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:

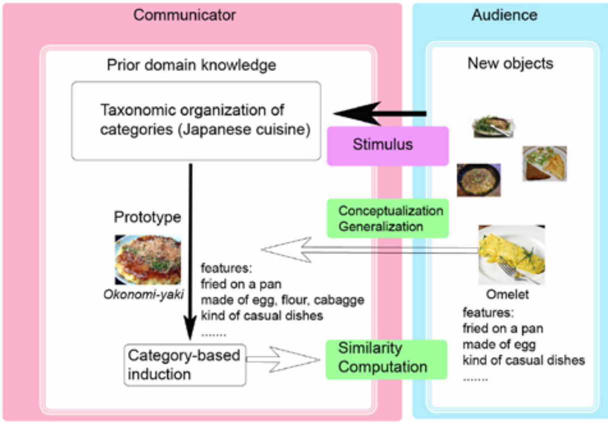

Figure 1: Asymmetrical coordination in a cross-cultural communication (when a Japanese communicator is learning Danish dishes as new objects)

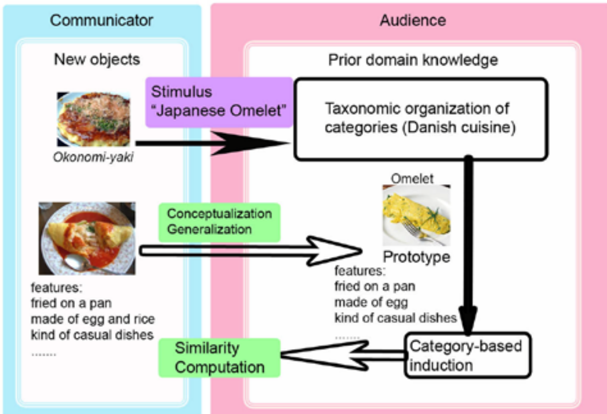

Figure 2: Asymmetrical coordination in a cross-cultural communication (when a Danish audience generalizes the meaning of "*okonomi-yaki*" from the stimulus "*omelet*")

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-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 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:

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

[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:

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

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

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

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 $\alpha=0$, $\beta=1$, 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].

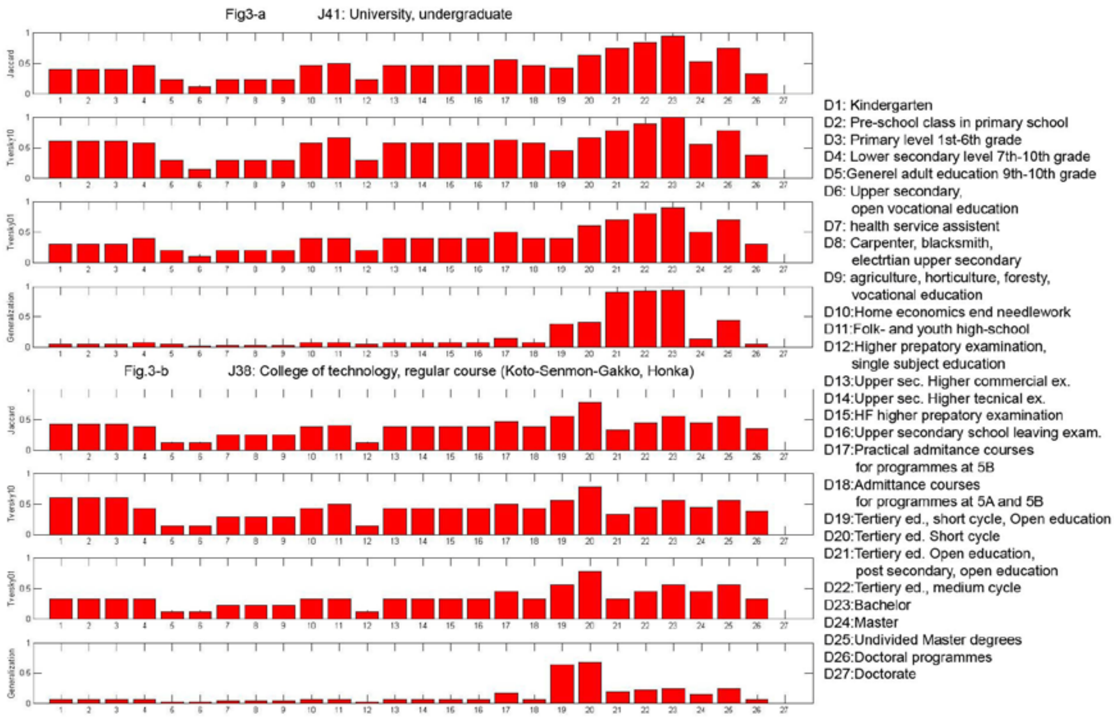

Figure 3: Similarity measure comparison (adapted from ref. [Glückstad 2012-b])

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).

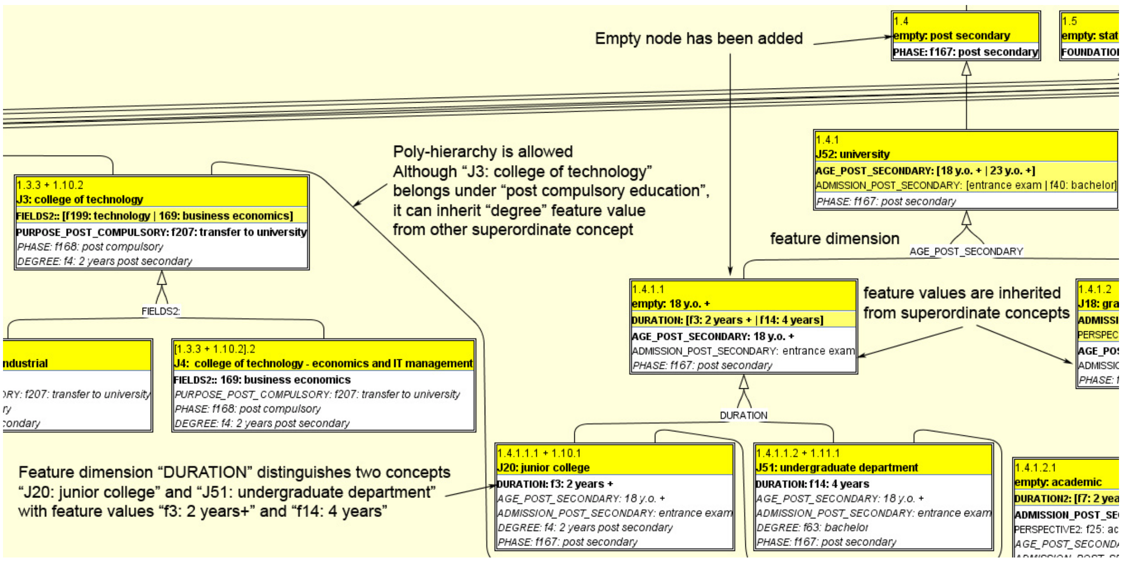

Figure 4: Terminological Ontology (adapted from ref. [Glückstad 2012-d])

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.

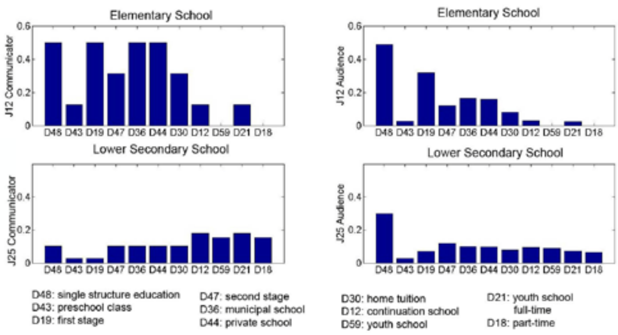

Figure 5: Cognitive simulation based on the BMG + IRM

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