RALF NAUMANN

Dynamics in the Brain and Dynamic Frame Theory for Action Verbs

Ralf Naumann naumann@phil.hhu.de
Heinrich-Heine-University Düsseldorf
Institut for Language and Information

Abstract

In this article we outline a theory of action verbs that combines a modality-independent (or abstract) conceptual component with a modality-specific one. Verbs as concepts are interpreted as ranked sets of nuclei structures in the sense of Moens and Steedman (1988). This information is stored in the middle temporal gyrus (Bedny and Caramazza 2011). Besides being amodal, this information is underspecified w.r.t. a particular way in which the action is executed (*grasp a needle* vs. *grasp a barbell*), i. e. it is not grounded in a particular situation. This underspecification can in general only be resolved if the type of object undergoing the change (*needle* vs. *barbell*) is known. Following Willems et al. (2009), this grounding is explained as an *implicit simulation* in premotor cortex, that is a preenactment of the action which makes it possible to predict the way in which the action evolves and which is distinct from explicit (motor) imagery.

1 Theories of grounded cognition: evidence and problems[1]

According to Zwaan and Kaschak (2008: 368), 'language is a sequence of stimuli that orchestrate the retrieval of experiential traces of people, places, objects, events, and actions.' They illustrate this view of language with an example taken from Barsalou (1999). When reading the sentence *John removed an apple pie from the oven*, a comprehender understands this sentence by retrieving past experiences involving persistent objects like apple pies and ovens as well as events of removing something, for instance, an apple pie from an oven. These traces usually include both motor experiences such as lifting the pie and feeling its weight and perceptual experience like seeing and smelling the pie and feeling the heat coming out of the oven. Similarly, when processing the verb *throw* or the sentence *Bill throws the ball*, a speaker mentally simulates an action of throwing (Pulvermüller 2005). On this view, 'the understanding of action-related sentences implies an internal simulation of the action expressed in the sentences, mediated

---

[1] The research was supported by the German Science Foundation (DFG) funding the Collaborative Research Center 991.

by the activation of the same motor representations that are involved in their execution' (Buccino et al. 2005: 361). On this view, understanding words and other linguistic items is based on the same neural substrate as imagining the actions and objects described by those linguistic expressions (Gallese and Lakoff 2005: 456). For example, Gallese and Lakoff argue that one can understand the sentence *Harry picked up the glass* only if one can imagine picking up a glass or seeing someone picking up a glass. This view is in line with the idea of Hebbian learning: neuronal correlation is mapped onto connection strength. As formulated by Hauk et al. (2004: 301): 'If word forms frequently co-occur with visual perceptions (object words), their meaning-related activity may be found in temporal visual areas, whereas action words frequently encountered in the context of body movements may produce meaning-related activation in the frontocentral motor areas'. If a verb refers to actions and events that are typically performed with the face, arm or leg, neurons processing the word and those processing the action described by that word frequently fire together and thus become more strongly linked. As a result, word-related networks overlap with motor and premotor cortex in a somatotopic fashion (Pulvermüller 1999). On this *semantic somatotopy* view of meaning, being able to simulate executing an action of the type denoted by the verb is constitutive of the verb's meaning.

Empirical evidence for theories of grounded (or embodied) cognition comes from neuroimaging studies using FMRI or ERP. When action words are processed, there is effector-specific activation of motor areas that is somatotopically organized. For example, a leg-related word like *kick* activates dorsal areas, where leg actions are represented and processed, whereas arm-related words such as *pick* or face-related words such as *lick* activate lateral or inferior frontal motor areas, respectively. Similarly, when reading or viewing the noun *hammer*, the hand and not the foot area of the motor system is activated.

Such theories of embodied cognition make a number of empirically testable predictions: (i) understanding an action verb and imagining performing that same action rely on the same neural tissue, in particular premotor cortex (Willems et al. 2009: 2388), (ii) understanding action verbs is primarily based on early, modality-specific, sensory-motor brain regions (Bedny and Caramazza 2011: 82) and (iii) these sensory-motor brain regions are automatically engaged during word comprehension (Bedny and Caramazza 2011: 82).

The first problem for theories of grounded cognition is that many neuroimaging studies failed to observe any increased activity for action-verbs anywhere in the motor system (Bedny and Caramazza 2011: 87). A notable exception is the study by Willems et al. (2009). In an fMRI study they examined whether implicit stimulations of actions during language understanding involve the same cortical motor regions as explicit motor imagery. The participants were presented with verbs that are either related to actions that are usually executed with the hand, like *throw*, or with verbs that are not related to this body part, like *kneel*. In order to control for spurious activation due to explicit imagery, there were two different tasks: participants either read the verbs (lexical decision task LD) or they actively imagined performing the actions denoted by these verbs (imagery task IM). Contrary to earlier results, they found a double dissociation. Primary motor cortex showed effector-specific activation during imagery, but not during the lexical decision task. For the premotor area they found out that there was effector-specific activation that distinguished between manual and non-manual verbs, both in LD and in IM. But importantly, there was no overlap or correlation between regions activated during the two tasks. More precisely, portions of BA6 and BA4 that were defined on the basis of effector-specific activity during the IM task showed no such activity during LD. Similarly, regions in BA4 and BA6 that showed effector-specific activity during LD showed no such activity during IM. The authors conclude: “These double dissociations show that implicit motor simulation and explicit motor imagery do not necessarily engage the same neural tissues in premotor and primary motor cortices and by inference may not include the same cognitive processes” (Willems et al. 2009: 2396).

Similar to the Willems et al. study, Postle et al. (2008) found effector-specific activity in premotor cortex only when participants viewed actions performed with hand, arm or foot. By contrast, when they silently read the corresponding verbs, there was only activation in premotor cortices. Importantly, premotor leg, arm and hand areas responded to *all* action-verbs in the same way, i. e. there was no somatotopical reaction. In addition, several of these premotor areas also responded to nouns and even non-words. These results constitute strong evidence against prediction (i) i. e. that understanding action verbs and imagining performing those actions rely on the same, or at least overlapping, neural tissues. Summarizing, one gets the following correlations:

- Primary motor cortex is active during motor imagery; during processing of action verbs this cortex is not active, provided no corresponding instructions are given.
- Premotor cortex areas are active during comprehension of action verbs; however, there is no overlap with areas in this cortex that are active during explicit imagery. In addition, there need be no effector-specific activity.

According to Bedny and Caramazza (2011: 87), results like the above raise the important question of whether such activity in left premotor areas is specific to action verb comprehension or whether this activity rather reflects a more general contribution of premotor cortex to language. Evidence for such a more general contribution comes from several studies. Graziano (2006) showed that activity in premotor areas is more sensitive to the behavioral context and possible goals and results brought about by an action.[2] Schluter et al. (1998) found that premotor cortex is involved in higher-order aspects of movement like sequencing and movement selection. Similarly, this cortex is involved in planning and predicting actions and sequentially structured events (Schubotz and von Cramon 2004). When taken together, one gets that the premotor cortex shares features with adjacent prefrontal cortex (Miller and Cohen 2001).

Evidence against prediction (ii) comes from studies involving the middle temporal gyrus (MTG).[3] There is more activity in MTG when participants generate action verbs than when they generate color names for visually presented nouns. MTG is more active when action verbs are processed compared to the processing of nouns for concrete objects and color adjectives. Furthermore, MTG response is equally high with action verbs like *run* and mental state verbs like *think* and it is equally low for nouns denoting animals like tigers which are rich in motion features and nouns like *rock* which are low in motion features. In addition, MTG responds more to verbs like *give* compared to verbs like *run*. This area responds to action verbs in the absence of a sentence context. Representations are neither visual nor motion related and regions in MTG that are activated during processing of action verbs do not overlap with visual-motion regions. Bedny and Caramazza (2011: 91) conclude that "these results argue that the MTG stores modality-independent representations that encode conceptual rather than perceptual properties. … Together, these results suggest that the MTG represents conceptual information about events or meaning-relevant grammatical information about verbs."

A key question with respect to prediction (iii) is: Do effector-specific activations show that they are used by speakers to *semantically* analyze the word or the words in a sentence? As first noted in Postle et al. (2008), this need not be the case. The motor activation can be an epiphenomenon of processing the word or the constituents in the sentence. The speaker semantically analyzes the expressions and simultaneously or subsequently (s)he mentally imagines executing a corresponding action or event. As

---

[2] This example as well as the following ones are taken from Bedny and Caramazza (2011).

[3] For details on the following, see the discussion in Bedny and Caramazza (2011) as well as the references cited therein.

noted by Bedny and Caramazza (2011: 83), language-perception interactions need not result because action-verb meanings are represented but rather because verb meaning representations prime visual motion representations during contemporaneous linguistic and perceptual tasks.[4]

2 Action, events and the dynamic structure of action verbs

When viewed from a linguistic, in particular semantic, viewpoint, a general weakness of most studies involving action verbs consists in the restriction to test isolated verb forms, in general infinitive forms like *kick* or *throw*.[5] However, what type of action or event is denoted by an expression, say a sentence, in which an action verb occurs, not only depends on the verb but also on its arguments and their semantic (or referential) properties. Consider, for instance, the German examples in (1).

(1) a. Hans lief (stundenlang im Park herum).
    b. Hans lief zum Bahnhof.
    c. Hans lief durch den Park.
    d. Hans lief zu Hochform auf.

Example (1a) is an activity expression admitting of modification with a *for*- but not with an *in*-adverbial. It describes an action as unbounded in the sense that no particular goal (say a destination to be reached) is specified.[6] By contrast, example (1b) describes a running that has an explicit goal: the station. The action is therefore bounded by this destination. Linguistically, this is reflected by the admissibility of modification with *in*-adverbials but not of that with *for*-adverbials. Example (1c) can be taken to either describe an unbounded or a bounded event. In the first case it corresponds to (1a) (*Hans ran across the park*), whereas in the second case it corresponds to the English translation *Hans crossed the park*. The last example differs from the preceding ones. Here *laufen*

---

[4] As noted by Willems et al. (2009: 2398), another reason why there is effector-specific activity in motor areas can be due to the fact that participants in those studies were not prevented from forming mental images. Furthermore, Postle et al. (2008) note that the positive results can be artifacts of differences in imageability between critical and control stimuli. For example, in the Hauk et al. (2004) study, action verbs were compared to hash-marks as lower-level control. As a result, effector-specific activity could have been triggered by increased imagery to concrete action language as compared with more abstract language (see also Willems et al. 2009: 2398).

[5] This limitation becomes even more apparent in languages like Dutch or German where the infinitive form is in general distinct from tensed forms, whereas in English the infinitive coincides with the present tense form.

[6] This does not mean that Hans didn't have a particular destination in mind; for example, the university which he was running to.

is used in an idiomatic and not in its literal sense. (1d) does not necessarily describe an event which involves a particular motor program involving the legs. For example, it can be used in a situation where Hans did a great job in convincing the audience during a talk he gave at the university.[7]

In order to explain these differences one has to take into consideration that events occur in time, in contrast to 'normal' objects like tables and trees which persist in time.[8] Furthermore, action and events have a particular temporal-causal or dynamic structure. This structure can be described in terms of a nucleus structure in the sense of Moens and Steedman (1988), which consists of a linearly ordered sequence of constituents or parts: a development process (DP), a culmination (Cul) and a consequent state (CS) (in Figure 1 $\alpha(e)$ and $\beta(e)$ are the beginning and end point of the event *e*, respectively).

The important point is that the examples in (1) describe different nuclei structures. The nucleus structure for (1a) consists of a DP only because no destination, and therefore no CS (*be at the destination)* is specified. For (1b) the nucleus structure is the one depicted in Figure 1. Here a destination is determined together with the CS *Hans is at the station.* (1c) has two corresponding nuclei structures, i. e. those of (1a) and (1b). These examples already make clear that a nucleus structure is underspecified in at least two respects if only the verb, say *laufen*, is taken into consideration. First, the sort, or type, of a possible goal is not (yet) determined. Second, the exact way in which the running is executed is not (yet) determined. The two kinds of underspecification are not unrelated. Consider the examples in (2).

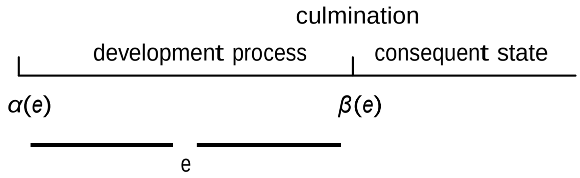

Figure 1: Nucleus structure for bounded processes bringing about a result

(2) a. Bill grasped the needle.
    b. Bill grasped the barbell.

The way the grasping is executed depends on the object that is grasped. As noted by Willems et al. (2009: 2307), very different action plans are necessary to successfully ex-

---

[7] Though this example can also be used to describe a perfect 100 m performance by Hans in athletics.

[8] Thus, for each time slice of a 'normal' object one always gets the complete object. By contrast, for actions and events one usually only gets a proper part.

ecute the two actions described by the sentences in (2). Similarly, throwing a Frisbee or a baseball requires different grips and different arm motions. These examples show that the sortal information provided by the direct object is, at least in general, important to resolve the underspecification with respect to the exact motor program to be executed. To make the fact of different nuclei structures determined by the same verb clearer, let us consider another set of examples involving the verb *kick*.

(3) a. John kicked Bill.
 b. John kicked Bill several times.
 c. John kicked the ball into the goal.
 d. John kicked the bucket.

Example (3a) can be used to describe a single (atomic) kicking, the corresponding nucleus structure of which consists of a Cul (without a CS, see Moens and Steedman 1988 and Naumann 2001 for details): NS_{Cul}. A sequence of such atomic kickings is described by (3b): NS_{Cul*}. The nucleus structure is complex because it consists of a sequence of nuclei structures having a Cul only. Sentence (3c) describes an event in which the kicking of the ball causes the latter's location to change: before the kicking it was not in the goal whereas it is in the goal as an effect of the kicking. In this case, two nuclei structures are related by a causal relation. The first nucleus structure consists of a Cul describing the kicking proper and the second is a nucleus structure consisting of a DP, a Cul and a CS describing the movement of the ball into the goal: NS_1 CAUSE NS_2. For (3d), the situation is different. In this sentence, *kick* is not used in its literal sense but it is used idiomatically. Since *kick the bucket* means *die*, the nucleus structure consists of a Cul together with a CS (*be dead*).

Reconsidering the examples in (3), one gets: after processing *John kicked*, which is common to all four sentences, a comprehender cannot (yet) know which of the four nuclei structures is described by the sentence. However, using linguistic knowledge/experience (e. g. frequency information) as well as world knowledge (what type of nucleus structure occurs most often in the context of a kicking), (s)he has a particular expectation about which nucleus structure is most likely be described. For example, the literal (non-idiomatic) uses are in general more expected than the idiomatic sense in (3d).[9] For the literal uses, a possible ordering can be $\mathrm{NS}_{Cul} < \mathrm{NS}_1$ CAUSE $\mathrm{NS}_2 < \mathrm{NS}_{Cul*}$, i. e. single kickings are most expected, followed by kickings that are used to obtain a

---

[9] However, in a context in which it is clear that John is going to die, (3d) can be the most expected continuation.

particular effect and sequences of atomic kickings are least expected.[10] In a particular context, this default ordering has to be changed. For example, upon listening to … *the ball into the goal* after processing *John kicked …* a comprehender comes to know that the kicking had a destination and that therefore this sentence isn't used to describe an action of the most expected nucleus structure (Cul) but of type NS_1 CAUSE NS_2. As a consequence, a less expected nucleus structure has to be chosen.

3 Interpreting action verbs in the brain

In our account, understanding the meaning of an action verb is in part determined by knowledge of (i) the set of possible nuclei structures which describe possible temporal-causal evolutions of actions and events denoted by the verb and (ii) the default ranking among the elements of this set. This information about the meaning of a verb is stored in MTG.

This knowledge is only necessary for grasping the (complete) meaning of such a verb because verbs with identical sets of nuclei structures and default ranking would have the same meaning.[11] However, they differ with respect to implicit simulations in premotor cortex in the sense of Willems et al. (2009). Implicit simulations are pre-enactments of potential future experiences, the principal function of which is the ability to make predictions about how exactly an event will evolve and what its possible consequences are. For example, a word like *grasp* can serve as a cue to activate neural circuits involved in partial preparation of an action of grasping something. As noted by the authors: "This schematic, unconscious, prospective activation of effector-specific regions in premotor cortex presumably facilitates further action planning if subsequent cues call for *grasping* to be executed or to be imagined explicitly" (Willems et al. 2009: 2388).[12]

Linguistically, the ranked set of nuclei structures corresponds to the level of verbs in the lexicon. Conceptually, it can be taken as a symbolic, amodal representation of the

---

[10] Again, it must be stressed that this ordering is to be determined empirically and that it is in general – at least in part – context dependent. For example, in case of a penalty kick during a football match, NS_1 CAUSE NS_2 is likely to be most expected.

[11] But see below for a refinement of this thesis.

[12] By contrast, explicit imagery is covert enactment of an action. Like overt motor execution, motor imagery may entail the generation of an action plan (inverse model) as well as a prediction of the action's sensory consequences (Willems et al. 2009: 2388). Its principle function is either reflective (i. e. covert reenactment of prior actions) or prospective (e. g. an athlete usually imagines the concrete motor program before starting his performance).

concept expressed by the verb that is independent of sensory and motor simulations. By contrast, implicit simulations correspond to projections of the verb like VP or sentences. To be precise, implicit simulations are triggered when a comprehender has enough information to determine a specific way or manner in which the action is executed. As shown in the previous section, this is in general the case if (s)he knows which object undergoes the change brought about by the action. Thus, implicit simulation corresponds to the choice of an appropriate activity, modulo the direct object of the verb.

When taken together, the meaning of a verb consists of two dimensions: a symbolic, amodal dimension and different ways in which these representations can be grounded to specific activities that are undertaken in a particular situation.

Dimension	Level of Abstraction	Reference	Neural Correlate	Function	Linguistic Level
conceptual	symbolic and amodal	ranked set of nuclei structures	MTG	determination of possible evolutions in terms of a temporal-causal structure	(isolated) verb in the lexicon
implicit simulation	grounded	instantiated nuclei structures	regions in premotor cortex	prediction and planning (preenactment of actions)	projections of the verb (VP and S)

At the conceptual dimension actions and events are taken as types (or schemes), whereas at the second dimension these types are instantiated in a particular situation in space and time, yielding an action or event token. This differentiation has the advantage of computational economy since it leads to a reduction on the requirement on storage. Different nuclei structures can be instantiated (or grounded) to various situations belonging to different action types. One has a small number of abstract, symbolic and amodal temporal-causal structures (nuclei structures) that can be instantiated in an indefinite number of concrete situations in space and time. In particular, a nucleus structure of a particular type, say the one depicted in Figure 1 consisting of a DP Cul CS, can be used for (i) different action verbs and (ii) different instantiations of the same type of action. An example for (i) are verbs like *eat* and *run*. Both *eat an apple* and *run to the station* are of type DP Cul CS. They differ with respect to (i) the place in the default ordering and (ii) the types of possible activities that can instantiate this structure. Whereas this nucleus structure is the most expected one for *eat*, this does not hold for *run*, which basically describes unbounded actions with no particular goal or destination. For *eat*, appropriate activities include putting food into the mouth using the hands, a fork or a spoon or, in the case of an animal, the lips and the tongue. By contrast, for running events appropriate activities are fast movements typically involving the legs.

If a verb is encountered, the set of possible nuclei structures in middle temporal gyrus is activated. In the absence of further information, a comprehender assumes that an event corresponding to the most expected nucleus structure (or the most expected nuclei structures) is (are) described. Accessing verb meanings therefore involves accessing the corresponding nuclei structures. The more complex a nucleus structure, the longer the time to access and/or activate this structure. Thus, there is a cost in processing time that depends on the complexity of the nucleus structure. For example, the most expected nucleus structure for an activity verb like *run* is of type DP. By contrast, for a verb like *give*, which expresses a causal relation involving two different nuclei structures, the most expected nucleus structure is more complex.[13]

The activation of the ranked set of nuclei structures *does involve no immediate activation* of premotor or primary motor areas since no particular implicit or explicit simulation can yet be determined because the choice depends on the argument denoting the object undergoing the change as well on the actor executing the action.[14] Rather, premotor areas related to implicit stimulations are activated only after the nuclei structures are instantiated. As noted above, this is the case for projections of the verb, in particular the VP and the sentence level.

3.1 Empirical evidence for our approach

From what has been said so far, the following predictions can be derived from our approach:

- There is only weak activation of primary and premotor areas upon processing of the verb. Activation of the motor system is possible only if the underspecification inherent in a nucleus structure has been removed. This is in general possible only if the type of the object undergoing the change is known.

---

[13] NS_1 : DP (action undertaken by the actor); NS_2 : Cul CS (the recipient gets the theme).

[14] Though this does not exclude the possibility that a comprehender activates a particular simulation intentionally or by convention. For example, a football player or a football fan might usually immediately engage in triggering simulations of a player kicking a football upon hearing or reading the verb *kick*. But such simulations are independent of understanding the meaning of the verb or the sentence in which it occurs. Rather, the meaning of the verb primes particular sorts of motor programs that can be used in executing the action or event type.

- Sentences with an idiomatic sense elicit stronger activation in MTG because a less expected nucleus structure must be chosen. This reordering triggers a higher processing load reflected by a stronger activation in MTG.
- Complex nuclei structures trigger stronger activation because e. g. different types of nuclei structures must be related to each other (e. g. in a causal relation). The general rule is: the more complex a nucleus structure, the stronger the activation.
- Implicit simulation depends on the expertise of the comprehender. For example, both experts (players and fans) and laymen understand sentences about hockey matches. However, players and fans are better able to implicitly simulate actions undertaken during a game. Thus, one expects the same activation in MTG but differences with respect to premotor activity.

Evidence for the truth of the first two predictions comes from an fMRI study by Boulenger et al. (2008). They examined how literal versus idiomatic sentences with action verbs referring either to the leg (*kick)* or the arm (*grasp*) are processed in the brain.

(4) a. He kicked the ball.
b. He kicked the bucket.
(5) a. He grasped the needle.
b. He grasped the idea.

Brain activity was measured at the onset of the critical word in the sentence (*He grasped the IDEA)* which disambiguated between a literal and an idiomatic reading (early analysis window) and three seconds after its end (late analysis window). They found that (i) a common network of cortical activity was triggered for both conditions in both analysis windows, with the idioms eliciting overall more distributed activity; (ii) primary and premotor cortices were activated both for idioms and non-idioms; (iii) activation of (frontocentral) primary and premotor areas was relatively weak both at action verb onset (and therefore upon processing the action verb) and at the onset of the critical word. However, it was strong after the offset of the critical word both for literal and idiomatic readings; (iv) sentences with literal meanings failed to elicit stronger activation than sentences with an idiomatic reading in any brain area; (v) in the late analysis, window cortical activity was greater in MTG and the cerebellum.[15]

In the present context findings (iii) and (v) are the most important ones. Finding (iii) shows that there is no instant spreading of activation to primary or premotor cortex

---

[15] Furthermore, there was stronger activation of idioms in inferior frontal gyrus in both windows.

during action verb processing. Rather, this activation is delayed until after the direct object has been processed. This is in contrast to the results for processing isolated action verbs. Finding (v) can be taken as providing evidence for our claim that in case a verb is used in an idiomatic sense the default ordering on the set of nuclei structures must be changed (i. e. there is a reordering of the elements of this set), resulting in a higher processing load, reflected in the higher activity in MTG.[16]

Evidence for the third prediction comes from two studies by Shetreet et al. (2007) and Van Dam and colleagues (2010), respectively. Shetreet and colleagues found that MTG responds more strongly to sentences with verbs that have more arguments, even when the sentences have the same overall length. For example, processing *John gave Mary the book* (three arguments) triggers stronger activity in MTG than the sentence *John ran to the station* (two arguments). In our approach, a verb like *give* is related to a complex nucleus structure consisting of two substructures that are linked by a causal relation. The first nucleus structure describes the action undertaken by the giver (actor) whereas the second nucleus structure describes the event of the recipient receiving (and thereby coming to possess) the theme, i. e. the object given. Van Dam and colleagues (2010) found that the processing of action verbs like *wipe* that denote events describing a particular way of moving part of the body triggers stronger inferior parietal activity than verbs like *clean* for which no such manner is determined. This finding can be explained as follows. Levin and Rappaport-Hovav (to appear) distinguish between verbs of manner and verbs of result. Manner verbs specify a particular way in which an action is executed. For example, *wipe* and *brush* determine a particular way of cleaning an object without imposing the constraint that the result be attained at the end of the event. By contrast, result verbs specify a particular end state of the action. For example, *clean* requires the object undergoing the change, say a table, to be clean as a result of the cleaning activity undertaken by the actor. However, no specific type of activity (or manner) by which this end state is achieved is determined by the verb. In our approach, manner verbs like *wipe* have a most expected nucleus structure of type DP, i. e. they are basically activity verbs that are usually used to describe unbounded events which need not bring about a particular result (similar to a verb like *run).* By contrast, a result verb like *clean* has a most expected nucleus structure of type DP Cul CS. However, for *clean* only the culmination is explicitly determined (the object has to be clean) but no particular activity.

---

[16] For details on how such orderings can be changed, see Naumann (2011, 2013, 2014).

In our approach, this means that the interpretation of a result verb is already determined by the ranked set of nuclei structures since the only constraint is the one imposed on the end state (*be clean*), which is already specified at the lexical level and which therefore is independent of the object undergoing the change. As a consequence, being able to implicitly simulate how the action can be executed is not part of the meaning of the verb. From this it does not follow that a comprehender does not engage in an implicit simulation (and, additionally, in explicit imagery). But in this case, (s)he plans or imagines an execution that can be described by another verb, say *wipe* as in *wipe the table clean*.

Further evidence for our analysis comes from a study by McKoon and Macfarland (2000). They showed that there are no differences in processing time between transitive and intransitive uses of so-called externally caused event verbs like *break* and *awake*.

(6) a. The fire alarm awoke the residents.
  b. The residents awoke.

By contrast, for internally caused event verbs like *bloom* and *wilt*, processing times are significantly shorter than those for externally caused event verbs.

(7) a. The bright sun wilted the roses.
  b. The roses wilted.

Again, there are no differences between the transitive and the intransitive form. These results therefore show that the processing time depends on the type of the (preferred or most expected) nucleus structure. Furthermore, these examples show that the cost in processing time is *independent* of the exact syntactic realization (transitive vs. intransitive). Rather, it only depends on the corresponding types of nuclei structures.

Similar results were obtained by Gennari and Poeppel (2003). They showed that processing non-stative verbs like *vanish* and *solve* takes longer than processing stative verbs like *love* and *exist* (about 25 ms), even if the argument structures are identical (e. g. *exist* and *vanish)*.

Evidence for the fourth prediction comes from a study by Beilock and colleagues (2008). They let hockey players, hockey fans and hockey novices listen to sentences about hockey-related actions. They found that both for hockey players and hockey fans there was an increased activity in dorsal premotor cortex compared to the activity in this area for hockey novices. Furthermore, this stronger activity was influenced by experience with hockey games but not necessarily by motor experience directly related to playing the sport. For example, dorsal activity was the same for hockey players and hockey fans. In addition, only for hockey novices the primary sensory-motor cortices were active and increased primary sensory-motor activity correlated negatively with action sentence comprehension.

The above empirical results can be taken as evidence for the following two hypotheses: (i) result verbs are not directly related to particular implicit simulations or motor programs and (ii) for result verbs, grounding of a corresponding nucleus structure is, at least in part, independent of their types. By contrast, manner verbs require (i) activation of the related ranked set of nuclei structures in MTG and (ii) an implicit simulation in premotor cortex (in order to distinguish say *brush* from *wipe*). These hypotheses raise the following questions: (i) what is the exact relation between the ranked set of nuclei structures and implicit simulations? And (ii) where is this relation stored in the brain (i. e. what is the neuronal correlate of this relation)? One answer to the first question is that the ranked set of nuclei structures for a verb in MTG primes certain implicit simulations in premotor cortex. To be more precise: both manner and result verbs are related to a set of appropriate activities. Information about these activities is stored in regions of premotor cortex. For manner verbs this set is more restricted than that for result verbs. Furthermore, and more importantly, the set of activities for manner verbs is ranked in the sense that not all elements in this set are equally expected. By contrast, for result verbs there is no ranking on this set. For example, for *wipe*, one has *rub with a cloth or one's hand* and for *brush*, *rub with a brush*. The set of appropriate activities for *clean* comprises those for *wipe* and *brush* (and those for other manner verbs which denote actions for cleaning something). A possible answer to the second question goes along Hebbian lines. Neuronal correlation is mapped onto connection strength. If an action verb frequently co-occurs with body movements that are executions of an action of the type denoted by the verb, this strengthens the connection between regions in MTG and regions in premotor cortex. There remain, of course, a number of open empirical questions, for example: *Where in the brain is the 'meaning assembly' between a verb and its arguments located, i. e. what is the exact relation between verbal (dynamic) and non-verbal (static) meanings?* and *How is the ranked set of nuclei structures acquired in the brain during language learning?*

Furthermore, the above results also show that the various dimensions are not independent of each other. When taken together, the findings of the empirical studies used in this article suggest the following relation. Both implicit and explicit simulations are functionally or causally dependent on the conceptual domain consisting of the ranked set of nuclei structures in MTG. Empirical evidence supporting this claim is: (i) MTG responds to verbs in isolation for sentences with transitive verbs and (ii) the motor system is activated only after the direct object has been processed. Thus, when processing a verb, regions in MTG are activated but no effector-specific activity in the motor system is (yet) triggered. Consequently, MTG is activated *prior* to the motor system. By itself, this temporal relation does not show that there is a functional or causal relation between those dimensions. However, both types of activity are directly related to processing the verb and therefore to understanding its meaning, which makes it likely that some functional relation is involved. Of course, this claim needs to be confirmed by further empirical investigations.

Finally, an important empirical question is this: is the ability to trigger implicit simulations in premotor cortex constitutive of grasping the meaning of (or to have the concept corresponding to) an action verb? In our approach the answer is negative for the following reason. The two dimensions in the meaning of an action verb correspond to different functions language and cognition have. The conceptual dimension is related to naming and recognizing objects of the given type. Evidence for this comes from studies of patients suffering from apraxia as well as from the discussion of the results about hockey obtained by Beilock and colleagues. This dimension is non-goal oriented in the sense that no implicit preenactment of a possible execution is involved.[17] The second dimension, i. e. implicit simulation, is related to reflecting, predicting and planning an action of the given type by selecting appropriate activities and inferring future consequences of executing this action. Possible questions are: *How can the goal be reached?*, *What is an appropriate activity to reach the goal or to execute the action?* and *What are possible consequences of executing the action?* This dimension therefore is goal-oriented at a theoretical level (i. e. it does *not* involve the ability to execute a motor program). This ability is a necessary condition for being able to attain a goal or result by executing an action of the given type. For example, in the case of eating one can use the hands or, alternatively, a fork and a knife. By contrast, explicit imagery corresponds to the ability of actually executing a motor program to attain the goal.[18] The inability to have implicit simulations impairs a speaker for this particular function. This is the case for patients

---

[17] Though it may involve naming the goal of a possible execution, e. g. making an object clean for the verb *clean* since involving a goal (Cul) is part of the most expected nucleus structure of this verb.

[18] Additional evidence for this analysis comes from studies of apraxia, i. e. the inability to perform particular activities as a result of brain damage. People suffering from this inability are impaired for using objects of a particular kind, say a hammer, though they are unimpaired for (i) naming those objects and (ii) recognizing pantomimes associated with uses of those objects. Thus, integrity of motor processes is not necessary in order for object naming and action recognition to be in the normal range; see Mahon and Caramazza (2008) for details.

suffering from apraxia. However, the Beilock et al. study shows that that the ability to have implicit simulations comes in degrees. Hockey novices do have activity in the motor system, though it is less strong than the activity triggered in hockey players and hockey fans.[19]

3.2 Comparison to theories of grounded cognition

Our approach differs from theories of grounded cognition in the following respects. First, 'automatic activation' does not mean that the motor system is immediately activated when a verb is processed in the brain, i. e. that linguistically processed input immediately results in activation of the motor and sensory systems. Rather, what is immediately activated is the ranked set of nuclei structures. Groundedness is not an attribute of the verb proper but rather a property of its projections like VP or S. The reason for this is that the conceptual level stored in MTG is impoverished in the sense that verbs which have the same ranked set of nuclei structures cannot be distinguished. This distinction is only made if a nucleues structure is instantiated. The neural correlate of this instantiation is an implicit simulation in premotor cortex.

It may seem that this view is contradicted by the results of Hauk et al. (2004) and others showing that the motor system is activated rather quickly. Recall that Hauk et al. found that when presented with the word *kick* the 'leg' region of the motor system is activated within a time span of about 200 ms. Yet those results do not provide counterevidence to our claims. First, those results were obtained for *isolated* verbs and not for sentences in which these verbs occur as a constituent and this fact was known to the participants. When taken in isolation, a verb like *kick* is interpreted by uniquely describing a nucleus structure consisting only of a Cul because a comprehender already knows that no further information, say about a goal of the kicking, is added, which may make it necessary to change the nucleus structure to one of type Cul CAUSE NS_2.

A second difference is that in our approach, following Willems et al. (2009), a distinction is made between implicit simulations and explicit imagery. Third, explicit imagery is an epiphenomenon of processing (and thereby understanding the meaning) of the verb. As pointed out in the previous section, a verb (or its corresponding ranked set of nuclei structures) primes certain ways in which an action denoted by the verb is executed. As a result, an implicit simulation can be triggered. This way of undertaking the

---

[19] However, it remains an open empirical question of whether this activity is related to *both* implicit simulation and explicit imagery or to only one of those activities in the motor system.

action may subsequently result in explicit imagery of the corresponding action. Fourth, and most importantly, a distinction between a symbolic and amodal dimension and a grounded dimension is made in the definition of the meaning of an action verb.

Summarizing, one can say that theories of grounded cognition only capture one particular dimension of a verb's meaning, i. e. that related to the motor system. However, they usually do not distinguish between implicit simulations and explicit imagery. In addition, if it is true that both of these activities in the motor system are functionally and causally dependent on a conceptual dimension, they fail to give a satisfactory account of how meanings are represented and accessed in the human brain. This failure is in large part due to the fact that most often only isolated verbs and not larger linguistic contexts, like sentences, in which those verbs occur are considered.

Another way of comparing theories of grounded cognition and ours is the following. Mahon and Caramazza (2008) distinguish four possibilities of how the motor system can be related to a conceptual dimension.

1. Processing the verb directly activates the motor system, with no intervening access to abstract conceptual content.
2. Processing the verb directly activates the motor system and in parallel activates abstract conceptual content.
3. Processing the verb directly activates the motor system and then subsequently activates an abstract conceptual representation.
4. Processing the verb directly activates an abstract conceptual representation and then activates the motor system.

Only on the fourth possibility is the conceptual dimension activated *before* the motor system, whereas in the other three possibilities the motor system is either independent of the conceptual dimension (1), works in parallel with it (2) or there is a cascading flow of information from the motor system to the conceptual dimension (3). The first three possibilities underlie the various forms of theories of grounded cognition: The motor system is never activated after the conceptual system (provided the latter is assumed at all). Our approach is characterized by the fourth possibility. First, the ranked set of nuclei structures in MTG is activated and subsequently implicit simulations in specific premotor areas are triggered by a spreading activation.

4 Comparison to other approaches

Similar to our approach, the grounding by interaction account proposed in Mahon and Caramazza (2008) distinguishes between an abstract or symbolic level of representation and its instantiation (or grounding) in a particular situation. The symbolic level is conceptual and characterized by various output modalities like being able to name an object or action falling under the given concept or knowing something about the way it is built up or construed. For example, in the case of a hammer this conceptual knowledge possibly involves being able to recount the history of the hammer as an invention, the materials of which the first hammer was made, or what hammers typically weigh (Mahon and Caramazza 2008: 67 f.). This conceptual information can apply to diverse sensory modalities like touch, vision or audition. What is missing from this level is the interaction with the world. Conceptual information is not isolated. Rather, it can be activated by events in the world that are processed by the sensory system. As an effect, the conceptual information gets instantiated in a particular situation. The specific sensory and motor information that is activated may change depending on the situation in which the abstract conceptual information is instantiated (Mahon and Caramazza 2008: 68). However, from this it does not follow that the sensory and motor information is constitutive of the concept. Rather, removing the sensory and motor system would result in impoverished and isolated concepts. Thus, the activation of sensory and motor processes contributes to the 'full' representation of the concept.

The approach presented here bears some similarity with constraint-satisfaction-based approaches, like that of Jurafsky (1996) for example. According to such accounts, the processing of a sentence first involves the activation of several possible interpretations. These interpretations are ranked according to a probability measure that is based, among other factors, on the likelihood of a particular word being used in a particular context or the likelihood of a verb to be used with a particular meaning. For example, the noun *nail* refers either to a body part (fingernail, toenail) or a metal fastener. Processing this word therefore involves activation of brain areas related to both meanings of the word.[20] This set of possible interpretations is narrowed down when further information in the sentence is processed: *The nail he used to put up the picture.*

---

[20] According to Zwaan and Kaschak (2008), from which this example is taken, the processing involves the activation of traces or mental simulations that are relevant to both senses of the word, in accordance with the embodiment thesis.

5 Conclusion

In this article we presented a theory of action verbs that combines an abstract, modality-independent component with a modality-specific component located in regions of premotor cortex. Semantically, this analysis is based on the observation that a verb like *kick* can be used to express different types of actions (*kick*/*kick the ball*/*kick the ball into the goal)* that differ with respect to parameters like telic/atelic, result/no_result or atomic/iteration. The conceptual information about events are the different types of nuclei structures and the meaning-relevant information about a verb is the ranked set of such structures that represents the conceptual dimension of its meaning. This information is amodal and concerns the temporal-causal structure of an action or event. It is stored in MTG, which has been shown to respond to the processing of verbs as opposed to nouns and adjectives.

This temporal-causal structure is underspecified with respect to the exact way or manner (motor program) an action of a particular type is executed because this way depends on the object undergoing the change. After combining with the direct object of the verb, this structure is grounded or instantiated by a spreading activation to premotor cortex leading to an implicit simulation which makes it possible to derive additional conclusions about this structure.

6 References

Barsalou, L. W. (1999). Perceptual symbol systems, Behavioral and Brain Sciences, 22, 577–660.

Beilock, S. L. et al. (2008). Sports experience changes the neural processing of action language, *Proceedings National Academy Sciences USA*, 105, 13269–13273.

Boulenger, V., O. Hauk and F. Pulvermüller (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension, *Cerebral Cortex*, 19, 1905–1914.

Buccino, G. L. Riggio, G. Melli, F. Binkofski, V. Gallese and G. Rizzolatti (2005). Listening to action related sentences modulates the activity of the motor system; a combined TMS and behavioral study, *Cognitive Brain Research*, 24, 355–363.

Gallese, V. and G. Lakoff (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge, *Cognitive Neuropsychology*, 22, 455–479.

Gennari, S. and D. Poeppel 2003. Processing correlates of lexical semantic complexity, *Cognition*, 89, B27–B41.

Graziano, M. 2006. The organization of behavioural representation in motor cortex, *Annual Review of Neuroscience*, 29, 105–134.

Hauk, O., L. Johnsrude and F. Pulvermüller (2004). Somatotopic representation of action words in human motor and premotor cortex, *Neuroscience*, 41, 301–307.

Jurafsky, D. 1996. A probabilistic model of lexical and syntactic access and disambiguation, *Cognitive Science*, 20, 137–194.

Levin, B. and M. Rappaport Hovav (to appear). Lexicalized meaning and manner/result complementarity, In: B. Arsenijević, B. Gehrke, and R. Marín, (Eds.), *Subatomic Semantics of Event Predicates*, Springer, Dordrecht.

Mahon, B. Z. and A. Caramazza (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content, *Physiology – Paris*, 102, 59–70.

McKoon, G. and T. Macfarland 2000. Externally and internally caused change of state verbs, *Language*, 76, 833–858.

Miller, E. K. and J. D. Cohen 2001. An integrative theory of prefrontal cortex function, *Annual Review of Neuriscience*, 24, 167–202.

Moens, M. and M. Steedman (1988). Temporal ontology and temporal reference, *Computational Linguistics*, 14: 2, 15–28.

Naumann, R. (2001). Aspects of changes: a dynamic event semantics. *Semantics* 18: 27–81.

Naumann R. (2011). Relating ERP-effects to theories of belief update and combining systems. In: M. Aloni et al. (Eds.), *Proc. 18th Amsterdam Colloquium*, LNCS, Springer, Berlin.

Naumann, R. (2013). Outline of a dynamic theory of frames. In: G. Bezhanishvili, S. Löbner, V. Marra, and F. Richter (Eds.), *Proceedings of the 9th International Tbilisi Symposium on Language, Logic and Computation*, Volume 7758 of Lecture Notes in Computer Science, pp. 115–137. Springer Berlin Heidelberg.

Naumann, R. (2014). *A dynamic update model of sentence processing*, ms, University of Düsseldorf.

Postle, N. et al. 2008. Action word meaning representations in cytoarchitectonically defined primary and premotor cortex, *Neuroimage*, 43, 634–644.

Pulvermüller, F. (1999). Words in the brain's language, *Behavioral Brain Science*, 22, 253–336.

Pulvermüller, F. (2005). Brain mechanisms linking language and action, *Nature Reviews Neuroscience*, 6, 576–582.

Schluter N. D. et al. 1998. Temporal interference in human lateral premotor cortex suggests dominance for the selection of movements. A study using transcranial magnetic stimulation, *Brain*, 121, 785–799.

Schubotz, R. I. and D. Y. von Cramon 2004. Sequences of abstract neurobiological stimuli share ventral premotor cortex with action observation and imagery, *Neuroscience*, 24, 5467–5474.

Shetreet, E. et al. (2007). Cortical representation of verb processing in sentence comprehension: number of complements, subcategorization and thematic frames, *Cerebral Cortex*, 17, 1958–1969.

Van Dam, W. O. et al. (2010). How specifically are action verbs represented in the neural motor system: an fMRI study, *Neuroimage*, 53, 1318–1325.

Willems, R. M. et al. 2009. Neural dissociations between action verb understanding and motor imagery, *Cognitive Neuroscience*, 22, 2387–2400.

Zwaan, R. and M. Kaschak (2008). Language in the brain, body, and world, In: M. Robben and M. Aydede (Eds.) *Cambridge handbook of situated cognition*, Cambridge: CUP, 368–381.