Breaking Down Finance

A method for concept simplification by identifying movement structures from the image schema PATH-*following*

Dagmar GROMANN [a,1], Maria M. HEDBLOM [b]

[a] *Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Spain*
[b] *Free University of Bozen-Bolzano, Italy*

Abstract. Image schemas provide preverbal conceptual structures and are suggested to be the conceptual building blocks from which cognitive phenomena such as language and reasoning are constructed. 'Motion along a path' is one of the first image schemas infants remember, making PATH-following one of the earliest cognitive building blocks. We are interested in the importance of this developmentally relevant image schema in abstract adult language. For this purpose, we propose a semi-automated method to extract image-schematic structures related to PATH-following from a multilingual financial terminology. Two major assumptions are that a linguistic mapping of image schemas facilitates the understanding of complex concepts and is persistent across languages. Our results show that complex textual representations can be made simpler to understand by extracting the underlying image schemas and that they are persistent across languages. Another result includes the identification of novel specifications of predefined image-schematic structures.

Keywords. Image schemas, information extraction, ontologies, lexico-syntactic patterns, terminological database, finance

1. Introduction

Image schemas provide a theory for concept formation based on sensory-motor experiences. As such image schemas represent pre-linguistic structures of (usually) spatio-temporal object relations. The common framework they provide for thought also manifests itself in natural language [13]. They may be employed to establish rigorous definitions that capture part of the meaning of natural language expressions. It is generally believed that analyzing natural language leads to a greater understanding of image schemas [6, 15].

Research on image schemas is performed in several disciplines; cognitive linguistics [5], developmental psychology [14] and more formal areas (e.g. [10, 23, 1]). It has been shown that while the conceptual notions of image schemas are mostly equivalent, the linguistic expression can vary slightly across languages [15]. Bennett and Cialone [1] strengthened the interrelationship between image schemas and natural language by analyzing CONTAINMENT in a biological textbook corpus. In their investigation they

---

[1]Corresponding Author: dgromann@iiia.csic.es

were able to identify eight different types of CONTAINMENT. Hedblom et al. [6] showed how the image schema PATH-following represents a family of theories rather than an individual schema and demonstrated how this can be used to ground abstract concepts. With examples, they demonstrated how image schemas capture the information skeleton in (some) linguistic metaphors. We are interested in the universal persistence of image schemas in abstract adult communication, i.e., finance, and across languages, i.e., English, Swedish, German, and Italian.

The two main assumptions this paper addresses are, first, the idea that image-schematic structures persist across languages and domains, and second, that these early developed image schemas shape abstract adult communication and conceptualization. To address these assumptions, we semi-automatically identify representations of the image schema family PATH-following in a multilingual financial terminology (extracted from IATE[2]). We extract financial terminological entries where English, Swedish, German, and Italian natural language descriptions are aligned. Thereby, we are able to see whether the image schemas involved in abstract concepts are consistent across languages. Secondly, we also show how complex financial concepts can be simplified when broken down to their image-schematic core. Furthermore, this experiment strengthens the link between language and image schemas as well as their relation to formal ontologies. Our results show that image-schematic structures occur with a high consistency within the same entries across all four languages, at times with slight variations though, e.g. omission of the SOURCE in the SOURCE_PATH_GOAL.

Since image schemas are multidisciplinary a clarification of their theoretical foundation is introduced in the next section. We continue by detailing the employed methodology, before we introduce the obtained results. The results are then compared to related work, followed by a discussion section. Finally, we provide some concluding remarks with a brief outlook to future work.

2. The theory of image schemas

The theory was introduced by Lakoff [11] and Johnson [8] in the late 1980s. Since then it has become an important theory to ground higher cognitive phenomena, such as language and reasoning, in the low-level sensations acquired from embodied experiences. Image schemas are defined as the abstract patterns derived from sensory-motor experiences found in embodied cognition [22]. Developed in early infancy, they are pre-linguistic conceptualisations that allow infants to make predictions about their surroundings [14]. Classic examples of image schemas include CONTAINMENT, SUPPORT, VERTICALITY, and SOURCE_PATH_GOAL.

Some important characteristics for image schemas are that they exist both as static and dynamic concepts [25], and both in simple and more complex form [15]. Additionally, there is no clear border for when one image schema becomes another and in language image schemas often appear in constellations of one or more image schemas combined [21], e.g. CONTAINMENT and PATH combined forms the conceptual structure in expressions such as *'get into trouble'*.

One use of image schemas is that they can act as an information skeleton in analogical transfer [9]. In infancy this means that when a child has learned that 'tables SUPPORT

---

[2]http://iate.europa.eu/

plates' they can infer that 'desks SUPPORT books'. As cognitive abilities become more complex and with the acquisition of the capacity for increasingly abstract and complex thinking in the early teens [20], this analogical transfer can help build conceptualisations of abstract concept. An example is *'to offer* SUPPORT *to a friend in need'*.

While some words and concepts cannot be described using image schemas, other abstract concepts can be. For instance, 'transportation' can be broken down into a combination of PATH and either SUPPORT or CONTAINMENT. This kind of combination is parallel in its constellation, but there are also combinations of image schemas that alter the nature of the image schema. For example, a common conceptualisation of the concept 'marriage' is as a LINKED_PATH. Here the components of the image schemas are merged rather than sequentially added. This illustrates the gestalt structure of image schemas, meaning that no component can be removed or added without changing the logics of the image schemas [12]. For example, it is not possible to remove the 'border' from the CONTAINMENT image schema, nor is it possible to speak of solely 'an inside' without at least implicitly considering a border and an outside as well. In natural language many image-schematic components are implicit, yet for formal analyses of image schemas these image schema components need to be considered more directly.

2.1. Ontology of PATH-following

Aiming to take the above mentioned aspects of image-schematic structures, components and combinatorial possibilities into account, Hedblom et al. [6] took a closer look at what in the literature is called the SOURCE_PATH_GOAL schema [12]. They presented a hierarchical structure, the PATH-following family (see Figure 1), that grew more specialised based on the addition of spatial primitives found in developmental psychology [15].

Their method took the image schema components into account and also considered concept integrations by introducing a graphical and logical representation for how image schemas occasionally 'share' components from different image-schematic notions. In the figure, some preliminary participants of the PATH-family were introduced. Their method also includes a more complete common logic formalisation for the graph, available in an Ontohub repository[3]. Important for this paper is to register how each node in the graph represent an individual image-schematic structure that can be mapped to natural language expressions and conceptualisation.

A CYCLE is an iterative temporal path. Hedblom et al. [6] argued that MOVEMENT_IN_LOOPS is the physical representation of the temporal CYCLE. In this paper we largely merge the temporal and the physical PATHs into one and look at cycle as a relative to the PATH-family. Consequently, a CYCLE is a specific manifestation where the SOURCE and the GOAL coincide and related to CLOSED_PATH_MOVEMENT.

3. Method

Our method relies on the ontology of PATH-following introduced in Section 2.1, which we utilize to identify linguistic manifestations of image-schematic structures. We extract potential candidate entries for the PATH schema in English by means of lexico-syntactic patterns and synonym sets. To retrieve the German, Swedish and Italian data we bene-

---

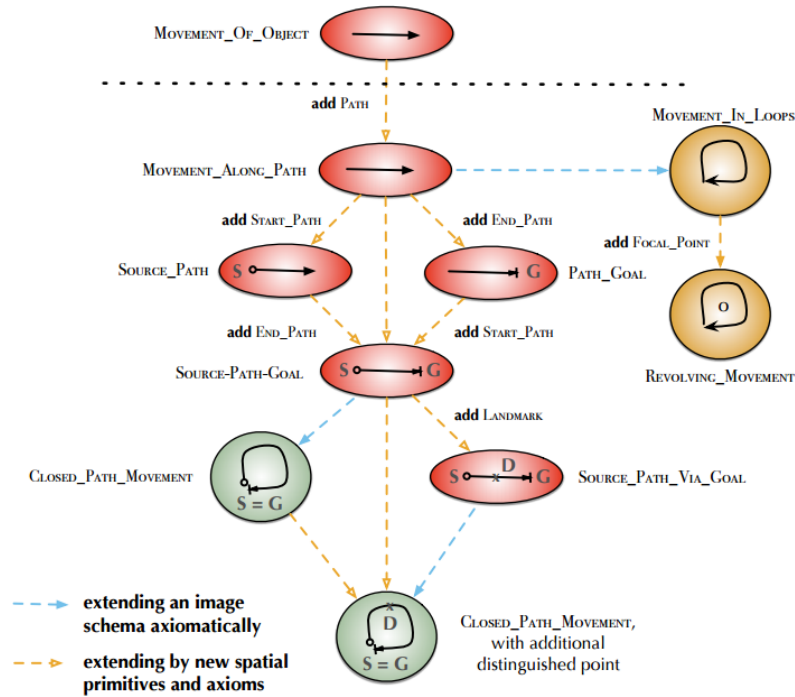[3]`https://ontohub.org/repositories/imageschemafamily/`

Figure 1. Path family as introduced by Hedblom et al. (2015)

fit from the alignment of multilingual data in the terminological database. The resulting corpus is manually analyzed by first language speakers to identify potential representations of PATH schemas. For this manual analysis we followed the structure of the utilized PATH-following ontology as well as a graphical representation method.

3.1. Financial terminological database

Concept-oriented terminological databases organize multilingual natural language data into terminological entries, so-called 'units of meaning'. A terminology seeks to mitigate ambiguity and polysemy of natural language by limiting its content to a specialized domain of discourse. The use of a given term is specified by means of its salient features and semantic type in a natural language definition. All natural language descriptions associated with the same entry are considered semantically equivalent. Such resources are typically applied to computer-aided translation, information extraction, machine translation, corporate terminology management, and many more.

Our data set for this experiment was extracted from the InterActive Terminology for Europe (IATE)[4], which classifies its 1.3 million entries in up to 24 European languages

---

[4]http://iate.europa.eu/

Pattern name	Content
From-to (PP, TO)	% from % to
Prepositions (PP)	around, across, through, behind, before, earlier
Movement (NN,NNS)	movement, track, path, transportation, transit, mobility, steps, passage
Process (NN, NNS)	process, operation, transfer, transferal
Development (NN, NNS)	development, evolution, progress, progress, progression, chance, migration
Cycle (NN, NNS)	cycle, course, chain, ring, rotation, circle, circuit, loop, sequel, orbit, wheel
Move (VB, VBG, VBZ)	move, transfer, drift, migrate, walk, drive, fly, proceed, etc.
Start (VB, VBG, VBZ)	start, commence, begin, etc.
End (VB, VBG, VBZ)	end, target, arrive, etc.

Table 1. Lexico-syntactic patterns and synonym sets for PATH following

by domain and sub-domain. For this experiment we only considered entries in the financial domain and its sub-domains and limited the extraction to entries with English, Swedish, German, and Italian natural language definitions. Since our examples only provide a small subset of the actual natural language descriptions associated with each term. such as context or language usage, we provide the IATE identifier for each example so that the full entry can be consulted online.

3.2. Lexico-syntactic patterns and entry extraction

We draw from research in pattern-based ontology development [24, 16] and metaphor identification [4] to align image schemas and their natural language representation. A widespread methodology for detecting metaphors is the initial identification of metaphoric expressions that are then automatically extracted and analyzed in their context [4]. Thereby, blended domains can be detected that provide candidates for metaphoric language. We adopt this idea and formulate linguistic expressions related to start, end, and movement along a path as lexico-syntactic patterns [24].

English lexico-syntactic patterns and synonym sets listed in Table 1 are employed to extract terminological entries that potentially contain the PATH schema from our financial terminological database. By extracting the entry based on the English definition only, we automatically obtain the definitions in other languages aligned with the same entry in the database. We assume that PATHs are internally structured in the sense that they have the structure SOURCE_PATH_GOAL, which could in language be modelled by indicating a trajectory 'from' a SOURCE 'to' a GOAL. This assumption is translated to the the first 'from-to' pattern as shown in Table 1. For additional patterns we utilized linguistic expressions related to PATH-following as defined by Hedblom et al. [6] and Mandler and Cánovas [15] and their synonym sets to establish a set of recurring linguistic structures as detailed in Table 1. A special case is a 'cycle' where start and end coincide and which requires its own pattern. Since a movement can also be abstractly defined by a 'development' we included it as a synonym set in the extraction process. Finally, BLOCKAGE is the hindered movement by an obstruction in the trajectory of the object from source to target, which is mainly represented by prepositions in the patterns, e.g. 'across'. All patterns and the POS tags of the morphological variants we considered are provided in Table 1.

3.3. Linguistic mapping of image-schematic structures

The manual mapping procedure was applied to the pattern-extracted entries per language. For each language one (for German and Swedish) or two native/fluent speakers (for English and Italian) identified image-schematic structures from the PATH-family presented in [6] on natural language definitions. Each candidate image schema was graphically represented, that is, diagrams were created to draw links and the objects moving between potential SOURCE and GOAL for each definition. We only considered them PATHs when the links defined actual movements over time. While following the general structure of the family, additional image-schematic components where considered in order to not only strengthen the PATH-family notion, but also by analysis improve the PATH-family to match natural language. This allowed for a freer interpretation of the terms which better mapped the intended content. At the end a comparison of all identified schemas allowed for an evaluation of their cross-linguistic persistence.

As this paper focuses on the PATH-following image schema family, one important aspect of the mapping method is to restrict the pattern-extracted entries to the terms that could be identified as a form of movement. This means that all terms referring exclusively to objects, both abstract and concrete (e.g. risks, credit cards), proper names (e.g. financial institutions), numbers and measurements are omitted. Terms that depict things like processes, events or changes over time are analysed further. In a financial terminology, several entries refer to processes that do not refer to movement or development over time, which were then not considered representing PATH schemas. The SOURCE and GOAL had to be explicitly expressed for the image schema structure to be defined as SOURCE_PATH_GOAL. If these parts were omitted by, for instance, using passive voice instead of active, the structure was reduced to SOURCE_PATH or PATH_GOAL, in order to achieve an improved correspondence between linguistic content and schema.

4. Results

Our analysis targeted the identification of image-schematic structures of PATH-following in natural language text across four natural languages. We were interested in the (a)symmetries of such structures across languages as well as the coverage of the predefined schematic structures (see Figure 1) within the domain of financial terminology. We first present our general image-schematic candidates and results before we investigate cross-linguistic divergences of the identified image-schematic structures.

We base our analysis on natural language definitions. Not all IATE entries contain natural language definitions of terms or definitions in the languages we desired. Limiting our extraction to the domain of finance with definitions in English, Swedish, German, and Italian resulted in 864 entries. The lexico-syntactic patterns and synonym sets from Table 1 were applied to those 864 entries, which further reduced our corpus to 190 entries. All 190 definitions for each language were analyzed manually by a first language/fluent speaker to find PATH schemas. The precision of the English patterns was unexpectedly low with only 57 English entries containing PATH following schematic structures, i.e., 30% in total. Judging from the number of identified image schemas for each pattern, nominal structures and prepositions returned most candidate entries.

A total of 67% of the 'cycle' synonym set and 52% of the 'process' nouns returned image-schematic structures, followed by 'from-to' with 29% of the 48 extracted entries.

The 37 extracted entries based on prepositions (across, through, around, etc.) and the 7 ones based on motion verbs resulted in image schema candidates in 30% of their cases. The 'end' pattern with 8 entries contained one schema, 'start' with three schemas candidates contained no actual schema at all. While the movement and development pattern extracted almost 20 entries each, only 19% in the former and 6% in the latter case contained PATH-related structures.

We partially attribute the low precision to the fact that there are a lot of general statements that do not relate to any movement in time or space. For instance, one part of the definition of 'central rate' states that 'Currencies have limited movement from the central rate according to the relevant band' (IATE:785015), which our 'from-to' and 'movement' patterns detected. However, neither the term nor the definition have any relation to PATH image-schematic structures. In contrast, 'capital outflow' which is defined as 'movement of assets out of a country...' (IATE:1104177) provides the kind of PATH-following we intended to find. Thus, the linguistic surface structure alone is not a sufficient indicator of movements along a path.

The results separated by language and structure are depicted in Table 2 as cumulative frequencies. Although one would expect there to be more PATHs because of transactions in finance, a majority of extracted entries could be discarded as object, institution, natural or legal person, strategies, techniques, or measures, that is, not related to any kind of PATH or movement over time. Events, processes, and actions provided excellent candidates for these image-schematic structures.

Image-Schematic Structure	English	Swedish	German	Italian	Total
LINK	2	4	2	2	10
PATH	3	7	6	7	23
SOURCE_PATH	7	6	7	11	31
PATH_GOAL	6	9	10	7	32
SOURCE_PATH_VIA_GOAL (SPVG)	3	2	3	1	9
PATH_VIA_GOAL	2	1		2	5
SOURCE_PATH_VIA	1				1
CAUSED_MOVEMENT	1				1
CLOSED_PATH_MOVEMENT	2		1	1	4
MOVEMENT_IN_LOOPS	1	1	1	1	4
PATH_SWITCHING	1		1	1	3
JUMPING	1	1			2
BLOCAKGE_AVOIDANCE	1		1	1	3
PATH_SPLITTING	4	3	4	3	14
SPG AND SPG	1	1	1	1	4
SPVG AND SPVG	1	1	1	1	4
SPG OR PATH_SPLITTING		1			1
SPG OR PATH			1		1
SPG OR LINK	1	1	1	1	4
Total	**57**	**54**	**58**	**54**	**224**

Table 2. Metrics for identified image-schematic structures across languages

All resulting image-schematic structures are ordered by approximated complexity in Table 2. Financial entries in our data set most frequently (30% of all cases) feature

a regular SOURCE_PATH_GOAL schema followed by the similar, yet simpler, pattern PATH_GOAL. On occasion, specific textual references concurrently defined two image-schematic structures that could equally be designated by the same given term. For such cases we opted for a representation with the logical operator "OR". For instance, an 'interlinking mechanism' (IATE:892281) can designate a cross-border payment procedure 'OR' a technical infrastructure, which we represent as SOURCE_PATH_GOAL 'OR' LINK.

We employed a graphical representation technique to identify the movements of objects between entities along PATHs for each definition in each language. It turned out that some of the identified image-schematic structures were not present in the predefined structures in Figure 1. From all languages four different scenarios depicted in Figure 3 could be identified by means of the graphical representation technique. Additionally, image-schematic structures of a 'double-way' SOURCE_PATH_GOAL movement could be observed in financial definitions. These movements were dependent on two variables: the number of PATHs and the number of OBJECTs that are moved along them. The four resulting image-schematic structures that are differentiated based on those two variables are depicted in Figure 2.
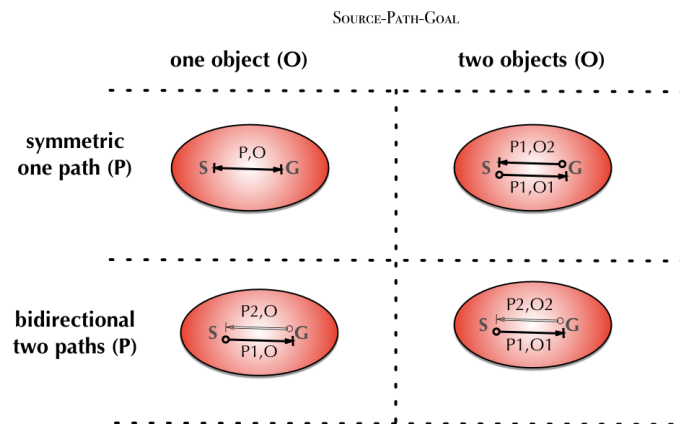

Figure 2. The returning object(s) problem

In a symmetric SOURCE_PATH_GOAL, one OBJECT moves or is being moved along one path until it returns to its starting point, potentially also passing a distinguishing point. For instance, taking out and repaying a loan is the transfer of money from the creditor to the debtor where the same object (money) can be returned on the same path (e.g. bank transfer) to the original source, that is, the creditor. Should the SOURCE and the GOAL coincide, the schema matches the CLOSED_PATH_MOVEMENT introduced in [6].

It is also possible, however, that the returning path differs from the initial one, in which case the schematic structure specifies two PATHs. If the same OBJECT moves from the SOURCE and back again on a different PATH, we consider this a bidirectional SOURCE_PATH_GOAL. In the event of SOURCE and GOAL being identical the PATH that returns to the SOURCE can either be equivalent to the initial PATH (symmetric) or differ from the original PATH (bidirectional). The latter would be considered a bidirectional

CLOSED_PATH_MOVEMENT. For instance, 'painting the tape' (IATE: 927775) is an example of several transactions (PATHs) being used in a CLOSED_PATH_MOVEMENT to create the impression of price movement of a financial instrument. Since this is a repeated cycle we even consider it a MOVEMENT_IN_LOOPS adding a temporal component. It could be argued that this image-schematic structure integrates other concepts or image-schematic structures, such as CONTAINMENT, however, for the purpose of this paper we are exclusively interested in variations and occurrences of SOURCE_PATH.

A second dimension we identified is whether the returning OBJECT is identical to the first outgoing one. In finance, often the returning object is different from the one initially moved along the PATH, basically capturing any kind of exchange or purchase. The SOURCE for one object becomes the GOAL for the second object, and vice versa. We refer to two different OBJECTs moving along the same path as poly-object symmetric SOURCE_PATH_GOAL. If two OBJECTs move along two different PATHs, we call this a poly-object bidirectional SOURCE_PATH_GOAL. A real life example is the exchange of shares (the first OBJECT) from the stock market (the first PATH) and money (the second OBJECT) from a bank transaction (the returning PATH) between a client and a broker.

We encountered four PATH-related structures in our sample that could not be explained by the predefined ones in Figure 1. To accommodate these structures with our approach, we decided to extend the PATH family by adding four structures, namely JUMPING, PATH_SWITCHING, PATH_SPLITTING, and BLOCKAGE_AVOIDANCE, which are depicted in Figure 3. The illustration of BLOCKAGE, itself an image schema, serves the sole purpose to clarify the movement involved in BLOCKAGE_AVOIDANCE.
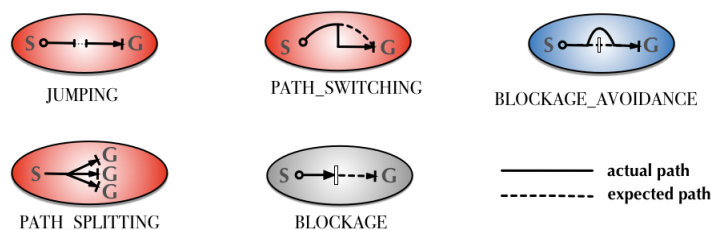

Figure 3. Four kinds of complex PATH structures extracted from the financial domain

First, JUMPING[5] represents a temporary or spatial discontinuity of a given PATH. For instance, 'bond washing' (IATE:3544441) is a method of obtaining tax-free capital profits by selling the bond immediately before the coupon pays and buying it back right thereafter to avoid tax payments. 'Bond washing' is a classical metaphor based on the notion of 'cleaning', which indeed captures important aspects of the term. However, when explaining the underlying process behind the term also the PATH-following family can be used. Considering ownership as the PATH from the initial acquisition of the bond (SOURCE) to the gains it generates (GOAL), 'bond washing' leads to this interruption of the PATH and can be seen as an example of JUMPING. While it may be argued that JUMPING is simply a sequential combination of two disjoint SOURCE_PATH_GOAL, JUMPING takes on its own logic as both paths are involved in one particular movement as demon-

---

[5]Jumping is not to be confused with the motion verb *to jump*. It refer to a jump in time or space, much like 'teleportation' rather than a temporary elevation.

strated in the conceptualisation example above. Therefore, we argue that JUMPING can be justified as a complex image schema in its own right.

Second, in case of PATH_SPLITTING one object is distributed along a path to several GOALs. It could be argued that this represents merely a type of cardinality. However, since the PATH can be asymmetrical or bidirectional, we consider it an image-schematic structure in its own right. For instance, in all kinds of 'tender procedures' (e.g. IATE:887199) the identical piece of information (a call) is sent to several parties, who return their individual pieces of information (the bids). Hence, this is an example of bidirectional PATH_SPLITTING. One example to account for this image-schematic structure in sensory-motor experiences would be the distribution of auditory information to several recipients with varying replies.

Third, in PATH_SWITCHING the expected PATH is fully discontinued and replaced by a new PATH. For instance, the definition of 'refinancing' (IATE:786103) specifies the extending of a new loan and a mutual agreement to discontinue the previous loan. Thus, the original loan PATH is switched to a new loan PATH with altered conditions. It is important to note that the definition clearly specifies the replacement of a debt obligation with a new one and not merely altering the conditions of an existing loan. This explicit switching of the agreed path is an excellent example of PATH_SWITCHING.

Finally, the active avoidance of a BLOCKAGE can be considered an image-schematic construction that combines a number of pre-existing structures and schemas. The course of the PATH is (intentionally) altered to prevent the discontinuation of the movement of the object due to a BLOCKAGE. A 'Paulian action' (IATE:822870) allows a creditor to take action to avoid potential fraudulent activities of an insolvent debtor, granting the former rights to have a debtor's transaction to that end reversed. Thus, the term as such represents an example of BLOCKAGE_AVOIDANCE. Here the connection to the physical world is the actual obstruction of the trajectory of an object and its alteration of the path to avoid any interruption of its course by the BLOCKAGE.

A slight asymmetry in the distribution of image-schematic structures across languages could be observed. In English and German definitions more structures could be identified than in Swedish and Italian as shown in Table 2. However, those quantified results fail to provide any insights into the differences across languages. In 55% of all cases the same image schema detected in English could also be found in the definitions of the other two languages. In 27% of the cases where the schemas were not identical, the differences arise from either an addition or omission of a SOURCE, GOAL, or VIA, while the general structure is that of a SOURCE_PATH_GOAL. Differences that arise from other sources can be pinned down to 10% of all entries. We could observe a slight preference of GOAL usage in Swedish and German as opposed to a heightened use of SOURCE in Italian in the reduced SOURCE_PATH_GOALs.

Our method deliberately relied on explicitly described content only. This means that omissions that arise from linguistic or grammatical differences across languages or stylistic choices effected the extraction result. For instance, differences can arise from a heightened use of passive constructions in one language, e.g. German, and an increased utilization of active SOURCEs and GOALs due to grammatical choices in another. One of the reasons for this choice was the intention to analyse linguistic consistency in relation to schematic persistence across languages.

We found in a final cross-linguistic analysis that most cross-linguistic differences in the identification of schematic structures arise from unnecessarily complicated descrip-

tions, or even inconsistencies, in one language. Semantically identical entries resulted in diverging image schemas for two major reasons: a) the difference in lexical or grammatical choices (e.g. passive vs. active voice), and b) the omission of salient features. All languages but English showed a heightened use of nominal constructions and passive voice, which led to the frequent omission of SOURCE and GOAL. For instance, 'selling ... by' in English is juxtaposed to 'Umwandlung von ...' (transformation of) in German and 'operazione che ...' (operation that) in Italian. When the passive voice was used in English, it was frequently supplemented with a 'by' and the subject or object of the sentence. Thus, the number of simple PATH schemas as opposed to the more complex SOURCE_PATH_GOAL schemas is much lower in English than in the other languages. The second set of differences refers to the features and differences in content. For instance, the number of explicitly mentioned GOALs is much higher in Swedish and German than in English and Italian, the latter of which focuses more on the SOURCE. For automated methods, both differences lead to a certain degree of difficulty. Our method could uncover inconsistencies across languages for both cases, which we consider an added benefit of the linguistic mapping of image schemas.

This approach equally uncovered conceptual inconsistencies across and within languages. For instance, 'equity capital' and 'equity financing' (IATE: 1119090) are modelled as synonymous where in fact the former refers to equity of the company while financing refers to the process of generating such capital. Thus, they should clearly be separated into two entries, a claim that is supported by the fact that the entry's definition consists of two sentences that define both concepts. In view of potentially automating the approach, we found, as can be expected, that a linguistic analysis of the specification's surface structure would definitely lead to misleading results. For instance 'lifecycling' (IATE: 3516328) describes a shift of a person's investment approach at a specific moment in life rather than a CYCLE as the term suggests. Furthermore, our manual approach and cross-linguistic analysis revealed (unintentionally) repeated definitions and entries, e.g. 'fine-tuning operation' (IATE: 111402 & 907147).

5. Related Work

From a top-down perspective, Kuhn [10] analyzed noun phrases in WordNet glosses and connects them with spatial abstractions that model image-schematic affordances. Particularly interesting is his analysis of nesting and combining image schemas in natural language to represent more complex concepts, e.g. 'transportation' brings together SUPPORT and PATH. One bottom-up approach that is very close to ours in methodology and objective is Bennett and Cialone [1] who investigated the construction of spatial ontologies from a biological textbook corpus by applying sense clusters. They exemplified their approach by using the image schema of CONTAINMENT. Lakust and Landua [13] investigated the linguistic encoding of PATH in English speaking children and adults and find an asymmetrically higher frequency of PATH_GOALs over SOURCE_PATHs. Participants were asked to verbalize visualizations, which also included finance-related events, such as change of possessions necessitating a transaction between agents.

Automated solutions to extracting spatial expressions from natural language corpora rely on machine learning for annotating text. Handcrafted rules for each language help to extract motion verbs across languages and named entities or predefined spatial ex-

pressions [16]. The extracted data are then qualitatively mapped to ontological formalizations. The idea of an embodied construction grammar [2] equally requires the manual crafting of a lexicon. Thus, the central issue we are facing, namely the mapping of identified spatial expressions to actual image schemas, persist in those approaches and no fully automated solution has been provided. Additionally, the size and specialized type of our data set rules out any machine learning approaches.

It has been supported that the conceptual system underlying image schemas changes in individual languages, even though the fundamental conceptual notions vary marginally cross-linguistically [15]. In Korean CONTAINMENT can only be expressed by differentiating whether it is tight or loose [17], which is not systematically encoded in English and thus an optional distinction. Papafragou et al. [19] found that English speakers more likely linguistically encode manner of motion information than Greek speakers. This was generalized to cross-linguistic asymmetries and the authors differentiated 'Manner languages' (e.g. German, Russian, Chinese) from 'Path languages' (e.g. French, Spanish, Turkish). Since SOURCE_PATH_GOAL schemas are not only spatial but also temporal, time has been frequently considered as an important aspect. Fuhrmann et al.[3] found that in Chinese a vertical representation of time is preferred over the English horizontal one. Núñez and Sweetser [18] found that the spatial construal of time can vary in the sense of whether the future is depicted as in front or behind the speaker.

6. Discussion

6.1. Method discussion

Lexico-syntactic patterns were applied to extract image-schematic candidates based on the English definition of terms. Our initial patterns resulted in more than 3000 extracted entries that at first analysis contained less image-schematic structures than we had expected and desired. A repeated tweaking of the patterns reduced this number to 190 pattern-extracted entries with a precision of only one third. Given the issues with our current approach discussed below, we abstained from creating a gold standard for this specific data set. Thus, we do not provide any numbers on the potentially missed image schemas here. However, we can definitely state that the start and end schemas were the least successful ones. The approach to extract from English definitions only, however, returned good results from our database since only two of the 57 resulting English definitions only contained an image schema in English and in no other language.

The low precision was mainly due to the chosen approach, which relied on the surface structure of linguistic expressions without considering their meaning in context. Additionally, the choice of patterns and linguistic expressions generally has a strong influence on the results [13]. Although in finance one would expect an abundance of PATH schemas because transactions are central to the domain, a surprisingly high number of abstract and concrete objects (e.g. bonds, debit cards), entities (e.g. institutions, agents), abstract strategies (e.g. hedging), measurements (e.g. exchange rate) among others were present in our data and identified by our patterns. Additionally, the type of transactions we found was very different as was the nature of the PATH-schematic structures they referred to. For instance, a simple transaction of buying and selling is very different from, e.g. 'painting the tape' (IATE:927775), a market manipulation strategy that utilizes a se-

ries of transactions, i.e., a MOVEMENT_IN_LOOPS, to influence price movements. We consider the analysis of the exact PATH schemas in natural language as useful and also identified new schematic structures presented above. However, for further experiments a more refined approach to extracting image schemas that goes beyond the surface structure is required. Alternatives to a pattern-based approach, such as a construction grammar for image schemas [2] or deep natural language analysis, will yield improved results. However, the size of the data set makes this scenario not a very good candidate for machine learning.

Low numbers of human judges are a common issue in semantic annotation tasks of any kind [16]. The low number of native speakers in our analysis might also have created an unwanted bias. Although we did specify basic criteria for definitions qualifying as image-schematic structures, the final decision might be subjectively biased due to the low number of judges. We did, however, evaluate the quality of the schema identification process by means of the final cross-linguistic comparison, which made us re-evaluate each individual schema candidate in each language. In this comparison the number of identical schemas that were detected across languages was rather high with more than 50% and of the non-identical ones the variation was frequently reduced to an omission of SOURCE or GOAL. One way to improve on the issue of the bias is to have a larger sample of analysts that perform the image-schematic mapping. This should primarily be a method to obtain a gold standard as at the same time a stronger level of automation for the actual method is needed.

6.2. Results discussion

A clear preference for the SOURCE_PATH_GOAL schema could be observed in all languages. In contrast, [15] claimed that PATH_GOAL is more important and in fact more prevalent in the (pre-linguistic) usage of schemas by adults and children, an argument that is supported by the findings of [13]. They presumed that children do not require SOURCEs to conceptualize a PATH_GOAL, which is why it is often omitted in cross-linguistic analyses of image schemas. Our experiment could not provide strong evidence for or against this claim. Although there is a slight increase of PATH_GOALs over SOURCE_PATHs, the predominant schema still explicitly contains the SOURCE. In fact, in Italian a predominance of SOURCE_PATH over PATH_GOAL could be observed but requires more extensive investigation.

The definition adopted here [15] is that image schemas are not just gestalts but conceptual structures. The omission and/or addition of a SOURCE or GOAL changes the perspective of the schema [13]. It is important to differentiate whether the description explicitly states that an agent transfers an OBJECT or that an OBJECT is being transferred to a beneficiary [13]. Along the same line of argumentation we claim that the directionality of the path as well as the number of paths and objects involved in a SOURCE_PATH_GOAL schema influence the perspective of the conceptualization. These two influential variables on the basic underlying schema as well as the four new image-schematic structures we identified can be considered specifications of the overall MOVEMENT_ALONG_PATH schema.

Some of the terms were defined as combinations of image schemas. While we here looked at only PATH-following, we noticed that many concepts would have been better described as combinations of a member of the PATH-following family and additional

image schemas or image-schematic structures such as SCALING or CONTAINMENT, so-called conceptual integrations [15]. Such integrations as well as conceptual blends [6, 10] repeatedly surfaced in our analysis as did different FORCES that might be exerted to a schema. We consider this point definitely important to investigate in future studies.

Our analysis revealed differences across the four languages which could partially be explained by grammatical decisions of the terminologists/experts, partially also by inconsistencies across languages. While the sample in our experiment is considerably too small for any generalized conclusions, the results hint at a high persistence of image schemas across languages. The exact nature of movement along a path can definitely be analyzed in more detail by for instance investigating whether financial descriptions consider the manner of movement, e.g. as done by [19] for a more general corpus.

Prepositions and verbs returned the most promising results in most bottom-up approaches [1, 7, 13], which we could not confirm in our experiment. Synonym sets of nouns returned most image-schematic candidates here. However, this might be attributed to our selection of prepositions and verbs rather than the domain and not represent a contradiction to previous findings.

7. Conclusion and future work

The presented method illustrates how some essential aspects of complicated terms and concepts can be described by using image schemas as a means for simplification. Our analysis contributes two dimensions and four specifications to the most central SOURCE_PATH_GOAL image-schematic structure. While in this study PATH-following was the only image schema considered, in future work more image schemas should be analyzed to better explain the concepts. In fact, conceptual blending and image-schematic integrations, such as PATH and CONTAINMENT repeatedly surfaced during the analysis and could be structured as a paper on their own.

For this first experiment, we exclusively focused on the natural language definitions associated with entries in four languages. In future work it would be interesting to evaluate the image-schematic consistency between the definition and the term that it defines. Additionally, the contrast of the definitions analyzed and the use of the terms in contexts of texts provided by financial experts might provide further interesting insights into the relation of natural language and image schemas. A comparison of our results to other domains of discourse could further strengthen our claim of a domain- and language-independent existence of image-schematic structures.

This approach not only contributes to image schema research by showing that the developmentally most relevant building blocks of our cognitive inventory are carried to abstract adult communication, but also strengthens the idea that image schemas are linguistically and cognitively universal since they exist across languages. The practical use of this approach not only lies in the relation of image schemas and natural language, but since the basis is provided by a formalized theory of PATH-following it also explores the relation between lexical and model-theoretic semantics. In this sense, we believe that this image-schematic method provides an interesting approach to learning spatial ontologies from multilingual text to be explored further in future experiments. Since manual ontology engineering is cumbersome and error prone, automated approaches are required.

We believe that the combination of linguistic and formal analysis of image-schematic structures across languages can allow for their more specialized use in automated approaches and computational systems. Thus, future work will focus on the automation of image-schematic extractions from multilingual textual evidence based on formalized theories. This also includes exploring interconnections of image schemas in form of integrations as well as conceptual blending.

References

[1] B. Bennett and C. Cialone. Corpus guided sense cluster analysis: a methodology for ontology development (with examples from the spatial domain). In P. Garbacz and O. Kutz, editors, *8th International Conference on Formal Ontology in Information Systems (FOIS)*, volume 267 of *Frontiers in Artificial Intelligence and Applications*, pages 213–226. IOS Press, 2014.

[2] B. Bergen and N. Chang. Embodied construction grammar in simulation-based language understanding. *Construction grammars: Cognitive grounding and theoretical extensions*, 3:147–190, 2005.

[3] O. Fuhrman, K. McCormick, E. Chen, H. Jiang, D. Shu, S. Mao, and L. Boroditsky. How linguistic and cultural forces shape conceptions of time: English and mandarin time in 3d. *Cognitive science*, 35(7):1305–1328, 2011.

[4] P. Group. Mip: A method for identifying metaphorically used words in discourse. *Metaphor and symbol*, 22(1):1–39, 2007.

[5] B. Hampe and J. E. Grady. *From perception to meaning: Image schemas in cognitive linguistics*, volume 29 of *Cognitive Linguistics Research*. Walter de Gruyter, Berlin, 2005.

[6] M. M. Hedblom, O. Kutz, and F. Neuhaus. Choosing the right path: image schema theory as a foundation for concept invention. *Journal of Artificial General Intelligence*, 6(1):22–54, 2015.

[7] M. Johanson and A. Papafragou. What does children's spatial language reveal about spatial concepts? Evidence from the use of containment expressions. *Cognitive science*, 38(5):881–910, June 2014.

[8] M. Johnson. *The body in the mind: the bodily basis of meaning, imagination, and reason*. The University of Chicago Press, Chicago and London, 1987.

[9] Z. Kövecses. *Metaphor:A Practical Introduction*. Oxford University Press, USA, 2010.

[10] W. Kuhn. An image-schematic account of spatial categories. In S. Winter, M. Duck-ham, L. Kulik, and B. Kuipers, editors, *Spatial information theory*, pages 152–168. Springer Berlin Heidelberg, 2007.

[11] G. Lakoff. *Women, fire, and dangerous things. what categories reveal about the mind*. The University of Chicago Press, 1987.

[12] G. Lakoff and R. Núñez. *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. Basic Books, 2000.

[13] L. Lakusta and B. Landau. Starting at the end: the importance of goals in spatial language. *Cognition*, 96(1):1–33, 2005.

[14] J. M. Mandler. *The foundations of mind : origins of conceptual thought: origins of conceptual though*. Oxford University Press, New York, 2004.

[15] J. M. Mandler and C. Pagán Cánovas. On defining image schemas. *Language and Cognition*, 0:1–23, may 2014.

[16] I. Mani and J. Pustejovsky. *Interpreting motion: Grounded representations for spatial language*. Number 5 in Explorations in Language and Space. Oxford University Press, 2012.

[17] L. McDonough, S. Choi, and J. M. Mandler. Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology*, 46(3):229–259, 5 2003.

[18] R. E. Núñez and E. Sweetser. With the future behind them: Convergent evidence from aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive science*, 30(3):401–450, 2006.

[19] A. Papafragou, C. Massey, and L. Gleitman. When english proposes what greek presupposes: The cross-linguistic encoding of motion events. *Cognition*, 98(3):B75–B87, 2006.

[20] J. Piaget. *The origins of intelligence in children*. NY: International University Press, New York, 1952. Translated by Margaret Cook.

[21] F. Santibáñez. The object image-schema and other dependent schemas. *Atlantis*, 24(2):183–201, 2002.

[22] L. Shapiro. *Embodied cognition*. New problems of philosophy. Routledge, London and New York, 2011.

[23] R. St. Amant, C. T. Morrison, Y.-H. Chang, P. R. Cohen, and C. Beal. An image schema language. In *International Conference on Cognitive Modeling (ICCM)*, pages 292–297, 2006.

[24] R. Stevens and N. Aussenac-Gilles. Ontoenrich: A platform for the lexical analysis of ontologies. *Knowledge Engineering and Knowledge Management: EKAW 2014 Satellite Events, VISUAL, EKM1, and ARCOE-Logic, Linköping, Sweden, November 24-28, 2014. Revised Selected Papers.*, 8982:172, 2015.

[25] M. Y. Tseng. Exploring image schemas as a critical concept: Toward a critical-cognitive linguistic account of image-schematic interactions. *Journal of Literary Semantics*, 36(2):135–157, 2007.