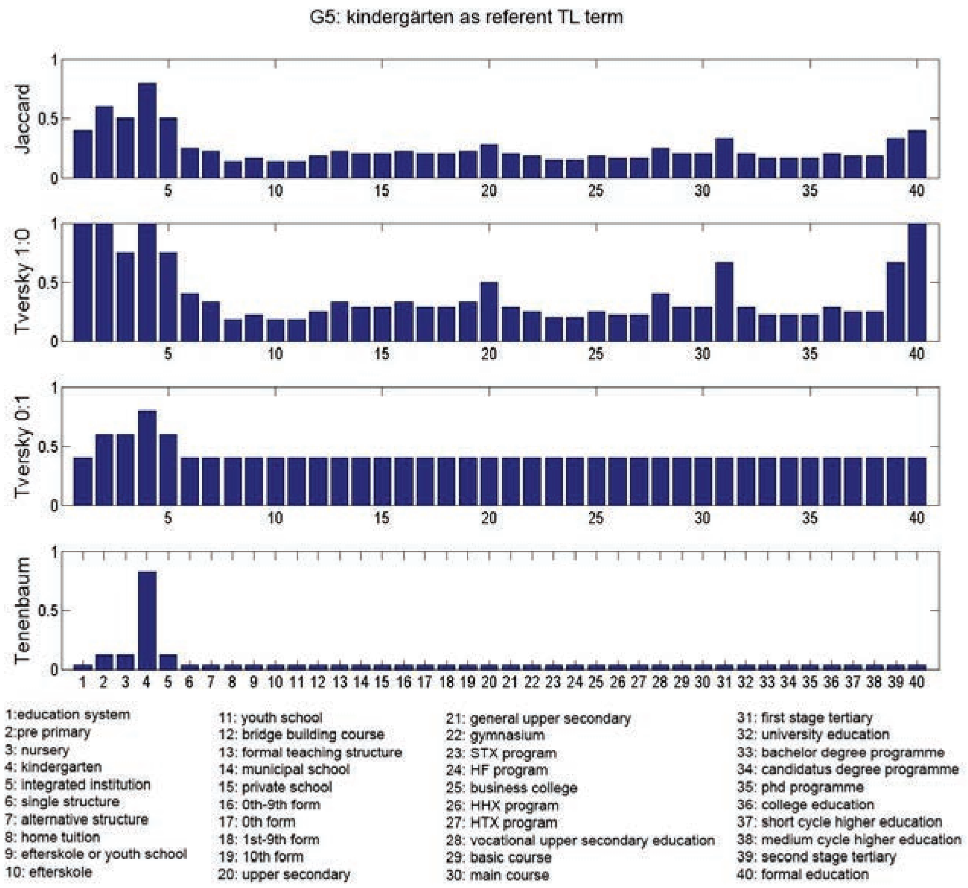


Figure 6-2: Probability of generalizing Danish "kindergarten" based on a German reader's prior knowledge

Although it is absolutely necessary in the future to conduct and verify whether the cognitive model proposed by Tenenbaum & Griffiths can be applied to real human cross-cultural communication, their arguments and theory beautifully explain the revised RTC framework in some of my empirical simulations employing corpora-based datasets.

A challenging issue is then – as posed on page 176 in sub-question **ii):** how contexts are chosen, i.e. how human prior knowledge is represented, and how appropriate feature-sets are obtained. I challenged this question by assuming that the TO method (Madsen et al., 2004) could potentially be used for representing prior-knowledge of culturally-specific domain knowledge. However, my seventh empirical study in Chapter 5 indicated that the TO method combined with the BMG, due to its strict principles, has a disadvantage of delimiting feature-dimensions that are potentially important for linking cross-cultural concepts. On the contrary, the direct application of the BMG to the loosely structured datasets considers all available feature-dimensions without restriction, which enabled the computation of the "fuzziness" inherent to the human mind, by reflecting prior knowledge of a learner who compares a new object with something he/she knows in advance. My empirical results seem to be consistent with the explanations of the *uni-dimensional strategy* and *multi-dimensional strategy* dealt with in Murphy (2004). For the uni-dimensional strategy, e.g. if pictures of objects are given as stimuli, subjects would first choose one dimension (e.g. size, shape, texture etc) and subsequently divide the shown items up into categories that differ in that dimension (large vs. small, square vs. round, etc.). A number of studies (Medin et al., 1987; Regehr and Brooks, 1995; Thompson, 1994; Ward et al., 1991) all show the very strong bias towards uni-dimensional categories. This approach is highly similar to some of the strict TO principles. However, Murphy (2004: 128-129) reflects on *why people make categories that are so different from those of everyday life*. His argument is that our concept of something, e.g. "birds", *is multidimensional and rich, which gives it inductive power. Uni-dimensional categories are obviously not rich and have no inductive power: Since you need to observe the single feature to classify the item, and there are no other features, there is no further information to be gained from categorization* (Murphy, 2004: 129). He explicitly points to the example of feature-feature correlations discussed in the category learning paragraph in Chapter 3 in this thesis, and emphasizes that multidimensionality of concepts inherently gives a much stronger induction power compared to that of the uni-dimensional.

In my eyes, this multi-dimensional view, naturally supports the IRM approach (Kemp et al., 2006) investigated in the sixth and seventh empirical studies in Chapter 5. Murphy (2004) argues that when several categorical features are related, the cluster of these features form a so-called family resemblance structure. *This can happen either through the relations of prior knowledge* (Ahn, 1990; Kaplan and Murphy, 1999; Spalding and Murphy, 1996) *or through induction that relate the features* (Lassaline and Murphy, 1996). Murphy and Allopenna (1994) shows that *when the features of a category formed a consistent set, the category was much easier to learn than when they were inconsistent or simply neutral.* Their findings in a way indirectly explain how the IRM played a role in the sixth empirical study in Chapter 5, where the following three strategies were tested: **1)** applying the IRM directly to two CSC-feature matrices respectively representing educational domain knowledge in Japan and Denmark for first categorizing them into categorical classes that are to be afterwards compared and aligned; **2)** applying the IRM directly to a matrix where the two CSC-feature matrices respectively representing the Danish- and Japanese educational domain knowledge are merged; and **3)** applying the BMG to directly compute similarity relations between CSCs in the two cultures, thereafter applying the IRM for clustering CSCs in the respective cultures into categorical classes.

The empirical results from the above strategies 1) and 2) could be interpreted based on Murphy and Allopenna's work, that, if a consistent set of features is formed among CSCs, a category has been successfully learned. On the other hand, if feature sets for CSCs were inconsistently structured, e.g. few relevant- and many irrelevant features, the clustering results have not been optimal. The question is then how to interpret the results obtained from the above strategy 3). It seems to me that the following suggestion by Murphy and Allopenna is key for answering this question. More specifically, they suggest that *knowledge helps learning because it relates the features in the category, rather than through the properties of the features themselves* (Murphy, 2004: 151). In addition, Kaplan and Murphy's (2000) work shows that when thematic features are present in categories, background knowledge is helpful even though the knowledge is incomplete or imperfect. Subjects are also able to ignore features that are inconsistent and still be able to use the most accurate knowledge. The third strategy of the sixth empirical study in Chapter 5 first applies the

BMG. This implies that the loosely structured sets of features representing culturally-specific domain knowledge, either of the Japanese or the Danish educational system, is considered as prior knowledge, and the other part is considered as new information compared against the prior knowledge. This process identifies all existing links between the Danish CSCs and the Japanese CSCs when they share at least one common feature. Thus, if a Japanese who has knowledge about a categorical class consisting of Japanese CSCs that provide "compulsory education", the IRM likely categorizes Danish CSCs that also provide "compulsory education" and creates a link at the categorical class level. Such links for co-clustering categorical classes in the two cultures are found in the η sorted graph in Figure 6-3.

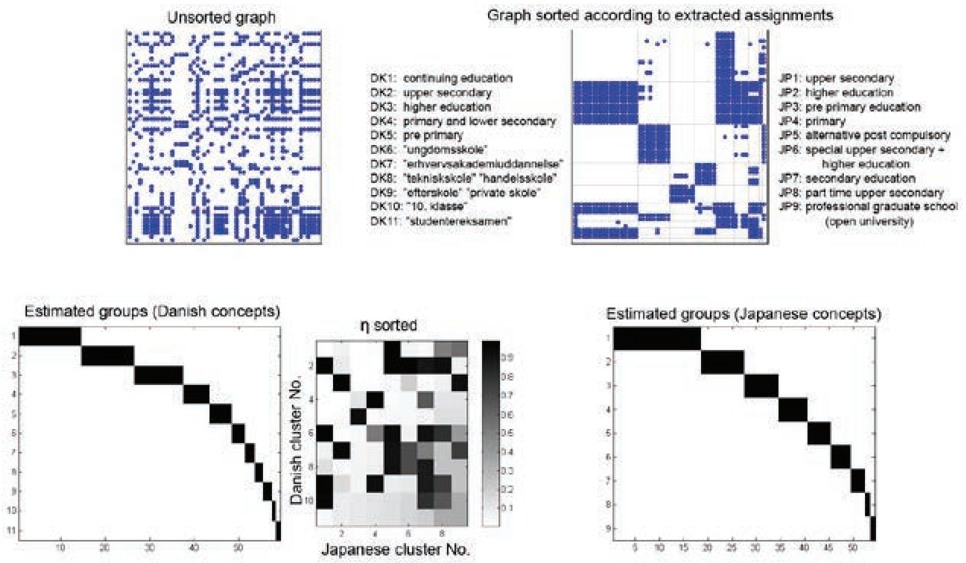


Figure 6-3: IRM clustering based on Strategy 3 (BMG + IRM)

In the seventh empirical study in Chapter 5, the two approaches - the TO+BMG and the BMG+IRM - are compared. From the viewpoint of the CSC mapping in cross-cultural communication, the BMG+IRM which is a cognitive model of generalization is designed to compute such fuzziness of the human mind by reflecting prior knowledge of a learner who compares a new object with something he/she knows in advance. On the other hand, an advantage of the TO+BMG approach is the clarification of domain knowledge in a specific knowledge community and the ability to compare the *hierarchical structures* across different communities. These two different views are also dealt in Murphy (2004) where it is hypothesized that a category structure is stored in the long-term memory in the form of a so-called categorical network. The categories located at the higher level in this hierarchy are called superordinate and the ones that are located at the lower level are referred to as subordinate categories. Category members are connected by the *set inclusion* relation. This relation is also referred to as an "is-a" relation in (Collins and Quillian, 1969). Murphy describes three basic principles for forming a hierarchical structure of categories. The first principle is that *any category can have only one immediate superordinate*. The second principle is that the "is-a" relation can only hold asymmetrically, so to speak. As an example of this asymmetry, all dogs are animals, but all animals are not necessarily dogs so that the "is-a" relation obviously only makes sense being applied as a uni-directional relation from dogs to animals. The third principle is that category relations are transitive so that properties of a superordinate category are inherited by sub-category members. Murphy emphasizes that these hierarchical descriptions enable us to generalize the properties possessed by mammals in many categories under "mammal" in the hierarchy. This ability is important for us *to immediately access knowledge about new entities that one hasn't had direct experience with* (Murphy, 2004: 202).

Although the first principle of Murphy differs from the TO principle which allows a poly-hierarchical structure, this hierarchical view is naturally in line with the TO+BMG approach. On the contrary, another view is that without storing the hierarchical links in the memory, information about category inclusion can be computed based on the properties that are known of a given category (Rips, Shoben, and Smith, 1973). In this view of *computed models*, the properties that are generally true of a category are also true of its subordinates in a hierarchy. It means that more specific categories inherently have the same features as more general categories but with one or more additional features. The way I have created the loosely structured datasets, which the BMG+IRM approach is applied to in the sixth and seventh empirical studies, are in fact based on this principle. Based on several studies e.g. Sloman (1998) discussed in Chapter 3, and Murphy (2004: 209) summarizes that *the feature model, which say that hierarchies are computed rather than pre-stored, have the edge in this battle.* However, Murphy also points out that *a list of features may also be insufficient to account for all the data. At least some category relations may be explicitly learned and represented.* Accordingly, a possible conclusion indicated by Murphy (2004: 209) is that *it seems most likely that some category relations are directly learned and stored, whereas oth-*

ers are computed as needed. Although perhaps inelegant, this kind of mixed model may be the best explanation of the jumbled set of facts that people know about the world. As it is indicated in my seventh empirical study of comparing the TO+BMG and the BMG+IRM, both approaches have some advantages and disadvantages. Thus a unique combination of the two approaches into a unified BMG+IRM+TO approach, could be the future potential that enables us not only to map CSCs by respecting nuances of each concept existing in each respective culture, but also to construct TOs that are cross-culturally interoperable as well as mono-culturally clarified. In any sense, how human knowledge should be represented in the COM framework is still open for discussion and most importantly requires human-based empirical studies in addition to the undertaken computer simulations.

3. Contrast to human mapping

My focus in this work has been on the employment of cognitive approaches and their theoretical arguments behind the selection of specific cognitive models for cross-cultural communication scenarios. Thus, it is still an open question, how to assess the effects of how well the meaning of an SL concept is optimally conveyed to a TL audience and how exactly a successful knowledge transfer or cross-cultural communication can be achieved. Strictly speaking, for assessing such effects, a thorough human-based assessment is required. Although I consider this as out-of-scope for this thesis work and naturally positioned as further research to be undertaken, I still try here to compare a small set of human mapping results with the COM results obtained from the series of empirical studies in Chapter 5. The purpose is not for assessing the COM results against human mapping (translation), but rather for opening a discussion on what kind of methodologies are potentially most suitable for conducting a human-based assessment in the future.

For example, Figure 6-4 shows, from top to bottom respectively, Danish concepts that are potentially corresponding to a Japanese concept “J41: public lower secondary school” computed with: 1) the BMG+IRM from a Japanese learner’s viewpoint; 2) the BMG+IRM from a Danish learner’s viewpoint; 3) the TO+BMG from a Japanese learner’s viewpoint; 4) the TO+BMG from a Danish learner’s viewpoint; and 5) a human mapping. As indicated above, the human mapping result is by no means scientifically representative but only included for a preliminary illustrative purpose. It is obtained from a single subject, a Japanese native speaker at CBS, who grew up in Japan and has taken a full-time master education at University of Copenhagen in Denmark, and has been living in Denmark for more than 30 years. This human mapping has been performed by the following procedure:

1) A list of 16 original Japanese concepts and their English translation has been provided to the subject.
2) The subject was asked to select corresponding Danish concept(s) from a list of 59 Danish terms and their English translation. The subject was allowed to select more than one concept.
3) As a next step, the definitions of each Japanese concept (list of features) were shown to the subject. The subject was allowed to change or add her selection of corresponding Danish concept(s) completed in the previous procedure.
4) As a last step, the definitions of each Danish concept (list of features) were shown to the subject. The subject was allowed to change or add her selection of corresponding Danish concept(s) completed in the previous procedure.

Through this process, the subject has first chosen, “D36: municipallity school (folkeskole)” as the most suitable corresponding concept to “J12: elementary school” and “J40: public elementary school”. She has also selected “D44: private school (frie grundskoler)” as corresponding concept to “J36: private elementary school”. However, she clearly expressed that she was not sure about whether this choice was appropriate. Interestingly, she was not able to select any of the Danish concepts corresponding to “J25: lower secondary school” and “J41: public lower secondary school” until procedure 4) described above. Figure 6-4 shows the results for this specific concept - “J41: public lower secondary school”. In my eyes, the top graph in Figure 6-4 (corresponding to the BMG+IRM approach from a Japanese audience’s viewpoint) indicates a scenario where the subject was not able to identify any corresponding Danish concept without knowledge about the definitions of Danish concepts and solely relied on her Japanese knowledge. On the other hand, the results shown in the second top graph of Figure 6-4 could be interpreted in the way that, after the subject has read through the definitions of Danish concepts and confirmed with her Danish educational knowledge, her perception has slightly shifted from a Japanese to a Danish view and she has eventually chosen "D48: primary and lower secondary education "10 års undervisningspligt" as the corresponding concept.

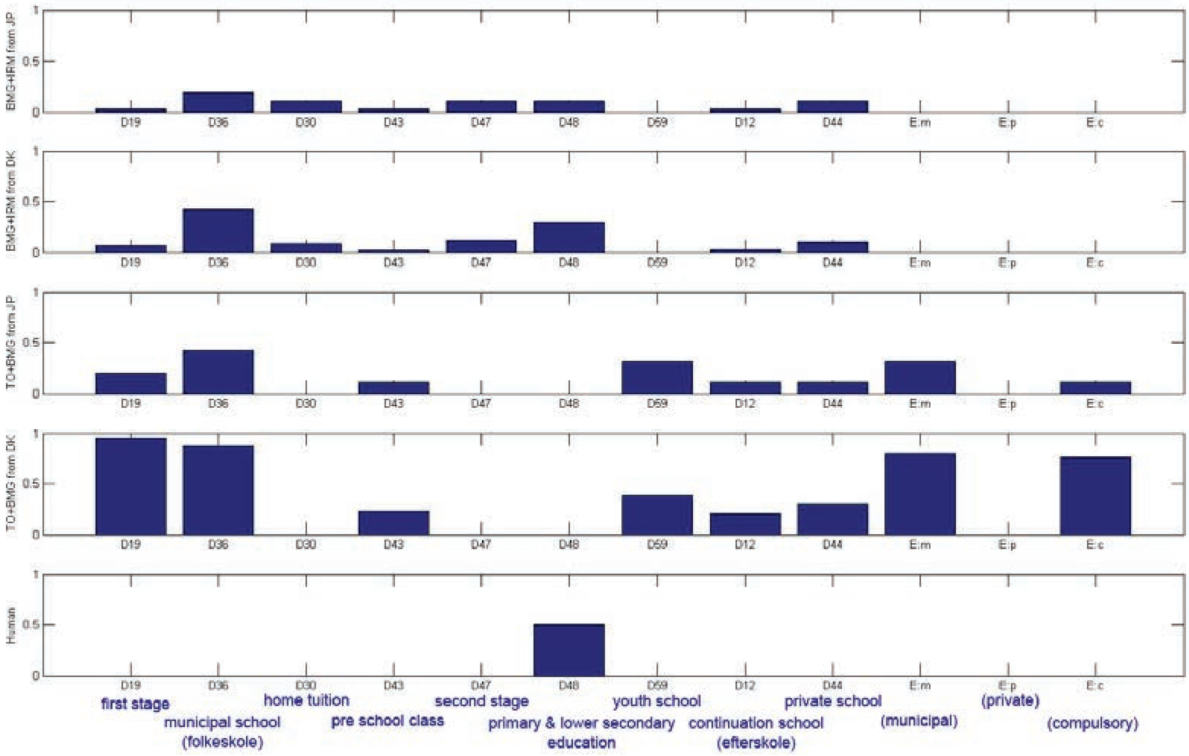


Figure 6-4: Human mapping - J12 elementary school

Although this final but very preliminary human assessment was quickly made, it provides me with some interesting clues on what should be pursued for future research undertakings. First of all, I have intentionally chosen a subject who has bi-cultural knowledge about Denmark and Japan. Identifying people who have in-depth knowledge of these remote cultures and countries are obviously very limited. In addition, I realized by discussing with her that contemporary conceptualization of Danish educational concepts is slightly different, probably due to the generation gap in terms of recently familiar Danish words such as "udskoling (referring to 7-9 grades)", "frie grundskoler (corresponding to private schools)". So considering these additional "time factors", bridging two remote cultures by appropriate translation seems like an almost impossible challenge to overcome. Thus, a thourough human-based assessment method should be process oriented – for example, on how people induce knowledge or imagine associative words from a given translated word, e.g. the feature listing method (Rosch, 1978), instead of comparing mapping results against a sort of gold-standard mapping, that has no guarantee of being correct. Other issues to be considered are whether subjects should have bi-cultural or mono-cultural backgrounds and whether subjects should assess English translations or only local expressions of original concepts. These issues have to be taken into careful consideration and any human assessment methodologies should be carefully planned accordingly for any future research undertakings.

4. Cognitive and linguistic relativities

In this final section, one of the possible considerations in terms of the human-based assessment will be discussed.

From the viewpoint of cognitive scientists, the so-called taxonomic organization of categories is inherently rooted in a child's first language- and concept acquisitions (Murphy, 2004: Chapter 7&9). This perspective is highly consistent with the Theory of Communicative Supertypes proposed by (Durst-Andersen, 2011a&b): *it is the mother tongue, and not a foreign language, that goes into our bodies and brains, thereby becoming internalised and automatised already at an early age*. When it comes to an adult's second language acquisition, e.g. related to a newly encountered object, this new information is aligned with his/her prior-knowledge, that is, the taxonomic organization of categories that has been developed since his/her childhood (Murphy, 2004; Sperber & Wilson, 1986). In (Durst-Andersen, 2011a&b), these disparate concepts and language acquisition phenomena are very illustratively defined as the *private voice of language* and as the *public voice,* respectively.

The Theory of Communicative Supertypes (Durst-Andersen, 2011a&b) defines that languages fall into three different supertypes according to choice of semiotic direction: (1) reality-oriented languages such as Chinese, Russian and Hindi that speak of reality through the situation being common to the speaker and the hearer; (2) speaker-oriented languages such as Japanese, Spanish and Turkish that speak of reality through the speaker's experience of the situation; and finally (3) hearer-oriented languages such as English, Danish and Swedish that speak of reality through the hearer's experience of it. According to Durst-Andersen (2011a & b), due to these completely different communication processes, misunderstandings are inevitable phenomena in situations where a non-native English speaker communicates in English (as *public voice*) based on the inner *private voice* which his/her mother tongue belongs to. Durst-Andersen (2011a & b) argues that, *due to the fact that each communicative supertype correlates with specific perceptual, cognitive, communicative, and pragmatic parameters, the effect is paramount on its users. People having native languages that belong to different supertypes will perceive, think, communicate, and behave differently: Chinese and English people will look differently at the same picture, will solve problems by using different strategies, will make different sorts of social contracts and exhibit different types of mentalities.*

With this mindset, a future research challenge could be to investigate how a human's taxonomic organization of category is influenced by the three different kinds of so-called communicative supertypes defined by Durst-Andersen (2011a&b). Assuming that the differences in the taxonomic organization among these three communicative supertypes are identified, a next challenge could be to identify how such differences in the taxonomic organization of categories (that are culturally- and linguistically rooted in one's mother tongue) influence a real cross-cultural communication scenario using English as second language. This future research would simultaneously indicate whether such empirical human data supports the results obtained from the cognitive model employed in the COM-framework.

Chapter 7.

Conclusions

Concluding remarks: Bridging Remote Cultures: Influence of cultural prior-knowledge in cross-cultural communication[15]

To summarize this thesis, I point the reader to my below paper to be presented at the International Organized Session celebrating that 2012 is the 100th anniversary of the "Alan Turing Year Special Session on AI Research that Can Change the World", The 26th Annual Conference of the Japanese Society for Artificial Intelligence, 12-15 June, 2012, Yamaguchi, Japan. This paper conveniently comprises a compressed summary of *all the empirical studies* in contrast to the COM framework that are based on the integration of the Relevance Theory of Communication theory (Sperber & Wilson, 1986) and the Knowledge Effects (Murphy, 2004). As indicated in the title of Alan Turing's special session, my humble dream is that the COM framework will in the future contribute to bridge remote cultures and to solve the inherently difficult challenge of EuroAsian communication which even skilled professional translators cannot easily solve in the contemporary globalized world!

[15] In: International Organized Session - Alan Turing Year Special Session on AI Research that Can Change the World, The 26th Annual Conference of the Japanese Society for Artificial Intelligence, 12-15 June, 2012, Yamaguchi, Japan

3P1-IOS-2a-6

Bridging Remote Cultures: Influence of Cultural Prior-Knowledge in Cross-Cultural Communication

Fumiko Kano Glückstad

Dept. of International Business Communication, Copenhagen Business School (CBS)

The role of ontology in a multilingual context is one of the emerging challenges in our modern information society. This work first explains different types of ontology applications in a multilingual context based on a number of dimensions defined in [Cimiano 2010]. These dimensions are useful for clarifying the role of ontologies depending on different types of cross-cultural communication scenarios. What is emphasized here is a new dimension in the ontology applications, namely the inherent asymmetric relation of communication between a communicator and an information receiver, which has been inspired by the pragmatic approach of the so-called Relevance Theory of Communication (RTC) [Sperber 1986]. Based on this ground theory, a new framework for simulating the cognitive processes involved in a cross-cultural communication is proposed.

1. Introduction

One of my Japanese acquaintances who has been living in Denmark for more than 40 years formulated his difficult mission of undertaking translation tasks in the following way: "*Once I deeply understood two cultures [in this Denmark and Japan] and cultural differences/nuances of conceptual meanings existing in the two countries, it became impossible for me to translate culturally-specific terms into the other language. Existing language resources (dictionaries etc.) are in this context useless*". What my acquaintance indicated is that it becomes virtually an impossible task to precisely translate or convey the meaning of a Culturally-Specific Concept (CSC) if no exact equivalent concept exists in the Target Language (TL) culture. Despite this inherent frustration, communicators or translators are still required to convey such CSCs into a TL in an optimal manner such that a TL reader can instantly infer the original meaning of a given Source Language (SL) concept. Is there some way to solve this inherently difficult challenge which even skilled human translators cannot easily cope with ?

This challenge of translating CSCs is not only caused by the absence of equivalent concepts in a TL culture, but also due to differences of the background knowledge possessed by the two parties involved in a cross-cultural communication scenario. The Relevance Theory of Communication (RTC) explains that *communication requires some degree of co-ordination between communicator and audience on the choice of a code and a context. The notion of mutual knowledge is used to explain how this co-ordination can be achieved: given enough mutual knowledge, communicator and audience can make "symmetrical" choices of code and context* [Sperber 1986]. However, the *symmetrical* choice of code and context is not sufficient for establishing optimal communication between communicator and audience. [Sperber 1986] emphasizes that, although *all humans live in the physical world*, mental representations are constructed differently due to differences in our close environment and our different cognitive abilities. Because people use different languages and have mastered different concepts, the way they construct representations and make inference is also dissimilar. [Sperber 1986] call this narrower physical environment a *cognitive environment*. In particular, [Sperber 1986] focuses on our conceptual cognitive abilities involved in communication and emphasize that *manifest facts* are important elements for conceptual cognition. Accordingly, human communication is viewed by [Sperber 1986] as such that a realistic notion of *mutual manifestness* is not strong enough to support the *symmetrical* choice of code and context. Since an individual possesses a total cognitive environment that is the set of facts based on his/her perceptual ability, inferential ability, actual awareness of facts, knowledge he/she has acquired and so on, it is much easier to achieve "*asymmetrical*" coordination between communicator and audience. This is in a way obvious because *human beings somehow manage to communicate in situations where a great deal can be assumed about what is manifest to themselves and others, but nothing can be assumed to be truly mutually known or assumed* [Sperber 1986].

This ground theory forms the basis of a framework that is to be introduced in this paper for the first time. As [Sperber 1986] stresses, a communication is an inferential process, hence, a cross-cultural communication is without doubt based on inferences. It is challenging but also fascinating to explore how an English word used by a person in one culture is perceived, understood and conceptualized by a person coming from another part of the globe. This is the motivation for my proposal of a new framework: Cognitive Ontology Mapping (COM).

In the following section, I first outline the view of different types of ontology applications in a multilingual context based on the categorization dimensions proposed by [Cimiano 2010]. This enables the identification of the position of the COM framework. Section 3 explains the theoretical framework of COM that is a combination of pragmatic theory and cognitive modelling. Section 4 further reviews algorithms of each component in the COM framework followed by a presentation of examples of simulations that have been achieved in the previous works in

Contact: Fumiko Kano Glückstad, CBS, DK-2000 Frederiksberg, Denmark, {fkg.ibc | fkg.isv}@cbs.dk

Section 5. In Section 6, I undertake discussions and future implications followed by conclusions in Section 7.

2. Ontology in a Multilingual Context

2.1 Categorization dimensions

[Cimiano 2010] defines a number of dimensions used in categorizing different types of *ontology localizations*. The term ontology localization is defined by [Cimiano 2010] as *the process of adapting a given ontology to the needs to a certain community, which can be characterized by a common language, a common culture or a certain geo-political environment.* [Cimiano 2010] views the ontology localization problems from two different aspects: a *lexical layer* and a *conceptual layer*. The lexical layer of an ontology refers to labels of concepts, properties and other elements used to describe concepts. On the other hand, the conceptual layer implies a conceptual structure itself that may need to be adapted due to *a different cultural or geo-political context*. [Cimiano 2010] emphasizes that *the adaptation of the conceptualization layer will be primarily driven by the inexistence of conceptual equivalents (or concepts with the same granularity level) in the target community whenever the final purpose of the ontology is to be equally valid in the source and target communities.* Accordingly, [Cimiano 2010] describes how the localization of the different layers (lexical and conceptual) interact with each other, and introduces different dimensions characterizing the localization process based on the outline defined in [Espinoza 2009].

***International (standardized) domain vs. culturally influenced domain*:** First of all, ontologies could be categorized based on characteristics of domains. Some domains are internationally standardized as they are identified in e.g. the technical domain or the financial report domain. On the other hand, some domains are culturally dependent. For example, domains such as related to legal systems, political systems and social welfare systems are all culturally and geo-politically influenced. An ontology developed for a culturally influenced domain is referred to as a *culturally-dependent ontology* which is one of the main focuses in the COM.

Functional vs. documental localization: Based on the so-called functionalist theories to translation by [Nord 1997], [Cimiano 2010] argue that an *ontology might be localized with different goals in mind*. Their idea of *functional localization* is that, if a domain is culturally influenced, e.g. a different geo-political reality for a target community, the original source ontology has to be adapted to the target community by maintaining similar functions in both communities. It means that it is necessary to change the conceptual structure in a source community to fulfill the requirements of a target community. Hence, functional localization implies the creation of a new ontology on the basis of the original one. Contrary, documental localization means that the original ontology can be supported by members of another linguistic community. Hence, what is required in this type of localization is to document the meaning of the original ontology for a different language community.

***Interoperable vs. independent ontology*:** Interoperable ontology means that a new target ontology and the original source ontology are to a certain extent interoperable with each other. Thus, any changes to the conceptual structure are restricted in order to ensure a certain degree of interoperability. On the other hand, independent ontology means that the target ontology corresponds to the source ontology only in a functional manner so that significant changes to the conceptual structure of the target ontology are acceptable in order to meet the needs and capture the specificities of the target community.

2.2 Position of the COM

The aforementioned dimensions effectively clarify the roles of ontologies in different scenarios. For example, the XBRL (eXtensible Business Reporting Language) - standardized domain-specific ontology - used in the financial business report domain across different cultures is the use case studied under the framework of the MONNET project on Multilingual Ontologies for Networked Knowledge [Declerck 2010]. The case of XBRL is representing a typical pattern of the *documental localization* in the *international domain* defined in [Cimiano 2010]. The MONNET project also employs other cases such as the public sector use case where the Customer Service Guide (NL: *Klantdienstwijzer*) of the Dutch Immigration and Naturalization Office (IND) is localized into different languages targeted for foreign immigrants and visitors to the Netherlands. This use case is obviously in a culturally influenced domain. However, it only deals with translation of the Dutch legal conceptual structure into other languages, i.e. there is no inherent need of aligning legal conceptual systems across different countries/cultures. Hence the localization of this application has a rather standardized aspect, and is considered as *documental localization.* On the other hand, let us assume that the Japanese- and the Danish governments are going to negotiate a pension treaty for the future. In such a case, the *documental localization* is no longer sufficient and applicable. Instead, this scenario requires the alignment of two *independent ontologies* in a *culturally-influenced domain*. This scenario implies that the two *independent ontologies* are functionally peculiar. The COM approach is challenging precisely this issue of how to link such two *independent ontologies* which are *functionally* distinctive in a *culturally-influenced domain*, for the purpose of improved cross-cultural communication.

2.3 Patterns of Cross-Cultural Communication

Although the aforementioned dimensions in [Cimiano 2010] explain these different scenarios in an effective manner, we realize that another dimension is required for explaining how ontologies in a multilingual context should be linked according to different communication patterns. For example the first- and second scenarios in the MONNET framework are both considered as *assimilative* communication. In the first scenario, the standardization consortium in a way forces each individual party to accept the way the consortium has decided to conceptualize the domain knowledge. The second scenario is the same on a smaller scale given that the Dutch government forces immigrants to accept the way the Dutch government is functioning. Both patterns represent unidirectional assimilative communication patterns. Thus, an one-way ontology localization is applicable. On the other hand, for the third pattern, the Japanese- and the Danish governments are positioned at an "equal level" and have to respect, interact and understand each other for establishing applicable mutual understandings. From this viewpoint, it should be noticed that the one-way ontology localization function in a different way than the dual-way ontology alignment, in a cross-cultural communication. In the COM approach, a cross-cultural communication is considered as a dual-way interactive communication pattern.

3. Theoretical Framework of the COM

As discussed in the previous, the COM is supposed to deal with the dual-way interactive cross-cultural communication by aligning independent ontologies that are functionally distinctive in a culturally-influenced domain. In here, two independent ontologies that are aligned are considered as *cognitive environments* possessed by the parties involved in a communication as defined in [Sperber 1986]. In the aforementioned third scenario, both a Danish- and a Japanese governmental officer, respectively, must possess different cognitive environments due to the use of different languages, different conceptualization of the society, different domain knowledge rooted in their respective cultures etc. Thus, although the purpose of communication is to achieve a mutual understanding, the *asymmetric* coordination of code and context between them is a realistic approach and view of communication [Sperber 1986]. How this asymmetric coordination of code and context can be reflected in the aforementioned ontology alignment is the challenge which the COM approach is trying to solve. COM challenges this by referring to the ground theories of concept learning and categorizations.

In the Cognitive Science, [Murphy 2004] states that *concepts are the glue that holds our mental world together. .... If we have formed a concept (mental representation) corresponding to that category (class of objects in the world), then the concept will help us understand and respond appropriately to a new entity in that category. Concepts are a kind of mental glue, then, in that they tie our past experiences to our present interactions with the world, and because the concepts themselves are connected to our knowledge structures.* This statement inherently indicates that the study of concepts is highly connected to the RTC [Sperber 1986] that has just been dealt with in the introduction. However, in the study of concepts, the focus is rather on a representation of conceptual knowledge – relationships between groups of objects and features possessed by these objects. The representation of conceptual knowledge enables humans to learn new concepts and to make category-based induction which forms the basis for cross-cultural communication.

Figures 1 and 2 show how the representation of conceptual knowledge, as well as the mechanisms of the concept learning and the category-based induction are integrated in the framework of cross-cultural communication based on the RTC. Figure 1 illustrates a scenario where a Japanese communicator intends to convey the meaning of the traditional Japanese dish "*Okonomi-yaki*" to a Danish audience, in English. However, the Japanese does not know how "*Okonomi-yaki*" should be referred to in English or in Danish for that matter. Here, we need to notice that the Danish audience is not native English speaking as well as the Japanese communicator. Thus, the Danish audience neither has perfect conceptual knowledge of British nor American cuisine expressed in English. Accordingly, we assume that the Japanese try to browse a cooking book of Danish cuisine translated into English. The Japanese is supposed to have the picture (conceptual knowledge) of the Japanese cuisine based on his/her prior experience as a Japanese living in Japan for many years. He/she knows that the main features of "*Okonomi-yaki*" are: "fried on a pan", "made of egg, flour, cabbage", "optional ingredients can be chosen" "kind of a casual dish" etc. In addition, he/she has some kind of idea of how this specific type of Japanese dish "*Okonomi-yaki*" is placed (categorized) in the entire picture of the Japanese cuisine. Based on this prior knowledge, he/she compares "*Okonomi-yaki*" with Danish dishes (new objects) found in a Danish cooking book and identifies the most similar concepts. Let us assume that the Japanese identifies "*omelet*" which possess the features: "fried on a pan" "made of egg" "optional ingredients can be chosen" "kind of a casual dish" as the most similar concepts in this scenario, as shown below:

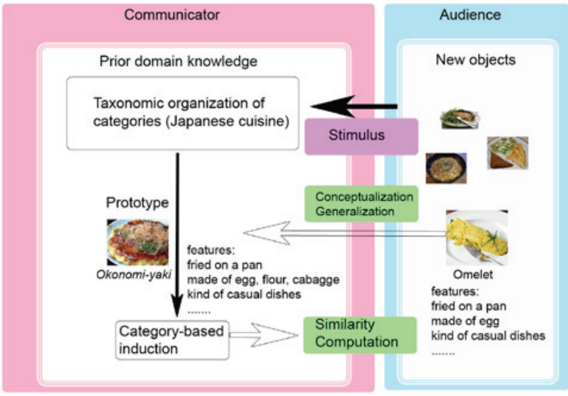


Figure 1: Asymmetrical coordination in a cross-cultural communication (when a Japanese communicator is learning Danish dishes as new objects)

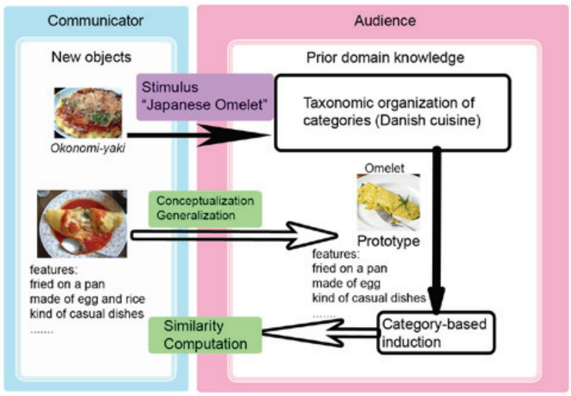


Figure 2: Asymmetrical coordination in a cross-cultural communication (when a Danish audience generalizes the meaning of "*okonomi-yaki*" from the stimulus "*omelet*")

Figure 2 illustrates how the Danish audience, who is told by the Japanese communicator that "Japanese omelet will be served today", will generalize the meaning of the original source concept "*Okonomi-yaki*" from the English stimulus "*Japanese omelet*". A Dane has no idea of how "*Okonomi-yaki*" looks like and only knows that the dish is referred to as "*Japanese omelet*" in English. His/her conceptualization of cuisine is rooted in the

Danish culture, hence, when he/she hears the word "*omelet*", he/she induces from this category label, how "*omelet*" looks like, i.e. "made of egg", "fried on a pan", "optional ingredients can be chosen" based on his/her prototype image of "*omelet*" as shown in Figure 2. Consequently, what he/she generalizes from the word "*omelet*" may be "*Japanese ome-rice*" which is quite different from what the Japanese communicator originally intented to convey i.e. the true meaning of "*Okonomi-yaki*". These scenarios shown in Figures 1 and 2 are hypothetical illustrations of the so-called "*asymmetric*" coordination in cross-cultural communication based on the RTC.

As briefly mentioned above, this cross-cultural communication framework integrates the mechanisms of the concept learning and the category-based induction as well as a system for representing conceptual knowledge. Here, diverse theories and studies that have been performed in the history of cognitive science play in. Within the cognitive science discipline, there have been two major views on concepts: the so-called prototype view by [Rosch 1975] and the exemplar view by [Medin 1978]. In the paradigm of [Rosch 1975], a concept is represented as features that are typically identified among the category members. Assuming that a feature list (a set of features) is a concept representation, the categorization process of a new object is the computation of similarity of the new item measured up against the existing feature list. It means that if a feature that is commonly possessed by both a new object and the representation, the feature in question receives "credit". On the contrary, if a feature is possessed only by the representation or by the object, the feature in question loses "credit" [Tversky 1977]. In the exemplar theory [Medin 1978], similarity plays an important role, too. Assuming that an individual's concept of dogs is the set of dogs that the person remembers, a new object, say a given animal, observed by this person should be weighted on the basis of how similar his/her memory of dogs is to the new given animal object. An important thing in this view is therefore to assess *how similar the object is to each memory* [Murphy 2004]. Based on these traditional views, a new approach called the *Knowledge approach* has more recently appeared as a reaction to the prototype- and exemplar approaches. The idea of the knowledge approach is that when people learn a new concept, for example related to animals, the new information about animals is integrated with one's prior knowledge about biology, animal behavior or other relevant domain knowledge. The relation between the new concept and the prior knowledge is bi-directional: i.e. the learning process of the new concept can be influenced by the prior knowledge, while the new information of this concept may also influence the prior knowledge. The knowledge approach considers concepts as part of our general knowledge of the world. It means that concepts should be consistent with whatever else we know. In order to maintain such consistency, people use their prior knowledge to reason about a new object and decide what category it is or to learn a new category. Unlike the prototype- and exemplar view, the knowledge view claims that *people do not rely on simple observations or feature learnings in order to learn new concepts. They pay attention to the features that their prior knowledge says are the important ones* [Murphy 2004], see also [Murphy 1994], [Markman 1997], [Wisniewski 1994], [Spalding 1996] [Lassaline 1996]. This *knowledge effect* is perfectly accommodated in the framework of an asymmetric cross-cultural communication scenario. Accordingly, what the COM approach is trying to encompass is to simulate this asymmetric cross-cultural communication framework accommodating the *knowledge effect* - on our computers. Thus in the following section, I review our empirical work for identifying the COM components that are suitable for realizing the cognitive simulation of asymmetric cross-cultural communication.

4. Components of the COM

4.1 Alignment component

As described in the previous, the key component of the COM approach is the mechanism of how a learner acquires a new object and categorizes it in connection with his/her prior knowledge. As the prototype theory explains, by assuming that a feature list (a set of features) is a concept representation, the categorization process of a new object is the computation of similarity of the new item measured up against the existing feature list possessed by the learner. Accordingly, Tversky's Contrast Model [Tversky 1977] which realizes a similarity computation based on the prototype viewpoint has been selected as the first candidate. [Tversky 1977] states that *similarity serves as an organizing principle by which individuals classify objects, form concepts, and make generalizations.* Tversky´s view of similarity is distinguished from the traditional theoretical analysis c.f. [Shepard 1987] on two key points: **1)** while the theoretical analysis of similarity relations has been dominated by the continuous metric space models, Tversky argues that *the assessment of similarity between objects may be better described as a comparison of features rather than as the computation of metric distance between points;* and **2)** although *similarity has been viewed by both philosophers and psychologists as a prime example of a symmetric relation*, the asymmetric similarity relation has been demonstrated in [Tversky 1977] based on several empirical evidences. Based on these points, [Tversky 1977] proposed a classic feature-set model of similarity, later coined Tversky's Ratio Model, as described by the following equation:

$$sim\,(y,x) = \frac{f(Y \cap X)}{f(Y \cap X) + \alpha * f(Y - X) + \beta * f(X - Y)} \qquad (1)$$

Here, X and Y are the feature sets of object x and object y, respectively. f denotes a measure over the feature sets. The function f is defined as additive in the series of our empirical studies [Glückstad 2012-a], [Glückstad 2012-b]. $(Y \cap X)$ represents the sets of features present in both X and Y, $(Y\text{-}X)$ represents the sets of features present in Y but not in X, and $(X\text{-}Y)$ represents the sets of features present in X but not in Y. Since the similarity score in equation (1) is normalized, the obtained score lies between 0 and 1. α and β are free parameters representing an asymmetric relationship between X and Y. Assignment of these parameters severely influences similarity measurements. When defining $\alpha = \beta = 1$, $sim\,(y, x) = f\,(Y \cap X) \,/\, f\,(Y \cup X)$ corresponds to the well-known algorithm: Jaccard's coefficient measure [Jaccard 1901]. When defining $\alpha = 1$ and $\beta = 0$, $sim\,(y, x) = f(Y \cap X) \,/\, f\,(Y)$ corresponds to what is found in e.g. [Bush 1951].

A noteworthy point explained in [Tversky 1977] is that *if sim(y,x) is interpreted as the degree to which y is similar to x, then y is the subject of the comparison and x is the referent. Hence the features of the subject are weighted more heavily than the features of the referent (i.e., α > β). Consequently, similarity is reduced more by the distinctive feature of the subject than by the distinctive features of the referent.* Based on this idea of asymmetric relation between a subject and a referent, as well as on the Relevance Theory of Communication that inherits the asymmetric co-ordination between communicator and audience on the choice of code and context, what is required in a cross-cultural communication is that a communicator should provide the set of assumptions that are adequately relevant to the audience, and the stimulus (that is English translation in here) produced by the communicator should be such that it avoids gratuitous inferential processing effort on the audience´s part. Considering that *similarity serves as an organizing principle by which individuals classify objects, form concepts, and make generalizations* [Tversky 1977], the most similar concept to a source concept, which is identified in the audience´s taxonomic organization of categories through the feature matching, would be the set of assumptions which are adequately relevant to the audience.

Interestingly, [Tenenbaum 2001] more recently demonstrated that Tversky´s model is remarkably similar to the Bayesian Model of Generalization (BMG) that is rooted in Shepard´s theory of the generalization problem. In [Tenenbaum 2001], three crucial questions of learning, after [Chomsky 1986], are addressed in order to explain the BMG: ***1)*** *what constitutes the learner´s knowledge about the consequential region;* ***2)*** *how the learner uses that knowledge to decide how to generalize; and* ***3)*** *how the learner can acquire that knowledge from the example encountered.* For instance, one example x of some consequence R is given. It is assumed that x can be represented as a point in a continuous metric psychological space and R corresponds to some region (referred to as *consequential region*) of this space. A task of the learner is to infer the probability that a newly encountered object y will fall within R given the observation of the example x. This conditional probability can be expressed as: $P(y \epsilon R|x)$. In order to compute the conditional probability, [Tenenbaum 2001] first answers the question: the learner´s knowledge about the consequential region that is represented as a probability distribution $p(h|x)$ over an *a priori*-specified hypothesis space $\mathcal{H}$ of possible consequential regions $h \epsilon \mathcal{H}$. Prior to observing x, this distribution is the prior probability $p(h)$, then becomes the posterior probability $p(h|x)$ after observing x. According to [Tenenbaum 2001], the learner uses this knowledge for generalization by summing the probabilities $p(h|x)$ of all hypothesized consequential regions that contain y, as follows:

$$P(y \epsilon R|x) = \sum_{h:y \in h} p(h|x) \qquad (2)$$

[Tenenbaum 2001] further describes how a rational learner arrives at $p(h|x)$ from $p(h)$, after the generalization, through the use of Bayes´ rule as follows:

$$P(h|x) = \frac{p(x|h)p(h)}{p(x)} \qquad (3)$$

According to [Tenenbaum 2001], *what likelihood function P(x|h) is determined by how we think the process that generated the example x relates to the true consequential region for R.* For example, Shepard's Universal Law of Generalization [Shepard 1987], the Bayesian analysis of inductive reasoning proposed in [Heit 1998] and the standard machine learning literature argue that *the example x and consequential region R are sampled independently, and x just happens to land inside R* [Tenenbaum 2001]. Thus, the likelihood is defined in a binary fashion in the following way:

$$P(x\,|h) = \begin{cases} 1 & if\ x \in h \\ 0 & otherwise \end{cases} \qquad (3)$$

Opposed to this, [Tenenbaum 1999] argues that *under many conditions, it is more natural to treat x as a random positive example of R, which involves the stronger assumption that x was explicitly sampled from R.* This argument leads to the "*strong sampling*" scheme defined as:

$$P(x\,|h) = \begin{cases} 1/|h| & if\ x \in h \\ 0 & otherwise \end{cases} \qquad (4)$$

Here, $|h|$ indicates the size of the region h [Tenenbaum 2001]. [Tenenbaum 2001] demonstrates that this cognitive model of learning can also be applied to Tversky's set-theoretic model. In order to demonstrate this, they have reformulated Tversky´s model (1) to the following formula:

$$sim(y,x) = 1/\,[1 + \frac{\alpha \cdot f(Y-X) + \beta \cdot f(X-Y)}{f(Y \cap X)}] \qquad (5)$$

This formula (4) is, according to [Tenenbaum 2001], mathematically equivalent to the re-formulation of (2) as follows:

$$P\,(y \in R|x) = 1/[1 + \frac{\sum_{h:x \in h, y \notin h} p(h,x)}{\sum_{h:x,y \in h} p(h,x)}] \qquad (6)$$

A key point here is that *the bottom sum ranges over all hypotheses that include both x and y, while the top sum ranges over only those hypotheses that include x but not y. If we identify each feature k in Tversky´s framework with a hypothesized subset h, where an object belongs to h if and only if it possesses feature k, and if we make the standard assumption that the measure f is additive, then the Bayesian model as expressed in equation (6) corresponds formally to Tversky´s Ratio Model (5) with asymmetric parameters α=0, β=1* [Tenenbaum 2001]. In equation (6), $P(h, x) = P(x|h)P(h)$ which represents the weight assigned to the hypothesis h in terms of the example x.

It means that, if the free parameters in equation (5) is set as *α=0, β=1,* this algorithm is formally corresponding to equation (6) of the BMG which compute *the conditional probability that y falls under R (Consequential Region) given the observation of the example x* [Tenenbaum 2001]. The consequential region R in our work based on the COM shown below indicates the categorical region where a subject y belongs. In equation (6), a hypothesized subset h is defined as the region where a concept belongs to h, if and only if, it possesses feature k [Tenenbaum 2001]. In the COM framework, the number of objects possessing the k^{th} feature in the referent ontology explained below is considered as the size of the region h. For example, if a feature "made of egg" is possessed by ten objects in a Danish cooking book (assuming that this cooking book is the entire domain knowledge of the Danish cuisine) the weight is assigned as 1/10.

This strong sampling can intuitively be illustrated in a situation where the feature "objects that have four legs" is given to us as an example. We immediately imagine that this object must be something related with an animal or possibly a piece of furniture. Hence, we unconsciously limit the hypothetical region to a narrower region in order to achieve a more effective generalization. Finally, [Tenenbaum 2001] explains that the prior $P(h)$ is not constrained in their analysis so that *it can accommodate arbitrary flexibility across contexts*. Hence in this work, we set $P(h) = 1$.

Accordingly, the BGM could be applicable to the scenario described in Figures 1 and 2. In case of Figure 1, y is implicitly considered as a newly encountered object existing in the target culture (the Danish cooking book) that should be compared with the source concept ("*Okonomi-yaki*") by the Japanese communicator. It means that by exchanging assignment of variables x and y, the algorithm defined in equation (6) also computes the probabilities that the Danish audience generalizes the source concept ("*Okonomi-yaki*") from a stimulus ("*Japanese omelet*") presented by the Japanese communicator, as shown in Figure 2.

Accordingly, Tversky's Ratio Model (equation 5) assigning different combinations of α and β parameters : **i)** α=1 and β=1: which corresponds to the Jaccard Similarity Coefficient representing a symmetric similarity relationship between objects x and y; **ii)** α=1 and β=0: which only computes distinctive features present in Y, not in X; and **iii)** α=0 and β=1: which corresponds to BMG (equation 6) without a strong sampling scheme and **iv)** the BMG with a strong sampling scheme (equation 4) have been applied to different datasets presented in [Glückstad 2012-a] [Glückstad 2012-b].

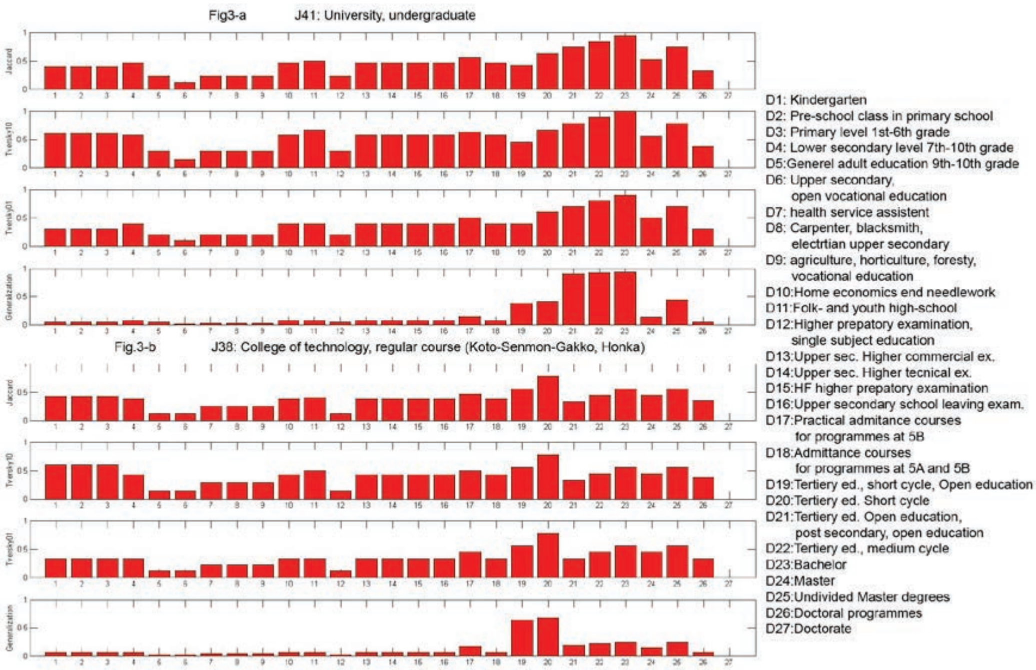


Figure 3: Similarity measure comparison (adapted from ref. [Glückstad 2012-b])

Figure 3 shows one of the comparative analyses of the aforementioned four similarity measures. In this empirical study [Glückstad 2012-b], the four similarity measures have been applied to standardized datasets based on the International Standard Classification of Education provided by the UNESCO Institute for Statistics. Both figures 3-a and 3b respectively present results obtained from the four similarity measures from top to bottom: **i)** α=1 and β=1: which corresponds to the Jaccard Similarity Coefficient representing a symmetric similarity relationship between objects x and y; **ii)** α=1 and β=0: which only computes distinctive features present in Y, not in X; and **iii)** α=0 and β=1: which corresponds to BMG without the strong sampling scheme and **iv)** the BMG with the strong sampling scheme. Figure 3-a illustrates a scenario where a Japanese learner who has Japanese domain knowledge of the Japanese educational system is learning about new objects existing in the Danish educational system domain. Figures 3-a and 3-b show that, in contrast to the first three similarity measures, the size principle in the fourth algorithm (BMG) effectively identifies specific concepts that are more similar than others. For example "D19: Tertiary, short cycle, open education" and "D20: Tertiary, short cycle education" are identified as the most similar concept to "J38: college of technology, regular course (高等専門学校本科: *Koto-Senmon-Gakko, Honka*)", and in the same way, "D21: Tertiary, post secondary open education", "D22: Tertiary, medium cycle education", and "D23: Bachelor" are identified as the most similar concepts to "J41: university, undergraduate (大学学部: *Daigaku Gakubu*)". On the other hand, the first to third similarity measures indicate that the aforementioned Danish concepts are only slightly more similar than the others. In addition, other Danish concepts referring to the pre-primary to lower secondary educations, i.e. D1-D4 are also considered slightly more similar than the others. In short, the fourth similarity measure, the BMG with the strong sampling scheme, sensitively reflects the feature structure of concepts in comparison, and both intuitively and effectively identifies the most similar concepts by reflecting on the learner's prior knowledge. This tendency has been confirmed also in other studies employing other type of datasets, i.e. datasets obtained from a strictly structured ontology called Terminological Ontology [Glückstad 2012-a] and loosely structured feature sets employed in [Glückstad 2012-c]. Further details of the qualitative data analyses are described in [Glückstad 2012-a] and [Glückstad 2012-b]. Although the evaluation is based only on qualitative analyses in these studies, the BMG with the strong sampling scheme seems to work based on its theoretical foundation on cognitive processing explained in [Tenenbaum 2001], when appropriate feature structures are employed as datasets. Thus the BMG equipped with the strong sampling scheme is not only theoretically but also empirically suitable for potentially realizing the simulation of the asymmetric cross-cultural communication framework accommodating the *knowledge effect* on our computers. The next question is then how such appropriate feature structures, that are applicable to the BMG algorithm, can be obtained.

4.2 Ontological component

The hypothesis started by the assumption that the Terminological Ontology (TO) method [Madsen 2004] is a suitable tool for the CSC mapping because the uniqueness of TO is feature specifications and subdivision criteria which enable us to construct concept representations based on well-organized feature structures.

The principles of TO have been developed in the research and development project called CAOS - Computer-Aided Ontology Structuring – where the aim has been to develop a computer system designed to enable a semi-automatic construction of ontologies [Madsen 2004]. The uniqueness of the TO is given by its feature specifications and subdivision criteria [Madsen 2004], [Madsen 2005]. A feature specification is presented as an attribute-value pair - for example as shown in Figure 4, [FIELDS: technology]. Thus, a representation of a whole concept is a feature structure, i.e. a set of feature specifications corresponding to the unique set of characteristics that constitutes that particular concept [Madsen 2004]. In Figure 4, each box that represents a particular concept is divided into three layers: **1)** top layer, lexical representation (term), **2)** middle layer, dimension specifications, and **3)** bottom layer, feature structure (set of feature specifications).

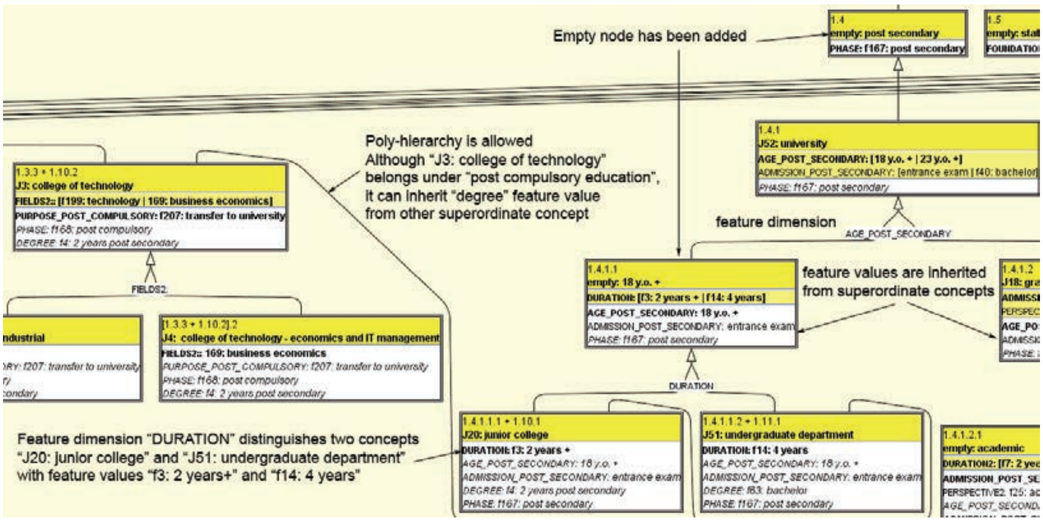


Figure 4: Terminological Ontology (adapted from ref. [Glückstad 2012-d])

The use of feature specifications is subject to principles and constraints described in detail in [Madsen 2004]. Most importantly, a concept automatically inherits all feature specifications of its superordinate concepts. This approach is fairly intuitive and reasonably consistent with the hierarchical structure of categories that are generally discussed by cognitive scientists [Murphy 2004]. However, the principles of TO also defines more strict rules derived from the traditional view of terminology that aims at proper standardization [ISO 2000]. For example, the principle of uniqueness of dimension defines that *a given dimension may only occur on one concept in an ontology*. [Madsen 2004] argues that *uniqueness of dimensions contributes to create coherence and simplicity in the ontological structure, because concepts that are characterized by means of a certain common dimension must appear as descendants of the same superordinate concept.* In the same way, [Madsen 2004] also defines the uniqueness of primary feature specifications as *a given primary feature specification can only appear on one of the daughters*. The argument is that these uniqueness principles *make it possible to a certain extent to carry out automatic placing of concepts into an ontology.* Another important principle is that the TO approach allows polyhierarchy structures so that one concept may be related to two or more superordinate concepts.

As a first attempt, [Glückstad 2012-a] applied the four similarity measures to datasets consisting of concepts (CSCs) and their respective features obtained from TOs respectively representing the Danish and the German educational systems. The CSCs and their respective features have manually been extracted from text corpora downloadable from the Eurydice web-site published by the Education, Audiovisual and Culture Executive Agency under the EU commission. The documents published by reliable authorities of each country describe the educational systems in a majority of the EU member countries both in English and in their native languages based on the UNESCO-ISCED classification template. The results obtained from the four similarity measures show the best performance with the BMG equipped with the strong sampling scheme. However, the results from that study also indicate that particularly strict rules for constructing TOs may risk causing the elimination of important features. It means that the original TO-approach may require a more flexible taxonomic organization of feature structures [Glückstad 2012-a] [Glückstad 2012-d].

Motivated by these results, [Glückstad 2012-c] investigates how the Infinite Relational Model (IRM) [Kemp 2006], a novel unsupervised machine learning method, can be applied to loosely-structured datasets consisting of CSCs and features that are manually extracted from text corpora. In [Glückstad 2012-c], three strategies have been tested in the experiments: **1)** applying the IRM directly to two concept-feature matrices, respectively representing the educational domain knowledge in Japan and Denmark for first categorizing them into categorical classes that are to be afterwards compared and aligned; **2)** applying the IRM directly to a matrix where the two concept-feature matrices, respectively representing the Danish- and Japanese educational domain knowledge are merged; and **3)** applying the BMG to directly compute similarity relations between concepts in the two cultures, thereafter applying the IRM for clustering concepts in the respective cultures into categorical classes (the BMG + IRM approach). The results from the three experimental strategies indicate that the third strategy - the BMG + IRM approach - seems to be the most effective approach for not only clustering CSCs into more specific and appropriate categorical classes but also for capturing complex relationships between each of the categorical classes existing in the two cultures. In addition, the direct application of the BMG to the datasets enables one to effectively analyze further specific similarity relations between category members existing in the two cultures. These results indicate that loosely-structured datasets are sufficient for mapping CSCs in a multi-cultural context. The details are described in [Glückstad 2012-c] presented at this conference. Furthermore, [Glückstad 2012-d] indicates, from the comparative analysis of the TO combined with BMG approach (the TO + BMG approach) and the BMG + IRM approach, that the integration of all methods, i.e. the BMG + IRM + TO approach could enable not only to be mapping CSCs by respecting nuances of each concept existing in different cultures but also to construct TOs that are cross-culturally interoperable as well as mono-culturally clarified.

An interesting point emphasized by [Kemp 2006] is that researchers who start with the viewpoint of - *how the semantic knowledge should be represented as a system of relations* - often devise complex representational schemes. On the contrary, [Kemp 2006] challenges this issue from the viewpoint of - *how representations of semantic knowledge are acquired*. When reviewing the three crucial questions of learning in [Tenenbaum 2001] addressed in order to explain the BMG: ***1)*** *what constitutes the learner's knowledge about the consequential region;* ***2)*** *how the learner uses that knowledge to decide how to generalize; and* ***3)*** *how the learner can acquire that knowledge from the example encountered*, the view by [Kemp 2006] on *how representations of semantic knowledge are acquired* is seamlessly integrated in the flow of the BMG by [Tenenbaum 2001]. Although it is still not clear which ontological component is the most suitable for simulating the asymmetric cross-cultural communication scenario, the integration of the IRM as a pre-process of the ontology construction seems to be not only theoretically but also empirically compelling as future research direction.

5. Cognitive Simulations

At present, the solution identified in [Glückstad 2012-c] has delivered the most optimal results. Accordingly, by using the results obtained from the BMG + IRM approach in [Glückstad 2012-c], Figures 5 demonstrate the simulation of the asymmetric cross-cultural communication patterns shown in Figures 1 and 2 based on the educational system domain knowledge in Denmark and Japan.

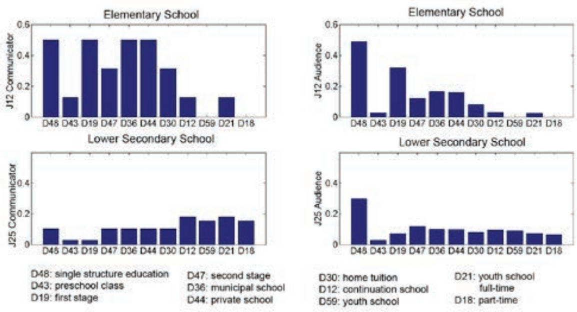


Figure 5: Cognitive simulation based on the BMG + IRM

In Figure 5, the left- and the right columns respectively represent the asymmetric cross-cultural communication illustrated in Figures 1 and 2. For example, the left-upper graph shows that a Japanese communicator who has prior knowledge of the Japanese educational system considers that "D48: single structure education", "D19: first stage", "D36: municipal school" and "D44: private school" are the most similar concepts to the Japanese elementary school. However, from the viewpoint of a Danish audience who has prior knowledge of the Danish educational system, D48 (Danish compulsory education consisting of primary and lower secondary levels) and D19 (the first part of the single structure corresponding to the primary education, however, this concept is not so common as the single structure system in Denmark) have higher relevance to the Japanese elementary school. Fascinatingly, the Japanese communicator in Figure 5 identifies "D12: continuation school (DA: *efterskole*)", "D21: youth school – full-time system" as the most similar concepts to the Japanese lower secondary school. In Denmark, the concept of "lower secondary school" does not exist, because the lower secondary level is included in the single structure education. The concepts which the Japanese identified are alternative educations targeted for young people in the age bracket of 14-17 years old. Thus, if the Japanese communicator selects "continuation school (DA: *efterskole*)" as translation for conveying meanings of the Japanese lower secondary school, the Danish audience might imagine other meanings than the ones the Japanese intended to convey. Contrary, the right-lower graph shows that "D48: single structure" is the most relevant concept to the Japanese lower secondary school from the viewpoint of the Danish audience. In this way, the cognitive simulation could potentially identify a translation candidate from an audience's viewpoint. Such a feedback function might be useful for, e.g. a pivot translation system employed for Machine Translation (MT) and Cross-Lingual Information Retrieval (CLIR).

6. Discussion and Future Implications

The COM framework introduced in this work is best described as a merger of the asymmetric cross-cultural communication scenario based on the Relevance Theory of Communication [Sperber 1986] and the Knowledge Approach [Murphy 2004] from the cognitive sciences. One question is how the domain knowledge possessed by people in different cultures, which may influence the similarity judgment and generalization process, should be defined. In the series of experiments introduced in this work, I considered the English text corpora describing the educational system in Denmark, Japan and Germany as the domain knowledge possessed by the average citizens residing in the respective countries. The argument here is that a person who has grown up in a specific country goes through the entire educational system in that particular country (although there may be exceptions due to the recent widespread globalization). Thus this person must know how the educational system in his/her specific country works. English text corpora (together with original texts in the respective local language) published by a reliable authority for each specific county must therefore be the most reliable and general prior knowledge of the average population of that country. The English text corpora produced by the respective countries are supposed to reflect their original culture. Here, we should be reminded that, although we usually use English as our second language to communicate with people outside of our own culture, our English production process is highly influenced by our mother tongue [Durst-Andersen 2011] and our cultural backgrounds [Murphy 2004], and therefore a communication is coordinated in an asymmetrical manner [Sperber 1986].

As briefly stated in the previous discussions, the question of *how representations of semantic knowledge are acquired* [Kemp 2006] is still an open question for our future research. From a cognitive scientists' viewpoint, the so-called taxonomic organization of categories is inherently rooted in a child's first language- and concept acquisitions [Murphy 2004]. This perspective is highly consistent with the Theory of Communicative Supertypes by [Durst-Andersen 2011] that *it is the mother tongue, and not a foreign language, that goes into our bodies and brains, thereby becoming internalised and automatised already at an early age*. When it comes to an adult's second language acquisition, e.g. about a newly encountered foreign object, this new information is aligned with his/her prior-knowledge, that is, the taxonomic organization of categories that has been developed since his/her childhood [Murphy 2004]. In [Durst-Andersen 2011], these disparate concepts and language acquisition phenomena are very illustratively defined as the *private voice* and the *public voice* of language.

The presented work at present is only based on the computational simulation of a cognitive framework of the asymmetrical cross-cultural communication. A future research challenge would be to investigate how a human's taxonomic organization of category is influenced by the three different kinds of so-called communicative supertypes that are inherently rooted in the grammatical type of languages defined in [Durst-Andersen 2011]. Assuming that the differences in the taxonomic organization among these three communicative supertypes are successfully identified, a second research challenge would be to identify how such differences in the taxonomic organization of categories, that are culturally rooted in one's previous experiences and linguistically rooted in one's mother tongue, influence a real cross-cultural communication scenario using English as second language or pivot. If the human data collected from the aforementioned investigations is applied to the COM framework, simulations of such cross-cultural communication scenarios will be feasible on our computers. Furthermore, it may eventually be feasible to link multi-lingual CSCs by aligning culturally-specific "*pivot*" ontologies and to realize human-intuitive MT/CLIR for bridging remote cultures. Although the BMG + IRM + TO approach is identified as the most optimal approach in our current studies, which specific components are the optimal algorithm for the COM approach is still an open question and requires further investigations.

7. Conclusions

In this paper, a new framework referred to as the Cognitive Ontology Mapping (COM) approach for ontology applications and for simulating the asymmetric relation of cross-cultural communication between a communicator and an audience, is introduced. The framework is based on two ground theories, the Relevance Theory of Communication [Sperber 1986] and the Knowledge Approach [Murphy 2004] from the cognitive science. The alignment component of the COM framework considers prior knowledge possessed by the parties involved in a cross-cultural communication as "cultural bias" based on the novel cognitive model, the Bayesian Model of Generalization (BMG) [Tenenbaum 2001]. The BMG component is theoretically and empirically integrated with the Infinite Relational Model (IRM) [Kemp 2006] in a seamless manner. The series of empirical studies introduced in this work indicates that the BMG + IRM + TO (Terminological Ontology) would be the most optimal approach not only to map CSCs by respecting nuances of each concept existing in the respective cultures, but also to construct TOs that are multi-culturally interoperable. This remains as a grand future challenge.

Acknowledgements

I would like to express my thanks to Prof. Jerker Denrell, Saïd Business School, University of Oxford, who provided me with valuable feedback which encouraged me to harness the ideas presented in this paper; Prof. Per Durst-Andersen, Copenhagen Business School for exciting and inspiring discussions in connection to the Relevance Theory of Communication and his Theory of Communicative SuperTypes; and finally Ass. Prof. Morten Mørup, Technical University of Denmark, who supported me with the technical and mathematical implementations as well as encouragements to employ the IRM in these new ways.

References

[Bush 1951] Bush, R. R., Mosteller, F. A Model for Stimulus Generalization and Discrimination., In Psychological Review, Vol., 58(6; 6), 413-423, 1951

[Chomsky 1986] Chomsky, N., Language and Problems of Knowledge: The Managua lectures. Cambridge, MA: MIT Press, 1986

[Cimiano 2010] Cimiano P., Montiel-Ponsoda E., Buitelaar P., Espinoza M., Gómez-Pérez A. A Note on Ontology Localization., Journal of Applied Ontology Vol. 5, No. 2, IOS Press, 2010.

[Declerck 2010] Declerck T., Krieger H.U., Thomas S.M., Buitelaar P., O'Riain S., Wunner T., Maguet G., McCrae J., Spohr D., Montiel-Ponsoda.E. Ontology-based Multilingual Access to Financial Reports for Sharing Business Knowledge Across Europe., In Internal Financial Control Assessment Applying Multilingual Ontology Framework, J. Rooz, J. Ivanyos, Eds. Budapest: HVG Press, 2010

[Durst-Andersen 2011] Durst-Andersen, P., Linguistic Supertypes: A Cognitive-semiotic Theory of Human Communication. Berlin-New York: De Gruyter Mouton, 2011

[Espinoza 2009] Espinoza, M., Montiel-Ponsoda, E., Gómez-Pérez, A. Ontology Localization. In Proc. the 5th International Conference on Knowledge Capture (KCAP) New York, NJ, USA, 2009

[Glückstad 2012-a] Glückstad F.K., Mørup M., Feature-based Ontology Mapping from an Information Receivers' Viewpoint, (under review)

[Glückstad 2012-b] Glückstad F.K., Cross-Cultural Concept Mappings of Standardized Datasets, (under review)

[Glückstad 2012-c] Glückstad F.K., Mørup M., Application of the Infinite Relational Model combined with the Bayesian Model of Generalization for Effective Cross-Cultural Knowledge Transfer., In Proc. The 26th Annual Conference of the Japanese Society for Artificial Intelligence, Japan, June 2012.

[Glückstad 2012-d] Glückstad F.K., Mørup M., Flexible- or Strict Taxonomic Organization?: Impact on culturally-specific knowledge transfer, In Proc. Terminology and Knowledge Engineering Conference 2012, Spain, June 2012.

[Heit 1998] Heit E., A Bayesian Analysis of Some Forms of Inductive Reasoning. In Rational Models of Cognition, Oaksford M., Charter N. Eds. Oxford: Oxford University Press, 248-274. 1998

[ISO 2000] ISO 704. Terminology Work — principles and methods, ISO, 2000

[Jaccard 1901] Jaccard, P., Étude comparative de la eistribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37: 1901

[Kemp 2006] Kemp, C., Tenenbaum, J.B., Griffiths, T.L. Yamada, T., Ueda, N., Learning Systems of Concepts with an Infinite Relational Model., In Proc. the 21st National AAAI Conference on 1:381-388, 2006.

[Lassaline 1996] Lassaline M. E., Murphy G. L., Induction and Category Coherence., In Psychonomic Bulletin & Review, 3, 95-99, 1996

[Madsen 2004] Madsen, B.N., Thomsen, H.E. and Vikner, C., Principles of a System for Terminological Concept Modelling. In Proc. the 4th International Conference on Language Resources and Evaluation. ELRA, 2004.

[Madsen 2005] Madsen, B.N., Thomsen, H.E. and Vikner, C., Multidimensionality in Terminological Concept Modelling. In Proc. of the 7th International Conference on Terminology and Knowledge Engineering, Copenhagen, 2005

[Markman 1997] Markman, A. B., Wisniewski, E. J., Similar and Different. In Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol., 23(1; 1), 54-70, 1997

[Medin 1978] Medin, D. L., Schaffer, M. M. Context Theory of Classification Learning., In Psychological Review, Vol., 85(3; 3), 207-238, May 1978

[Murphy 1994] Murphy,G. L., Allopenna, P. D. The Locus of Knowledge Effects in Concept Learning., In Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol., 20(4; 4), 904-919, 1994

[Murphy 2004] Murphy,G. L., The Big Book of Concepts. Cambridge, Massachusetts: The MIT Press, 2004

[Nord 1997] Nord, C., Translating as a Purposeful Activity. Functionalist Approaches Explained. Manchester: St. Jerome, 1997

[Rosch 1975] Rosch, E., Mervis, C. B., Family Resemblances: Studies in the Internal Structure of Categories., In Cognitive Psychology, Vol.7(4; 4), 573-605, 1975

[Shepard 1987] Shepard R.N., Towards a Universal Law of Generalization for Psychological Science. Science, 237, 1987.

[Spalding 1996] Spalding, T. L., Murphy, G. L., Effects of Background Knowledge on Category Construction., In Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol., 22(2; 2), 525-538, 1996

[Sperber 1986] Sperber, D., Wilson, D., Relevance: Communication and Cognition. Oxford: Blackwell, 1986

[Tenenbaum 1999] Bayesian Modeling of Human Concept Learning. In Advances in Neural Information Processing Systems 11, Kearns M.S., Solla S. A., Cohn D. A. Eds., Cambridge, MA: MIT Press. 59-65, 1999

[Tenenbaum 2001] Tenenbaum, J. B., & Griffiths, T. L., Generalization, Similarity, and Bayesian Inference. In Behavioral and Brain Sciences, Vol.24(4; 4), 2001

[Tversky 1977] Tversky, A., Features of Similarity. In Psychological Review, Vol., 84(4; 4), 1977

[Wisniewski 1994] Wisniewski E. J., Medin, D. L. On the Interaction of Theory and Data in Concept Learning. In Cognitive Science, 18, 221-281, 1994

English Summary

A Japanese acquaintance who has been living in Denmark for more than 40 years formulated his difficult mission of undertaking translation tasks in the following way: "*Once I deeply understood the two cultures [Denmark and Japan] and the cultural differences/nuances of conceptual meanings existing in the two countries, it became impossible for me to translate culturally-specific terms into the other language. Existing language resources [dictionaries etc.] are in this context useless*". What he was frustratingly expressing is that it becomes virtually an impossible task to precisely translate or convey the meaning of a Culturally-Specific Concept (CSC) if no exact equivalent concept exists in the Target Language (TL) culture. Despite this inherent frustration, communicators or translators are still required to convey such CSCs into a TL in an optimal manner such that a TL reader can instantly infer the original meaning of a given Source Language (SL) concept. In short, the key issue is whether there can be found a way to solve this inherently frustrating situation which even skilled human translators cannot easily cope with ?

The challenge of translating CSCs from an SL is not only caused by the absence of equivalent concepts in a TL culture, but also due to differences of the background knowledge possessed by the two parties involved in a cross-cultural communication scenario. Sperber & Wilson (1986) emphasize that, although *all humans live in the physical world*, mental representations are constructed differently due to differences in our close environment and our different cognitive abilities. Because people use different languages and have mastered different concepts, the way they construct representations and make inference is also dissimilar. Since an individual possesses a total cognitive environment that is the set of facts based on his/her perceptual ability, inferential ability, actual awareness of facts, knowledge he/she has acquired and so on, it is much easier to achieve a so-called "*asymmetric*" coordination between communicator and audience (Sperber & Wilson, 1986).

This ground theory forms the basis of the framework that is to be introduced in my thesis. As Sperber & Wilson (1986) stress, a communication is an inferential process, hence, a cross-cultural communication is with even stronger reasons based on inferences. It is challenging but also fascinating to explore how an English word used by a person in one culture is perceived, understood and conceptualized by a person coming from another part of the globe. This is the motivation for my proposal of a new framework in this thesis referred to as Culturally-specific Ontology Mapping (COM).

The uniqueness of the COM approach is to apply mathematical models derived from cognitive science to the issue having been considered as one of the most challenging topics within the research domain of Semantic-Web and multi-linguality, that is, linking of CSCs that are semantically inconsistent. The theoretical approach taken in this thesis is based on **1)** Sperber & Wilson´s Relevance Theory of Communication which considers communication as an inferential process as briefly stated above, and **2)** the Knowledge Effects proposed in the area of concept learning and categorization research in the cognitive science. In the Relevance Theory of Communication (Sperber & Wilson, 1986), an information receiver's prior-knowledge is considered as *context*, and an inferential process of new information *P* and the context *C* is called *contextualization*. In here, new knowledge generated from the union of *P* and *C* is called *contextual effect* which is the necessary condition for *relevance*. The context is selected from a set of background assumptions that are organized in an individual's *encyclopaedic memory*. In the Knowledge Effects and taxonomic organization of categories in the cognitive sciences, people *use their prior-knowledge to reason about a new object in order to decide what category it is, or in order to learn a new category* (Murphy, 2004: 60-61). Such a new category is stored in a person's mind as a concept and a plurality of such concepts are most likely connected in hierarchical networks in the brain. These hierarchical network connections are used for making inductive inferences and categorization judgments (Murphy, 2004: Chapter 7). The cognitive theories imply that an information receiver's prior-knowledge is used for his/her inferential process of learning new information and such prior-knowledge is organized in a taxonomic hierarchy in his/her memory. Supported by these ground theories, I employ a novel mathematical cognitive model called the Bayesian Model of Generalization (BMG) at the second-half of a series of empirical studies in this thesis. The BMG is originally proposed by the cognitive scientists (Tenenbaum & Griffiths, 2001) and is based on the traditional theories of the cognitive sciences, among others, Shepard's Universal Law of Generalization (Shepard, 1987) and Tversky's set-theoretic model (Tversky, 1977) and addresses crucial questions of learning processes suggested by (Chomsky, 1986). In order to apply the BMG, I consider a scenario of bidirectional communication: **a)** from an SL communicator to a TL information receiver: a translated term of an SL concept as a new object that appears to a TL receiver´s taxonomically organized prior-knowledge; and **b)** an SL communicator identifies a stimulus (that is translation) optimally relevant to a concept existing in the TL receiver´s taxonomy.

My hypothesis starts from the viewpoint that Terminological Ontology (TO), the method that has been introduced by (Madsen et al., 2004) at

CBS, contributes to identify an optimally relevant translation by assisting one to systematically organize conceptual features from domain knowledge (corresponding to the taxonomic organizations). Such systematically organized features can be used for linking an SL concept with a TL concept based on a plurality of cognitive models, among others the BMG, as algorithms of aligning two culturally-dependent taxonomies.

TO is a domain-specific ontology used for knowledge sharing, which normally is applied in terminology work within the domain of languages for special purposes. The unique points of TO that differentiate it from other types of ontologies are its feature specifications and subdivision criteria. A feature specification consists of a feature dimension and its value. Thus, a representation of a whole concept is a feature structure, i.e. a set of feature specifications corresponding to the unique set of characteristics that constitutes that particular concept. Terminological ontologists argue that concepts are defined in a language-dependent context, and therefore, TO is language- or culturally dependent. TO is developed within a knowledge sharing community, then dynamically updated and validated. If it is necessary to share knowledge with other communities, TOs developed in different communities should be compared, aligned and merged as needed. Hence, as a starting point of my work in this thesis, two culturally-dependent TOs are developed and aligned for identifying corresponding concepts existing in the two cultures under consideration.

To summarize, the initial framework of the COM approach proposed in this thesis is assumed to consist of the following modules: **1)** identification of domain specific content-aligned (or parallel) corpora, consisting of SL-English and English-TL language combinations; **2)** terms and features extractions from content-aligned English corpora which describe domain specific terms and their definitions in English for the respective domain-knowledge in the SL- and TL cultures; **3)** construction of ontologies for the respective domain-knowledge in the SL- and TL countries based on the information extracted in 2); **4)** creation of feature structures for each concept for the respective domain-knowledge in the two countries, and standardization of feature labels used in the two country-specific ontologies; **5)** alignment of the two structured feature sets based on feature-based ontology mapping algorithms applying cognitive models, i.e. Tversky´s set-theoretic model (Tversky, 1977) and eventually the BMG (Tenenbaum & Griffiths, 2001); and **6)** identification of corresponding translation candidates from the content-aligned corpora consisting of SL-English and English-TL language combinations.

With this framework as starting point, I investigate **how the background knowledge possessed by an SL communicator and a TL reader should be represented and linked in light of various cognitive processes involved in cross-cultural communication.** This research issue is considered from the viewpoint of two dimensions: **a)** what algorithms are suitable for mapping TOs? and **b)** within the defined framework of COM, what are the advantages or disadvantages of using the TO approach compared with other approaches for systematically representing and organizing conceptual features as culturally-dependent prior-knowledge?

In order to address the first sub-question posed in a), I implement several preliminary studies of mapping two culturally-dependent TOs. The empirical studies start from a very primitive and manual feature-mapping approach, gradually applying a rather basic cognitive model, Tversky's set-theoretic model (Tversky, 1977) by assigning three different patterns of parameter settings, and eventually identifying a further advanced algorithm that has been well-recognized as a novel cognitive model, the BMG (Tenenbaum & Griffiths, 2001) as the best possible algorithm for the COM framework. The fourth and fifth empirical studies in Chapter 5 indicate that the BMG with the "strong sampling scheme" sensitively reflects the feature structure of concepts in comparison, and both intuitively and effectively identify the most similar concepts by reflecting on the learner's prior knowledge.

For addressing the second sub-question posed in b), I implement two types of comparative analyses by applying the Infinite Relational Model (IRM) (Kemp et al., 2006), a novel unsupervised machine learning method, to loosely structured datasets consisting of CSCs and features that are manually extracted from text corpora. In the sixth empirical study in Chapter 5, three strategies of applying the IRM are first being tested: **1)** applying the IRM directly to two CSC-feature matrices, respectively representing the educational domain knowledge in Japan and Denmark for first categorizing them into categorical classes that are to be subsequently compared and aligned; **2)** applying the IRM directly to a matrix where the two CSC-feature matrices, respectively representing the Danish- and Japanese educational domain knowledge are merged; and **3)** applying the BMG to directly compute similarity relations between CSCs in the two cultures, thereafter applying the IRM for clustering CSCs in the respective cultures into categorical classes (the BMG + IRM approach). The results from the three experimental strategies indicate that the third strategy - the BMG + IRM approach - seems to be the most effective approach for not only clustering CSCs into more specific and appropriate categorical classes but also for capturing complex relationships between each of the categorical classes existing in the two cultures. In addition, the direct application of the BMG to the datasets enables me to effectively analyze further specific similarity relations between category members existing in the two cultures. These results indicate that loosely-structured datasets seem to be sufficient for mapping CSCs in a multi-cultural context. The seventh empirical study further analyze what are the advantages and disadvantages of representing semantic knowledge in accordance with rather strict principles of constructing ontologies for the purpose of cross-cultural knowledge transfer. More specifically, CSC mapping results obtained from the application of the BMG to datasets obtained from TOs (the TO + BMG approach) are compared against the direct application of the BMG to datasets obtained from text corpora, of which results are subsequently clustered by the IRM application (the BMG + IRM approach). The results indicate that the TO + BMG approach, due to its strict principles, has a disadvantage of delimiting the dimensions that may potentially be important for the CSC mapping. Although some feature specifications are considered as very important in several categorical classes, the TO principles allow me to use them only once in the ontology. For example, features defining the academic degree system such as "bachelor" and "master" are very common and decisive features that could be useful for mapping CSCs existing in different cultures. However, although there are several types of Bachelor- or Master degree programmes provided as e.g. "university education" under "tertiary education" and as "open education" under "continuing education for adults" in the Danish educational system, the TO principles allow me to use this only once in the ontology. Terminological ontologists argue that this problem is solved by creating empty concepts (nodes) respectively possessing attribute-value pairs "degree: Bachelor" and "degree: Master" at a higher level of the ontology, from which the two "Bachelor" concepts belonging to different superordinate concepts can inherit "degree: Bachelor" in a poly-hierarchical structure. In addition, it should be emphasized that one of the advantages of TO is the clarification of domain knowledge in a specific knowledge community and the ability to compare the *hierarchical structures* across different communities. The aforementioned principle of poly-hierarchical structures enables me to generate a new concept in one community which exists in the other community. Such concept generation is not feasible based on the similarity computation approach, i.e. the BMG + IRM approach. On the contrary, the BMG combined with the IRM, both of which originate from the cognitive sciences, is designed to compute the fuzziness of the human mind by reflecting prior knowledge of a learner who compares a new object with something he/she knows in advance.

To sum up, the series of empirical studies shown in this thesis identify the BMG (Tenenbaum & Griffiths, 2001) as the most suitable alignment component for the COM framework, which considers prior knowledge possessed by the parties involved in a cross-cultural communication as "cultural bias". The BMG component is theoretically and empirically integrated with the IRM (Kemp et al., 2006) in a seamless manner. In addition, the IRM can effectively be used for the automatic construction of feature-based ontologies. The series of empirical studies introduced in this thesis indicates that the BMG + IRM + TO is potentially the most optimal approach not only to map CSCs by respecting nuances of each concept existing in the respective cultures, but also to construct TOs, each of which clarifies knowledge in a specific knowledge community in a multi-culturally interoperable manner.

Kort dansk resumé

En af mine japanske bekendte, der har boet i Danmark i mere end 40 år, formulerede sine vanskeligheder når han foretager oversættelses-opgaver på følgende måde: "Efter jeg omsider havde opnået en dybere forståelse for de to kulturer [Danmark og Japan] og forstod grundlæggende kulturelle forskelle og nuancer og deres konceptuelle betydninger i de to lande, blev det næsten umuligt for mig at oversætte noget specifikt kulturelt betinget til det andet sprog. Eksisterende oversættelses-ressourcer [såsom ordbøger m.m.] er i denne sammenhæng nærmest ubrugelige". Hvad han frustreret prøvede at tilkendegive er, at det bliver en nærmest umulig opgave at præcist oversætte eller formidle betydningen af et kulturelt specifikt begreb (CSC), hvis ikke der forefindes et nøjagtigt tilsvarende begreb i målsproget (TL). På trods af denne betydelige frustration, er disse oversættere stadig forpligtet til at formidle et sådant CSC til et TL på en hensigtsmæssig måde, således at en TL-modtager med det samme kan udlede den oprindelige betydning af et givent kildesprogs- (SL) begreb. Kort sagt, er det centrale spørgsmål, hvorvidt der kan findes en måde at løse denne iboende og stærkt frustrerende situation, hvor selv en garvet og erfaren oversætter nærmest må give op?

Udfordringen i at oversætte CSC'er fra et kildesprog er ikke kun forårsaget af fraværet af tilsvarende begreber i en TL-kultur, men også forårsaget af de mange iboende forskelle i baggrundsviden og miljø mellem de kommunikerende parter. Sperber & Wilson (1986) understreger i deres artikel, at selv om vi alle lever i den samme fysiske verden, er vores lokale mentale repræsentationer opbygget forskelligt på grund af nuancerede forskelle i vores nær-miljø og vores forskellige kognitive forudsætninger. Eftersom forskellige kulturer har forskellige sprog og forskellige begreber er den måde man konstruerer repræsentationer og foretager slutninger på naturligvis også forskellig. Da den enkelte besidder et samlet kognitivt udfaldsrum der er baseret på hans/hendes perceptuelle forudsætninger, empiriske evner, bevidsthed om fakta, baggrundsviden osv., er det mere hensigtsmæssigt at opnå en såkaldt "asymmetrisk" koordination mellem afsender og modtager iflg. (Sperber & Wilson, 1986).

Denne basis-teori danner grundlag for de forskningsrammer, der bliver introduceret i min afhandling. Som Sperber & Wilson understreger i deres berømte bog, er enhver kommunikation eller overførsel af information, en empiriske proces, og dermed er enhver tværkulturel kommunikation utvivlsomt baseret på følgeslutninger. Det er udfordrende, men også fascinerende, at undersøge, hvordan et engelsk ord eller begreb, anvendt af en person i én kultur, opfattes, fortolkes og begrebsliggøres af en person fra en helt anden del af verden. Dette er hoved-motivationen for mit forslag i denne afhandling om et nyt forskningstema, som jeg har betegnet kulturspecifik ontologi-mapning (COM).

Et af de centrale forsknings-spørgsmål i denne afhandling er hvordan man repræsenterer baggrundsviden for både afsender og modtager i lyset af de forskellige kognitive processer, der er involveret i en tværkulturel kommunikation. Udgangs-hypotesen starter med antagelsen om at en terminologisk ontologi (TO) (Madsen et al., 2004) potentielt kan bruges som baggrunds-viden for kulturelt bestemt domæne-viden, som er nødvendig for følgeslutnings-processer i et tværkulturelt kommunikations-scenarie. Dernæst betragtes flere forskellige kognitive modeller der stammer fra forskning i kategorisering og koncept-læring som potentielle ontologi-mapningssalgoritmer, der afspejler følgeslutnings-processerne involveret i tværkulturel kommunikation.

Rækken af empiriske undersøgelser i denne afhandling identificerer metoden Bayesian Model for Generalisering (BMG) (Tenenbaum & Griffiths, 2001), som den bedst egnede justerings-teknik til COM-metoden, der indbefatter og forholder sig til forudgående viden som "kulturel bias" for de involverede parter i et tværkulturelt kommunikations-scenarie. BMG er teoretisk og empirisk integreret med den såkaldte uendelige relationelle model (IRM) (Kemp et al., 2006) på en hensigtsmæssig og problemfri måde. Rækken af empiriske undersøgelser viser således, at kombinationen BMG + IRM + TO potentielt giver den mest optimale tilgang for at adressere udfordringerne i denne afhandlings problemformulering. Fremgangsmåden respekterer ikke blot nuancer i ethvert koncept, der forefindes i de respektivt betragtede kulturer, men tilvejebringer endvidere terminologiske ontologier, der er multi-kulturelt interoperable.

Bibliography

Aguado-de-Cea, G. Gómez-Pérez, A., Montiel-Ponsoda, E. & Vilches-Blázquez, L.M. (2010). Rivière or Fleuve? Modelling Multilinguality in the Hydrographical Domain. *Proceedings of the Workshop on the Multilingual Semantic Web at the 19th International World Wide Web Conference,* Raleigh, NC, USA.

Ahn, W. (1990). Effects of Background Knowledge on Family Resemblance Sorting. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum. 149-156.

Bush, R. R. & Mosteller, F. (1951). A Model for Stimulus Generalization and Discrimination. *Psychological Review, 1951, Vol., 58*(6; 6), 413-423.

Cabré Castellví, M. T. (2000). Elements for a Theory of Terminology, Towards an Alternative Paradigm. *Terminology / 2000, Vol.6 , Nr.1,* Amsterdam, Netherlands: John Benjamins Publishing Company.

Cimiano, Philipp, Montiel-Ponsoda, Elena, Buitelaar, Paul, Espinoza, Mauricio & Gómez-Pérez Asunción. (2010). A Note on Ontology Localization. *Journal of Applied Ontology* Vol. 5, No. 2, IOS Press, 127-137.

Chomsky, N. (1986). *Language and Problems of Knowledge: The Mnagua Lectures.* Cambridge, MA: MIT Press.

Collins, Allan M. and Quillian & Ross M. (1969). Retrieval Time from Semantic Memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 241-248.

Cuadros Montse, Laparra Egoitz, Rigau German, Vossen Piek & Bosma Wauter. (2010). Integrating a Large Domain Ontology of Species into WordNet. *Proceedings of the 7th international conference on Language Resources and Evaluation (LREC 2010)*, N.Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias, Eds. European Language Resources Association (ELRA), 2310-2317.

Declerck, T., Krieger, H.U., Thomas, S.M., Buitelaar, P., O´Riain, S., Wunner, T., Maguet, G., McCrae, J., Spohr, D. & Montiel-Ponsoda, E. (2010) Ontology-based Multilingual Access to Financial Reports for Sharing Business Knowledge Across Europe. *Internal Financial Control Assessment Applying Multilingual Ontology Framework*, J. Rooz, J. Ivanyos, Eds. Budapest: HVG Press, 67-76.

de Souza K.X.S. & Davis J. (2004). Aligning Ontologies and Evaluating Concept Similarities, *On the Move to Meaningful Internet Systems 2004: Lecture Notes in Computer Science, Volume 3291*, Springer.

Durst-Andersen, P. (2011). *Linguistic Supertypes: A Cognitive-semiotic Theory of Human Communication*. Berlin-New York: De Gruyter Mouton.

Durst-Andersen, P. (2011). *Bag of Sproget: Det kulturalmentale univers i sprog og kommunikation*. Copanhagen, Denmark: Hans Reitzels Forlag.

Espinoza, M., Montiel-Ponsoda, E., & Gómez-Pérez, A. (2009). Ontology Localization. *Proceedings of the 5th International Conference on Knowledge Capture (KCAP)* New York, NJ, USA, 33-40.

Euzenat, J. & Shvaiko, P. (2007). *Ontology Matching*. Berlin Heidelberg: Springer-Verlag.

Fellbaum, C. (Ed.) (1998). *WordNet: An Electronic Lexical Database.* Cambridge, MA: MIT Press.

Frank, M. C., Goodman, N. D., Lai, P. & Tenenbaum, J. B. (2009). Informative Communication in Word Production and Word Learning. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society.*

Gluck, M. A. & Bower, G. H. (1988). From Conditioning to Category Learning: An Adaptive Network Model. *Journal of Experimental Psychology: Genral, 117,* 227-247.

Glückstad, F.K. (2011) Application of Classical Psychological Theory to Terminological Ontology Alignment. *Proceedings of NLPCS 2011*, August 20-21, Copenhagen Business School. / *Copenhagen Studies in Language*, Nr. 41, 2011, 227-238.

Glückstad, F.K. (2011). Alignment of Remote Cultures. *Proceedings of 2nd International Conference on Culture and Computing*. / red. Juan E. Guerrero. Los Alamitos: IEEE Computer Society Press, 21-26.

Glückstad, F.K. (2011). Asymmetric Similarity and Cross-Cultural Communication Process. *Proceedings of the 9th International Conference on Terminology and Artificial Intelligence*. 8-10 November 2011, Paris, France. Paris : Institut National des Langues et Civilisations Orientales 59-65.

Glückstad F.K. & Mørup M. (2012). Feature-based Ontology Mapping from an Information Receivers' Viewpoint. *Proceedings of NLPCS 2012 at the 14th International Conference on Enterprise Information Systems (ICEIS),* 28 July, Wroclaw, Poland.

Glückstad F.K. (2012). Cross-Cultural Concept Mappings of Standardized Datasets. *Proceedings of Workshop of Challenges to Knowledge Representation in Multilingual Contexts at the Terminology and Knowledge Engineering Conference 2012*, Madrid, Spain, June 19-22, 2012.

Glückstad F.K. & Mørup M. (2012) Application of the Infinite Relational Model Combined with the Bayesian Model of Generalization for Effective Cross-Cultural Knowledge Transfer. *Proceedings of the 26th Annual Conference of the Japanese Society for Artificial Intelligence*, June 2012, Yamaguchi, Japan.

Glückstad F.K. & Mørup M. (2012). Flexible- or Strict Taxonomic Organization?: Impact on Culturally-Specific Knowledge Transfer. *Proceedings of the 10th International Conference on Terminology and Knowledge Engineering 2012*, June 2012, Madrid, Spain.

Glückstad F.K. (2012). Bridging Remote Cultures: Influence of Cultural Prior-Knowledge in Cross Cultural Communication. *Proceedings of the 26th Annual Conference of the Japanese Society for Artificial Intelligence*, June 2012, Yamaguchi, Japan.

Goodman, N. (1965). *Fact, Fiction, and Forecast* (2nd ed.). Indianapolis: Bobbs-Merrill.

Grice, H. P. (1957). Meaning. *Philosophical Review* 66: 377-88. Reprinted in Steinberg and Jakobovits 1971: 53-9 and Grice 1989: 213-23.

Guut, E.A. (1990). A Theoretical Account of Translation – without a translation theory. *Target: International Journal of Translation Studies*, 2/2, 135-164.

Hansen T.J., Mørup M., & Hansen L. K. (2011). Non-parametric Co-clustering of Large Scale Sparse Bipartite Networks on the GPU, *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, IEEE.

Haussler, D., Kearns, M. & Schapire, R. E. (1994). Bounds on the Sample Complexity of Bayesian Learning Using Information Theory and the vc Dimension. *Machine Learning, 14,* 83-113.

Heit, E. & Rubinstein, J. (1994). Similarity and Property Effects in Inductive Reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 1994, Vol., 20*(2; 2), 411-422.

Heit, E. (1998). A Bayesian Analysis of Some Forms of Inductive Reasoning. In: M. Oaksford & N. Chater (Eds.), *Rational Models of Cognition.* Oxford: Oxford University Press. 248-274.

Huang H.H. & Kuo Y.H. (2010). Cross-lingual Document Representation and Semantic Similarity Measure: A fuzzy set and rough set approach, *IEEE Transaction on Fuzzy Systems, Volume 18:6*, IEEE.

Ishida, T. (2006) Language Grid: An Infrastructure for Intercultural Collaboration. *Proceedings of IEEE/IPSJ Symposium on Applications and the Internet* SAINT 06, 96-100.

ISO 704. (2000). *Terminology Work — Principles and Methods*.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles* 3**7**: 547–579.

Kageura K. (2002). *Dynamics of Terminology*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Kaplan, A. S. & Murphy, G. L. (2000). Category Learning with Minimal Prior Knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 2000, Vol., 26*(4; 4), 829-846.

Keil F. C. (1989). *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.

Kemp, C., Tenenbaum, J.B., Griffiths, T.L. Yamada, T. & Ueda, N. (2006). Learning Systems of Concepts with an Infinite Relational Model. *The Twenty-First National Conference on Artificial Intelligence, 2006.*

Kruschke, J. K. (1996). Base Rates in Category Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 1996, Vol., 22*(1; 1), 3-26.

Lassaline, Mary E & Murphy G. L. (1996). Induction and Category Coherence. *Psychonomic Bulletin & Review, 3*, 95-99.

Lassen, T. (2012). A Corpus Compilation and Processing Prototype for Terminology Work. *Proceedings of the 10th International Conference on Terminology and Knowledge Engineering 2012*, Madrid, Spain.

Madsen, B.N., Thomsen, H.E. & Vikner, C. (1999). The Project 'Computer-Aided Ontology Structuring' (CAOS). In: Steffen Leo Hansen (ed.): *Copenhagen Studies in Language 23, World Knowledge and Natural Language Analysis*. Samfundslitteratur, 9-38.

Madsen, B.N., Thomsen, H.E. & Vikner, C. (2004a). Principles of a System for Terminological Concept Modelling. *Proceedings of* *the 4th International Conference on Language Resources and Evaluation*. ELRA, 15-19.

Madsen, B.N., Thomsen, H.E. & Vikner, C. (2004b). Comparison of Principles Applying to Domain Specific Versus General Ontologies. *Proceedngs of Ontologies and Lexical Ressources in Distributed Environments 2004*. ELRA, 90-95.

Madsen, B.N., Thomsen, H.E. & Vikner, C. (2005). Multidimensionality in Terminological Concept Modelling. *Proceedings of the 7th International Conference on Terminology and Knowledge Engineering*, Copenhagen, 161-173.

Madsen N.B., Thomsen H.T., Halskov J. & Lassen T. (2010). Automatic Ontology Construction for a National Term Bank, *Proceedings of Terminology and Knowledge Engineering Conference 2010,* Dublin, Ireland.

Markman, A. B. & Wisniewski, E. J. (1997). Similar and Different. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 1997, Vol., 23*(1; 1), 54-70.

Masolo, C., Borgo, S., Gangemi, A., Guriano, N. & Oltramari, A. (2003). *WonderWeb Deliverable D18: Ontology Library,* ISTC-CNR, Treno, Italy.

Medin, D. L. & Edelson, S. M. (1988). Problem Structure and the Use of Base-Rate Information from Experience. *Journal of Experimental Psychology: General, 1988, Vol., 117*(1; 1), 68-85.

Medin, D. L. & Schaffer, M. M. (1978). Context Theory of Classification Learning. *Psychological Review, May, 1978, Vol., 85*(3; 3), 207-238.

Medin, D. L., Wattenmaker, W. D. & Hampson, S. E. (1987). Family Resemblance, Conceptual Cohesiveness, and Category Construction. *Cognitive Psychology, 1987, Vol.19*(2; 2), 242-279.

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communication of the ACM* 38:39-4

Mitchell, T.M. (1997). *Machine Learning*. New York: McGraw Hill.

Murphy,G. L. (2004). *The Big Book of Concepts*. Cambridge, Massachusetts: The MIT Press.

Murphy, G. L. & Allopenna, P. D. (1994). The Locus of Knowledge Effects in Concept Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 1994, Vol., 20*(4; 4), 904-919.

Murphy, G. L. & Medin, D. L. (1985). The Role of Theories in Conceptual Coherence. *Psychological Review, 1985, Vol., 92*(3; 3), 289-316.

Murphy, G. L. & Wisniewski, E. J. (1989). Feature Correlations in Conceptual Representations. In Tiberghien G. (Ed.), *Advances in cognitive science* (Vol. 2: Theory and Applications ed., pp. 23-45). Chichester: Ellis Harwood.

Mørup M., Madsen K.H., Dogonowski A.M., Siebner H. & Hansen L.K. (2010) Infinite Relational Modeling of Functional Connectivity in Resting State fMRI., In *Neural Information Processing Systems 2010.*

Niles, I. & Pease, A. (2001). Formal Ontology in Information Systems. *Proceedings of the International Conference on Formal Ontology in Information Systems, Vol. 2, No. 2,* 159-179.

Nord, C. (1997). *Translating as a Purposeful Activity. Functionalist Approaches Explained.* Manchester: St. Jerome.

Nosofsky, R. M. (1998). Similarity, Frequency, and Category Representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 1988, Vol., 14*(1; 1), 54-65.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A. & Shafir, E. (1990). Category-Based Induction. *Psychological Review, 1990, Vol., 97*(2; 2), 185-200.

Pitman J. (2002). Combinatorial Stochastic Processes. *Notes for Saint Flour Summer School.*

Regehr, G. & Brooks, L. R. (1995). Category Organization in Free Classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 1995, Vol., 21*(2; 2), 347-363.

Rips, L. J. (1975). Inductive Judgments about Natural Categories. *Journal of Verbal Learning and Verbal Behavior, 14.* 665-681.

Rips, L. J., Shoben, E. J. & Smith, E. E. (1973). Semantic Distance and the Verification of Semantic Relations. *Journal of Verbal Learning and Verbal Behavior, 12*(1), 1-20.

Rosch, E. & Mervis, C. B. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology, 1975, Vol.7*(4; 4), 573-605.

Rosch, E., Simpson, C. & Miller, R. S. (1976). Structural Bases of Typicality Effects. *Journal of Experimental Psychology: Human Perception and Performance, 1976, Vol.,2*(4; 4), 491-502.

Rosch, E. (1978). Principles of Categorization. In: E. Rosch and B.B. Lloyd (Eds.), *Cognition and Categorization*. Hillsdale, NJ:Erlbaum, 27-48.

Shannon, C. & Weaver, W. (1949). *The Mathematical Theory of Communication.* Urbana, IL: University of Illinois Press.

Shepard, R. N. (1987). Towards a Universal Law of Generalization for Psychological Science. *Science, 237*, 1317-1323

Sloman, S. A. (1998). Categorical Inference is not a Tree: The myth of inheritance hierarchies. *Cognitive Psychology, 1998, Vol.35*(1; 1), 1-33.

Spalding, T. L. & Murphy, G. L. (1996). Effects of Background Knowledge on Category Construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 1996, Vol., 22*(2; 2), 525-538.

Sperber, D. & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Suárez-Figueroa, M. & Gómez-Pérez, A. (2008). First Attempt Towards a Standard Glossary of Ontology Engineering Terminology. *Proceedings of the 8th International Conference on Terminology and Knowledge Engineering (TKE 2008),* Copenhagen, Denmark.

Tan P.N., Steinbach M. & Kumar V. (2006). *Introduction to Data Mining*. Boston, MA: Pearson Education, Inc.

Temmerman, R. (2000). *Towards New Ways of Terminology Description, the Sociocognitive Approach*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Tenenbaum, J. B. (1999). Bayesian Modeling of Human Concept Learning. In: Kearns, M. J., Solla, S. A., and Cohn,D. A. (Eds.), *Advances in Neural Information Processing Systems.*

Tenenbaum, J. B. & Griffiths, T. L. (2001). Generalization, Similarity, and Bayesian Inference. *Behavioral and Brain Sciences, 2001, Vol.24*(4; 4), 629-640.

Thompson, L. A. (1994). Dimensional Strategies Dominate Perceptual Classification. *Child Development, Dec., 1994,Vol.65*(6; 6), 1627-1645.

Tversky, A. (1977). Features of Similarity. *Psychological Review, Vol., 84*(4; 4), 327-352.

Vossen, Piek. (2004). EuroWordNet: A multilingual database of autonomous and language-specific wordnets connected via an interlingual index. *International Journal of Lexicography* 17.2, 161-173.

Vossen, Piek., Agirre, Eneko., Calzolari, Nicoletta., Fellbaum, Christiane., Hsieh, Shu-kai., Huang, Chu-Ren., Isahara, Hitoshi., Kanzaki, Kyoko., Marchetti, Andrea., Monachini, Monica., Neri, Federico., Raffaelli, Remo., Rigau, German., Tescon, Maurizio. & VanGent, Joop. (2008). KYOTO: A system for mining, structuring and distributing knowledge across languages and cultures. *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Morocco, 1462-1469.

Vossen, Piek, & Rigau, German (2010). Division of Semantic Labour in the Global WordNet Grid. *Proceedings of the 5th Global WordNet Conference GWC'10*.

Ward, T. B., Becker, A. H., Hass, S. D. & Vela, E. (1991). Attribute Availability and the Shape Bias in Children's Category Generalization. *Cognitive Development, 6*, 143-167.

Wisniewski,Edward J. & Medin, D. L. (1994). On the Interaction of Theory and Data in Concept Learning. *Cognitive Science, 18*, 221-281.

Wüster E. (1959/1960). Das worten der weld, schaubildlich und terminologisch dargestellt. *Sprachforum, 3*(3), 183-204.

Xu Z., Tresp V., Yu K. & Kriegel H.P. (2006). Infinite Hidden Relational Models, *Proceedings of 22nd Conf. on Uncertainty in Artificial Intelligence (UAI'06)* Cambridge, MA.

Enlarged figures

Enlarged figures

Figure 2-1: Dimensions of ontology localization

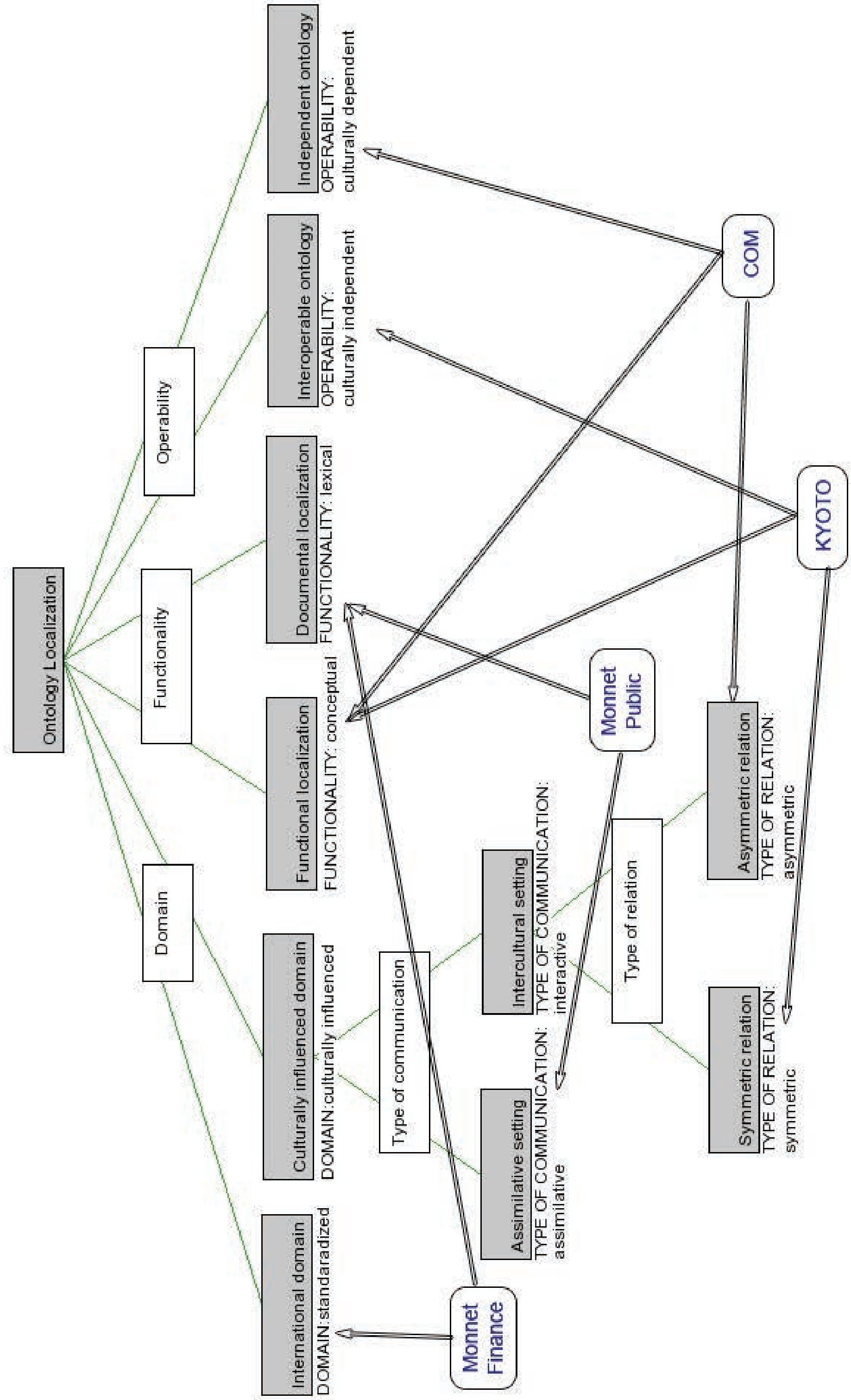


Figure 3-4: Image of ontology mapping: the SL-oriented translation

Figure 3-6: Image of ontology mapping: the TL-oriented translation

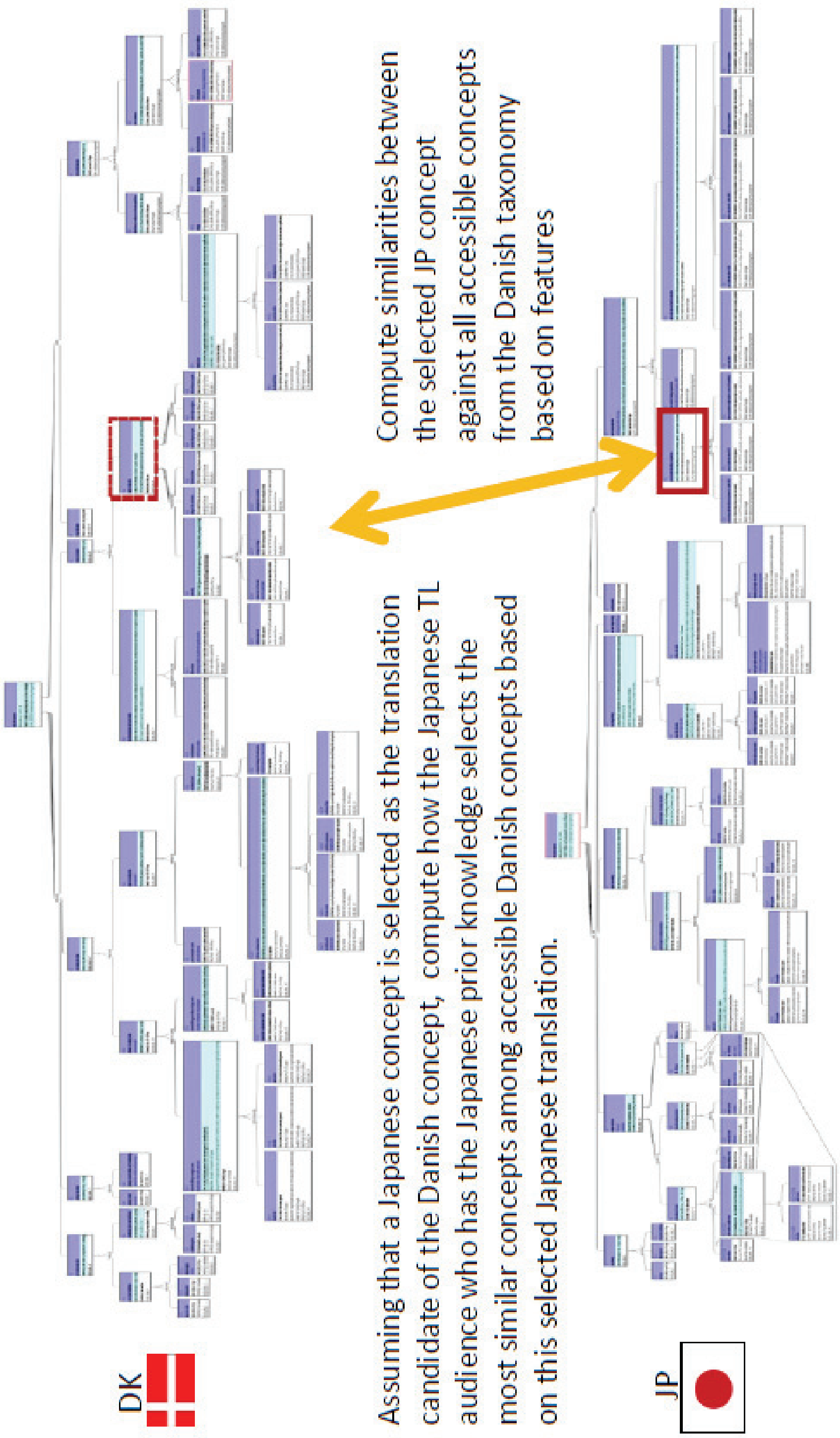


Figure 4-3: ISCED categorization (UK/DK/GE)

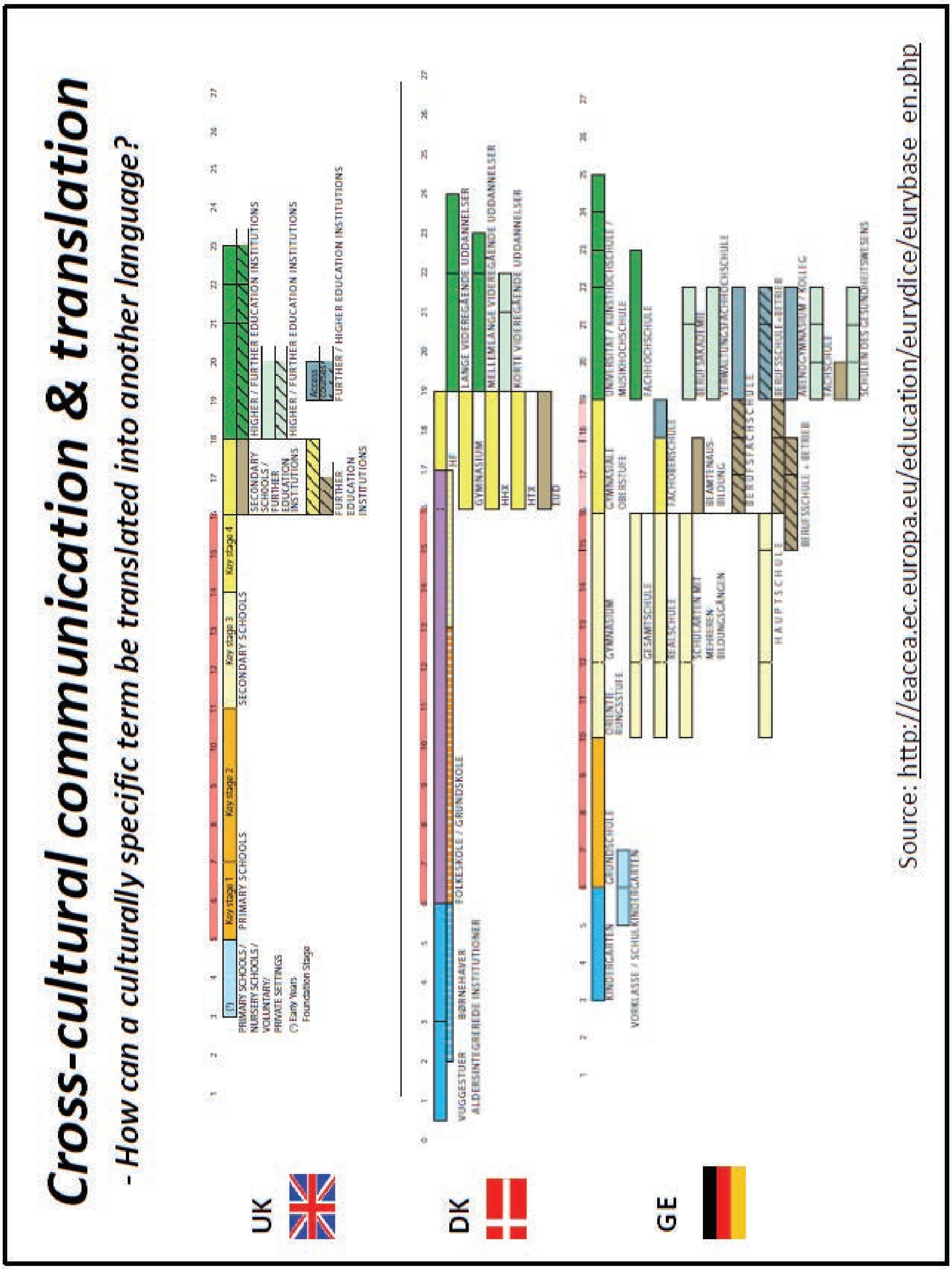


Figure 4-2: Terminological Ontology example

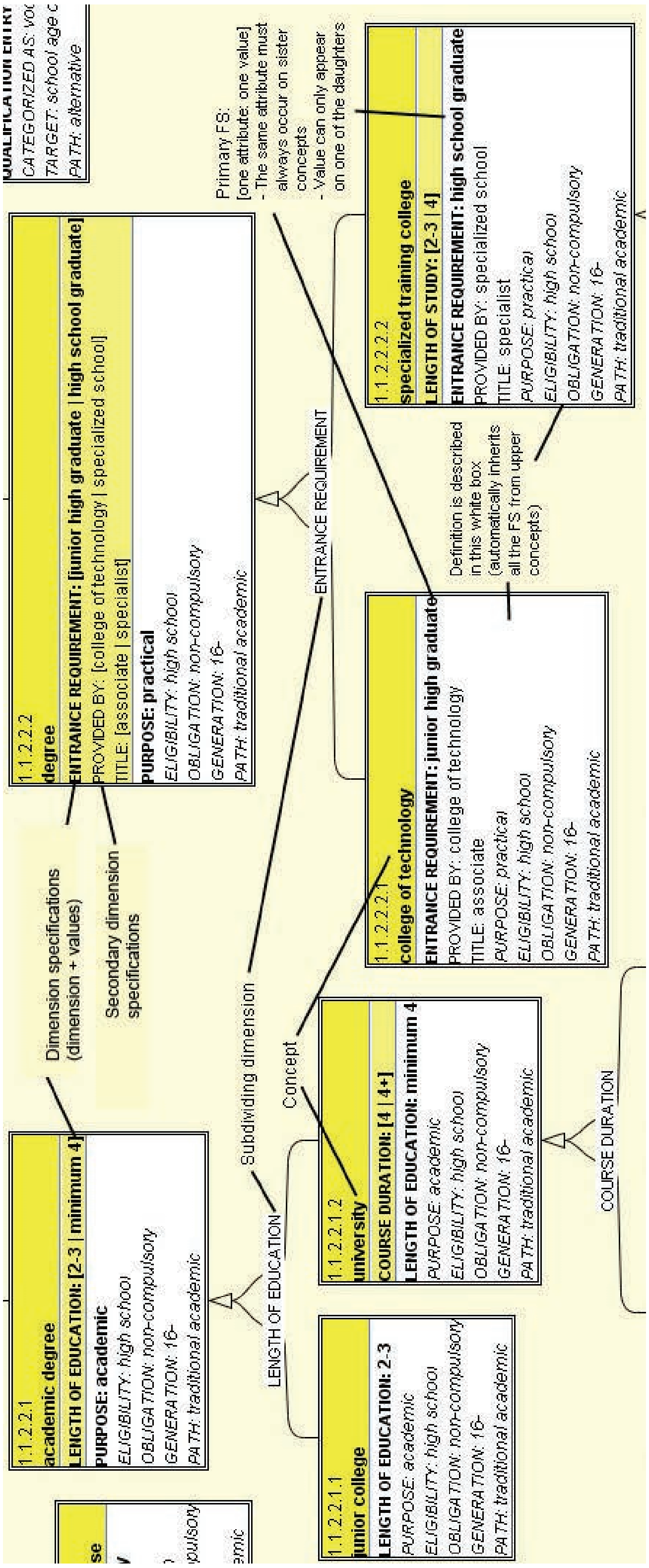


Figure 6-2: Probability of generalizing Danish "kindergarten" based on a German reader's prior knowledge

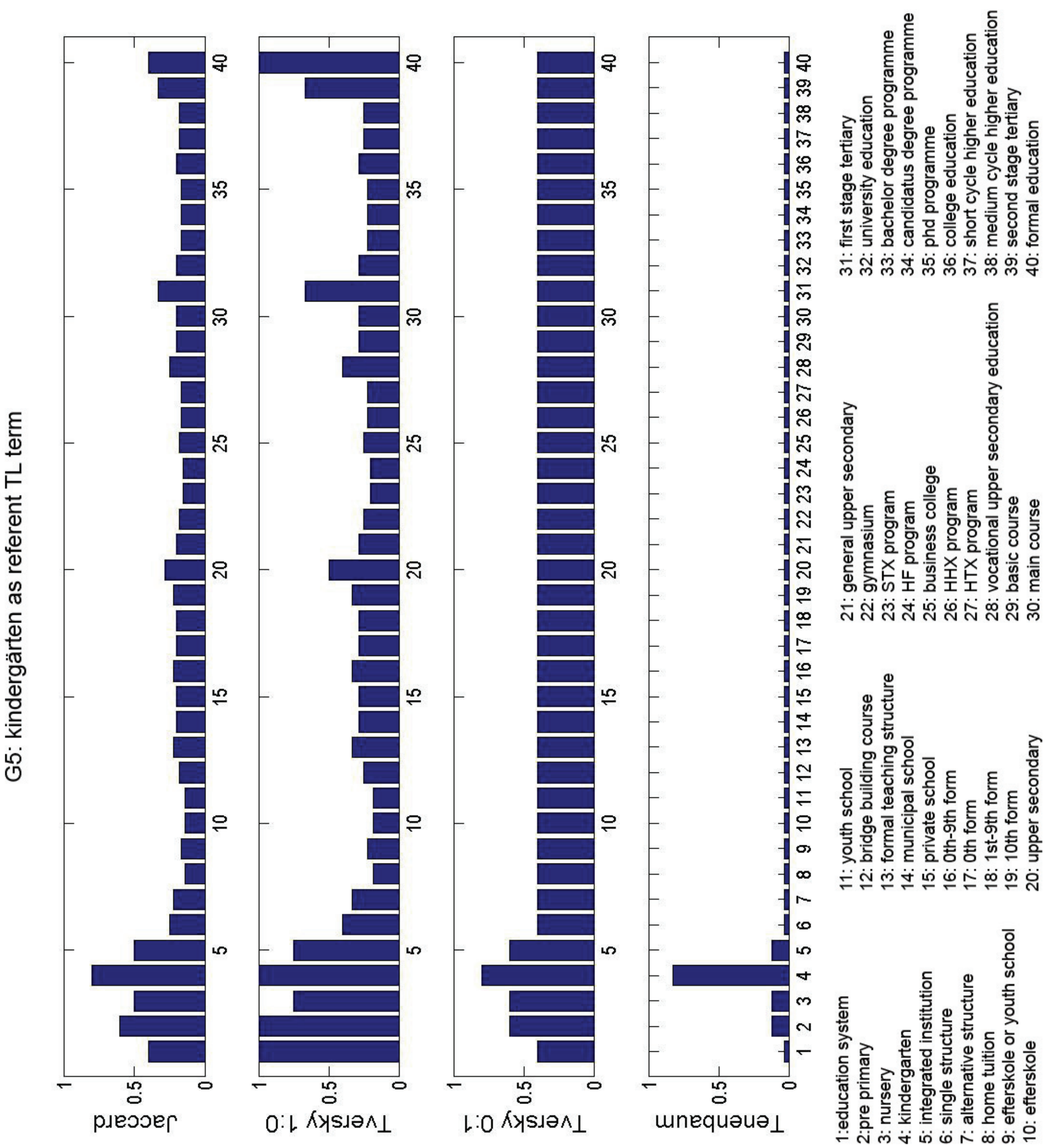
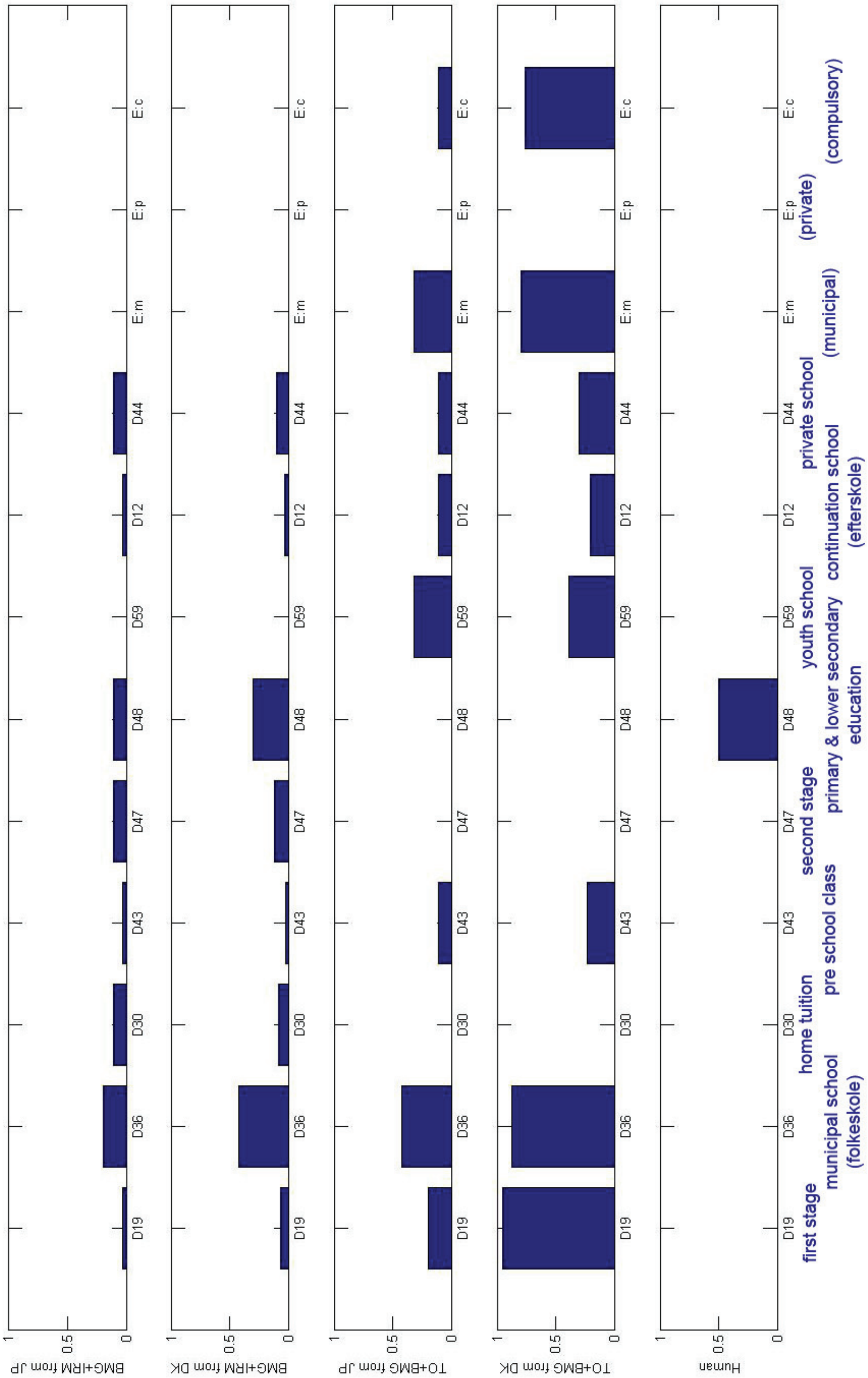


Figure 6-4: Human mapping - J12 elementary school



TITLER I PH.D.SERIEN:

2004

1. Martin Grieger
Internet-based Electronic Marketplaces and Supply Chain Management

2. Thomas Basbøll
LIKENESS
A Philosophical Investigation

3. Morten Knudsen
Beslutningens vaklen
En systemteoretisk analyse of moderniseringen af et amtskommunalt sundhedsvæsen 1980-2000

4. Lars Bo Jeppesen
Organizing Consumer Innovation
A product development strategy that is based on online communities and allows some firms to benefit from a distributed process of innovation by consumers

5. Barbara Dragsted
SEGMENTATION IN TRANSLATION AND TRANSLATION MEMORY SYSTEMS
An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process

6. Jeanet Hardis
Sociale partnerskaber
Et socialkonstruktivistisk casestudie af partnerskabsaktørers virkelighedsopfattelse mellem identitet og legitimitet

7. Henriette Hallberg Thygesen
System Dynamics in Action

8. Carsten Mejer Plath
Strategisk Økonomistyring

9. Annemette Kjærgaard
Knowledge Management as Internal Corporate Venturing
– a Field Study of the Rise and Fall of a Bottom-Up Process

10. Knut Arne Hovdal
De profesjonelle i endring
Norsk ph.d., ej til salg gennem Samfundslitteratur

11. Søren Jeppesen
Environmental Practices and Greening Strategies in Small Manufacturing Enterprises in South Africa
– A Critical Realist Approach

12. Lars Frode Frederiksen
Industriel forskningsledelse
– på sporet af mønstre og samarbejde i danske forskningsintensive virksomheder

13. Martin Jes Iversen
The Governance of GN Great Nordic
– in an age of strategic and structural transitions 1939-1988

14. Lars Pynt Andersen
The Rhetorical Strategies of Danish TV Advertising
A study of the first fifteen years with special emphasis on genre and irony

15. Jakob Rasmussen
Business Perspectives on E-learning

16. Sof Thrane
The Social and Economic Dynamics of Networks
– a Weberian Analysis of Three Formalised Horizontal Networks

17. Lene Nielsen
Engaging Personas and Narrative Scenarios – a study on how a user-centered approach influenced the perception of the design process in the e-business group at AstraZeneca

18. S.J Valstad
Organisationsidentitet
Norsk ph.d., ej til salg gennem Samfundslitteratur

19. Thomas Lyse Hansen
Six Essays on Pricing and Weather risk in Energy Markets

20. Sabine Madsen
Emerging Methods – An Interpretive Study of ISD Methods in Practice

21. Evis Sinani
The Impact of Foreign Direct Investment on Efficiency, Productivity Growth and Trade: An Empirical Investigation

22. Bent Meier Sørensen
Making Events Work Or, How to Multiply Your Crisis

23. Pernille Schnoor
Brand Ethos
Om troværdige brand- og virksomhedsidentiteter i et retorisk og diskursteoretisk perspektiv

24. Sidsel Fabech
Von welchem Österreich ist hier die Rede?
Diskursive forhandlinger og magtkampe mellem rivaliserende nationale identitetskonstruktioner i østrigske pressediskurser

25. Klavs Odgaard Christensen
Sprogpolitik og identitetsdannelse i flersprogede forbundsstater
Et komparativt studie af Schweiz og Canada

26. Dana B. Minbaeva
Human Resource Practices and Knowledge Transfer in Multinational Corporations

27. Holger Højlund
Markedets politiske fornuft
Et studie af velfærdens organisering i perioden 1990-2003

28. Christine Mølgaard Frandsen
A.s erfaring
Om mellemværendets praktik i en transformation af mennesket og subjektiviteten

29. Sine Nørholm Just
The Constitution of Meaning
– A Meaningful Constitution?
Legitimacy, identity, and public opinion in the debate on the future of Europe

2005

1. Claus J. Varnes
Managing product innovation through rules – The role of formal and structured methods in product development

2. Helle Hedegaard Hein
Mellem konflikt og konsensus
– Dialogudvikling på hospitalsklinikker

3. Axel Rosenø
Customer Value Driven Product Innovation – A Study of Market Learning in New Product Development

4. Søren Buhl Pedersen
Making space
An outline of place branding

5. Camilla Funck Ellehave
Differences that Matter
An analysis of practices of gender and organizing in contemporary workplaces

6. Rigmor Madeleine Lond
Styring af kommunale forvaltninger

7. Mette Aagaard Andreassen
Supply Chain versus Supply Chain
Benchmarking as a Means to Managing Supply Chains

8. Caroline Aggestam-Pontoppidan
From an idea to a standard
The UN and the global governance of accountants' competence

9. Norsk ph.d.

10. Vivienne Heng Ker-ni
An Experimental Field Study on the

Effectiveness of Grocer Media Advertising
Measuring Ad Recall and Recognition, Purchase Intentions and Short-Term Sales

11. Allan Mortensen
Essays on the Pricing of Corporate Bonds and Credit Derivatives

12. Remo Stefano Chiari
Figure che fanno conoscere
Itinerario sull'idea del valore cognitivo e espressivo della metafora e di altri tropi da Aristotele e da Vico fino al cognitivismo contemporaneo

13. Anders McIlquham-Schmidt
Strategic Planning and Corporate Performance
An integrative research review and a meta-analysis of the strategic planning and corporate performance literature from 1956 to 2003

14. Jens Geersbro
The TDF – PMI Case
Making Sense of the Dynamics of Business Relationships and Networks

15 Mette Andersen
Corporate Social Responsibility in Global Supply Chains
Understanding the uniqueness of firm behaviour

16. Eva Boxenbaum
Institutional Genesis: Micro – Dynamic Foundations of Institutional Change

17. Peter Lund-Thomsen
Capacity Development, Environmental Justice NGOs, and Governance: The Case of South Africa

18. Signe Jarlov
Konstruktioner af offentlig ledelse

19. Lars Stæhr Jensen
Vocabulary Knowledge and Listening Comprehension in English as a Foreign Language
An empirical study employing data elicited from Danish EFL learners

20. Christian Nielsen
Essays on Business Reporting
Production and consumption of strategic information in the market for information

21. Marianne Thejls Fischer
Egos and Ethics of Management Consultants

22. Annie Bekke Kjær
Performance management i Proces-innovation
– belyst i et social-konstruktivistisk perspektiv

23. Suzanne Dee Pedersen
GENTAGELSENS METAMORFOSE
Om organisering af den kreative gøren i den kunstneriske arbejdspraksis

24. Benedikte Dorte Rosenbrink
Revenue Management
Økonomiske, konkurrencemæssige & organisatoriske konsekvenser

25. Thomas Riise Johansen
Written Accounts and Verbal Accounts
The Danish Case of Accounting and Accountability to Employees

26. Ann Fogelgren-Pedersen
The Mobile Internet: Pioneering Users' Adoption Decisions

27. Birgitte Rasmussen
Ledelse i fællesskab – de tillidsvalgtes fornyende rolle

28. Gitte Thit Nielsen
Remerger
– skabende ledelseskræfter i fusion og opkøb

29. Carmine Gioia
A MICROECONOMETRIC ANALYSIS OF MERGERS AND ACQUISITIONS

30. Ole Hinz
Den effektive forandringsleder: pilot, pædagog eller politiker?
Et studie i arbejdslederes meningstilskrivninger i forbindelse med vellykket gennemførelse af ledelsesinitierede forandringsprojekter

31. Kjell-Åge Gotvassli
Et praksisbasert perspektiv på dynamiske
læringsnettverk i toppidretten
Norsk ph.d., ej til salg gennem Samfundslitteratur

32. Henriette Langstrup Nielsen
Linking Healthcare
An inquiry into the changing performances of web-based technology for asthma monitoring

33. Karin Tweddell Levinsen
Virtuel Uddannelsespraksis
Master i IKT og Læring – et casestudie i hvordan proaktiv proceshåndtering kan forbedre praksis i virtuelle læringsmiljøer

34. Anika Liversage
Finding a Path
Labour Market Life Stories of Immigrant Professionals

35. Kasper Elmquist Jørgensen
Studier i samspillet mellem stat og erhvervsliv i Danmark under 1. verdenskrig

36. Finn Janning
A DIFFERENT STORY
Seduction, Conquest and Discovery

37. Patricia Ann Plackett
Strategic Management of the Radical Innovation Process
Leveraging Social Capital for Market Uncertainty Management

2006

1. Christian Vintergaard
Early Phases of Corporate Venturing

2. Niels Rom-Poulsen
Essays in Computational Finance

3. Tina Brandt Husman
Organisational Capabilities, Competitive Advantage & Project-Based Organisations
The Case of Advertising and Creative Good Production

4. Mette Rosenkrands Johansen
Practice at the top
– how top managers mobilise and use non-financial performance measures

5. Eva Parum
Corporate governance som strategisk kommunikations- og ledelsesværktøj

6. Susan Aagaard Petersen
Culture's Influence on Performance Management: The Case of a Danish Company in China

7. Thomas Nicolai Pedersen
The Discursive Constitution of Organizational Governance – Between unity and differentiation
The Case of the governance of environmental risks by World Bank environmental staff

8. Cynthia Selin
Volatile Visions: Transactons in Anticipatory Knowledge

9. Jesper Banghøj
Financial Accounting Information and Compensation in Danish Companies

10. Mikkel Lucas Overby
Strategic Alliances in Emerging High-Tech Markets: What's the Difference and does it Matter?

11. Tine Aage
External Information Acquisition of Industrial Districts and the Impact of Different Knowledge Creation Dimensions

A case study of the Fashion and Design Branch of the Industrial District of Montebelluna, NE Italy

12. Mikkel Flyverbom
Making the Global Information Society Governable
On the Governmentality of Multi-Stakeholder Networks

13. Anette Grønning
Personen bag
Tilstedevær i e-mail som inter-aktionsform mellem kunde og med-arbejder i dansk forsikringskontekst

14. Jørn Helder
One Company – One Language?
The NN-case

15. Lars Bjerregaard Mikkelsen
Differing perceptions of customer value
Development and application of a tool for mapping perceptions of customer value at both ends of customer-supplier dyads in industrial markets

16. Lise Granerud
Exploring Learning
Technological learning within small manufacturers in South Africa

17. Esben Rahbek Pedersen
Between Hopes and Realities: Reflections on the Promises and Practices of Corporate Social Responsibility (CSR)

18. Ramona Samson
The Cultural Integration Model and European Transformation.
The Case of Romania

2007

1. Jakob Vestergaard
Discipline in The Global Economy
Panopticism and the Post-Washington Consensus

2. Heidi Lund Hansen
Spaces for learning and working
A qualitative study of change of work, management, vehicles of power and social practices in open offices

3. Sudhanshu Rai
Exploring the internal dynamics of software development teams during user analysis
A tension enabled Institutionalization Model; "Where process becomes the objective"

4. Norsk ph.d.
Ej til salg gennem Samfundslitteratur

5. Serden Ozcan
EXPLORING HETEROGENEITY IN ORGANIZATIONAL ACTIONS AND OUTCOMES
A Behavioural Perspective

6. Kim Sundtoft Hald
Inter-organizational Performance Measurement and Management in Action
– An Ethnography on the Construction of Management, Identity and Relationships

7. Tobias Lindeberg
Evaluative Technologies
Quality and the Multiplicity of Performance

8. Merete Wedell-Wedellsborg
Den globale soldat
Identitetsdannelse og identitetsledelse i multinationale militære organisationer

9. Lars Frederiksen
Open Innovation Business Models
Innovation in firm-hosted online user communities and inter-firm project ventures in the music industry
– A collection of essays

10. Jonas Gabrielsen
Retorisk toposlære – fra statisk 'sted' til persuasiv aktivitet

11. Christian Moldt-Jørgensen
Fra meningsløs til meningsfuld evaluering.
Anvendelsen af studentertilfredshedsmålinger på de korte og mellemlange videregående uddannelser set fra et psykodynamisk systemperspektiv

12. Ping Gao
Extending the application of actor-network theory
Cases of innovation in the telecommunications industry

13. Peter Mejlby
Frihed og fængsel, en del af den samme drøm?
Et phronetisk baseret casestudie af frigørelsens og kontrollens sameksistens i værdibaseret ledelse!

14. Kristina Birch
Statistical Modelling in Marketing

15. Signe Poulsen
Sense and sensibility:
The language of emotional appeals in insurance marketing

16. Anders Bjerre Trolle
Essays on derivatives pricing and dynamic asset allocation

17. Peter Feldhütter
Empirical Studies of Bond and Credit Markets

18. Jens Henrik Eggert Christensen
Default and Recovery Risk Modeling and Estimation

19. Maria Theresa Larsen
Academic Enterprise: A New Mission for Universities or a Contradiction in Terms?
Four papers on the long-term implications of increasing industry involvement and commercialization in academia

20. Morten Wellendorf
Postimplementering af teknologi i den offentlige forvaltning
Analyser af en organisations kontinuerlige arbejde med informationsteknologi

21. Ekaterina Mhaanna
Concept Relations for Terminological Process Analysis

22. Stefan Ring Thorbjørnsen
Forsvaret i forandring
Et studie i officerers kapabiliteter under påvirkning af omverdenens forandringspres mod øget styring og læring

23. Christa Breum Amhøj
Det selvskabte medlemskab om managementstaten, dens styringsteknologier og indbyggere

24. Karoline Bromose
Between Technological Turbulence and Operational Stability
– An empirical case study of corporate venturing in TDC

25. Susanne Justesen
Navigating the Paradoxes of Diversity in Innovation Practice
– A Longitudinal study of six very different innovation processes – in practice

26. Luise Noring Henler
Conceptualising successful supply chain partnerships
– Viewing supply chain partnerships from an organisational culture perspective

27. Mark Mau
Kampen om telefonen
Det danske telefonvæsen under den tyske besættelse 1940-45

28. Jakob Halskov
The semiautomatic expansion of existing terminological ontologies using knowledge patterns discovered

on the WWW – an implementation and evaluation

29. Gergana Koleva
European Policy Instruments Beyond Networks and Structure: The Innovative Medicines Initiative

30. Christian Geisler Asmussen
Global Strategy and International Diversity: A Double-Edged Sword?

31. Christina Holm-Petersen
Stolthed og fordom
Kultur- og identitetsarbejde ved skabelsen af en ny sengeafdeling gennem fusion

32. Hans Peter Olsen
Hybrid Governance of Standardized States
Causes and Contours of the Global Regulation of Government Auditing

33. Lars Bøge Sørensen
Risk Management in the Supply Chain

34. Peter Aagaard
Det unikkes dynamikker
De institutionelle mulighedsbetingelser bag den individuelle udforskning i professionelt og frivilligt arbejde

35. Yun Mi Antorini
Brand Community Innovation
An Intrinsic Case Study of the Adult Fans of LEGO Community

36. Joachim Lynggaard Boll
Labor Related Corporate Social Performance in Denmark
Organizational and Institutional Perspectives

2008

1. Frederik Christian Vinten
Essays on Private Equity

2. Jesper Clement
Visual Influence of Packaging Design on In-Store Buying Decisions

3. Marius Brostrøm Kousgaard
Tid til kvalitetsmåling?
– Studier af indrulleringsprocesser i forbindelse med introduktionen af kliniske kvalitetsdatabaser i speciallægepraksissektoren

4. Irene Skovgaard Smith
Management Consulting in Action
Value creation and ambiguity in client-consultant relations

5. Anders Rom
Management accounting and integrated information systems
How to exploit the potential for management accounting of information technology

6. Marina Candi
Aesthetic Design as an Element of Service Innovation in New Technology-based Firms

7. Morten Schnack
Teknologi og tværfaglighed
– en analyse af diskussionen omkring indførelse af EPJ på en hospitalsafdeling

8. Helene Balslev Clausen
Juntos pero no revueltos – un estudio sobre emigrantes norteamericanos en un pueblo mexicano

9. Lise Justesen
Kunsten at skrive revisionsrapporter.
En beretning om forvaltningsrevisionens beretninger

10. Michael E. Hansen
The politics of corporate responsibility: CSR and the governance of child labor and core labor rights in the 1990s

11. Anne Roepstorff
Holdning for handling – en etnologisk undersøgelse af Virksomheders Sociale Ansvar/CSR

12. Claus Bajlum
Essays on Credit Risk and Credit Derivatives

13. Anders Bojesen
The Performative Power of Competence – an Inquiry into Subjectivity and Social Technologies at Work

14. Satu Reijonen
Green and Fragile
A Study on Markets and the Natural Environment

15. Ilduara Busta
Corporate Governance in Banking
A European Study

16. Kristian Anders Hvass
A Boolean Analysis Predicting Industry Change: Innovation, Imitation & Business Models
The Winning Hybrid: A case study of isomorphism in the airline industry

17. Trine Paludan
De uvidende og de udviklingsparate
Identitet som mulighed og restriktion blandt fabriksarbejdere på det aftayloriserede fabriksgulv

18. Kristian Jakobsen
Foreign market entry in transition economies: Entry timing and mode choice

19. Jakob Elming
Syntactic reordering in statistical machine translation

20. Lars Brømsøe Termansen
Regional Computable General Equilibrium Models for Denmark
Three papers laying the foundation for regional CGE models with agglomeration characteristics

21. Mia Reinholt
The Motivational Foundations of Knowledge Sharing

22. Frederikke Krogh-Meibom
The Co-Evolution of Institutions and Technology
– A Neo-Institutional Understanding of Change Processes within the Business Press – the Case Study of Financial Times

23. Peter D. Ørberg Jensen
OFFSHORING OF ADVANCED AND HIGH-VALUE TECHNICAL SERVICES: ANTECEDENTS, PROCESS DYNAMICS AND FIRMLEVEL IMPACTS

24. Pham Thi Song Hanh
Functional Upgrading, Relational Capability and Export Performance of Vietnamese Wood Furniture Producers

25. Mads Vangkilde
Why wait?
An Exploration of first-mover advantages among Danish e-grocers through a resource perspective

26. Hubert Buch-Hansen
Rethinking the History of European Level Merger Control
A Critical Political Economy Perspective

2009

1. Vivian Lindhardsen
From Independent Ratings to Communal Ratings: A Study of CWA Raters' Decision-Making Behaviours

2. Guðrið Weihe
Public-Private Partnerships: Meaning and Practice

3. Chris Nøkkentved
Enabling Supply Networks with Collaborative Information Infrastructures
An Empirical Investigation of Business Model Innovation in Supplier Relationship Management

4. Sara Louise Muhr
Wound, Interrupted – On the Vulnerability of Diversity Management

5. Christine Sestoft
Forbrugeradfærd i et Stats- og Livsformsteoretisk perspektiv

6. Michael Pedersen
Tune in, Breakdown, and Reboot: On the production of the stress-fit self-managing employee

7. Salla Lutz
Position and Reposition in Networks – Exemplified by the Transformation of the Danish Pine Furniture Manufacturers

8. Jens Forssbæck
Essays on market discipline in commercial and central banking

9. Tine Murphy
Sense from Silence – A Basis for Organised Action
How do Sensemaking Processes with Minimal Sharing Relate to the Reproduction of Organised Action?

10. Sara Malou Strandvad
Inspirations for a new sociology of art: A sociomaterial study of development processes in the Danish film industry

11. Nicolaas Mouton
On the evolution of social scientific metaphors:
A cognitive-historical enquiry into the divergent trajectories of the idea that collective entities – states and societies, cities and corporations – are biological organisms.

12. Lars Andreas Knutsen
Mobile Data Services:
Shaping of user engagements

13. Nikolaos Theodoros Korfiatis
Information Exchange and Behavior
A Multi-method Inquiry on Online Communities

14. Jens Albæk
Forestillinger om kvalitet og tværfaglighed på sygehuse
– skabelse af forestillinger i læge- og plejegrupperne angående relevans af nye idéer om kvalitetsudvikling gennem tolkningsprocesser

15. Maja Lotz
The Business of Co-Creation – and the Co-Creation of Business

16. Gitte P. Jakobsen
Narrative Construction of Leader Identity in a Leader Development Program Context

17. Dorte Hermansen
"Living the brand" som en brandorienteret dialogisk praxis:
Om udvikling af medarbejdernes brandorienterede dømmekraft

18. Aseem Kinra
Supply Chain (logistics) Environmental Complexity

19. Michael Nørager
How to manage SMEs through the transformation from non innovative to innovative?

20. Kristin Wallevik
Corporate Governance in Family Firms
The Norwegian Maritime Sector

21. Bo Hansen Hansen
Beyond the Process
Enriching Software Process Improvement with Knowledge Management

22. Annemette Skot-Hansen
Franske adjektivisk afledte adverbier, der tager præpositionssyntagmer indledt med præpositionen à som argumenter
En valensgrammatisk undersøgelse

23. Line Gry Knudsen
Collaborative R&D Capabilities
In Search of Micro-Foundations

24. Christian Scheuer
Employers meet employees
Essays on sorting and globalization

25. Rasmus Johnsen
The Great Health of Melancholy
A Study of the Pathologies of Performativity

26. Ha Thi Van Pham
Internationalization, Competitiveness Enhancement and Export Performance of Emerging Market Firms:
Evidence from Vietnam

27. Henriette Balieu
Kontrolbegrebets betydning for kausativalternationen i spansk
En kognitiv-typologisk analyse

2010

1. Yen Tran
Organizing Innovationin Turbulent Fashion Market
Four papers on how fashion firms create and appropriate innovation value

2. Anders Raastrup Kristensen
Metaphysical Labour
Flexibility, Performance and Commitment in Work-Life Management

3. Margrét Sigrún Sigurdardottir
Dependently independent
Co-existence of institutional logics in the recorded music industry

4. Ásta Dis Óladóttir
Internationalization from a small domestic base:
An empirical analysis of Economics and Management

5. Christine Secher
E-deltagelse i praksis – politikernes og forvaltningens medkonstruktion og konsekvenserne heraf

6. Marianne Stang Våland
What we talk about when we talk about space:
End User Participation between Processes of Organizational and Architectural Design

7. Rex Degnegaard
Strategic Change Management
Change Management Challenges in the Danish Police Reform

8. Ulrik Schultz Brix
Værdi i rekruttering – den sikre beslutning
En pragmatisk analyse af perception og synliggørelse af værdi i rekrutterings- og udvælgelsesarbejdet

9. Jan Ole Similä
Kontraktsledelse
Relasjonen mellom virksomhetsledelse og kontraktshåndtering, belyst via fire norske virksomheter

10. Susanne Boch Waldorff
Emerging Organizations: In between local translation, institutional logics and discourse

11. Brian Kane
Performance Talk
Next Generation Management of Organizational Performance

12. Lars Ohnemus
Brand Thrust: Strategic Branding and Shareholder Value
An Empirical Reconciliation of two Critical Concepts

13. Jesper Schlamovitz
Håndtering af usikkerhed i film- og byggeprojekter

14. Tommy Moesby-Jensen
Det faktiske livs forbindtlighed
Førsokratisk informeret, ny-aristotelisk ἦθος-tænkning hos Martin Heidegger

15. Christian Fich
Two Nations Divided by Common Values
French National Habitus and the Rejection of American Power

16. Peter Beyer
Processer, sammenhængskraft og fleksibilitet
Et empirisk casestudie af omstillings-forløb i fire virksomheder

17. Adam Buchhorn
Markets of Good Intentions
Constructing and Organizing Biogas Markets Amid Fragility and Controversy

18. Cecilie K. Moesby-Jensen
Social læring og fælles praksis
Et mixed method studie, der belyser læringskonsekvenser af et lederkursus for et praksisfællesskab af offentlige mellemledere

19. Heidi Boye
Fødevarer og sundhed i sen-modernismen
– En indsigt i hyggefænomenet og de relaterede fødevarepraksisser

20. Kristine Munkgård Pedersen
Flygtige forbindelser og midlertidige mobiliseringer
Om kulturel produktion på Roskilde Festival

21. Oliver Jacob Weber
Causes of Intercompany Harmony in Business Markets – An Empirical Investigation from a Dyad Perspective

22. Susanne Ekman
Authority and Autonomy
Paradoxes of Modern Knowledge Work

23. Anette Frey Larsen
Kvalitetsledelse på danske hospitaler
– Ledelsernes indflydelse på introduktion og vedligeholdelse af kvalitetsstrategier i det danske sundhedsvæsen

24. Toyoko Sato
Performativity and Discourse: Japanese Advertisements on the Aesthetic Education of Desire

25. Kenneth Brinch Jensen
Identifying the Last Planner System
Lean management in the construction industry

26. Javier Busquets
Orchestrating Network Behavior for Innovation

27. Luke Patey
The Power of Resistance: India's National Oil Company and International Activism in Sudan

28. Mette Vedel
Value Creation in Triadic Business Relationships. Interaction, Interconnection and Position

29. Kristian Tørning
Knowledge Management Systems in Practice – A Work Place Study

30. Qingxin Shi
An Empirical Study of Thinking Aloud Usability Testing from a Cultural Perspective

31. Tanja Juul Christiansen
Corporate blogging: Medarbejderes kommunikative handlekraft

32. Malgorzata Ciesielska
Hybrid Organisations.
A study of the Open Source – business setting

33. Jens Dick-Nielsen
Three Essays on Corporate Bond Market Liquidity

34. Sabrina Speiermann
Modstandens Politik
Kampagnestyring i Velfærdsstaten.
En diskussion af trafikkampagners styringspotentiale

35. Julie Uldam
Fickle Commitment. Fostering political engagement in 'the flighty world of online activism'

36. Annegrete Juul Nielsen
Traveling technologies and transformations in health care

37. Athur Mühlen-Schulte
Organising Development
Power and Organisational Reform in the United Nations Development Programme

38. Louise Rygaard Jonas
Branding på butiksgulvet
Et case-studie af kultur- og identitets-arbejdet i Kvickly

2011

1. Stefan Fraenkel
Key Success Factors for Sales Force Readiness during New Product Launch
A Study of Product Launches in the Swedish Pharmaceutical Industry

2. Christian Plesner Rossing
International Transfer Pricing in Theory and Practice

3. Tobias Dam Hede
Samtalekunst og ledelsesdisciplin
– en analyse af coachingsdiskursens genealogi og governmentality

4. Kim Pettersson
Essays on Audit Quality, Auditor Choice, and Equity Valuation

5. Henrik Merkelsen
The expert-lay controversy in risk research and management. Effects of institutional distances. Studies of risk definitions, perceptions, management and communication

6. Simon S. Torp
Employee Stock Ownership:
Effect on Strategic Management and Performance

7. Mie Harder
Internal Antecedents of Management Innovation

8. Ole Helby Petersen
Public-Private Partnerships: Policy and Regulation – With Comparative and Multi-level Case Studies from Denmark and Ireland

9. Morten Krogh Petersen
'Good' Outcomes. Handling Multiplicity in Government Communication

10. Kristian Tangsgaard Hvelplund
Allocation of cognitive resources in translation - an eye-tracking and key-logging study

11. Moshe Yonatany
The Internationalization Process of Digital Service Providers

12. Anne Vestergaard
Distance and Suffering
Humanitarian Discourse in the age of Mediatization

13. Thorsten Mikkelsen
Personligsheds indflydelse på forretningsrelationer

14. Jane Thostrup Jagd
Hvorfor fortsætter fusionsbølgen udover "the tipping point"?
– en empirisk analyse af information og kognitioner om fusioner

15. Gregory Gimpel
Value-driven Adoption and Consumption of Technology: Understanding Technology Decision Making

16. Thomas Stengade Sønderskov
Den nye mulighed
Social innovation i en forretningsmæssig kontekst

17. Jeppe Christoffersen
Donor supported strategic alliances in developing countries

18. Vibeke Vad Baunsgaard
Dominant Ideological Modes of Rationality: Cross functional

integration in the process of product innovation

19. Throstur Olaf Sigurjonsson
Governance Failure and Icelands's Financial Collapse

20. Allan Sall Tang Andersen
Essays on the modeling of risks in interest-rate and inflation markets

21. Heidi Tscherning
Mobile Devices in Social Contexts

22. Birgitte Gorm Hansen
Adapting in the Knowledge Economy
Lateral Strategies for Scientists and Those Who Study Them

23. Kristina Vaarst Andersen
Optimal Levels of Embeddedness
The Contingent Value of Networked Collaboration

24. Justine Grønbæk Pors
Noisy Management
A History of Danish School Governing from 1970-2010

25. Stefan Linder
Micro-foundations of Strategic Entrepreneurship
Essays on Autonomous Strategic Action

26. Xin Li
Toward an Integrative Framework of National Competitiveness
An application to China

27. Rune Thorbjørn Clausen
Værdifuld arkitektur
Et eksplorativt studie af bygningers rolle i virksomheders værdiskabelse

28. Monica Viken
Markedsundersøkelser som bevis i varemerke- og markedsføringsrett

29. Christian Wymann
Tattooing
The Economic and Artistic Constitution of a Social Phenomenon

30. Sanne Frandsen
Productive Incoherence
A Case Study of Branding and Identity Struggles in a Low-Prestige Organization

31. Mads Stenbo Nielsen
Essays on Correlation Modelling

32. Ivan Häuser
Følelse og sprog
Etablering af en ekspressiv kategori, eksemplificeret på russisk

33. Sebastian Schwenen
Security of Supply in Electricity Markets

2012

1. Peter Holm Andreasen
The Dynamics of Procurement Management
- A Complexity Approach

2. Martin Haulrich
Data-Driven Bitext Dependency Parsing and Alignment

3. Line Kirkegaard
Konsulenten i den anden nat
En undersøgelse af det intense arbejdsliv

4. Tonny Stenheim
Decision usefulness of goodwill under IFRS

5. Morten Lind Larsen
Produktivitet, vækst og velfærd
Industrirådet og efterkrigstidens Danmark 1945 - 1958

6. Petter Berg
Cartel Damages and Cost Asymmetries

7. Lynn Kahle
Experiential Discourse in Marketing
A methodical inquiry into practice and theory

8. Anne Roelsgaard Obling
Management of Emotions in Accelerated Medical Relationships

9. Thomas Frandsen
Managing Modularity of Service Processes Architecture

10. Carina Christine Skovmøller
CSR som noget særligt
Et casestudie om styring og meningsskabelse i relation til CSR ud fra en intern optik

11. Michael Tell
Fradragsbeskæring af selskabers finansieringsudgifter
En skatteretlig analyse af SEL §§ 11, 11B og 11C

12. Morten Holm
Customer Profitability Measurement Models
Their Merits and Sophistication across Contexts

13. Katja Joo Dyppel
Beskatning af derivater
En analyse af dansk skatteret

14. Esben Anton Schultz
Essays in Labor Economics
Evidence from Danish Micro Data

15. Carina Risvig Hansen
"Contracts not covered, or not fully covered, by the Public Sector Directive"

16. Anja Svejgaard Pors
Iværksættelse af kommunikation
- patientfigurer i hospitalets strategiske kommunikation

17. Frans Bévort
Making sense of management with logics
An ethnographic study of accountants who become managers

18. René Kallestrup
The Dynamics of Bank and Sovereign Credit Risk

19. Brett Crawford
Revisiting the Phenomenon of Interests in Organizational Institutionalism
The Case of U.S. Chambers of Commerce

20. Mario Daniele Amore
Essays on Empirical Corporate Finance

21. Arne Stjernholm Madsen
The evolution of innovation strategy
Studied in the context of medical device activities at the pharmaceutical company Novo Nordisk A/S in the period 1980-2008

22. Jacob Holm Hansen
Is Social Integration Necessary for Corporate Branding?
A study of corporate branding strategies at Novo Nordisk

23. Stuart Webber
Corporate Profit Shifting and the Multinational Enterprise

24. Helene Ratner
Promises of Reflexivity
Managing and Researching Inclusive Schools

25. Therese Strand
The Owners and the Power: Insights from Annual General Meetings

26. Robert Gavin Strand
In Praise of Corporate Social Responsibility Bureaucracy

27. Nina Sormunen
Auditor's going-concern reporting
Reporting decision and content of the report

28. John Bang Mathiasen
Learning within a product development working practice:
- an understanding anchored in pragmatism

29. Philip Holst Riis
Understanding Role-Oriented Enterprise Systems: From Vendors to Customers

30. Marie Lisa Dacanay
Social Enterprises and the Poor
Enhancing Social Entrepreneurship and Stakeholder Theory

31. Fumiko Kano Glückstad
Bridging Remote Cultures: Cross-lingual concept mapping based on the information receiver's prior-knowledge

TITLER I ATV PH.D.-SERIEN

1992

1. Niels Kornum
Servicesamkørsel – organisation, økonomi og planlægningsmetode

1995

2. Verner Worm
Nordiske virksomheder i Kina
Kulturspecifikke interaktionsrelationer ved nordiske virksomhedsetableringer i Kina

1999

3. Mogens Bjerre
Key Account Management of Complex Strategic Relationships
An Empirical Study of the Fast Moving Consumer Goods Industry

2000

4. Lotte Darsø
Innovation in the Making
Interaction Research with heterogeneous Groups of Knowledge Workers creating new Knowledge and new Leads

2001

5. Peter Hobolt Jensen
Managing Strategic Design Identities
The case of the Lego Developer Network

2002

6. Peter Lohmann
The Deleuzian Other of Organizational Change – Moving Perspectives of the Human

7. Anne Marie Jess Hansen
To lead from a distance: The dynamic interplay between strategy and strategizing – A case study of the strategic management process

2003

8. Lotte Henriksen
Videndeling
– om organisatoriske og ledelsesmæssige udfordringer ved videndeling i praksis

9. Niels Christian Nickelsen
Arrangements of Knowing: Coordinating Procedures Tools and Bodies in Industrial Production – a case study of the collective making of new products

2005

10. Carsten Ørts Hansen
Konstruktion af ledelsesteknologier og effektivitet

TITLER I DBA PH.D.-SERIEN

2007

1. Peter Kastrup-Misir
Endeavoring to Understand Market Orientation – and the concomitant co-mutation of the researched, the re searcher, the research itself and the truth

2009

1. Torkild Leo Thellefsen
Fundamental Signs and Significance effects
A Semeiotic outline of Fundamental Signs, Significance-effects, Knowledge Profiling and their use in Knowledge Organization and Branding

2. Daniel Ronzani
When Bits Learn to Walk Don’t Make Them Trip. Technological Innovation and the Role of Regulation by Law in Information Systems Research: the Case of Radio Frequency Identification (RFID)

2010

1. Alexander Carnera
Magten over livet og livet som magt
Studier i den biopolitiske ambivalens