A conceptual analysis of verbs of pushing and pulling

Anton Benz, Torgrim Solstad, Oliver Bott, Martin Kahnberg, Andrea C. Schalley*

Abstract. Although verbal expressions of caused motion, such as *push* and *pull*, have been extensively studied within linguistics, semantic dimensions beyond path and manner of motion have received less attention. This pilot study aims to identify such dimensions involved in the expression of caused motion in German, focusing on observable properties of pushing and pulling events that determine the selection of verbs to describe events of caused motion. Using 3D graphical modeling, participants were presented with video clips of a computer-animated agent moving a barrel, thereby allowing for a systematic manipulation of properties and hence dimensions. We investigated four dimensions to assess their impact on verb selection: (i) angle of contact, (ii) movement of the agent relative to the barrel, (iii) the agent's orientation/facing, and (iv) the force employed. Cluster and principal component analyses were conducted on the collected linguistic data. Verbs were represented by five-dimensional vectors capturing correlations with the cosine and sine of the angle, and marginal probabilities in conditions of instantaneous movement, forward facing, and heavy force. Our findings indicate that conceptually distinguishable verb clusters are primarily defined by the movement feature – that is, whether the agent moves together with the barrel or not – and the cosine of the angle. Contrary to theoretical predictions, little evidence was found supporting the categorization of verbs based on the force applied to the barrel. These results suggest that the movement and position of the agent relative to the moved object are key determinants in the production of verbal descriptions of caused motion events.

Keywords. verbs of pushing and pulling; caused motion; conceptualization; lexicalization; German

1. Introduction. Dating back to seminal work by Talmy (cf. e.g., Talmy 1985), verbal expressions of caused motion (e.g., *put*, *push*) have been studied extensively within linguistics from the perspective of linguistic typology, theoretical linguistics, language acquisition and processing (cf. e.g., Allen et al. 2007, Goldberg 1995, Hendriks et al. 2008, Margetts et al. 2022). Whereas research in the tradition of Talmy has investigated the realization of motion and path components in terms of satelite- and verb-framed languages, another important line of research has focused on the argument-structural realization and alternations associated with these verbs (e.g., Levin 1993, Goldberg 1995; and much subsequent work).

However, less attention has been directed towards semantic dimensions beyond path and manner of motion at play in these verbs. Thus, while *push* and *pull* may both be considered to be

*Special thanks to our student assistants Sophia Höbel who annotated the data and to Henry Salfner for assisting us in the creation of the animations. Torgrim Solstad and Oliver Bott gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) to subproject B01, CRC 1646, project number 512393437. Authors: Anton Benz, Leibniz-Centre General Linguistics, ZAS (benz@leibniz-zas.de) & Torgrim Solstad, Bielefeld University (torgrim.solstad@uni-bielefeld.de) & Oliver Bott, Bielefeld University (oliver.bott@uni-bielefeld.de) & Martin Kahnberg, Karlstad University (kahnberg@gmail.com) & Andrea C. Schalley, Karlstad University (andrea.schalley@kau.se).

verbs of caused motion, they differ in the relation of the agent and the moved object. For *pull*, a force is exerted in the direction of the agent, whereas for *push*, the force is directed away from the agent. What is more, some verbs seem to encode a stronger application of force than others, as witnessed by the difference between *pull*, *drag* and *haul*. In this paper, we present a pilot study using 3D graphical modelling to elicit linguistic productions of relevance to some of the dimensions involved. This pilot study is part of a larger endeavour in which we aim to identify more precisely the semantic dimensions involved in the expression of caused motion across languages. More specifically, we ask which observable properties of pushing and pulling events determine which verb is best used to describe them, and how they can be represented. What are the observable properties of agent, moved object and their spatial relation?

One goal of our investigation is to model the semantic dimensions of caused motion verbs as geometric structures within the *Conceptual Spaces* framework (Gärdenfors 2000). In Gärdenfors' theory, concepts are analysed as regions in multi-dimensional spaces which are derived from (fine-grained) semantic dimensions that to a large extent are based on perception (Gärdenfors 2000, Gärdenfors & Warglien 2012). The geometric structures provide an objective, language-independent measure of semantic similarity. Previous research has provided profound evidence for a geometrical organization of concepts in the (direct) sensory domain, such as colour (Roberson et al. 2005), olfaction (Majid et al. 2018), static spatial relations (Levinson & Wilkins 2006), and even prototypical instances of motion events (Giese et al. 2008, Malt et al. 2014). Recent work (Gärdenfors 2020, Gärdenfors et al. 2018, Warglien et al. 2012, Wolff 2007) extended Gärdenfors' approach in the verbal domain and provided ideas for integrating structural accounts. However, less progress has been made in the conceptual space of caused motion events involving both an agent and a patient as for verbs like *push* and *pull*.

Past studies in the domain of caused motion have used 2D videos to elicit descriptions of basic pushing and pulling events, focusing on the difference between verb- and satellite-framed languages (e.g. Hickmann et al. 2018, Montero-Melis 2021). Based on a production experiment using short 3D video clips, the present study instead aimed at assessing in more detail which (observable) semantic dimensions make out the domain of pushing and pulling, which we see as a fundamental domain of physical interaction between agents and patients. Furthermore, previous studies focused on prototypical events, using human actors (e.g. Malt et al. 2014) or idealized animations (e.g. Montero-Melis 2021). However, one needs peripheral event instances in order to pinpoint conceptual boundaries. A systematic manipulation of several dimensions that moves from prototypical to peripheral event instances leads to a large number of combinatorial possibilities to be tested. We addressed this challenge by presenting participants with video clips in which a computer-animated agent moved a barrel over a short distance, thereby allowing for fine-grained adjustments of potentially impactful properties.

It should be noted that the main research goal in this pilot study was to determine the predictors that trigger the production of different verbs and to classify the verbs in semantic verb clusters. The role of modifiers, or — in Talmy's terms — satellites, of various types is not discussed in this paper, but is part of a more detailed analysis that is still on-going.

2. The experiment. Our experiment investigates how physical properties of pushing and pulling events influence the speaker's choice of verb (and modifiers). As mentioned before, force is a significant concept in cognitive linguistics, particularly in Talmy's work. However, it is a multifaceted concept rooted in physics. In physics, force is a quantity that has both magnitude and direction and is represented by a vector. In caused motion, force is exerted over a period of time along a part or the entire length of the path taken by the moving object. Therefore, when conceptualizing force in a caused motion event, we can consider the *strength* (magnitude) of the force that an agent applies to an object, and the length of time during which this force is applied—is it *instantaneous* or *continuous* over the entire path length? For the perception of force, the physical orientation of the agent relative to the object is also a relevant aspect that can determine how the event is verbalized. In terms of physical orientation, we distinguish between the angle between the agent and the object, and whether the agent's body is oriented towards the object or in the direction of movement. We, therefore, identified four dimensions to test their impact on verb selection.

The first dimension concerns the **Angle** of contact between the agent and the barrel's direction of movement. We expected verb meanings to be sensitive to the agent's relative position to the barrel. In prototypical pushing events, the agent is positioned directly behind the moving object (0° angle). In prototypical pulling events, on the other hand, the agent is directly in front of the object (180° angle). To obtain a fuller picture of verb choice, we added intermediary angles in 45° intervals. Based on native speaker intuitions, we expected speakers to switch from push to pull type verbs in the range of 90° to 135°. For this reason, we divided this range more finely into 15° intervals. Consequently, our study included seven different angles: 0°, 45°, 90°, 105°, 120°, 135°, and 180°. Figure 1 illustrates four of these angles (180°, 105°, 135° and 0°). Examples of selected videos illustrating this and the other dimensions discussed below can be found in the associated OSF archive.[1]

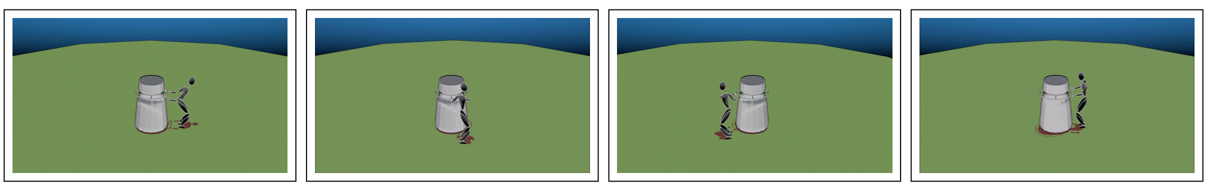

Figure 1: Stills for 180°/105°/135°/0° with left/right movement and agent facing forwards/to object

The second dimension involves the **Movement** of the agent, who can either remain in place, that is, only move their arms – resulting in *instantaneous* contact with the object – or move along with the object, leading to *continuous* contact. For example, a prototypical instance of *schubsen* 'shove' would be instantaneous, while *gehen mit* 'walk (along) with' would involve continuous movement.

The third dimension we manipulated is **Force**, which can be *light* or *heavy*. We expected this dimension to distinguish between uses of the German *push* type verbs, with *schubsen* 'push, shove' being light in strength and *stoßen* 'thrust, shove' being heavy.

We characterized the fourth dimension as the agent's **Facing**. We manipulated two different agent postures: One where the agent faces towards the object, and one where the agent faces in the direction of the barrel's movement. The intuition behind this was that the body posture of the agent would reflect their attention and therefore be relevant for the perception of the agent's effort.

[1] https://osf.io/mfstn/?view_only=4d5736ce36c547df94278e8d5188ac28

Finally, as a counter-balancing factor, we also created two version of all videos, one in which the **Direction** of the barrel's movement across the screen was manipulated, either to the *right* or to the *left* on the screen, as seen from participant's perspective.

It should be noted that we chose to use an inanimate object as the moved object to avoid any complications with regard to the interaction between agent and patient. Ultimately, however, it would be desirable to include objects that have varying properties of 'self-propelling', such as objects on wheels or even another animate entity.

2.1. METHODS.

2.1.1. PARTICIPANTS. We recruited 81 German native speakers (36 male, 45 female; mean age: 24.5 years) via the platform Prolific. Participants were directed to our server where the experiment was hosted.

2.1.2. DESIGN. Based on the above dimensions, the experiment employed a $7 \times 2 \times 2 \times 2$ within-participants and within-items factorial design manipulating the following factors: ANGLE between the agent and the barrel's direction of movement (seven angles: 0°, 45°, 90°, 105°, 120°, 135°, 180°), MOVEMENT type (*continuous* vs. *instantaneous*), FACING (*towards barrel* vs. *forwards in direction of movement*), FORCE applied (*light* vs. *heavy*). In addition, DIRECTION of the barrel's movement was included as a counter-balancing factor (*to the right* vs. *to the left* on the screen). This design yields 56 distinct main conditions (excluding DIRECTION). However, since there is no physical distinction between facing forward and facing the object for the 0° angle, we obtained a total of 52 visually discernible main conditions. The dependent variable reported in this paper was the verb produced.

2.1.3. STIMULI. Including the *left* vs. *right* manipulation involved for DIRECTION, we created 104 animated video clips, each approximately 3 seconds long, depicting a human-like agent pushing or pulling a barrel (see Tab. 1 and Figure 1). The videos were generated using a state-of-the-art physics engine (Blender, `https://www.blender.org/`). The animations were distributed on two lists such that all 52 (distinguishable) conditions involving ANGLE, MOVEMENT, FORCE and FACING were in both lists. Half of the left or right DIRECTION conditions were included in one list, the rest in the other. The order of conditions was pseudo-randomized, seeking to minimize repetition across factors. For example, two subsequent trials always differed by at least 45° in ANGLE to make videos more easily discernible. Finally, a reversed order version of each of these lists was created, bringing the number of experimental lists to four.

2.1.4. PROCEDURE. The experiment was implemented using the free Onexp software (version 1.3.1).[2] After reading short written instructions, participants proceeded to a short practice of two trials, upon which they received the experiment in 52 trials in a single block. Participants watched video clips of approximately 3 seconds each, presented individually on separate pages, showing an agent pushing or pulling a barrel. The videos were displayed centrally on their screen, accompanied by the question "What does the person do with the barrel?". Below this, a sentence frame with a text field was provided for them to complete, see (1). Their task was to describe as spontaneously as possible what the agent was doing. Before starting the main experiment, participants

[2] See http://onexp.textstrukturen.uni-goettingen.de

were informed that their descriptions would later be used to characterize the individual video clips. Therefore, it would be important that their descriptions be specific enough for another person to assign them to the correct video clips. They were also instructed that each video clip was independent, featuring a different animated person, and that the barrel might behave differently in each clip.

(1) The person [ ] (*original in German*)

2.2. DATA AND CODING. We gathered a corpus of 4,212 descriptions (word length range: 3–70, mean 8.7). We annotated >20 features of the descriptions (not yet finalised). In the following analysis, we concentrated on 3 features: (1) the main verbs in the matrix clauses that express movement of the barrel, (2) whether the main verb is a particle verb or not, and (3) whether the barrel (the moved object) is the direct object or embedded in a prepositional phrase (PP). We found 95 different matrix verb constructions with 9 matrix verbs that have a frequency $> 0.5\%$, see (2).[3]

(2)

ziehen 'pull'	1635	*gehen* 'walk'	173
schieben 'push'	1156	*bewegen* (REFL) 'move (oneself)'	102
drücken 'press'	195	*bewegen* 'move'	71
schubsen 'push', 'shove'	195	*laufen* 'walk'	29
stoßen 'thrust, shove'	176		

2.3. DESCRIPTIVE RESULTS. Based on introspection, we identified four types of event descriptions, see Figure 2. The first two types conceptualize the movement event as direct causation involving a force exerted on the barrel. They are distinguished by the direction of the force vector – either towards or away from the agent. The third type describes the agent's movement, with the barrel following as a satellite. The fourth type describes the movements of the agent and the barrel in two separate clauses, conceptualizing their respective movements as two parallel events. Examples for all types are shown in (3) (originals in German).

(3) The person . . .
  a. simply pushes the barrel to the left.
  b. pulls the barrel behind him with one arm.
  c. walks with the barrel to the right.
  d. walks to the left and pushes the barrel next to him.

Inspection of the frequency data (see Table 1) reveals further regularities. Verbs of the 'walk-with' type occur only with continuous movement and are most frequent at intermediate angles. The 'shove' type verbs *schubsen* and *schieben* occur only with instantaneous movements and are most frequent at lower angles. Other verbs show a steady increase or decrease in frequency with angle and can occur with both instantaneous and continuous movement.

2.4. ANALYTIC RESULTS. For the cluster analysis, each verb was represented by a 5-dimensional vector, capturing its correlation with the cosine and sine of the angle, and its marginal probabili-

[3]See the OSF archive https://osf.io/mfstn/?view_only=4d5736ce36c547df94278e8d5188ac28 for a list with data.

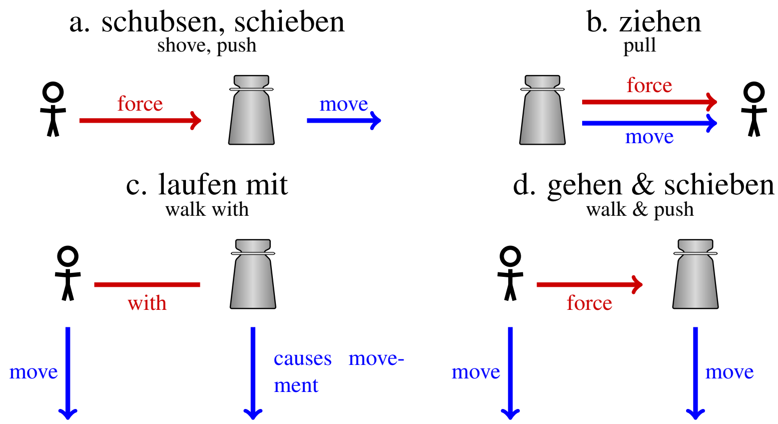

Figure 2: Classification of event descriptions

ties of being produced in three conditions: instantaneous Movement (MeanMov), forward Facing (MeanFacing), and heavy Force (MeanForce). All features were coded as binary variables (1 indicating presence). A positive correlation with cosine indicates that verb frequency decreases as the angle increases, whereas a negative correlation signifies an increase in frequency as the angle decreases. A positive correlation with sine suggests that the verb is most frequently used for angles near $90°$.

A principal component analysis (PCA) was conducted to examine the underlying structure of the dataset. According to Kaiser's criterion (Eigenvalue > 1), the optimal dimensionality would be 2 components. However, adopting a more lenient threshold (Eigenvalue > 0.7), 3 dimensions are justified. We chose to retain 3 principal components, which together explain 91% of the total variance (with the first 2 components explaining 73%). The factor loading plot and a score plot of individual verbs for the first two components are presented in Figure 3. The factor loading plot indicates that the first component is primarily influenced by correlations with sine, cosine, and MeanMov. Additionally, a heatmap depicting clusters based on these features shows that verbs are predominantly grouped according to their correlation with these three variables, see Figure 4.

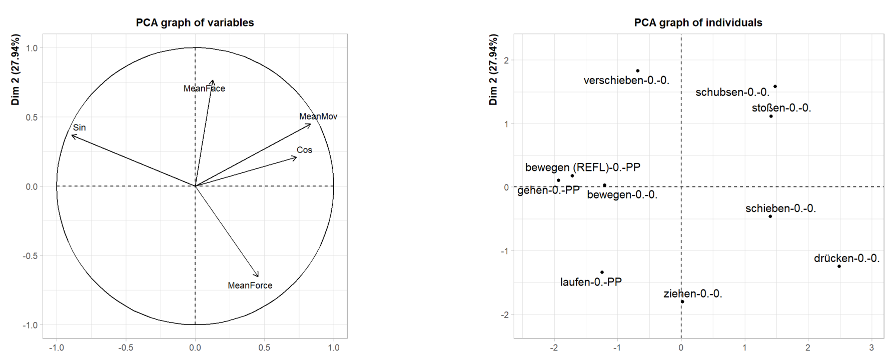

Figure 3: Factor loading plot and score plot for dimensions 1 & 2 of the PCA analysis.

K-means clustering ($k = 3$) was performed for binary feature combinations (see Figure 4). The Movement feature (*continuous* vs. *instantaneous*) identified three verb clusters: verbs such as *bewegen (refl.)* 'move', *gehen* 'walk', and *laufen* 'walk' were associated with the +continuous

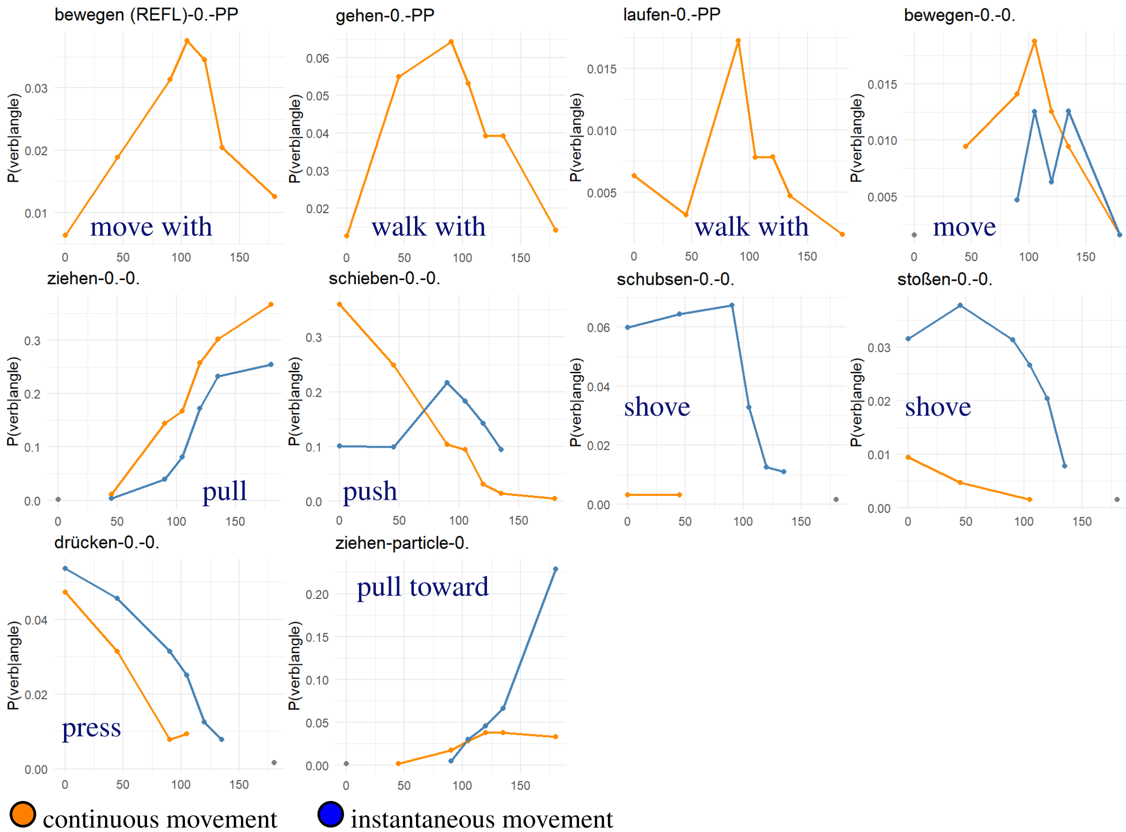

Table 1: Production probabilities of frequent verbs conditional on Movement features. Probabilities $P(verb|angle)$: proportion of descriptions with matrix *verb* given the *angle* ($\mathbf{0°}$, $\mathbf{45°}$, $\mathbf{90°}$, $\mathbf{105°}$, $\mathbf{120°}$, $\mathbf{135°}$, $\mathbf{180°}$). No line or grey dot: *angle* not produced for *angle* and Movement. Coding of verb features: '*verb-0.-0.*': no particle verb, barrel direct object; '*verb-0.-PP*': no particle verb, barrel realized in a PP

feature, while *schubsen* 'push' and *stoßen* 'thrust' were marked as +instantaneous. The remaining verbs were unmarked with respect to the continuous/instantaneous distinction (Figure 4c). For all −continuous verbs, the barrel was realized as the direct object, whereas for all +continuous verbs, it was embedded in a prepositional phrase (e.g., *'move oneself with the barrel'*). A corresponding cluster analysis based on sine and cosine showed that the +continuous verbs formed a cluster positively correlated with sine, while the +instantaneous verbs formed a cluster positively correlated with cosine. Verbs unmarked with respect to MeanMov, with the exception of *ziehen* 'pull', were grouped in the +cosine cluster. *Ziehen* was isolated and formed its own −cosine cluster. The features Facing and Force did not result in clearly delineated clusters. The results of the cluster analyses are summarised in Table 2.

The cluster analysis confirmed the conceptual classification of event descriptions based on introspection and semantic intuition shown in Table 2. It added a finer distinction in the class of verbs expressing direct causation with a force applied in the direction of movement by differentiating between verbs that only occur in +instantaneous contexts, and verbs that are unmarked with

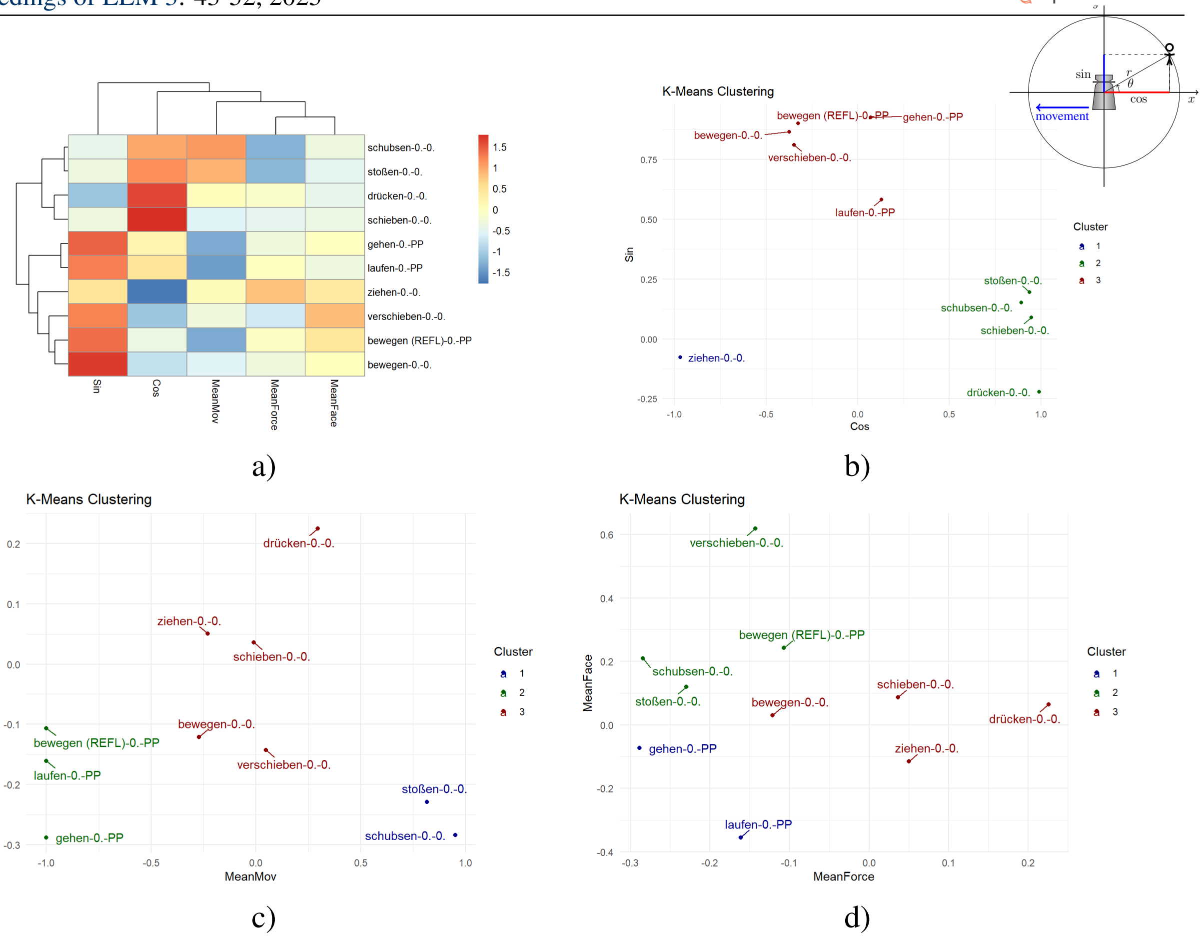

Figure 4: Cluster Analyses: (a) Heatmap showing clustering of verbs based on their correlations with sine, cosine, MeanMov, MeanForce, and MeanFace. (b) – (d) K-means clustering results ($k = 3$): (b) clusters based on correlations with sine and cosine only; (c) clusters based on MeanMov and MeanForce values; (d) clusters based on MeanForce and MeanFace values. Clear separation of clusters is observed in (b) and (c), with no separation visible in (d)

respect to Movement.

For verbs produced by at least 15 participants, we fitted linear mixed effect models with Cos_Angle (Cos), Movement,[4] Force, and Facing as fixed effects and by-participant random intercepts. Predictors vary for individual verbs. We found the following patterns for Cos: Cos was no significant predictor for +continuous-verbs; all other verbs were either positively or negatively correlated with Cos (Table 2), except *bewegen* 'move', which also did not correlate with Cos. The other predictors may correlate with individual verbs, but we found no general pattern correlated to verb clusters.

3. Discussion. In this paper, we presented the results of a pilot study in which we elicited verbal descriptions for 3D video animations of caused motion events involving an animate agent and an inanimate patient (a barrel). We manipulated four factors in the animations, (i) the angle between

[4] We dropped Movement for +cont- and +inst-verbs. To resolve convergence issues, Angle was transformed using the cosine function. The sine of the angle was excluded from the list of fixed effects because the unequal number of conditions at lower and higher angles led to spurious significant results in the linear model due to data asymmetry.

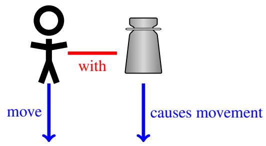	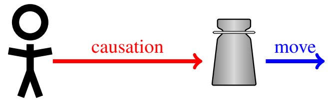		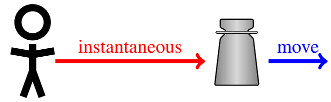
+ continuous	unmarked		+ instantaneous
+ Sin_Angle	+ Cos_Angle	− Cos_Angle	+ Cos_Angle
bewegen (REFL)-PP (*move self with*)	schieben (*push*)	ziehen (*pull*)	stoßen (*push*)
gehen-PP (*walk with*)	drücken (*press*)		schubsen (*push*)

Table 2: Verb clusters: semantic feature (red), use correlated (blue)

agent and patient, (ii) the continuity of the agent's Movement, (iii) the Force employed, (iv) and the agent's Facing, that is, the direction of their attention. For these factors, conceptually clearly distinguishable verb clusters can only be defined by the Movement (continuity) feature, which tells us whether the agent moves together with the barrel (+cont) or remains in place (+inst), and the Cos of the angle. Interestingly, the results provide little evidence that verbs are categorized according to the Force applied to the barrel (as predicted by Gärdenfors & Warglien's 2012 theory). It is rather the movement and position of the agent in relation to the barrel that determine production of verbal descriptions.

Our ongoing coding efforts will allow us to assess the influence of the manipulated factors on a number of modifiers, including – but not limited to – the force of the agent, the source, path, and goal of the movement, or its manner. Going beyond German, by testing English, Italian, Russian, Swedish and Persian, we will also develop a basis for a more broad typological investigation of the semantic dimensions of caused motion.

References

Allen, Shanley, Aslı Özyürek, Sotaro Kita, Amanda Brown, Reyhan Furman, Tomoko Ishizuka & Mihoko Fujii. 2007. Language-specific and universal influences in children's syntactic packaging of manner and path: A comparison of english, japanese, and turkish. *Cognition* 102(1). 16–48. https://doi.org/10.1016/j.cognition.2005.12.006.

Giese, Martin A., Ian Thornton & Shimon Edelman. 2008. Metrics of the perception of body movement. *Journal of Vision* 8(9). 13–13. https://doi.org/10.1167/8.9.13.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Gärdenfors, Peter. 2000. *Conceptual spaces: The geometry of thought*. Cambridge, MA: The MIT Press.

Gärdenfors, Peter. 2020. Events and causal mappings modeled in conceptual spaces. *Frontiers in Psychology* 11. https://doi.org/10.3389/fpsyg.2020.00630.

Gärdenfors, Peter, Jürgen Jost & Massimo Warglien. 2018. From actions to effects: Three constraints on event mappings. *Frontiers in Psychology* 9. https://doi.org/10.3389/fpsyg.2018.01391.

Gärdenfors, Peter & Massimo Warglien. 2012. Using Conceptual Spaces to Model Actions and Events. *Journal of Semantics* 29(4). 487–519. https://doi.org/10.1093/jos/ffs007.

Hendriks, Henriette, Maya Hickmann & Annie-Claude Demagny. 2008. How adult english learners of french express caused motion: A comparison with english and french natives. *Acquisition et interaction en langue étrangère* 27(15-41).

Hickmann, Maya, Henriëtte Hendriks, Anne-Katharina Harr & Philippe Bonnet. 2018. Caused motion across child languages: a comparison of English, German, and French. *Journal of Child Language* 45(6). 1247–1274.

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.

Levinson, Stephen C. & David P. Wilkins (eds.). 2006. *Grammars of space: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.

Majid, A., S. G. Roberts, L. Cilissen, K. Emmorey, B. Nicodemus, L. O'Grady, B. Woll, B. LeLan, H. de Sousa, B. L. Cansler, S. Shayan, C. de Vos, G. Senft, N. J. Enfield, R. A. Razak, S. Fedden, S. Tufvesson, M. Dingemanse, Ö. Öztürk, P. Brown, C. Hill, O. Le Guen, V. Hirtzel, R. van Gijn, M. A. Sicoli & S. C. Levinson. 2018. Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences U.S.A.* 115(45). 11369–11376.

Malt, Barbara C., Eef Ameel, Mutsumi Imai, Silvia P. Gennari, Noburo Saji & Asifa Majid. 2014. Human locomotion in languages: Constraints on moving and meaning. *Journal of Memory and Language* 74. 107–123.

Margetts, Anna, Sonja Riesberg & Birgit Hellwig (eds.). 2022. *Caused accompanied motion: Bringing and taking events in a cross-linguistic perspective* Typological Studies in Language. Amsterdam / Philadelphia: John Benjamins. https://doi.org/10.1075/tsl.134.

Montero-Melis, Guillermo. 2021. Consistency in motion event encoding across languages. *Frontiers in Psychology* 12(625153).

Roberson, D., J. Davidoff, I.R.L. Davies & L.R. Shapiro. 2005. Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology* 50(4). 378–411.

Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms. In Timothy Shopen (ed.), *Language typology and syntactic description iii: Grammatical categories and the lexicon*, 57–149. Cambridge: Cambridge University Press.

Warglien, Massimo, Peter Gärdenfors & Matthijs Westera. 2012. Event structure, conceptual spaces and the semantics of verbs. *Theoretical Linguistics* 159–193.

Wolff, Phillip. 2007. Representing causation. *Journal of Experimental Psychology: General* 136(1). 82–111.