Comparing and Combining Semantic Verb Classifications[†]

Oliver Čulo[*], Katrin Erk[°], Sebastian Padó[+],
Sabine Schulte im Walde[#]
[*]*Inst. of Applied Linguistics, University of Mainz, Germersheim, Germany*
[°]*Dept. of Linguistics, University of Texas at Austin*
[+]*Dept. of Linguistics, Stanford University*
[#]*Inst. for Natural Language Processing, University of Stuttgart, Germany*

Abstract. In this article, we address the task of comparing and combining different semantic verb classifications within one language. We present a methodology for the manual analysis of individual resources on the level of *semantic features*. The resulting representations can be aligned across resources, and allow a contrastive analysis of these resources. In a case study on the Manner of Motion domain across four German verb classifications, we find that some features are used in all resources, while others reflect individual emphases on specific meaning aspects. We also provide evidence that feature representations can ultimately provide the basis for linking verb classes themselves across resources, which allows us to combine their coverage and descriptive detail.

Keywords: lexical semantics, verb classes, semantic resources, semantic features, resource linking

1. Introduction

Semantic verb classifications, groupings of verbal predicates according to semantic properties, are of great interest to both theoretical and computational linguistics.[1] Examples of such properties are, among others, common meaning components (Koenig and Davis, 2001), or shared argument structure (Levin, 1993). In computational linguistics, verb classifications have emerged as a central tool for generalisation in the face of the ubiquitous sparse data issue. Example applications range from word sense disambiguation (Dorr and Jones, 1996; Kohomban and Lee, 2005), to information access tasks such as query generalisation (Navigli and Velardi, 2003), question answering (Burke et al., 1997; Shen and Lapata, 2007), machine translation (Dorr, 1997; Prescher et al., 2000; Koehn and Hoang, 2007), psycholinguistic modelling (Padó

---

[1] In the context of this article, we will use the more general terms *verb classifications* and *verb classes* to refer specifically to *semantic* verb classifications.

et al., 2006) and statistical lexical acquisition in general (Merlo and Stevenson, 2001; Korhonen, 2002; Schulte im Walde, 2006).

In consequence, there has been continuous interest in the practical construction of broad-coverage semantic verb classifications. Today, such resources exist for many languages; for the most-studied languages, such as English, French, and German, there are even multiple classifications. This wealth paradoxically turns out to be quite problematic, since potential users have to make a choice from a potentially large field of contenders. For English, the options include, among others, WordNet (Fellbaum, 1998), FrameNet (Fillmore et al., 2003), Roget's Thesaurus (Chapman, 1977), HECTOR (Atkins, 1992), VerbNet (Kipper et al., 2000), and OntoNotes (Hovy et al., 2006). Even though all of these resources group the same kinds of objects, their results differ substantially, and often with respect to quite fundamental distinctions. For example, Schulte im Walde (2003) defines semantic classes for German verbs by similar criteria as the German version of FrameNet (Erk et al., 2003); however, while Schulte im Walde classifies the *manner of motion (MOM)* verbs *eilen* and *hasten* (both meaning: 'to rush, to hurry') into a MOM subclass RUSH, FrameNet does not distinguish speed of motion into a separate class and groups these verbs with other SELF_MOTION verbs.

Even though it is possible to contend that there is one correct classification, and that all others err at some point, this position seems rather extreme. In this article, we argue for an alternative standpoint that attributes the differences to fundamental *design decisions*: Semantic classification is a complex task, and resources differ in the *importance* which they attach to different semantic criteria or *features* which determine class assignment. In the German motion verb example from above, one classification concentrates on the type of motion (*to rush* vs. *to move*), the other on the *agentive mover* property. The picture that emerges is one of different resources on an equal footing. (Of course, this does not exclude the possibility that resources also contain individual wrong classifications.)

This situation is rather puzzling for the use of semantic verb classes in computational linguistics, both with respect to their acquisition and their application. As for the *acquisition of verb classes*, automatic methods such as those suggested by Korhonen et al. (2003), Schulte im Walde (2006), or Joanis et al. (2008) have the potential of reducing the prohibitive cost of manual methods. However, they require decisions about both the experiment setup (with regard to feature selection) and the choice of a manually constructed gold standard for evaluation. How can we judge the quality of verb classes learnt from corpora, when a single authoritative gold standard does not exist? With respect to the

application of lexicon resources, NLP approaches (such as the above examples) rely on the verb classifications and the relations (mostly inheritance and part-whole) between the semantic classes for inferences and generalisation. However, all existing lexical resources have their problems, like coverage gaps or variations in granularity. So it is unclear which resource is the most appropriate to use for any given purpose.

In this article, we suggest that a deeper analysis of the resources is required that addresses the following two questions:

(1) How did the differences between classifications arise, and how can we characterise them?

(2) Is there a way to bridge the differences between the resources and even combine their respective strengths?

An investigation of these questions has both theoretical and practical benefits. On the theoretical side, we obtain a better understanding of the *design decisions* present in the process of constructing verb classes by hand. On the practical side, we develop a methodology for *combining* resources in a principled fashion. Ultimately, this should enable more inferences through more relations between semantic classes. For example, in the *eilen* (*to rush*) example above, this would amount to knowing that the verb pertains to both manner and the speed of motion, rather than just one of the two.

We begin by addressing question (1). We perform a data-driven analysis of four individual verb classifications in German within the limited domain of *manner of motion* verbs. We investigate the design decisions of the four classifications, and we compare and combine them using two new methodologies that we introduce: First, we derive *semantic features* used to structure the different resources; second, we induce links between the features of different resources.[2] The comparison of these features allows us to identify the individual weaknesses and strengths of the classifications, and to assess their degree of (dis-)similarity.

The feature-based analysis then allows us to approach question (2), since it results in a list of central features that all resources consider important, in contrast to features that only individual resources have picked up. It thus maps all resources onto a common set of semantic features (to the extent that this is possible). We can use this mapping to construct a *manual linking* between the semantic verb classes of the classifications. We evaluate different techniques for such a linking, and we provide a methodology that can be transferred to similar resources and other languages.

---

[2] We view features as describing *prototypical* rather than necessary and sufficient properties (Taylor, 1989; Hampton, 1993). See the discussion in Section 3 for details.

We study the manual, rather than automatic, linking of the classification resources because our focus is on understanding which types of inter-resource linkings work and which do not, by a careful manual analysis of the properties of various classifications. Automatic procedures (Shi and Mihalcea, 2005; Giuglea and Moschitti, 2006; Chow and Webster, 2007) cannot currently provide linkings with the level of detail and accuracy required for human interpretation. Subsequent work will apply our insights to the acquisition and usage of verb classes.

The article is structured as follows. Section 2 introduces the four semantic classifications of German verbs under consideration. Section 3 performs a manual comparison of the four classifications with respect to their underlying semantic features, Section 4 presents a manual linking of the features across the verb classifications, and Section 5 uses the feature links to induce links between complete verb classes. Finally, Section 6 presents conclusions, including discussions of reusability, the manual effort involved, and prospects for automation.

2.   Description of Verb Classifications

This section introduces four manually constructed semantic classifications of German verbs. One of the earliest extensive verb classifications in German – which is at the same time a rather idiosyncratic one – is the process-based classification by Ballmer and Brennenstuhl (1986), henceforth **BB**. In addition, we chose the two classifications that had most impact in natural language processing, representing major lexical resources not only in German but cross-linguistically: the semantic taxonomy GermaNet (**GN**), cf. Hamp and Feldweg (1997); Kunze (2000), and the FrameNet classes (Fillmore et al., 2003) in their German version compiled by the SALSA project (**SALSA**), cf. Erk et al. (2003). Finally, the semantic classes by Schulte im Walde (2003) (**SIW**) were created specifically for evaluation purposes, as a gold standard resource for an automatic acquisition of semantic classes. BB and SIW are original classifications of German verbs, whereas GN and SALSA are based on existing English resources.

We describe the resources with respect to (1) the motivations and goals of their work, (2) their overall structure, that is, the organisation of the classes and the relations between the classes, and (3) the general decision criteria applied in verb sense distinction and grouping verbs into classes. Steps (2) and (3) are then described in more detail with respect to a selected extract of the classifications, the *manner of motion (MOM)* domain. We conclude this section by describing and comparing how the different resources analyse a sample of verbs.

2.1. A Process-based Classification (**BB**)

Ballmer and Brennenstuhl (1986) classify 8,000 common German verbs (non-prefix verbs only) according to their meaning. Their goal is to build a complete thesaurus of German verbs. Verbs are grouped into classes, which are formed by paraphrasing based on a set of 10 elementary verbs; if verbs agree in central parts of their paraphrases, they are grouped together. Example classes are MOVING ONESELF AWAY FROM A PLACE, with the verbs *sich distanzieren, sich entfernen* (both meaning "distance oneself"), *wegfahren* "drive away" and *verschwinden* "disappear"; or the class paraphrased as SOMEBODY TRANSPORTING SOMETHING FROM A PLACE, USING AN INSTRUMENT/VEHICLE with verbs like *karren* "cart", *schiffen* "ship", and *schaufeln* "shovel".

The verb classes are then organised into *process models*. For example, the process model FORTBEWEGUNG, "moving ahead", contains the verb classes for resting, wanting to move, raising, starting to move, moving ahead, moving in a circle, moving as a passenger, accompanying, getting lost, arriving, stopping, etc. Within a process model, one category stands for each phase of the process, that is, an initial situation, a transition from initial to end situation, an end situation, precondition, result, or consequence. Where one phase of the process can be instantiated by various kinds of movements, there are subcategories. For instance, in the AKTIVBEWEGUNGSMODELL "active motion model", during the ABLAUFPHASE, "active phase", one could make a turning motion, a forward motion, a backward motion etc.; each of these possibilities would be described by a subcategory represented as a verb class. The classes that belong to the same process model are linked by semantic relations such as temporal ordering, causation or implication.

The classification contains five motion-related processes, one describing non-agent, inchoative motion (BEWEGUNGSMODELL: EIGENVERÄNDERUNGEN VON INDIVIDUEN/OBJEKTEN IM RAUM), "self change of individuals/objects in space", one for motion in place with an agent (AKTIVBEWEGUNG) "active motion", one for agent motion with change of place (FORTBEWEGUNG), "moving ahead", one for transport (TRANSPORT), and one for movement with control over a vehicle (FREMDBEWEGUNG), "external motion". The processes all include non-movement as beginning and end state as well as preparatory and wrap-up phases, such as orienting oneself in agentive models, or packing and unpacking in the transport model.

2.2. WordNet/GermaNet (**GN**)

WordNet is a lexical semantic taxonomy developed at the University of Princeton (Miller, 1990; Fellbaum, 1998). The lexical database is inspired by psycholinguistic research on human lexical memory. The resource organises English nouns, verbs, adjectives and adverbs into classes of synonyms (*synsets*), which are connected by semantic relations such as hyponymy, hypernymy, meronymy, etc. The hypernym-hyponym relation imposes a multi-level hierarchical structure on the taxonomy. Words with several senses are assigned to multiple classes. The decision on synonymy is mainly based on substitution tests in prototypical contexts.

The method of WordNet has been transferred to languages other than English. The University of Tübingen is developing the German version of WordNet, *GermaNet* (Hamp and Feldweg, 1997; Kunze, 2000). An example verb in GermaNet is *eilen*, "rush", which is assigned to a common synset with the verbs *sputen, beeilen*, "hurry", and *pressieren*, "be under pressure". The hypernym synsets of the verb class are (bottom-up) *spezielle Geschwindigkeit* (specific speed), *spezielle Bewegart* (specific kind of moving), *fortbewegen* (move ahead), *bewegen* (move), and *lokalisieren* (locate).

The motion verbs in GermaNet are arranged in accordance with the position verbs, below *lokalisieren* "localise"; in fact, *bewegen* "move" and *Position einnehmen* (gloss: "something is or is located or is being located in space") are the only hyponyms of (this sense of) *lokalisieren*, so GermaNet also establishes a close relation between position and motion. Even more, the hyponyms of *Position einnehmen* are position verbs in different stages (partly similar to BB processes) of getting into vs. being in a position. In addition, further down in the is-a hierarchy of *Position einnehmen* are verbs where an agent causes motion, such as *tragen* "carry", *werfen* "throw", *bringen* 'bring', *lehnen* "lean", which again would be motion verbs in BB. But unlike in BB, the position verbs are not part of the motion verbs. The motion verbs themselves subsume the specific verb synsets *regen, rühren* "move slightly" and *rühren* "stir", but also the coarse categories *bewegen auf Stelle* "move in place", two senses of *fortbewegen* ("moving away from source" and "moving ahead with direction"), and *transportieren* "transport". Inchoative vs. causative motion is therefore not a criterion on high-level GermaNet, but change of place and means for movement are. Criteria such as specific kinds of movement and agentivity are distinguished further down in the hierarchy.

2.3. FrameNet/SALSA (**SALSA**)

FrameNet (Fillmore et al., 2003) is based on Fillmore's frame semantics (Fillmore, 1982) and thus describes *frames*, the background and situational knowledge needed for understanding a word or expression. Each frame provides its set of semantic roles, the participants and properties of the prototypical situation. For example, the Motion frame is introduced as following: *Some entity (Theme) starts out in one place (Source) and ends up in some other place (Goal), having covered some space between the two (Path).* To construct frames, FrameNet uses semantic properties both of the target words to be classified and of their semantic roles (Ellsworth et al., 2004). The criteria for sense distinction also lead to a consistent separation of causative, inchoative and static uses into different frames.

Links between the frames of FrameNet are described using a set of currently eight relations. We consider three of these as central, both conceptually and quantitatively (they account for over 85% of the frame-to-frame relations in FrameNet). *Inheritance* is an is-a relation between a parent frame and a child frame that includes full inheritance of semantic roles. *Subframe* is used for linking a scenario frame to its subevents; they may be temporally ordered (in which case scenarios are like BB's processes). *Using* expresses deep conceptual relatedness, as well as a weaker relation of presupposition, and does not require a full mapping of all semantic roles.

The Berkeley FrameNet project is building a dictionary which links frames to the words and expressions that introduce them, illustrating them with example sentences from the British National Corpus. Frames may be evoked by verbs as well as nouns, adjectives, prepositions, adverbs, and multi-word expressions. The SALSA project (Erk et al., 2003) is annotating the German TIGER corpus (Brants et al., 2002) with frames and frame-semantic roles. Its aim is to construct a large, semantically annotated corpus resource as a reliable basis for the large-scale acquisition of word-semantic information. In the course of the annotation, the project builds a German FrameNet, linking the (English) frames to German target expressions. As the FrameNet hierarchy is still being constructed, we can only describe those parts that are actually present. We use the current SALSA snapshot from the outset of our study (early 2004) as basis for our analysis. The resource has grown substantially since then.

SALSA's motion-related classes are not organised in a single contiguous inheritance hierarchy but all point to the central Motion class via the *Using* relation. Motion is unspecified with respect to the type of mover; only its child frame Self_Motion, which also inherits from

INTENTIONALLY_ACT, requires an animate mover. A further area of
motion frames contains CAUSE_MOTION, CARRYING and SENDING,
which all inherit from or use INTENTIONALLY_AFFECT. A "process"
of motion (in BB's terms) is described in the scenario frame MO-
TION_SCENARIO with the sub-situations DEPARTING, MOTION, and
ARRIVING.

2.4. GOLD STANDARD FOR AUTOMATIC CLASS ACQUISITION (**SIW**)

The semantic classification of Schulte im Walde (2003) contains 168
verbs. The purpose of the classification is not lexicographic (that is, to
be exhaustive), but to provide a standard for evaluating the reliability
and performance of clustering experiments that seek to automatically
acquire semantic verb classes. The basis of class creation is subjective
conceptual knowledge, monolingual and bilingual dictionary entries and
corpus search. Verbs are assigned to classes according to their similarity
in meaning, and each verb class is assigned a semantic class label. Some
classes are arranged into a common larger group that again bears a
label, yielding a rather flat hierarchy of only two levels. For example,
the coarse label MANNER OF MOTION is sub-divided into the finer la-
bels LOCOMOTION, ROTATION, RUSH, VEHICLE, FLOTATION. The class
description is closely related to FrameNet: Each verb class is given
a conceptual scene description which captures the common meaning
components of the verbs. Annotated corpus examples illustrate the
combinations of verb meaning and conceptual constructions, to capture
the variants of verb senses.

Since it is intended to represent the gold standard for a statistical
task, the choice of verbs is based on empirically relevant demands. The
classes include both high and low frequency verbs, in order to exercise
the clustering technology in both data-rich and data-poor situations:
the frequencies of the verbs in a large corpus range from 8 to 71,604.
Because any bias in the classification could influence the evaluation of
clustering methods, the classification was constructed to be as unbiased
as possible. Factors that were controlled for include verb frequency,
ambiguity, and semantic domain.

The classification by SIW contains 18 motion verbs in five motion
subclasses: LOCOMOTION contains agentive verbs of forward movement,
ROTATION refers to verbs expressing that specific kind of movement,
not distinguishing agentive vs. inchoative characteristics, RUSH relates
to the specific hurry in motion, FLOTATION to the floating of objects,
and VEHICLE to motion with a vehicle, subsuming both agentive and
participant roles. Verbs denoting the start or the end of a motion
"process" (in BB's terms), such as EXISTENCE verbs, ASPECT verbs, or

POSITION verbs, are assigned to a separate top-level class, not related to motion. Some agentive transport verbs are subsumed under TRANSFER OF POSSESSION.

2.5. INSPECTION OF VERBS ACROSS CLASSIFICATIONS

Having introduced the four resources and their respective design decisions, we now present a few examples to illustrate how resources may differ in their classification of individual verbs. Table I shows a sample of verbs and their classes. While some verbs are analysed quite similarly across resources, the analyses deviate considerably for others.

An example of a very parallel analysis is *wimmeln* "to swarm", which is the same in all resources that include it (though BB adds more specific information). The verb is analysed in all resources by reference to the GROUP MOTION class.

Some differences between resources arise for the case of *sitzen*, "to sit", which in German, like in English, has one reading referring to body posture ("She was sitting on the chair") and another one that expresses a position but remains neutral with respect to posture ("The sensors were sitting in a thermos flask."). GN, SALSA and SIW include the verb in a class describing *position*. In addition, both GN and BB describe the verb as related to non-motion. SALSA is the only resource to analyse *sitzen* as a body posture predicate. All three grouping criteria, or semantic features – *position*, *non-motion* and *body posture* – seem reasonable for *sitzen*. Note that these features are not mutually exclusive (in fact, GN lists *position* as a sub-criterion of *non-motion*), but place emphasis on different features of the meaning.

A similar point can be made for the verb *fallen*, "to fall": GN focusses on the direction of the motion, which the verb shares with *herunterbewegen* "to move downwards", SALSA stresses the physical force causing the motion, and BB notes that *fallen* is usually done inadvertently. In combination, these features form a good description of *fallen*, but each resource concentrates on a different one.

The case of *einpacken*, "to pack", is different in that not all resources list it as related to motion. The verb is related to motion in BB due to its organisation in PROCESS MODELS, which include the preparation and wrap-up phases as well as the "active" phases: *einpacken* is listed for the preparation phase of TRANSPORT. The other resources do not relate *einpacken* to motion because their organisation principles differ.

Similarly, *anschauen*, "to watch", is in a motion class only in BB. However, in this case it is unclear why the verb was classified as related to MOTION. One possible interpretation is that the authors are inspired by an analysis of linguistic expressions in terms of local and spatial

Table I. A cross-resource inspection of some verbs in terms of verb classes.

anschauen "look at" – In SALSA PERCEPTION ACTIVE; in GN hyponym of perception verb *sehen* "see". In BB classification into active motion model AKTIVBEWEGUNG in subclass BEMUSTERN "judge".

ausdehnen "expand" – BB lists *ausdehnen* in the non-agent movement model as well as in the agent movement model. In SALSA, the verb is in a frame describing an item changing its physical size. In GN, *ausdehnen* is below spatial *erstrecken, spannen* "span", causative change (of plans) *verschieben* "postpone" and the change of state verbs *vergrößern* "enlarge" (inchoative) and *verformen* "deform" (both causative and inchoative). So SALSA and GN mainly refer to state change, but not to motion.

einatmen "breathe in" – In BB an agent moving in place. In GN, SIW not related to motion. In SALSA frame BREATHING, which uses FLU-IDIC_MOTION.

einpacken "pack" – In BB preparation phase of TRANSPORT process. In GN, SIW, SALSA not related to motion.

fahren "drive, ride" – In SALSA three classes: riding a vehicle (*ride*), driving a vehicle (*drive*) and transpor-

tation (*drive*); in addition, a class encompassing both *drive* and *ride*. The SALSA annotation found the driver/passenger distinction problematic, since German *fahren* does not differentiate between the focal participant being driver or passenger. However the same distinction is made in GN and BB, two resources developed on German data. In SIW simple LOCOMOTION verb.

fallen "fall" – In BB either just MOTION or ERRONEOUS MOTION. In GN MOTION WITH PATH specified as vertical. SALSA has separate class for *motion by gravity*.

sitzen "sit" – In SALSA POSTURE describing stable body posture of agent, as well as BEING_LOCATED, describing the (geographic) position of an object. In GN *position* verb under *rest.* In BB REST PHASE in motion models. In SIW position verb BE IN POSITION.

wimmeln "swarm" – In SALSA MASS_MOTION; in GN similar class GROUP MOTION. In BB active motion model AKTIVBEWEGUNG in subclass OSZILLIEREN IM KOLLEKTIV "oscillate collectively", which refers both to group motion (as in SALSA and GN) and also to the kind of movement.

patterns (see for example Gruber, 1965). Nevertheless, we tend to see classifications such as *anschauen* in BB as erroneous.

Summarising our observations, we find four categories of the assignment of verbs to classes: (1), verbs for which the resources agree (*wimmeln*); (2), cases where the resources stress different, but compatible semantic features of a verb's meaning (*sitzen, fallen*); (3), idiosyncrasies of individual resources (*einpacken*); (4), cases where we would argue that one of the resource is in error (*anschauen*).

3. Deriving lexical semantic features

In the previous section, we have described the backgrounds and structures of the four verb classifications under consideration, how they were constructed and what purpose they were constructed for. Our analysis in Section 2.5 showed that, as a consequence of these underlying differences, the resources can deviate considerably in the analysis of individual verbs.

Nevertheless, it is necessary to find a *common lexical-semantic representation* for all resources in order to arrive at a precise understanding of these differences and to provide a principled analysis of them. Since this representation is supposed to model the differences between individual classification decisions, it must arguably be more fine-grained than the classes themselves. We adopt the level of *semantic features* as these common properties.

Importantly, the features that we will construct should not be understood as semantic atoms (Katz and Fodor, 1964), that is, basic building blocks into which the meaning of any word can be decomposed, nor as necessary and sufficient conditions that clearly delineate word meanings in the sense of Aristotelian concepts. We conceive of semantic features as reflections of *prototypical* semantic properties (Taylor, 1989; Hampton, 1993) that are (implicitly or explicitly) used by the different resources to group verbs into classes and thus can be used to formally compare and contrast these resources. They are thus related to feature norms in psycholinguistics (Vigliocco et al., 2004; McRae et al., 2005), where features elicited from humans (for example, by asking *What makes an apple an apple?*) have found wide application in explaining phenomena in human language processing.

To see the virtue of a feature-based representation, recall our analysis of the verb *sitzen* "to sit" from Section 2.5. For this verb, *position*, *non-motion* and *body posture* are a good start. Provided that we can obtain such features for the verb, we can explain differences between classifications through differences in the importance that the different resources attach to these features. For example, SALSA concentrates on the *body posture* feature, while BB focuses on the *non-motion* aspect. More generally, features allow us to rephrase the divergence typology we introduced in Section 2.5: the resources can choose to use the same features as central and therefore perform identical class assignments (case 1); the resources happen to choose different features as central, not because one or the other of them misclassified the verb (case 2); a resource chooses idiosyncratic but plausible features (case 3); or a resource chooses idiosyncratic and implausible features (case 4).

The merit of a feature-based analysis of verb classification rests centrally on a clear methodology for deriving these features from the resources. Therefore, we continue in Section 3.1 by defining guidelines for this task, exploiting both direct and indirect sources of information. Section 3.2 reviews the results of this process and shows how this representation already allows for some comparison across different verb classifications.

3.1. A Method for Feature Derivation

Our method for feature derivation is strictly data-driven. Features are derived from resources individually, resulting in a separate set of semantic features for each resource. This is an important methodological point: pre-defining a feature ontology for use with all classifications would force us to adjust the ontology each time a new feature comes up, and would carry the danger of overlooking important aspects not anticipated in the pre-defined ontology. In contrast, features derived from resources provide an "unbiased" view by reproducing more faithfully the distinctions made by each resource.

For each verb class in a given resource, we identify the pieces of information by means of which we can reconstruct what semantic aspects led to the creation of a class, both on its own and in contrast to neighbouring classes. The result of this process is a characterisation of each verb class in a resource in terms of a set of lexical semantic features that apply to all verbs in the class.

The information sources that we can use fall into two general categories. As *direct* sources, we use definitions, verb class names and example sentences given for a verb class, which provide direct insight on the lexical semantic content of a verb class. In addition, there are *indirect* sources, for example structural information like inheritance hierarchies from which we can draw conclusions about the distribution of lexical semantic features within the resource. Deduced features are assigned a two-part name, consisting of a label for the resource and a short mnemonic for the feature. For example, *gn.movement* designates a GermaNet feature called *movement*.

3.1.1. *Direct sources*
In BB, paraphrases and labels are the most important source of feature information. Each verb class is accompanied by a paraphrase that describes the most general common meaning of the verbs in that class. Consider the class *starten jemand 1 etwas 2*, which can be translated into English as "start somebody 1 something 2", where the numbers 1

and 2 denote nominative and accusative case in German, respectively. This class is characterised by the following paraphrase:

(1) machen jemand 1 daß anfangen es daß sich fortbewegen jemand 1
     'make   someone that begin   it that move   someone'

This gloss indicates a movement (*bb.movement*) that is caused by some person (*bb.caused_motion*, *bb.agent*). The aspect of the motion being caused by some outer force and the motion being started (rather than being in progress) is expressed by the class label FD^B, where FD stands for FREMDBEWEGUNG "external motion" and superscript B for the beginning phase of the process.

In SALSA, much information can be drawn from the frame definitions. For example, the frame FLUIDIC_MOTION in SALSA is defined as:

(2)   A Fluid moves from a Source to a Goal along a Path or within an Area.

There are at least two semantic features contained in this definition. The verbs relate to a motion (*salsa.movement*) that involves a FLUID as the moving object (*salsa.first_mover_fluid*). In this case, the definition gives no further specifications on SOURCE, PATH and GOAL (in contrast for example to the frame ARRIVING, in which GOAL is profiled), so they are not used as features. Note that in SALSA, features are often derived from properties of roles (like SOURCE or THEME) because of FrameNet's focus on semantic roles.

In GN, each synsets is characterised by a so-called *gloss*, a short description of the verb class content, and/or by example sentences that illustrate typical uses of the verbs in the corresponding synset. Such a gloss is for example *sich leicht, ein wenig bewegen* "to move slightly, a little" for the synset *regen, rühren* "to budge, to stir". This gloss leads us to two features, namely *gn.movement* and, as it is only a light movement, *gn.a_little*.

SIW provides definitions for each verb class, together with subcategorisation frames and corpus examples, illustrating the frame patterns for each verb. The class BRING INTO POSITION is defined as follows:

(3)   [A person or some circumstances]$_{\text{MOVER}}$ bring
        [something]$_{\text{PATIENT}}$ into
        [a spatial configuration]$_{\text{CONFIGURATION}}$.

From this definition we deduce the feature *siw.movement* for the movement that happens before the patient reaches the spatial configuration. As this class characterises some sort of accompaniment, where two

or more movers share part of the same way but end up in different
locations, we also introduce the feature *siw.theme_split*. Furthermore,
the two features *siw.caused_motion* relating to the fact that the mover
causes the patient to move and *siw.result_static_location* for the result
state of the patient at the end of the motion process are used to describe
this class.

3.1.2. *Indirect sources*

This category comprises features that are not explicit in the information
pertaining to individual classes, but can be inferred in other ways. In
particular, relations between classes in the classifications can be used to
deduce features. Since all classifications we consider are tree- or graph-
shaped, two prominent types of indirect sources are (a), inheritance
of features from a more general parent to the more specific children
classes, and (b), contrastive features between sister classes (classes with
a common parent class). Also, an overly unspecific class definition may
need to be enriched with common features of the verbs in the class, a
process that clearly involves world knowledge.

BB contains flat hierarchies, where the parent categories denote
complete processes, and the child categories describe the different phases
of this process. We found that features can be reliably percolated from
parent to the child categories which belong to the *Ablaufphase* "active
phase" of a model. The BB classes in the other phases are additionally
ordered according to temporal criteria within the process models; how-
ever, since the relationship between temporal order and the existence
of specific features is quite complex, we ignored the temporal order for
the current study.

GN defines a much deeper hierarchy for its semantic classes than
the other three classifications. When deriving semantic features from
the hierarchy, we confer to a verb class all features of its parent class,
unless inspection shows that a feature should not be passed on (which
happens only in a small minority of cases). For example, the direct
information in the GN gloss for the verb synset *schwimmen* "swim"
provides the example sentence *Er hat in der Schule schwimmen gelernt*
"He learned swimming in school". This sentence already hints out to
an "active" swimming, as opposed to the non-agentive swimming in the
sentence *Ein Stück Holz schwimmt auf dem See* "A piece of wood swims
on the lake". This assumption is confirmed in the indirect information
provided by the class hierarchy: *schwimmen* is a child of the sense of
fortbewegen that refers to a directed movement. Taking this indirect
information into account, we interpret the movement as volitional,
adding the semantic feature *gn.agent*. Since we know that swimming

usually happens in water, we also add the feature *gn.medium_water* for this synset.[3]

In SALSA, the hierarchy is much sparser than in GN. Nevertheless, the structure of the resource allows for a contrastive analysis of classes. The classes which we consider as MoM classes are the MOTION frame itself, plus classes related to the MOTION frame by the relations INHERITANCE or USING.[4] We consider MOTION to be an unspecific base class, only characterised by the feature *salsa.movement*. All other classes from the MoM domain share the feature *salsa.movement* but differ in at least one additional feature from MOTION. For example, the frame MASS_MOTION differs from MOTION in that the theme of the motion consists of more than one entity (*salsa.first_mover_many*). In addition to contrasting frames with the neutral class MOTION, we also compare sister classes directly with each other. For example, if we compare OPERATE_VEHICLE and BODY_MOVEMENT, both involve a mover that is sentient (*salsa.first_mover_sentient*), but there is a second moving entity in both cases: for OPERATE_VEHICLE it is a vehicle that is controlled by the first, sentient mover (*salsa.second_mover_vehicle*), for BODY_MOVEMENT the second moving entity is a body part of the sentient mover (*salsa.second_mover_bodypart*).

In the SIW classification, indirect feature sources are rare, the only one available being the flat hierarchy of depth two. In the MoM domain, there are two top-level classes, POSITION and MOTION, which are defined by only one feature each, namely *siw.result_static_location* and *siw.movement*, respectively. These features can be passed on to the children classes.

3.2. ANALYSIS OF MoM FEATURES FOR INDIVIDUAL RESOURCES

This section presents the results of applying the feature derivation method presented in Section 3.1 to the Manner of Motion domain in the semantic classifications from Section 2.5. The complete feature data is freely available for academic research and can be downloaded from `http://loci.macbay.de/Laden/verbClasses.zip`.

Quantitative analysis.   Table II provides quantitative overview, listing statistics about the number of verbs, classes and features per classification. The first row shows the total number of classes in the MoM domain of each resource. The second row lists the number of non-empty

---

[3] While a feature *gn.medium_fluid* might have been preferable, given that swimming can take place in other fluids besides water, the feature *fn.medium_water* describes the *prototypical* case of swimming.

[4] Compare the discussion of the FrameNet relations in FrameNet in Section 2.

Table II. Statistics for the MoM domain of the four verb classifications.

	BB	GN	SALSA	SIW
classes	95	58	21	9
non-empty classes	84	30	17	7
verbs	1215	41	112	24
avg. verbs / non-empty class	14.5	1.4	6.6	3.4
distinct features	68	49	22	9
avg. features assigned to non-empty class	6.0	3.2	4.8	2.6

classes: Some of the resources, in particular GermaNet, use classes that do not contain any verbs for structuring reasons. We see that the MoM domains are of very different sizes: the largest classification, BB, contains ten times as many classes as the smallest one, SIW. Rows three (number of verbs for each resource) and four (average number of verbs per class) indicate that this is not only a consequence of finer-grained distinctions. The process-based BB resource simultaneously covers the highest number of verbs, and has the largest classes. At the other end of the scale, we find that SIW, which was originally designed as an evaluation resource, has the lowest verb coverage; GN, with its highly hierarchical structure, has the smallest classes. SALSA occupies an intermediate position with respect to coverage and class size.

The last two rows provide statistics on the derived features. Row 5 lists the number of distinct features that were derived for each resource; note that these figures show a reasonably close correspondence to the number of classes of this resource, indicating that the resources with more classes in fact make use of a larger inventory of semantic properties. The last row lists the average number of feature instances assigned to a non-empty class. While these figures still show variation between resources, there is much less difference at this level than at the level of classes. The highest number of features per class was found again in the BB classification, which results mainly from the very detailed definition of the process classes (cf. Section 2.1).

Most frequent features. To illustrate the nature of the derived features, Table III lists the most frequently used features for each resource. The BB features reflect the top level of the resource's hierarchy with its five different MoM processes: Movements are classified as agentive vs. non-

Table III. Most frequently used features.

BB	bb.dislocation, bb.agent, bb.controlled, bb.theme_split, bb.movement_in_area
GN	gn.dislocation, gn.agent, gn.obj_moved, gn.caused_motion, gn.theme_not_many
SALSA	salsa.common_path, salsa.path_not_profiled, salsa.caused_motion, salsa.goal_profiled
SIW	siw.caused_motion, siw.result_static_location, siw.theme_split

agentive and controlled vs. uncontrolled, reflected in the frequent use of the features *bb.agent* and *bb.controlled.* Also, BB distinguishes between movement in place and movement with dislocation, as shown by the frequent use of the features *bb.dislocation* and *bb.movement_in_area*, the latter one being very important in the transportation model, where there are several different categories for local and global transport.

In GN, the strongly hierarchical and more lexically driven structure of the resource makes it less obvious than in BB that there are features that are used throughout the MoM domain. Nevertheless, we found a number of such features, such as the distinction between movements in place and those involving a dislocation (frequent use of *gn.dislocation*). There are also frequent distinctions between agentive self-motion and patients being moved, as is reflected by the fact that both *gn.agent* and *gn.obj_moved* as well as *gn.caused_motion* belong to the most often used features in the resource GN, as is shown in table III.

The SALSA classification puts its emphasis on the properties of the source, path, and goal of the movement. This is a result of FrameNet's general semantic role-based classification scheme (cf. Section 2.3) when applied to the MoM domain. In consequence, three out of the four listed SALSA features deal with source-path-goal: *salsa.common_path*, *salsa.path_not_profiled* and *salsa.goal_profiled.*

Finally, the SIW classification is both the smallest and uses the lowest numbers of features; we identified only 3 features which appear more than once. One of them, *siw.result_static_location*, reflects the top-level distinction between positioning classes on the one hand and motion classes in general on the other hand; the others *siw.caused_motion* and *siw.theme_split* express aspectual information (caused motion vs. self-motion vs. transportation).

Detailed example analyses. We return to two of the verbs we analysed in Section 3.1, namely *fallen*, "fall" and *wimmeln*, "swarm" and analyse

them in terms of on the semantic features we have derived for their verb classes. Table IV shows the verb classes containing these two verbs along with their semantic features.

The verb *wimmeln* is analysed as monosemous in BB, GN and SALSA. SIW does not cover this verb. In BB, GN and SALSA it receives a more or less uniform analysis: the meaning aspects of movement and collective movement are encoded in all three resources. In addition, GN and BB state that *wimmeln* does not involve overall dislocation. BB, which has the most fine-grained analysis in this case, additionally indicates the agentivity of the mover and the controlledness and, in terms of BB's description, oscillatory nature of the movement.

The situation is more complex for the verb *fallen*, which occurs in three different classes in BB, in one class only in GN and SALSA, and is again not covered by SIW. The first BB class (class $BW_{0.5}^A$, cf. table IV) covers the meaning of "to fall over", that is, a change of orientation without change of location. The BB analysis stresses the erroneous, uncontrolled aspect of the motion. The second class ($BW_{2.1}^A$) simply states that *fallen* can also be a dislocating event, again in an uncontrolled manner. The third reading given in BB ($FB_{0.3}^A$) seems somewhat unusual. It describes *fallen* as a controlled motion caused by something. The class also contains other intransitive verbs such as *versinken* "sink" and is described by the gloss "*something is being moved by something*". We speculate that BB considers gravity as the underlying cause of the movement. The GN analysis contains neither of the two aspects of controlledness or causation, but stresses the downward direction of the movement. SALSA focuses on the non-agentive character of the motion (*salsa.no_self_mover*) and specifies that gravitational forces control the path of the motion (*salsa.path_gravity*).

Conclusions.   Our analyses show that a semantic feature-based representation for verb classes is able to characterise these classes simply and intuitively and provides a suitable common level for comparing and contrasting the resources.

Note that due to our choice of deriving features first within individual resources, the resulting representations are by no means equivalent. We tend to see this as an advantage, since the various representations provide us with a more complete picture of potential aspects of the verb classes' meanings, seen from different angles. However, it is a valid question how independent representations can be contrasted across resources. This is unproblematic when equivalences or contradictions are "read off" the feature names manually, as we did in the examples above. The next section will provide a more general method for determining the relationship of arbitrary feature pairs from two resources.

Table IV. A cross-resource inspection of two verbs in terms of semantic features.

BB	GN	SALSA
fallen		
$BW_{0.5}^{A}$, *'uncontrolled, erroneous movement'*: bb.uncontrolled, bb.no_dislocation, bb.erroneous $BW_{2.1}^{A}$, *'to dislocate'*: bb.uncontrolled, bb.dislocation $FB_{0.3}^{A}$, *'be made to move'*: bb.cause, bb.caused_motion, bb.controlled, bb.dislocation, bb.movement;	{*fallen*} gn.direction, gn.dislocation, gn.movement, gn.path_vertical, gn.path_down	MOTION_DIRECTIONAL: salsa.no_self_mover, salsa.movement, salsa.path_gravity
wimmeln		
$BW_{2.1}^{A}$, *'oscillate collectively'*: bb.movement, bb.oscillation, bb.theme_many, bb.agent, bb.controlled, bb.no_dislocation	{*wimmeln*} gn.dislocation, gn.theme_many	MASS_MOTION: salsa.first_-mover_many, salsa.movement

4. Linking features across resources

In this section, we discuss the question of identifying correspondences between features from different resources in more detail (Sec. 4.1). The resulting method will then enable us to construct a comparison of the resources on the level of individual semantic features (Sec. 4.2).

4.1. FEATURE LINKING

The most straightforward relation between features is *equivalence.* As an example consider *gn.theme_many*, *bb.theme_many*, and *salsa.first-mover_many*, all of which have the interpretation of the mover consisting of multiple entities. All of these features have been assigned to the verb class containing *wimmeln* "swarm". Similarly, the features *bb.rapid* and *gn.rapid*, both assigned to verb classes containing *eilen* "rush" and *spurten* "sprint", are equivalent.

Features can be equivalent even if their names are not the same across resources: *gn.cause_onset* and *bb.initial_impulse* both describe cases of motion caused by an initial impulse where the motion continues

after the impulse has ceased acting on the theme. (The opposite would be motion that is kept up by a continued force.)

However, equivalence is not the only possible relation between features. Consider the features *salsa.first_mover_fluid* and *siw.theme_or_medium_fluid*. The first feature describes motions of fluids and is used for the FrameNet/SALSA class Fluidic_motion, which contains words such as *sprühen* "spray" or *sich verteilen* "spread", and *to bubble*, *to cascade*, and *to flow* for English. In contrast, the feature *siw.theme_or_medium_fluid* can describe either motion *of* fluids or movements of solid entities *in or on* fluids. Thus, the SIW class Flotation characterised by this feature contains *fließen*, "flow", but also *gleiten*, "glide", and *treiben*, "float".

In this case, one feature (here *siw.theme_or_medium_fluid*) is more general than another one (*salsa.first_mover_fluid*). We will call the latter a *subfeature* of the former. The subfeature relation between two features is established by verifying that the more general feature applies to all verbs in all classes for which the more specific feature is listed, but not the other way around.

Subfeature relations arise naturally from situations where features differ in their granularity. For example, the features *salsa.movement*, *siw.movement*, *gn.movement* and *bb.movement* all describe verb classes that involve motion. This feature applies to most of the verb classes included in this study, though there are exceptions like *ruhen* (*to rest*), which describes non-movement. The BB feature *bb.movement_in_area* is a subfeature of these four general motion features: It describes motion verbs that profile the area or territory traversed, and applies to verbs such as *durchkreuzen* (*to cross*).

Using these two relations (*equivalence* and *subfeature*), the features of the four resources we are considering were linked manually. The resulting links are also included in the archive mentioned above in Section 3.2.

4.2. Cross-resource comparison of features

The cross-resource feature links now allow us to compare resources according to the features they use. Section 2 has shown that the four resources that we study differ considerably, in their overall design decisions, hierarchical structure, scope of verbs listed for MoM, and grouping of verbs into classes. An obvious question to ask is to what extent they differ with respect to the features they employ to structure the MoM domain.

One main method we used in feature definition was to contrast verb classes, in particular sister classes. In this sense we can view the features

as the sense distinctions that each resource makes. Hence, a comparison of resources by their features should help answer the question of whether the central sense distinctions that the resources draw are as idiosyncratic as their hierarchical structures, or whether there are some central sense distinctions that would appear in all or almost all resources.

Table V shows five large groups of features, each of which is present in all four resources to some extent. The first group describes the mover, a feature group with a strong presence across all resources. All four resources distinguish a class in which the mover is a fluid. In addition, three out of four resources have features for an agentive (volitional) mover, motion by a group of entities, and the movement of a human body part. In contrast, only GN lists *pflanzen*, "to plant", among the MoM verbs and thus has a feature for the moved object being a plant.

The second group contains features pertaining to source, path or goal of the motion. While three out of four resources have one or more features specifying the path, only GN and SIW use rotation to distinguish verb classes. (SIW refers only to a movement in place.)

The third group, *Instrument*, shows that motion by way of a vehicle occurs as a feature in all four resources. The fourth group has again some common and some idiosyncratic features: All resources distinguish some form of caused motion; only two of them have a separate feature for onset impulse, that is, a force causing the beginning of the motion as opposed to driving it continually. Motion for social reasons (as in *flanieren*, "to stroll") is again an idiosyncratic BB feature.

The last group shows a collection of features describing properties of the motion itself. Again, we see that three of these features occur in all or almost all groups: speed of motion, motion accompanied by noise, and non-motion (standing still). Decreasing the speed of a motion event (*bremsen*, "to brake"), idiosyncratic for BB, is typical of its process scheme that includes a starting and an end phase for each process.

Summing up, we see that while the resources differ considerably in their treatment of individual verbs, the semantic features they use are not all idiosyncratic. Rather, we can identify *central* semantic features by virtue of their appearance in all or nearly all resources. It is worth noting that these central features contain no surprises (maybe with the exception of motion accompanied by noise): They are what one would expect a resource to use for grouping MoM verbs. These findings correspond to the findings of Schulte im Walde and Erk (2005).

Table VI gives some quantitative information about feature links between the resources. The resource with the highest number of features is BB; still, it does not have as many linked features as GN (24 vs. 26), though GN has little more than half as many features. One possible

Table V. Occurrence of sample criteria in the four resources

	BB	GN	SIW	SALSA
Type of mover				
motion of fluid	+	+	+	+
agentive mover	+	+	-	+
mass motion	+	+	-	+
movement of body part	+	+	-	+
moved object: plant	-	+	-	-
Source/Path/Goal				
path specified	+	+	-	+
source specified	+	+	-	+
rotation	-	+	+	-
Instrument				
motion by vehicle	+	+	+	+
Causation				
caused motion	+	+	+	+
onset impulse	+	+	-	-
motion cause: social	+	-	-	-
Properties of the event				
rapid motion	+	+	+	+
motion with noise	+	+	-	+
non-motion	+	+	+	-
decrease speed	+	-	-	-

reason is that BB includes many verbs that other resources do not list under MoM, resulting in idiosyncratic features such as motion for social reasons, or decreasing speed of motion (see Table V). As for the

Table VI. Statistics about features per resource and feature links per resource.

	features	linked feat.	percentage
BB	68	24	35.3
GN	40	26	65.0
SALSA	22	17	77.3
SIW	9	8	88.9

percentage of linked features, we note that in the case of SIW all but one features are linked. As it has the fewest features of the four resources, they seem to correspond to generally accepted "basic" distinctions in the MOM domain.

5. Using feature links to derive links between verb classes

We now come to the second step of the roadmap we proposed in Section 1: bridging the differences between classifications. This involves linking the resources' verb classes themselves rather than just linking their semantic features. Our goal is not to develop a full-fledged computational model for mapping verb classes, but to analyse the virtues and limitations of feature linking as basis for this task. Section 5.1 begins by revisiting the intuition behind our approach and discussing methodological issues. Sections 5.2 and 5.3 then analyse the outcome of a naive feature-based verb class mapping. We again use the MoM data for concrete illustration.[5]

5.1. From feature links to verb class links

In theory, the situation is simple: Overlap in semantic features between two classes indicates commonality in meaning. Thus, two classes that share all features are equivalent. However, in practice, each resource introduces idiosyncratic features that cannot be linked to other resources at all (e.g., features referring to properties of the process model in BB, such as *bb.controlled*). This leads to the question of what exactly

---

[4] Movement of round objects

[5] In this section, we use "feature" to refer to resource-independent features which are linked as described in the previous section.

constitutes equivalence between verb classes from different resources, given that these resources take different perspectives on conceptualising the world. If classes cannot be expected to share all features, how many features are "enough"? To explore the relationship between feature links and verb class links, we performed a small experiment on the MoM domain. We manually created a gold standard of verb class links, and checked how well feature links predicted verb class links.

The gold standard was constructed by compiling all links between verb class pairs which we have judged by manual inspection to be either equivalent or in a subclass/superclass relation. We found 145 such links. While this number appears small compared to the total number of possible links (10,750), there is a certain number of classes which can properly be called identical across resources. In a precision-oriented approach like the one we take in this paper, such links are highly valuable. For example, they allow us to combine the lexical coverage of different resources, and to profit from complementary descriptions and inheritance links of verb classes in different resources (cf. Section 3.2).

To predict verb class links on the basis of feature links, we will consider two decision procedures, which correspond to two extreme positions: (a), two classes are linked when they share at least one feature (minimal evidence); (b), two classes are linked when they share all features (maximal evidence). For procedure (a), we disregard the feature *xx.movement* for the linking, as it would have provided links between virtually all classes within the motion domain. On the other hand, a feature like *xx.non_movement* is clearly of interest for the minimal evidence procedure. While both procedures are obviously empirically inappropriate, they allow us to isolate specific error types and to gauge the fundamental usefulness of feature links for deriving verb class links.

5.2. Quantitative Analysis

Table VII shows the number of verb class links that result from the two decision procedures. The "correctly generated" row shows the number of gold standard links that are generated by feature-driven linking (true positives). "Missing" is the number of gold standard links not generated by the feature-driven linking (false negatives); and "overgenerated" is the number of links that are falsely created (false positives).

The left hand side of the table shows the results for the "minimal evidence" mapping. We see that almost all links in the gold standard have been created (the number of missing links is very small). This bolsters the first part of our hypothesis: there are almost no verb class links which are not warranted by feature links. However, the presence of one matching feature alone is a very weak indicator of a class mapping.

Table VII. Number of links in different classes, depending on two feature-driven linking schemes

Link types	Minimal evidence	Maximal evidence
Correctly generated	134	0
Missing	11	145
Overgenerated	1435	0

Around 15% of all possible verb class pairs share one feature, which gives us a large number of spurious links (thus, overgenerated).

The right hand side of the table shows the results for the "maximal evidence" mapping, for which we require that all features of A are mapped onto B and vice versa. In this case, there are no overgenerated links. However, as predicted, none of the class pairs meets the maximal evidence condition at all: all 145 links are missing.

5.3. ERROR ANALYSIS

When we analysed the erroneous and missing links predicted by the two feature-based decision procedures, we found that the GermaNet class PFLANZEN "plant" participated in a particularly high number of errors. In the gold standard, this class is linked (in the subclass relation) to exactly one other class, namely BRING INTO POSITION from the SIW resource. Using the "minimal evidence" mapping, however, PFLANZEN is linked to 57 other classes, a massive overgeneration problem. This can be traced back to the presence of a very general feature, *gn.obj_moved*, which is shared by many classes in all resources, and which is sufficient to trigger the linking. In the "maximal evidence" case, on the other hand, not a single link is found for this class (missing link error). This is a result of its feature *gn.obj_plant*, which describes the moved object as being a plant, and which has not been linked to any features from other classes, including BRING INTO POSITION. The representation of the latter class does not contain any feature that describes that the moved object may be anything (e.g., *siw.obj_top*) and that could be linked to the *gn.obj_plant* in the subfeature relation.

The case of PFLANZEN exemplifies a first problem of feature-based verb class linking, namely *inadvertent gaps in feature representations*. Very general features are particularly vulnerable in this respect: they may be taken for granted during the analysis and not represented explicitly. For example, the SIW class ROTATION, which has the feature *siw.rotation*, only concerns movements in place, but not movements in circular paths. However, this is only implicitly stated in the class de-

scription. Consequently there is no feature like *siw.movement_in_place*, which leaves a gap in the feature representation.

Another problem is that links may be overgenerated when classes share so many features that they cannot be distinguished well, even though they are ultimately not equivalent. For example, consider the SALSA class BRINGING and the SIW class BRING INTO POSITION. Both classes involve *caused motion* and two movers (*siw.theme_split* vs. *salsa.has_second_mover*). Therefore, these classes will be linked by the "minimal evidence" decision procedure, and even for considerably stricter decision procedures. However, BRING INTO POSITION describes a situation in which an object is positioned while the agent moving the object does not change their overall position, and the result of the action is a static location of the object, hence the feature *siw.result_-static_location*. BRINGING, in contrast, describes a situation where two movers traverse a common path and end up in at the same location, so it has the feature *salsa.common_path*. This means that the two classes should not be linked. What appears to be missing in our present framework is a way of explicitly representing information about *feature contradiction*. We only encode "positive" information about the similarity of features (via either equivalence or subfeature). However, the above analyses show the need for representing "negative" information: if two classes have contradictory features, they should not be linked.

Finally, recall that our present study had to exclude the feature *movement* from consideration, since in the MoM domain virtually all classes share this feature. This indicates that our model should benefit from the inclusion of feature weights that model their respective *informativity*: Very general features, which convey little information, should not serve as sole evidence for links between verb classes.

5.4. DISCUSSION

This section investigated the usefulness of semantic features for the task of linking verb classes across resources. Our general impression is that this approach is promising. Notably, virtually all verb class links in the gold standard are warranted by feature links. Where this is not the case, it usually results from errors in the manually created features or feature links.

However, the decision procedures that we proposed for linking verb classes are far from usable in practice. Our error analysis indicates that this is due to the limits of our present representation. An important avenue of research is therefore the development of richer feature representations, for example in terms of feature importance, or compatibility between features. Work on this topic has to address two separate (but

related) levels. The first level is the linguistic one, where it must be decided what *types* of feature information have a bearing on links between verb classes. As discussed above, we think that feature informativity, encoded as weights, and incompatibility information are both promising candidates in this respect. However, feature weighting in particular requires more investigation, since results from cognitive psychology indicate that there is no unique "correct" way of weighting features (Sloman et al., 1998). The second level is the operational one, which is concerned with obtaining these types of information automatically from data. It will be discussed below in more detail.

6. Conclusion

This article has been concerned with the comparison and combination of semantic verb classifications. Faced with the availability of competing resources, we first asked whether some classifications are "more correct" than others. In an informal analysis, we found that the relation between classifications usually is one of complementarity: Different resources emphasise different meaning aspects, and thus arrive at different segmentations of the semantic space. Often, analyses are equally plausible, and a combination of the aspects from all analyses appears to yield a more complete picture of the meaning aspects of a verb.

In order to put the comparative analysis of verb classifications on a more principled basis, we have fleshed out the notion of meaning aspect in the form of *lexical semantic features*, and have described a method to derive these features from the classifications themselves. We have demonstrated the feasibility of our approach by analysing the *manner of motion (MoM)* domain of four German semantic verb classifications.

The resulting feature representations allow us to compare the verb classifications both quantitatively and qualitatively. We have presented statistics on the number and distribution of lexical semantic features within the classifications, giving an overview of how fine-grained and detailed the analyses in the different classifications are with respect to particular meaning aspects. In addition, by linking features of different resources, we have been able to distinguish between semantic aspects of general relevance for the MoM domain, and aspects specific to particular resources (cf. Table V).

In a second step, we have gauged the usefulness of using features for linking verb classes across resources. To do so, we have interpreted the features of a verb class as a description of its semantic content, in order to identify pairs of classes whose semantics are near-equivalent or stand in a clear superfeature–subfeature relation. While we have only

presented some first steps in this direction, we think that the feature-based combination of verb classes has a considerable potential, both with respect to combining semantic information across resources, and for improving the coverage of resources.

The feature-based comparison and combination of verb classes provides a number of distinct benefits. In contrast to current automatic approaches (Shi and Mihalcea, 2005; Giuglea and Moschitti, 2006; Chow and Webster, 2007), our approach takes into account all sources of information that are available from a resource, and is thus less reliant on the redundancy of the encoded information. As we argued above, it can, for example, establish links between verb classes even when there is no overlap in verbs. Also, it provides a greater degree of control over the analysis process, since its results and their justifications are human-readable. They are thus amenable to human inspection and verification, which in turn allows us to obtain high-quality links as well as insights into the structure of the resources.

To conclude, we will discuss three central issues arising from this article, namely reusability, manual cost, and automation.

Reusability. The question of the usefulness of our method beyond the present study can be split into two more precise questions: (1), Can the verb class features that we list for FN, BB, GN, SIW be used to describe other resources? (2), Can the technique described in this article be applied to other domains, resources, and languages?

With respect to (1), we expect at best a limited reusability of the feature list. While some of the features that we identified were used in all four resources, others were highly resource-specific (cf. the idiosyncratic classes of BB). Thus, the analysis of new resources would almost certainly require the addition of features. The same holds for the analysis of new languages; for example, Japanese and Romance languages show a conceptualisation of the motion domain that differs from Germanic languages (Ohara, 2004). As for (2), we have tested the reusability of the technique proposed in this paper by performing a small case study that applied the scheme developed for MoM verbs to the Perception domain of our four resources. We did not find any problems in deriving features and linking verb classes. However, when new resources are analysed, the methods for eliciting features have to be adapted to their respective structures. Finally, we also expect the technique to be largely language-independent. Evidence in this direction is provided by SALSA, one of the four resources we have considered. SALSA is the German version of the FrameNet resource originally developed for English, and retains the original English class structure, solely replacing English by German lemmas. In this sense,

our feature analysis of SALSA should apply equally to the English FrameNet verb classes of the MoM domain.

Manual Cost. Next, there is the issue of the manual effort required by the techniques we have proposed in this paper. The creation and assignment of features for individual classes is comparatively straightforward, requiring a few minutes per class. However, the time required for a subsequent consistency check, as well as the determination of relations between features, is less predictable. The consistency checking step, which involved detecting mistakenly omitted feature assignments as well as merging near-synonymous features, turned out to be important for the coherence and completeness of the feature space, but required considerable time.

Automation. Finally, it is an important question to what extent the individual steps of comparing and combining lexical resources can be automated. The difficulty of the first step, the automatic induction of features that describe the lexical resources, strongly depends on the structure of the resources. Direct sources (such as glosses, and examples) are easier to exploit than indirect sources (such as the hierarchical structure), and we believe that the amount of manual interaction depends on the direct accessibility of properties. However, there are encouraging results concerning the acquisition of semantic features and semantic relations from corpora (Hearst, 1992; Lin, 1998; Maedche and Staab, 2000; Cimiano, 2006; Nastase et al., 2006; Pantel and Pennacchiotti, 2006; Baroni and Lenci, 2008). Relying on such lexical acquisition methods, automatically induced features and relations can complement directly accessible ones. In a second step, it will often be necessary to bridge non-identical but related properties; this task can be approached using standard approaches to semantic similarity, such as distributional measures (Dagan et al., 1999; Curran, 2004; Weeds and Weir, 2005; Budanitsky and Hirst, 2006; Padó and Lapata, 2007). Once a set of (comparable) rich features is available, we assume that an automation of the verb class linking is relatively straightforward.

Acknowledgements

References

Atkins, S.: 1992, 'Tools for corpus-aided lexicography: The HECTOR project'. *Acta Linguistica Hungarica* **41**, 5–72.

Ballmer, T. T. and W. Brennenstuhl: 1986, *Deutsche Verben. Eine sprachanalytische Untersuchung des Deutschen Verbwortschatzes*. Tübingen, Germany: Gunter Narr Verlag.

Baroni, M. and A. Lenci: 2008, 'Concepts and properties in word spaces'. *Italian Journal of Linguistics*. To appear.

Brants, S., S. Dipper, S. Hansen, W. Lezius, and G. Smith: 2002, 'The TIGER Treebank'. In: *Proceedings of the Workshop on Treebanks and Linguistic Theories*. Sozopol, pp. 24–41.

Budanitsky, A. and G. Hirst: 2006, 'Evaluating WordNet-based Measures of Lexical Semantic Relatedness'. *Computational Linguistics* **32**(1), 13–47.

Burke, R. B., K. J. Hammond, and V. A. Kulyukin: 1997, 'Question answering from frequently-asked question files: Experiences with the FAQ finder system'. *AI Magazine* **18(2):57**(2), 57–66.

Chapman, R. (ed.): 1977, *Roget's International Thesaurus*. New York: Harper and Row, 4th edition.

Chow, I. and J. Webster: 2007, 'Integration of linguistic resources for Verb Classification: FrameNet Frame, WordNet Verb and Suggested Upper Merged Ontology'. In: *Proceedings of the Eighth International Conference on Intelligent Text Processing and Computational Linguistics*. Mexico City, pp. 1–11.

Cimiano, P.: 2006, *Ontology Learning and Population from Text*, Chapt. Learning Attributes and Relations. Springer.

Curran, J.: 2004, 'From distributional to semantic similarity'. Ph.D. thesis, University of Edinburgh. URL: http://hdl.handle.net/1842/563.

Dagan, I., L. Lee, and F. Pereira: 1999, 'Similarity-Based Models of Word Cooccurrence Probabilities'. *Machine Learning* **34**(1-3), 43–69. Special Issue on Natural Language Learning.

Dorr, B. J.: 1997, 'Large-Scale Dictionary Construction for Foreign Language Tutoring and Interlingual Machine Translation'. *Machine Translation* **12**(4), 271–322.

Dorr, B. J. and D. Jones: 1996, 'Role of Word Sense Disambiguation in Lexical Acquisition: Predicting Semantics from Syntactic Cues'. In: *Proceedings of the 16th International Conference on Computational Linguistics*. Copenhagen, Denmark, pp. 322–327.

Ellsworth, M., K. Erk, P. Kingsbury, and S. Padó: 2004, 'PropBank, SALSA, and FrameNet: How Design Determines Product'. In: *Proceedings of the LREC Workshop on Building Lexical Resources From Semantically Annotated Corpora*. Lisbon, Portugal.

Erk, K., A. Kowalski, S. Padó, and M. Pinkal: 2003, 'Towards a Resource for Lexical Semantics: A Large German Corpus with Extensive Semantic Annotation'. In: *Proceedings of the 41st Annual Metting of the Association for Computational Linguistics*. Sapporo, Japan, pp. 537–544.

Fellbaum, C. (ed.): 1998, *WordNet – An Electronic Lexical Database*, Language, Speech, and Communication. Cambridge, MA: MIT Press.

Fillmore, C. J.: 1982, 'Frame Semantics'. *Linguistics in the Morning Calm* pp. 111–137.

Fillmore, C. J., C. R. Johnson, and M. R. Petruck: 2003, 'Background to FrameNet'. *International Journal of Lexicography* **16**(3), 235–250.

Giuglea, A.-M. and A. Moschitti: 2006, 'Semantic Role Labeling via FrameNet, VerbNet and PropBank'. In: *Proceedings of the Joint Annual Meeting of the Association for Computational Linguistics and International Conference on Computational Linguistics*. Sydney, Australia, pp. 929–936.

Gruber, J. S.: 1965, 'Studies in lexical relations'. Ph.D. thesis, Massachusetts Institute of Technology.

Hamp, B. and H. Feldweg: 1997, 'GermaNet – a Lexical-Semantic Net for German'. In: *Proceedings of the ACL Workshop on Automatic Information Extraction and Building Lexical Semantic Resources for NLP Applications*. Madrid, Spain, pp. 9–15.

Hampton, J.: 1993, 'Prototype models of concept representation'. In: I. V. Mechelen, J. Hampton, R. S. Michalski, and P. Theuns (eds.): *Categories and concepts: Theoretical views and inductive data analysis*. London: Academic Press.

Hearst, M.: 1992, 'Automatic Acquisition of Hyponyms from Large Text Corpora'. In: *Proceedings of the 14th International Conference on Computational Linguistics*. Nantes, France, pp. 539–545.

Hovy, E., M. Marcus, M. Palmer, S. Pradhan, L. Ramshaw, and R. Weischedel: 2006, 'OntoNotes: the 90% solution'. In: *Proceedings of the Joint Human Language Technology and North American Association of Computational Linguistics Conference*. New York, pp. 57–60.

Joanis, E., S. Stevenson, and D. James: 2008, 'A General Feature Space for Automatic Verb Classification'. *Natural Language Engineering* **14**(3), 337–367.

Katz, J. J. and J. A. Fodor: 1964, 'The structure of a semantic theory'. In: J. J. Katz and J. A. Fodor (eds.): *The Structure of Language*. Prentice-Hall.

Kipper, K., H. T. Dang, and M. Palmer: 2000, 'Class-Based Construction of a Verb Lexicon'. In: *Proceeedings of the Seventeenth National Conference on Artificial Intelligence*. Austin, TX, pp. 691–696.

Koehn, P. and H. Hoang: 2007, 'Factored Translation Models'. In: *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Prague, Czech Republic, pp. 868–876.

Koenig, J.-P. and A. R. Davis: 2001, 'Sublexical Modality and the Structure of Lexical Semantics'. *Linguistics and Philosophy* **24**, 71–124.

Kohomban, U. S. and W. S. Lee: 2005, 'Learning semantic classes for word sense disambiguation'. In: *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Ann Arbor, MI, pp. 34–41.

Korhonen, A.: 2002, 'Subcategorization Acquisition'. Ph.D. thesis, University of Cambridge, Computer Laboratory. Technical Report UCAM-CL-TR-530.

Korhonen, A., Y. Krymolowski, and Z. Marx: 2003, 'Clustering Polysemic Subcategorization Frame Distributions Semantically'. In: *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*. Sapporo, Japan, pp. 64–71.

Kunze, C.: 2000, 'Extension and Use of GermaNet, a Lexical-Semantic Database'. In: *Proceedings of the 2nd International Conference on Language Resources and Evaluation*. Athens, Greece, pp. 999–1002.

Levin, B.: 1993, *English Verb Classes and Alternations*. Chicago, IL: University of Chicago Press.

Lin, D.: 1998, 'Automatic Retrieval and Clustering of Similar Words'. In: *Proceedings of the 17th International Conference on Computational Linguistics*. Montreal, Canada, pp. 768–774.

Maedche, A. and S. Staab: 2000, 'Discovering Conceptual Relations from Text'. In: *Proceedings of the 14th European Conference on Artificial Intelligence*. Berlin, Germany, pp. 321–325.

McRae, K., G. S. Cree, and M. S. S. C. McNorgan: 2005, 'Semantic feature production norms for a large set of living and nonliving things'. *Behavior Research Methods* pp. 547–559.

Merlo, P. and S. Stevenson: 2001, 'Automatic Verb Classification Based on Statistical Distributions of Argument Structure'. *Computational Linguistics* **27**(3), 373–408.

Miller, G. A. (ed.): 1990, *WordNet: An On-line Lexical Database*. Oxford University Press. Special issue of the International Journal of Lexicography **3**(4).

Nastase, V., J. Sayyad-Shiarabad, M. Sokolova, and S. Szpakowic: 2006, 'Learning Noun-Modifier Semantic Relations with Corpus-based and WordNet-based Features'. In: *Proceedings of the 21st National Conference on Artificial Intelligence*. Boston, MA, pp. 781–787.

Navigli, R. and P. Velardi: 2003, 'An Analysis of Ontology-based Query Expansion Strategies'. In: *Proceedings of the ECML Workshop on Adaptive Text Extraction and Mining*. Cavtat-Dubrovnik, Croatia, pp. 42–49.

Ohara, K. H.: 2004, 'Manner of Motion in English and Japanese'. In: *Proceedings of the 3rd International Conference on Construction Grammar*. Marseille, France.

Padó, S. and M. Lapata: 2007, 'Dependency-based Construction of Semantic Space Models'. *Computational Linguistics* **33**(2), 161–199.

Padó, U., M. Crocker, and F. Keller: 2006, 'Modelling Semantic Role Plausibility in Human Sentence Processing'. In: *Proceedings of the 17th Meeting of the European Chapter of the Association for Computational Linguistics*. Trento, Italy, pp. 345–352.

Pantel, P. and M. Pennacchiotti: 2006, 'Espresso: Leveraging Generic Patterns for Automatically Harvesting Semantic Relations'. In: *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*. Sydney, Australia, pp. 113–120.

Prescher, D., S. Riezler, and M. Rooth: 2000, 'Using a Probabilistic Class-Based Lexicon for Lexical Ambiguity Resolution'. In: *Proceedings of the 18th International Conference on Computational Linguistics*. Saarbrücken, Germany, pp. 649–655.

Schulte im Walde, S.: 2003, 'Experiments on the Automatic Induction of German Semantic Verb Classes'. Ph.D. thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. Chapter 2.

Schulte im Walde, S.: 2006, 'Experiments on the Automatic Induction of German Semantic Verb Classes'. *Computational Linguistics* **32**(2), 159–194.

Schulte im Walde, S. and K. Erk: 2005, 'A comparison of German semantic verb classifications'. In: *Proceedings of the Sixth International Workshop on Computational Semantics*. Tilburg, Netherlands.

Shen, D. and M. Lapata: 2007, 'Using Semantic Roles to Improve Question Answering'. In: *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. pp. 12–21.

Shi, L. and R. Mihalcea: 2005, 'Putting Pieces Together: Combining FrameNet, VerbNet and WordNet for Robust Semantic Parsing'. In: *Proceedings of the 6th International Conference on Intelligent Text Processing and Computational Linguistics*. Mexico City, Mexico, pp. 100–111.

Sloman, S. A., B. Love, and W. Ahn: 1998, 'Feature centrality and conceptual coherence'. *Cognitive Science* pp. 189–228.

Taylor, J. R.: 1989, *Linguistic categorization*. Oxford University Press.

Vigliocco, G., D. P. Vinson, W. Lewis, and M. F. Garrett: 2004, 'Representing the meanings of object and action words: The featural and unitary semantic space hypothesis'. *Cognitive Psychology* **48**, 422–488.

Weeds, J. and D. Weir: 2005, 'A Flexible Framework for Lexical Distributional Similarity'. *Computational Linguistics* **31**(4), 439–476.