Assesing the Feature-Driven Nature of Similarity-based Sorting of Verbs

Pinar Öztürk, Mila Vulchanova, Christian Tumyr, Liliana Martinez, and David Kabath

Abstract—The paper presents a computational analysis of the results from a sorting task with motion verbs in Norwegian. The sorting behavior of humans rests on the features they use when they compare two or more words. We investigate what these features are and how differential each feature may be in sorting. The key rationale for our method of analysis is the assumption that a sorting task rests on a similarity assessment process. The main idea is that a set of features underlies this similarity judgment, and similarity between two verbs amounts to the sum of the weighted similarity between the given set of features. The computational methodology used to investigate the features is as follows. Based on the frequency of co-occurrence of verbs in the human generated cluster, weights of a given set of features are computed using linear regression. The weights are used, in turn, to compute a similarity matrix between the verbs. This matrix is used as an input for the agglomerative hierarchical clustering. If the selected/projected set of features aligns with the features the participants used when sorting verbs in groups, then the clusters we obtain using this computational method would align with the clusters generated by humans. Otherwise, the method proceeds with modifying the feature set and repeating the process. Features promoting clusters that align with human-generated clusters are evaluated by a set of human experts and the results show that the method manages to identify the appropriate feature sets. This method can be applied in analyzing a variety of data ranging from experimental free production data, to linguistic data from controlled experiments in the assessment of semantic relations and hierarchies within languages and across languages.

Index Terms—Verb features, verb sorting, similarity.

I. Introduction

SORTING tasks are a popular knowledge elicitation technique used in psychology and cognitive studies [1], [2]. In a typical sorting task participants are asked to sort in groups items in a particular domain. This kind of task rests on the common assumption that, in categorization processes, humans rely on specific features that differentiate one group of objects from another, and that these features characterize and define the group in a broader domain [3].

We designed a sorting task to study the semantic domain of verbs of human locomotion below the basic level ([4], [5], [6]). Specific verbs of locomotion include words, such as English *strut*, *stroll*, *gambol*, *hop*, and the like.

Our main assumption is that the way speakers group those verbs is revealing about the semantic structure of this field.

Our hypothesis is that the size (how many) and constitution (what verbs) of these groups can be used to derive the semantic features that characterize both individual lexical items and the domain as a whole. We investigated whether and how it is possible to discover such relations and patterns for the set of motion related verbs, based on verb clusters provided by the human subjects. The paper presents a computational method that aims to discover the most salient features and their degree of saliency.

The approach adopted in this paper resembles vector-based semantic space models which rely on patterns of word co-occurrence to derive similarity estimates ([7], [8]). The difference from such approaches is that they aim to extract information either from the broader lexical or from the syntactic context of the target word, while our approach targets groupings based on closer semantic similarity within a well-defined conceptual and semantic domain (e.g., words describing human locomotion). In our formalisation, both the columns and the rows in the raw matrix are target words, i.e. it is a verb-verb matrix. Even though this approach might appear narrow and highly restricted to the domain it applies to, it is justified on the basis of research and intuitions in lexical semantics, as well as human categorization. Thus, studying the grouping of words that are partially synonymous with each other and can be subsumed under the same superordinate term, can be used to reveal the underlying features that characterize this semantic field and the basic (superordinate) term. Moreover, Semantic space models have been criticized exactly on the grounds of not being able to address the nature of the semantic relationship that underlies proximity of words in the semantic space [7]. We address this shortcoming by using a feature-verb matrix to estimate the weighting of features.

Another difference between the current approach and existing approaches in cognitive science and psychology is that, while the latter have used human elicitation to verify the findings from semantic space models [9], we adopt a parallel experimental strategy: we seek to find out the extent to which a computational model based on human data can improve by using featural data elicited from the human data.

The outline of the paper is the following. We first introduce the human sorting task experiment and its linguistic background in the next section. We then proceed, in section III with the computational method for computing feature weights and the clusters based on various combinations of the features.

Section IV presents the computational experiments and the results of applying the computational method. We discuss the obtained results in section V and conclude with a summary and future directions in section VI.

II. HUMAN EXPERIMENTS

Germanic languages are characterized by a rich system of specific verbs describing locomotion, and the distinctions among the items in this domain are not always very clear. Furthermore, little is known about the way native speakers of these languages acquire such highly specific vocabulary, and whether they use salient perceptual features of the actions these words denote, and then map these features onto the lexical items at hand or simply rely on the linguistic contexts in which they first encounter these verbs [10].

As a first step in studying the native speakers's knowledge of specific locomotion verbs, we asked native speakers of Norwegian to group 41 verbs that were selected through a 3 step process, a semantic recall task, an elicitation task, with results from both being compared to a comprehensive list compiled on the basis of dictionary information [6].

The verbs appeared on small paper cards and participants were asked to sort them in groups by similarity. Participants are then asked to describe what features they have used in the grouping process. All the features mentioned by one or several subjects constitute the candidate feature set including 15 features. Using the computational method described in the next section, we tried to select the subset from this candidate set of the features that were most influential in the overall sorting experiment.

To avoid confounding of the results, the human subjects were given the opportunity of placing verbs whose meaning they did not know or, for some reason, whose placement they felt uncertain about, in a separate group labeled "out", which indicated exclusion from the sorting. Verbs excluded in this way are considered as a negative contribution and were excluded from further analyses. A total of three verbs were excluded by more than two subjects and were removed from the dataset for analysis.

The groups for each participant were photographed by digital camera, and the results for all participants were manually entered in an excel file and consequently converted into a *verb co-occurence* matrix of which each cell indicates how many of the subjects put the two corresponding verbs into the same group. These raw data served as the input for agglomerative hierarchical clustering. This matrix constitutes also the input to the computational method described in the next section.

III. THE COMPUTATIONAL METHOD

The inputs to the method are a feature-verb matrix representing subjects' description of which features were taken into consideration when grouping verbs, and the verb co-occurence matrix prepared after the human experiment. In this paper, the feature-verb matrix has size 15 x 41, while the verb co-occurence matrix is of size 41 x41.

The overall method is summarized in Algorithm 1 where S_{human} is a verb-verb matrix, i.e., the co-occurence matrix generated by accumulating the sorting data provided by the subjects. $S_{human}(v_i, v_j)$ represents the number of subjects who put verbs v_i and v_j into the same group. It is considered to represent the human judgment of similarity between the verbs. S_{comp} is the computed (more precisely, to be computed) feature-based verb similarity matrix.

Algorithm 1 Method

1: $C_{human} \leftarrow$ Cluster data based on S_{human}
2: Matrix A $\leftarrow$ human description of feature-verb relations
3: **repeat**
4:    Compute feature weights $\mathbf{W}$ (as described in algorithm 2)
5:    Generate weighted feature-based verb similarity matrix S_{comp} using $\mathbf{W}$ and A (details described in algorithm 3)
6:    $C_{comp} \leftarrow$ Cluster the data based on S_{comp}
7:    Evaluate alignment between C_{human} and C_{comp}
8:    **if** $C_{comp} \not\simeq C_{human}$ **then**
9:       Remove the feature with the lowest weight
10:   **end if**
11: **until** $C_{comp} \simeq C_{human}$ **or** # of features < 2

The algorithm describes the process of evaluating the calculated feature weights with regard to the grouping data provided by the human subjects. The grouping data are clustered (the result is denoted as C_{human} in Algorithm 1) using agglomerative hierarchical clustering. After the weights of the features are computed as explained in section III-A, a weight-based verb similarity matrix S_{comp} is computed (explained in section III-B) using these weights. Then, the verbs are clustered again using the same clustering methods, this time using the new similarity matrix S_{comp}. These clusters are depicted as C_{comp} in Algorithm 1.

If the computed clusters C_{comp} and human based clusters C_{human} align, i.e. are fairly similar (depicted as $C_{comp} \simeq C_{human}$), the features and weights are considered to indicate what the human subjects based their clustering of the verbs on. If the clusters do not align, some features are removed from the set of features and C_{comp} is computed anew, and the process is repeated until an alignment has been achieved.

A. Computation of Weights

A central idea underlying the proposed method is that similarity between two verbs is equal to the weighted sum of the similarities between the involved features, which is defined by Equation 1 where w_n is the weight of feature a_n.

$$S(v_i, v_j) = w_1 f(a_{1i}, a_{1j}) + w_2 f(a_{2i}, a_{2j}) + \cdots + w_n f(a_{ni}, a_{nj}) \quad (1)$$

Values $f(a_i, a_j)$ represent the similarity between verbs v_i and v_j computed applying the similarity metric f on the cells of the feature-verb matrix A which captures subjects' description of which features were used when placing each verb in a group. The f function uses one of the well-known similarity measures for binary vectors [11].

In addition to the rationale captured by Equation 1, Equation 2 conveys another central assumption in our method:

$$S_{comp}(v_i, v_j) = S_{human}(v_i, v_j) \qquad (2)$$

The instantiation of equations 1 and 2 for all verbs yields the following linear system of equations, which, when solved, provide values for the weights $w_{1...n}$ for the features.

$$\begin{bmatrix} f(a_{11}, a_{12}) \\ f(a_{11}, a_{13}) \\ f(a_{11}, a_{14}) \\ \vdots \\ f(a_{nm}, a_{nm-1}) \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} S_{human}(v_1, v_2) \\ S_{human}(v_1, v_3) \\ S_{human}(v_1, v_4) \\ \vdots \\ S_{human}(v_m, v_{m-1}) \end{bmatrix}$$

Algorithm 2 describes the process of calculating the weights of the 15 features in the candidate feature set. It uses the feature-verb matrix A and the human generated verb co-occurrence matrix S_{human} to calculate the weights; both are data from human experiments. The value of feature a_n for a verb v_j is denoted as a_{nj} and is found in the feature-verb matrix A. The similarity of feature a_n between verb v_i and v_j is computed by $f(a_{ni}, a_{nj})$, and w_n is the weight or importance of feature a_n. The value of the weights are determined by solving the set EQ of $\frac{i^2}{2}$ linear equations where m denotes the number of verbs. $S_{human}(v_i, v_j)$ denotes the number of subjects having placed the verbs i and j in the same group. A similar approach is taken in [12] where the concerned items are movies and similarity between two movies is associated with the number of persons who rated both of these movies.

Algorithm 2 Calculation of weights

1: $n \leftarrow$ Number of features
2: $EQ \leftarrow$ Empty set of linear equations
3: **for each** verb v_i **do**
4:    **for each** verbs $(v_j), j = (i+1)$ **do**
5:       Add $w_1 f(a_{1i}, a_{1j}) + \cdots + w_n f(a_{ni}, a_{nj}) = S_{human}(v_i, v_j)$ to EQ
6:    **end for**
7: **end for**
8: Solve EQ for $\mathbf{W}$
9: **return** $|\mathbf{W}|$

B. Computation of Feature-based Verb Similarity Matrix

Algorithm 3 describes the process of calculating the similarity between verbs based on the feature weights which

were computed using algorithm 2. The similarity between two verbs i and j, denoted as $S_{comp}(v_i, v_j)$ is then computed using Equation 1. This process generates the feature-based similarity matrix S_{comp}.

Algorithm 3 Calculation of verb similarity based on weights

1: $n \leftarrow$ Number of features
2: **for each** verb v_i **do**
3:    **for each** verb $(v_j), j = (i+1)$ **do**
4:       $S(v_i, v_j) = \sum_{k=1}^{n} w_k f(a_{ki}, a_{kj})$
5:       $S_{comp}(i, j) = S(v_i, v_j)$
6:    **end for**
7: **end for**
8: **return** S_{comp}

IV. EXPERIMENTS AND RESULTS

We have conducted a set of experiments to see how the different distance metrics would affect the clustering performance, and the effect of the different linkage methods in hierarchical clustering of the verbs. Another set of experiments were devoted to investigating which features are most salient in the clustering. For this purpose we used the algorithm 2 to determine the weights of features and algorithm 3 to compute the distance matrix. Then we applied hierarchical clustering, again using different linkage methods.

Regarding the human clustering, we have experimented with the distance metrics provided by MATLAB such as *Jaccard*, *Correlation*, *Euclidean*, *Minkowski*, *Cosine*, *Chebychev* etc. In addition, we have implemented the Multiset distance metric [1] which has proven appropriate in previous analyses of verb similarity (Dimitrova-Vulchanova et al., in press). As to linkage methods, MATLAB provides several methods including *Centroid*, *Median*, *Single*, *Average*, and *Complete*. The best clustering tree of human grouping data was found to be provided by *Euclidean* as the distance metric and *Average* as the linking method. Euclidian Average has proven useful in plotting cross-linguistic differences and similarities in the naming of cutting and breaking scenes in a representative sample of world's languages (Majid et al. 2008), and our results confirm the advantages of the method in similar tasks. Figure 1 illustrates Jaccard-Average combination while Figure 2 shows the cluster tree when Euclidean-Average combination is used.

We have identified a set of features to have a role, in various degrees, in the human grouping process (referred to as feature-verb matrix above). Our anticipation is heavily based on the descriptions provided by the subjects who participated in the experiments. However, we have also supplemented these

---

[1]Calculated as

$$d(S_i, S_j) = 1 - \frac{\sum_{o \in O} min(n_{0, S_i}, n_{0, S_j})}{\sum_{o \in O} max(n_{0, S_i}, n_{0, S_j})}$$
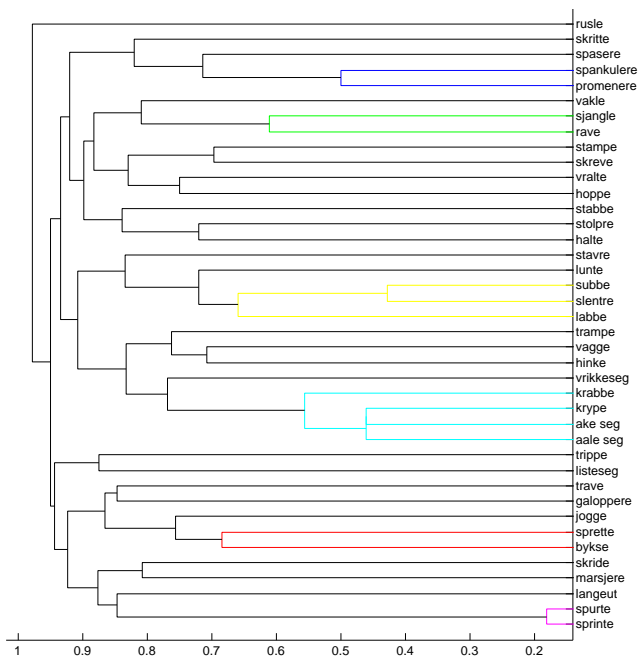
Fig. 1. Clustering of human grouping data using Jaccard metric and Average link.
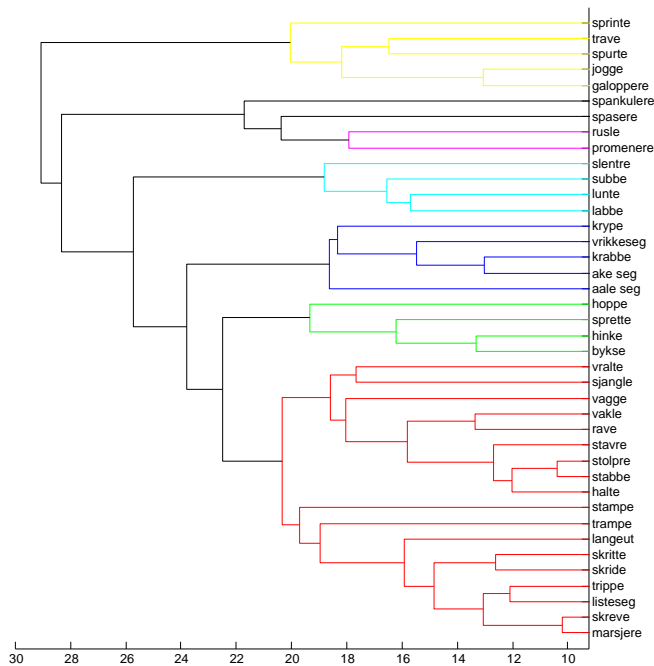

Fig. 2. Clustering of human grouping data using Euclidean metric and Average link.

data with information from the dictionary definitions of the verbs, as well as human expert judgments. Using the method presented in section III we have estimated the weights (i.e., salience) of the features in the grouping process. As a next step we computed the distance matrix (i.e, the verb-verb matrix) to be used as input for the clustering. In this process we

have experimented with different distance metrics and linking methods, as already described above.

Initially we had 15 features: *contact* (with substrate), *limbs* (body parts involved in moving), *propulsion*(pattern), *position* (of parts of the body not involved in the motion), *symmetrical (motion pattern)*, *sideways* (motion pattern), *stride* (length), *(typical)* agent, *cause*, *sound*, *speed*, *effort*, *agility*, *social* (context), *purpose*. The computed weights of these features are shown in figure 3.
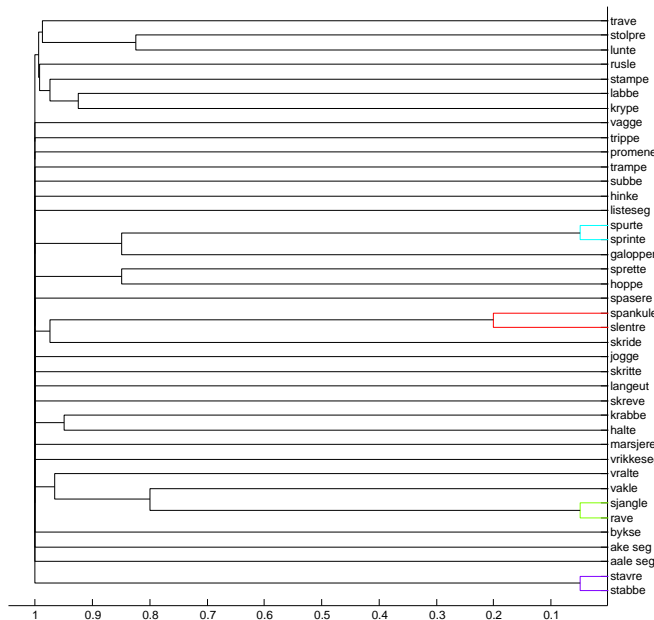

Fig. 4. Clusters based on 15 features, Jaccard metric and Average linkage was used.

Using these weights we studied the hierarchical trees of the verbs. The Euclidean-Average combination has shown the best performance, according to human expert judgments. Two of the hierarchical trees based on these weights are shown in Figures 4 (using Jaccard metric and Average linkage method) and 5 (Euclidean and Average). As can be seen in the sorted feature set according to weights (see Figure 3), some of the weight values are significantly lower than others. Moreover, both the Jaccard Average and the Euclidian Average clusters based on all 15 features were not particularly successful in capturing the structure of the semantic field and deviate substantially from the human data cluster, as judged by human experts. This deviation from the human data cluster may suggest that either a/some features are irrelevant, and/or b/ not all features capture adequately the semantic relationship in the semantic field at hand. Therefore we have analyzed different and fewer numbers of feature combinations. The feature weights showed the same trend, while clustering performance varied depending on the number of features and which features were chosen. In general removal of the two low-weight features *propulsion* and *position*, the two middle
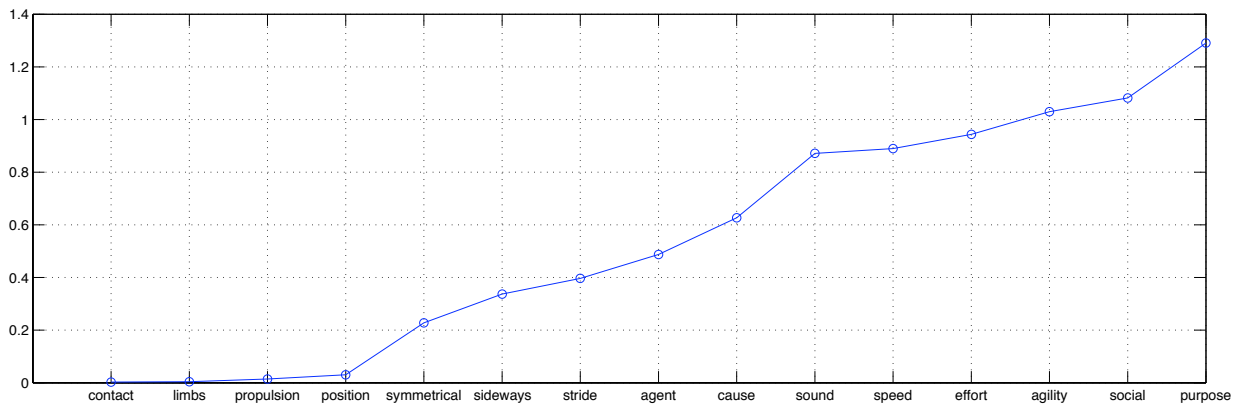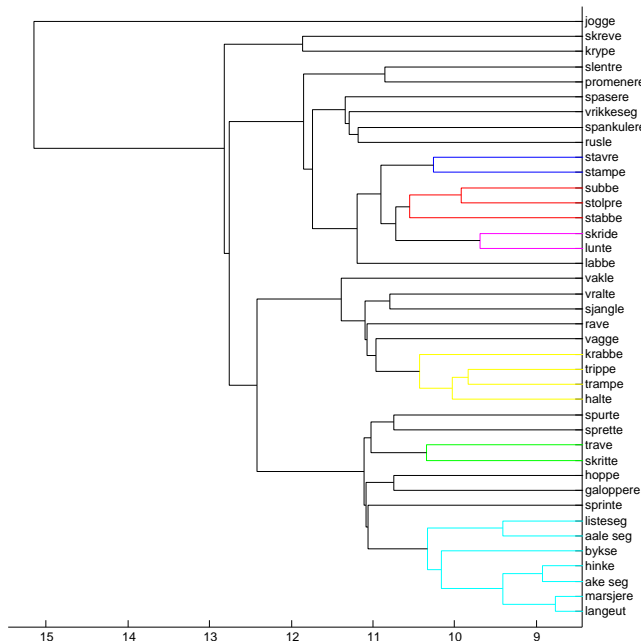
Fig. 3. Weight values of the 15 features.

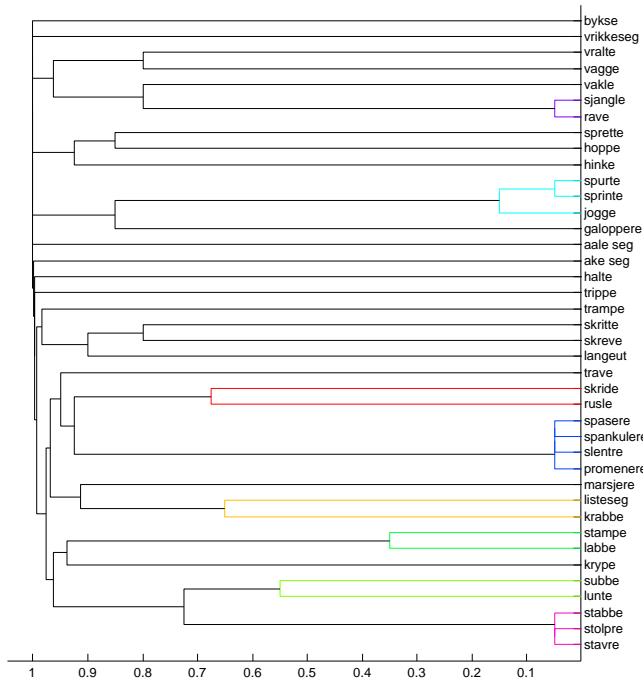Fig. 5. Clusters based on 15 features, Euclidean metric and Average linkage was used.

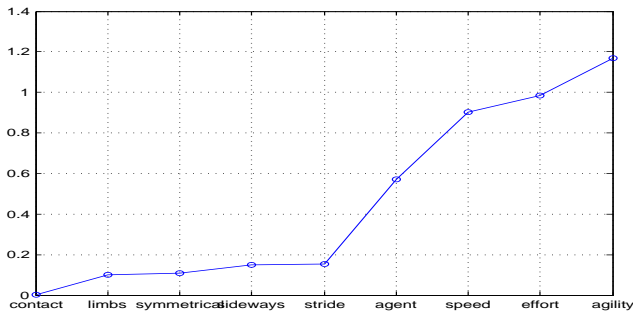Fig. 7. Clusters based on 9 features, with Jaccard metric and Average linkage.

Fig. 6. Weights for 9 features.

The clusters based on these 9 features are illustrated in Figures 7 (Jaccard-Average combination) and 8 (Euclidean-Average). Corresponding feature weights are shown in Figure 6. Figure 9 illustrates the clusters for the following 8 features: 'contact', 'limbs', 'symmetrical', 'sideways', 'stride', 'agent', 'speed', and 'agility' where Euclidean metric and Average linking is used. In this experiment the feature 'effort' has been removed, while Figure 10 illustrates the clusters when the feature 'contact' is removed instead.

V. DISCUSSION

The results from the computational method employed have highlighted a number of interesting features of this kind of research. Firstly, they have underscored the validity
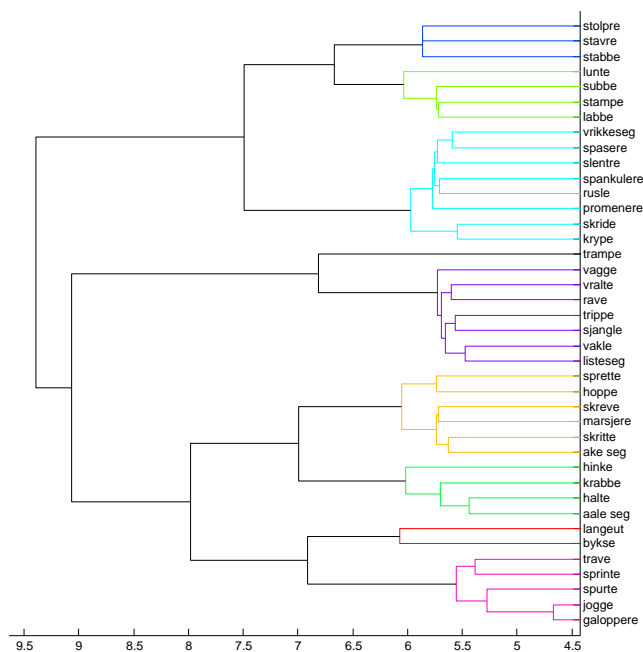
features *cause* and *sound*, and the two high-weight features *social* and *purpose* produced a balanced effect.

Fig. 8. Clusters based on 9 features, Euclidean metric and Average linkage was used.
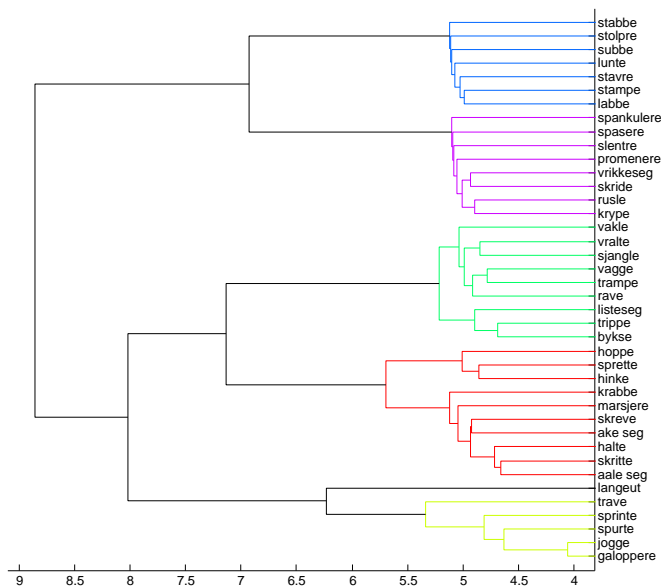

Fig. 9. Clusters based on 8 features, Euclidean metric and Average linkage is used. "Effort" is removed.

of combining human data analyses with computational methods (see also McDonald 2000). In addition, they have demonstrated that computer modelling of the data can provide useful insights for the underlying semantic similarities, as well as complement or even supplement the human analysis. Concerning the distance and linkage methods, the Euclidian Average has proven most useful in representing the underlying similarities in the data, as well as in visualising the structure

of the semantic field of more specific verbs of locomotion. In contrast, the Jaccard distance metric does not seem to capture the structure of the field, and the clusters created by this method appear ad hoc and largely accidental, where one finds words describing very different kinds of motion on the same branch (e.g., vralte (move slowly swinging from side to side) and hoppe (jump)), while similar words appear on very distant nodes (e.g., the high velocity verbs). This is confirmed in our previous work as well, whereby Jaccard plots, while not particularly revealing, were good at capturing subtle details of specific similarities between isolated items.

The method of feature weighting has also proven successful and the removal of features has produced neat and succinct clusters. It is worth mentioning that feature removal has a negative side to it, since it increases the weights of certain features, while removing other features which might be relevant for an in-depth detailed analysis. Furthermore, there is a risk of capturing only the overall and more general tendencies in the structure of the semantic field at hand, while missing more subtle aspects of semantic similarity. Our tentative conclusion at this stage is that a set of 9 or 8 features is within the comfortable zone in this respect. The weighted feature cluster with 8 features is most representative of this method and reveals a graded structure of the field of locomotion, with clear-cut clusters defined on a continuum from low-speed, heavy (longer stride), non-agile motion patterns to high-velocity, agile and effort-demanding locomotions. The middle clusters reflect the importance of contact with the substrate, limb alternation, which are features carrying less weight in the 8-feature plot. This kind of graded structure has, in fact, been mentioned in the descriptions provided by the participants in the study. Some have indicated that, when arranging the groups, they have been guided by an inner structure in terms of slow effortful movement to high speed agile motion. Even though there is no exact match between the cluster obtained from the human sorting data C_{human} and the feature-weighted cluster C_{comp}, both reveal the most salient semantic features relevant for the grouping, such as speed, effort, agility, contact with the substrate. We also hypothesise, based on these results that the cluster based on the human data, reflects the individual differences and variation in what features individual speakers find most relevant for the grouping. We further hypothesise that these features are perceptual in nature and may vary according to the specific contexts in which these lexical items were acquired. For instance, for verbs that denote unsteady/swinging gaits, other factors (e.g., speed or effort) may be found irrelevant. In contrast, the cluster obtained by computer modeling and feature-weighting is based on features that the participants mentioned in the subsequent interview session and dictionary definitions of the verbs, and as such, are the result of deliberate conceptualisation. This finding is interesting in its own right and confirms usage-based accounts of language acquisition as tightly temporally and spatially-bound ([13], [14], [15], [16]).

It is worth noticing that the feature-weighted clusters based on fewer features (8 and 9) still display some anomalies. For instance, verbs like *krype* (creep), *krabbe* (crawl for human infants), *ake seg* (move butt-scooting) and *aale seg* (slither, creep like a snake) all belong in different and not immediately coherent clusters, while in the human data cluster they appear on the same branch. What these verbs share, and what is reflected in the human sorting, is the fact that all of these types of locomotion are non-default (for humans), presuppose greater contact with the substrate, in the case of aale seg, full body contact with the ground, and the use of more limbs than just the legs. We propose that the feature-weighted cluster does not reflect this similarity properly as the result of removing some of the features that underlie the similarity among the above verbs, most notably the two low-weight features: "propulsion" and "position". As observed above, this is one of the down-sides of feature removal and modeling.
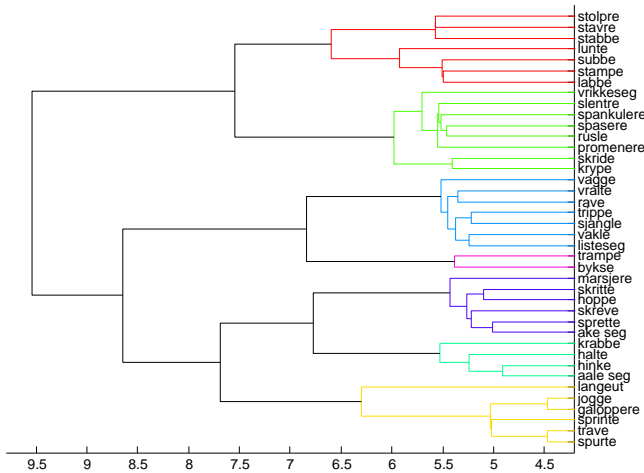

Fig. 10. Clusters based on 8 features, Euclidean metric and Average linkage is used. 'Contact' is removed.

Overall, the removal of different features from the feature-weighted plots has proven a successful step in the attempt to model similarity representations and to detect what features appear to be more important for describing similarities among verbs of motion at the specific level. In the 9 feature set, the removal of the features "contact" and "effort" has less of an influence on the clustering of verbs, and both clusters are close to the one based on the original human data. This is true, in particular, of the effect of removing "contact" from the feature set (see Figure 10). This can be taken as indication that this feature is less relevant for identifying special patterns of locomotion and the distinctions among them, and similarly for the corresponding verbs. There is a natural explanation for this fact: very few verbs in the set describe unsupported gaits (these are the hopping/jumping verbs, and in part, the running verbs),so capturing similarities or differences will not reside directly in the feature of contact. The removal of "effort" has greater, albeit inessential, consequences for the similarity

plot. We observe that removing that feature has the effect of reducing the distances within the smaller clusters, while retaining the overall "bigger" similarity clusters, e.g., among the walking gait verbs, and in general, between walking and running verbs. In contrast, the removal of the feature "agility" has drastic consequences for the similarity plot and produces an altogether non-coherent clustering. An additional effect is reducing, or rather removing, the distinctions especially within the walking verbs group. We conclude that agility is an important feature in capturing similarities/differences among more specific verbs of locomotion which are below the basic level.

In conclusion, we have seen that the model is successful in identifying features relevant for the clustering of specific verbs of locomotion. In addition, we observe that the features which play a role in describing verbs of motion at the basic level, such as "contact", "speed", "effort" [5], while still relevant for the specific verbs, do not help so much in distinguishing among those verbs. It is other features, such as, most notably "agility", which are good candidates for capturing the underlying structure of the field. This result is very satisfactory, and confirms that humans resort to different sets of features in categorising the world at different levels of categorization, which differ in degree of granularity and detail (e.g., the basic level vs. the level below the basic level). This finding also aligns with recent trends in cognitive science to look at the various grain-levels of categorisation and their linguistic encoding ([17] and the papers therein).

VI. CONCLUSION

The results from the computational method employed have highlighted a number of interesting features of this kind of research. Firstly, they have underscored the validity of combining human data analyses with computational methods. In addition, they have demonstrated that computer modeling of the data can provide useful insights for the underlying semantic similarities, as well as complement or even supplement the human analysis. Data from applying this design to more languages is needed in order to assess fully its applications and use.

REFERENCES

[1] E. S. Cordingley, "Knowledge elicitation techniques for knowledge-based systems," in *Knowledge elicitation: principle, techniques and applications*. New York, NY, USA: Springer-Verlag New York, Inc., 1989, pp. 87–175.

[2] J. Geiwitz, J. Kornell, and B. P. McCloskey, "An expert system for the selection of knowledge acquisition techniques," Santa Barbara, CA: Anacapa Sciences, 1990, technical Report 785-2.

[3] D. Roberson, I. R. L. Davies, G. G. Corbett, and M. Vandervyver, "Free-sorting of colors across cultures: Are there universal grounds for grouping?" *Journal of Cognition and Culture*, vol. 5, no. 3, pp. 349–386, 2005.

[4] E. Rosch, C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Bream, "Basic objects in natural categories," *Cognitive Psychology*, vol. 8, pp. 382–439, 1976.

[5] M. Vulchanova, L. Martinez, and O. Edsberg, "A basic level category for the encoding of biological motion," in *Conceptual Spaces and the Construal of Spatial Meaning. Empirical evidence from human communication*, J. Hudson, C. Paradis, and U. Magnusson, Eds. Oxford: Oxford University Press, in press.

[6] K. Coventry, M. Vulchanova, T. Cadierno, L. Martinez, and R. Pajusalu, "Locomotion below the basic level: Sorting verbs across languages," in preparation.

[7] S. Padó and M. Lapata, "Dependency-based construction of semantic space models," *Computational Linguistics*, vol. 33, no. 2, pp. 161–199, 2007.

[8] M. Baroni, B. Murphy, E. Barbu, and M. Poesio, "Strudel: A corpus-based semantic model based on properties and types," *Cognitive Science*, vol. 34, no. 2, pp. 222–254, 2010.

[9] S. McDonald, "Environmental determinants of lexical processing effort," PhD thesis. University of Edinburgh, 2000.

[10] E. Dabrowska, "Words as constructions," in *New Directions in Cognitive Linguistics*, E. Vyvyan and S. Pourcel, Eds. Amsterdam: John Benjamins, 2009.

[11] B. Zhang and S. Srihari, "Binary vector dissimilarity measures for handwriting identification," in *Proceedings of SPIE*, vol. 5010, 2003, pp. 28–38.

[12] S. Debnath, N. Ganguly, and P. Mitra, "Feature weighting in content based recommendation system using social network analysis," in *WWW '08: Proceeding of the 17th international conference on World Wide Web*. New York, NY, USA: ACM, 2008, pp. 1041–1042.

[13] L. B. Smith, "Action alters shape categories," *Cognitive Science*, vol. 29, no. 4, pp. 665–679, 2005.

[14] K. Coventry and S. Garrod, "Saying, seeing and acting: The psychological semantics of spatial prepositions." Hove: Psychology Press, 2004.

[15] K. Abbot-Smith and M. Tomasello, "Exemplar-learning and schematization in a usage-based account of syntactic acquisition," *Linguistic Review*, vol. 23, no. 3, pp. 275–290, 2006.

[16] E. Dabrowska, "The mean lean grammar machine meets the human mind: Empirical investigations of the mental status of rules." in *Cognitive foundations of linguistic usage patterns. Empirical approaches*, H. Schmid and S. Handl, Eds. Berlin: Mouton de Gruyter., 2010.

[17] "Motion encoding in language," Oxford University press, in press.