Semantic Clusters Combined with Kinematics: The Case of English and Modern Greek Motion Verbs

Marietta Sionti¹, Leonardo Claudino², Yiannis Aloimonos², Carolyn P. Rose³, Stella Markantonatou⁴

¹University of Athens, ²University of Maryland, ³Carnegie Mellon University, ⁴Institute for Language and Speech Processing/’R.C.Athena’
msiodi@phil.uoa.gr, claudino@cs.umd.edu, yiannis@cs.umd.edu, cprose@cs.cmu.edu, marks@ilsp.athena-innovation.gr

Abstract
We combine corpus driven linguistic knowledge with experimentally obtained sensorimotor data in an effort to better specify the minimum conceptual representation of a motion event that distinguishes it from all other events, which still are largely presented in a vague and not objectively calculated quantitative method. We use American English and Modern Greek data as a case study, in order to focus on the clustering of motor actions and its correspondence to previous linguistic classifications of both languages.

1. Introduction

In the present work, we attempt to ground linguistic notions to sensorimotor data, in order to specify both linguistic and psychological terms based on embodied characteristics. Sensorimotor data are collected by measuring the performance of Modern Greek and American English native speakers.

Based on these results, we attempt to create monolingual semantic classifications of verbs drawing on sensorimotor data and classify verbal semantic and syntactic properties. Moreover, the desired outcome of this work is a perceptual grounding of abstract linguistic notions, such as directionality and transitivity, which have traditionally been used in the linguistic studies of motion verbs. Robust classification serves both as an effective monolingual and crosslingual study of motion verbs of American English and Modern Greek.

In the remainder of the paper, we describe the classifications of Modern Greek (Antonopoulou, 1987) and American English (Levin, 1993; FrameNet) motion verbs. We then discuss the verb collection criteria, the method for
the identification of the English equivalents of selected Greek motion verbs, the sensorimotor experimental procedure and its results. Implications for methodological considerations are also addressed. We conclude with our future work.

2. Motivation

2.1. Why Verb Classes?

Verb classes are well organized sets of semantically-related verbs sharing a range of linguistic properties (Levin), which frequently reflect corpus based properties. These classifications facilitate the investigation of the relations between the ‘representative element’ (e.g., walk/meander/…) of a verb class and its members, while crosslingual study of classes of verbs that are grounded on the same set of sensorimotor data are likely to reflect typological differences across languages.

2.2. Why Sensorimotor Data?

Cognitive studies have shown the relation between embodied experience and language, since image schemas are learnt as a sequence of interrelated sensorimotor patterns (Lakoff & Johnson, 1999; Arbib, 2008; Kemmerer, 2006; Fadiga et al, 2006).

2.3. Why do Theoretical Linguists Care?

Motor data of an event tend to be independent of existing divergences in its cross-linguistic representation. Consequently, a motor data representation suggests itself as a concrete and objective quantitative method for depicting cross-language lexical semantic relations. Furthermore, the distinction between “argument” and “adjunct” whose binary nature has been questioned across theories (Galen, Grenager & Manning, 2004) can probably be better understood by delineating conceptual representations of motion events in this way (both linguistic and motor).

2.4. Why do Computational Linguists Care?

NLP tasks, such as machine translation, construction of knowledge bases from text (Fahlman, 2006) and word sense disambiguation do profit from sound
3. Classification of Modern Greek Verbs

The classification of Modern Greek Motion Verbs (MGMV) (Antonopoulou, 1987) exhibits an overall semantic structure found with motion verbs of several Indo-European languages and, at the same time, present certain aspeutal idiosyncrasies. In contrast, Antonopoulou (1987) adopted prototype theory as the most suitable method for the investigation for MGMV; prototype theory is by default closer to the cognitive approach adopted here.

Antonopoulou's (1987) taxonomic sets were defined with the use of two groups of criteria. The first group comprises criteria such as transitivity, causativity, agentivity, intentionality and aspect that are non-measurable with sensorimotor methods at the moment. The second group comprises change-of-location, directionality, path, dependent motion, as well as change of orientation, manner, medium and instrumentality, whose allocation on the sensorimotor data is feasible.

4. Classification of American English Verbs

Though the difference in perspective of Levin’s (1993) English Verb Classes and Framenet’s categorization is well attested (Baker & Ruppenhofer, 2002), both these classifications are important for this work. Levin’s classes are based on semantic grouping and valence alternations. Very much like Antonopoulou, Levin offers a rich anthology of verbs enriched with syntactic information that is crucial for our long term goal, namely distinguishing between argument vs. adjunct. On the other hand, Framenet’s grouping of words-according to conceptual structures can easily be matched to Antonopoulou’s second group of criteria and, finally, to sensorimotor data.

5. Sensorimotor Experiment

5.1. Verb Collection Criteria

Twenty five predicates (Table 1) were selected from Antonopoulou’s long list of MG verbs of motion and position (1987). Both intransitive and transitive verbs
were included because we assumed zero knowledge of verbal valence. The selected verbs fulfilled requirements imposed by the nature of sensorimotor experiments and lab limitations. Each action should be performed by one human; for instance, *kalpazo* (*gallop*) is left out because it applies only to horses, while *akolutho* (*follow*) was excluded since it violated the one participant prerequisite.

We tried to identify the equivalent English verbs that would be performed in the same manner as the Greek verbs. However, we realized that the dictionary-based translation would not be the preferred one, in terms of everyday communication. English translational equivalents of Greek verbs traditionally provided by Greek-English dictionaries were either obsolete or did not align one to one with their suggested Greek equivalents. Therefore, at an early stage, we adopted the following method for identifying the English equivalents of selected Greek motion verbs: First, we videotaped the Greek participant who performed the selected set of actions. Then, we split the video into segments,
<table>
<thead>
<tr>
<th>Verb Class</th>
<th>Greek (Antonopoulou, 1987)</th>
<th>English (Levin, 1993)</th>
<th>English (FrameNet)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perpato verbs (crew can be the head of its own class)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perpato</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vimezito</td>
<td>2</td>
<td>March</td>
<td>Run</td>
</tr>
<tr>
<td>Pisopo</td>
<td>3</td>
<td>Step back</td>
<td>Run</td>
</tr>
<tr>
<td>Triklizo</td>
<td>4</td>
<td>Stagger</td>
<td>Run</td>
</tr>
<tr>
<td>Dhashkelizo</td>
<td>5</td>
<td>Stride/step over</td>
<td>Run</td>
</tr>
<tr>
<td>Biskilo</td>
<td>6</td>
<td>Crawl</td>
<td>Run</td>
</tr>
<tr>
<td>Trevo</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aneveno verbs (upward motion)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aneveno</td>
<td>8</td>
<td>Go up (step)</td>
<td>Go up (step)</td>
</tr>
<tr>
<td>Anforiso</td>
<td>9</td>
<td>Go up (ramp)</td>
<td>Go up (ramp)</td>
</tr>
<tr>
<td>Kateveno verbs (downward motion)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kateveno</td>
<td>10</td>
<td>Go down (step)</td>
<td>Go down (step)</td>
</tr>
<tr>
<td>Katiforizo</td>
<td>11</td>
<td>Go down (ramp)</td>
<td>Go down (ramp)</td>
</tr>
<tr>
<td>Hamlono (only with the body)</td>
<td>12</td>
<td>Crouch</td>
<td>Crouch</td>
</tr>
<tr>
<td>Protop verbs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picho (epi topou)</td>
<td>13</td>
<td>Jump/hop</td>
<td>Pick up and put on (lower)</td>
</tr>
<tr>
<td>Picho (apo apo)</td>
<td>14</td>
<td>Jump over</td>
<td>Pick up and put on (lower)</td>
</tr>
<tr>
<td>Picho (apo kapiou)</td>
<td>15</td>
<td>Jump down</td>
<td>Pick up and put on (lower)</td>
</tr>
<tr>
<td>Katevazo</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Katevazo verbs (downward motion)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anevaizo</td>
<td>17</td>
<td>Pick up and put on (higher)</td>
<td>Pick up and put on (higher)</td>
</tr>
<tr>
<td>Anevaizo verbs (upward motion). Though sikono can form its own class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sikono</td>
<td>18</td>
<td>Lift/raise</td>
<td>Lift/raise</td>
</tr>
<tr>
<td>Ipmono</td>
<td>19</td>
<td>Lift high</td>
<td>Lift high</td>
</tr>
<tr>
<td>Jirizo (antithetic katefthinsi)</td>
<td>20</td>
<td>Turn around</td>
<td>Turn (as verb of changing direction)</td>
</tr>
<tr>
<td>Jirizo (g giro apo karekia)</td>
<td>21</td>
<td>Circle (e.g. chair)</td>
<td>Circle (e.g. chair)</td>
</tr>
<tr>
<td>Jirizo (e.g. selida)</td>
<td>22</td>
<td>Turn (e.g. page)</td>
<td>Turn (as verb that cause to move in place)</td>
</tr>
<tr>
<td>Jirizo verbs (cause to turn)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piriastro</td>
<td>23</td>
<td>Rotate</td>
<td>Rotate</td>
</tr>
<tr>
<td>Anapodojirizo</td>
<td>24</td>
<td>Turn over</td>
<td>Turn over</td>
</tr>
<tr>
<td>Nio</td>
<td>25</td>
<td>Roll</td>
<td>Roll</td>
</tr>
</tbody>
</table>
each of which corresponded to a single action according to Greek semantics. Next, we presented the segments to ten American English native speakers each one of which was asked to name the performed action according to his/her intuition. Interestingly enough, the responses to seventeen videos were not homogeneous. In order to overcome this impasse, we designed a multiple choice questionnaire. Each question had three choices: two of them were the most frequent answers from the previous phase and the third one was selected from WordNet’s lemma. A new group of ten native speakers were shown each video segment before filling in the new questionnaire. Finally, we collected the frequent verbs for each action from both phases and we formed our English verb group (Table 1). These verbs were performed by eight American English native speakers—according to the following described experimental procedure—but still in case of ambiguity the participants of the sensorimotor experiment were asked to act according to their flair.

5.2. Experimental Procedure

5.2.1. Method

The detailed sensor data were analyzed to identify latent factors that represent stable patterns across the many dimensions of low level data. These factors appear as discrete sets (synergies) of joint angles and orientations associated with each action.

5.2.2. Equipment

A full body Moven system contains 16 inertial motion trackers. Each sensor module comprises 3D gyroscopes, 3D accelerometers and 3D magnetometers. Using advanced sensor fusion algorithms (Moven Fusion Engine) the inertial motion trackers give absolute orientation values which are used to transform the 3D linear accelerations to global coordinates which in turn give the translation of the body segments. The advanced articulated body model (23 segments and 22 joints biomechanical model) implements joint constraints to eliminate any integration drift or foot-sliding (X Moven Suit’s manual).

5.2.3. Phases

The sensorimotor experiment is divided in two phases; capturing of the main dataset for Greek—described in this presentation—and for English (8 subjects each)—to be done later.
5.2.4. Subjects

The age range is 25-30 years old, 5 men and 3 women. All subjects were native Greek and American English speakers, so they were encouraged to implement each meaning according to their intuition.

5.2.5. Action Performance

Each verb was uttered by the experimenter and then the subject performed the corresponding action. When the verb was performed only with the body of the subject, the action was limited to a floor area restricted by a quadrangle. In order to normalize the distance, subjects were encouraged to start acting at a specific corner of the quadrangle (Fig.1). Effort was made to involve as few objects as possible. Still, the subjects asked for items that could be found in the lab:

- a step (verbs 8, 10, 15),
- a ramp (9, 11),
- one or several balls (5, 14)
- table, book, cylinder, chair (22-25)
- chair (20, 21)
In order to standardize the procedure, the same objects were used throughout the experiments (whenever an object was required).

5.2.6. Analysis

In the raw data matrix, each verb is related with one row only, while each row is a set of uniformly sampled poses. For one subject, individual poses are sets of joint angles for all \( J \) joints measured by the Xsens Moven suit, with the angles being described in terms of directions \( x, y \) and \( z \):

\[
data_{\text{matrix}} = \{ \text{verb}_1; \text{verb}_2; \ldots; \text{verb}_M \}\]

where each row is:

\[
\text{verb}_v = [ \text{pose}_1, \text{pose}_2, \ldots, \text{pose}_T ]
\]

and each pose is horizontally expanded as:

\[
\text{pose}_t: [ \text{angle}_\text{joint}_1_x(t), \text{angle}_\text{joint}_1_y(t), \text{angle}_\text{joint}_1_z(t), \ldots, \text{angle}_\text{joint}_J_x(t), \text{angle}_\text{joint}_J_y(t), \text{angle}_\text{joint}_J_z(t) ]\]

considering that there are \( J \) joints (e.g. head, elbow, knee, ...) and \( 0 \leq t \leq T - 1 \).

When more subjects are considered, each verb (row) is computed as the average performance across all subjects, after length normalization. Principal Component Analysis (PCA) is computed on that matrix as in (Santello, 1998), so the two eigenvectors used for projection reflect 3-D and time information, allowing at the same time a two-dimensional visualization of the action scatter (Fig. 3, 4).

For Fig. 5-10 the matrix is (considering \( N \) performing subjects):

\[
\text{verb}_v_{\text{data\_matrix}} = \{ \text{subj}_1; \text{subj}_2; \ldots, \text{subj}_N \},
\]

each subject’s performance is \( \text{subj}_s = [ \text{pose}_1, \text{pose}_2, \ldots, \text{pose}_T ] \) and each pose is horizontally expanded as in the overall data matrix (described
above), but with 0≤t≤T′-1. Here, T′ is the number of time samples uniformly selected from each participant’s performance of verb v. T′ is not necessarily the same as the T defined in the previous matrix, which was related to average verb performances instead.

5.3. Methodological Considerations

We preferred the PCA statistical approach on the grounds of objectivity and familiarity with the linguistic methodology, as opposed to mathematical models and optimization techniques used for similar research. Moreover, the correct use of necessary objects received much consideration. We avoided the usage of several different objects because the posture of the hand, the arm and the body would change the outcome significantly, e.g. when grasping. Therefore we used the most unmarked objects and the same objects were used by all subjects. In this way we standardized the procedure without focusing to the objects used. Ideally, we would like to capture every verb with a relatively large set of the objects that are denoted by the verb’s (syntactic) dependents; this is left to future research.

6. Results

Concerning the two following types of plots for all verbs:

Fig 5, 7. displays the result of projecting each average verb in the previously defined data_matrix onto the first two eigenvectors computed by PCA on that matrix. In other words, each numbered dot corresponds to a mapping of a verb’s time sequence of poses into a single 2-D point.

Fig.6, 8 illustrate the first 10 eigenvectors computed by PCA (rows) in terms of the weights it assigns to each joint-direction-time. Each small rectangle corresponds to a specific joint-direction (e.g., head-x), each column in that rectangle refers to a t-time instant, 0≤t≤T-1. If a column is bright at time t within the head-x rectangle, it means that there is something that the testing subjects do with the head at that direction at time t, which varies significantly among the testers and thus is picked up by PCA. Following that reasoning, by inspecting sets of joint-directions highlighted by the main principal components, one can draw a better understanding of why the scatter at Fig. 5 and 7 appear in this form.

The two main visual groups of actions are:

• the rectangular: it includes walking like actions (leg-related)
• the blue ellipsis: it includes manipulation of an object (arm-related) (Fig.5, 7). It is projected on both PC1 (distinct use of hips) and PC2 (emphases on knees and shoulders) (Fig. 4). Therefore, all these actions have approximately the same profile in terms of joints-angles.
Actions 6 (busulo-crawl) and 7 (trexo-run) are relatively apart from the leg-related group. However, they appear to be virtually projected on the same coordinate on Principal Component 1 (PC1), the one that relates to hips x (displacement on the sagittal plane). At the same time they are distant when they are projected on PC2, which involves arms as well. This is only natural because when a subject crawls, it moves its arms, hips and legs more than in actions of the green ellipsis. Action 7 (trexo-run), on the other hand, seems to be closer to the rectangular group because the displacement of the above mentioned parts encloses more energy than the rest leg-related actions but still less than action 6 (busulo-crawl). Furthermore, the fact that run-trexo and crawl-busulo are projected on different PC2 narrows down to (i) the height of the knees (upward-downward and forward-backward respectively) and (ii) the frequent movement

---

**Figure 5.** Average Actions Greek Verbs (the numbers are explained in Table1)

---

**Figure 6.** Weights of each joint-direction-time feature according to the 10-first principal components, including time, for the Greek language.
of shoulders in all directions. Action 7 differs from the rest in terms of velocity and this is incorporated in the depiction of Time in the plot.

The visualization presented here supports the results of brain imaging studies; the schematic of the distributed semantic representation in the brain of action verb processing is based on the body parts performing them (Wermter et al., 2005), e.g. arm-related and leg-related.

Comparing this visualization with linguistic classifications (Antonopoulou, 1987; Levin, 1993; FrameNet) we see similarities. Antonopoulou’s classes are more fine-grained than the two big categories in Fig.5; here two of the Antonopoulou’s classes appear in the same scheme – rectangular or ellipsis. In particular, the members of the walk-group, *pido*-group and *kaitevno*-group are coiled together (green ellipsis). The same occurs for *jirizo2*-group, *kaitevazo* and

![Figure 7. Average Actions English Verbs (the numbers are explained in Table 1)](image)

![Figure 8. Weights of each joint-direction-time feature according to the 10-first principal components, including time, for the English language](image)
Of particular interest is the fact that the actions 13, 14, and 15 (\textit{pidho-jump}) share the same morphological representation \textit{pidho} but not the same performance; eventually they form one tight ‘group’.

Actions 20, 21 (\textit{jirizo1}) and 22, 23, 24, 25 (\textit{jirizo2}) also share the same morphological representation \textit{jirizo}. These actions present themselves close to each other on the PC1 projection but are separated on the PC2 projection. The first group is mainly about leg motion and the second one about object manipulation that necessarily involves arm displacement as is clearly indicated on PC2. In English all \textit{roll} verbs belong to the same class.

Though the plots were based on similar or even identical actions, certain divergences occurred. For instance, the verb \textit{dhraskelizo} has traditionally been translated as \textit{stride}, since both share longer steps. The Greek subjects always needed a small obstacle, such as a ball or a hole on the ground, to perform longer steps. On the contrary, the English subjects clearly distinguished between \textit{stride} and \textit{step over} (although WordNet assigns this meaning to \textit{stride} as well). Similarly, we would expect \textit{vimatizo} to be closer to \textit{pace} rather than \textit{march}, but it should be noted that the majority of the English participants were unsure for the exact representation of \textit{pace}. When we compare the signals of \textit{vimatizo} and \textit{march}, we see significant similarity in the manner of stepping, while English subjects also emphasize the movement of the arms.

Of particular interest in the Greek plot is the distance of \textit{trexo-run} and \textit{busulo-crawl} from the rest leg-related actions (especially for English, \textit{march} and \textit{crawl} show the same behavior). Although we would expect \textit{run} to resemble \textit{trexo}, differences occurred due to two reasons: (i) the English subjects tented to use their arms less than the Greek ones, and (ii) each group performed \textit{march} in a different way; this time, the English subjects used their arms more than the Greek ones (as opposed to the performance of \textit{run}). The blue and green ellipsis are projected both on the same PC1 (highlighted hips) and 2 (highlighted knees and shoulders).

Furthermore, the reason why \textit{march, trexo} and \textit{crawl-busulo} are projected on different PC2 narrows down to (i) the height of the knees (upward-downward and forward-backward respectively) and, (ii) the frequent movement of shoulders in all directions. But still, these verbs are projected on the same PC1. PC1 is about the forward and backward displacement of the hips and emphasizes on the leg related actions, namely the walking like actions, which, in turn, is considered to be the actions’ common linguistic characteristic.

The above findings are still consistent with the aforementioned linguistic analysis. The Greek \textit{trexo} can form its own class in terms of velocity according to (Antonopoulou, 1983). At the same time, in Fig.5, 7, action 7 differs from all the other actions performed because of velocity. This fact is incorporated in the
depiction of time in Fig. 6. Probably Levin (1993) gives us a hint that these verbs need special treatment, since she enlists them under both the meander verbs and the run class (it must be kept in mind that Levin’s classification has taken into account the criterion of intentionality; however, intentionality is still not measurable with sensorimotor techniques).

7. Conclusions and Current Directions

In this paper we have presented work to date on corresponding linguistics terms -extracted from classifications- to sensorimotor clusters. Specifically, we explored their differences and resemblances based on totally unsupervised methods. We demonstrated the difficulties in aligning Greek and English motion verbs, while our future work aims to extend this approach. The analysis of each motion verb and the combination of detailed linguistic and motor entries that would appear as the verb’s “identification card” is feasible and helpful for sound semantic classifications.

Acknowledgement

We are grateful to reviewers and audience for their feedback at the 20th International Symposium on Theoretical and Applied Linguistics in Thessaloniki (2011), the Embodied and Situated Language Processing Workshop in San Diego (2010), and the Interdisciplinary Workshop on Verbs, the Identification and Representation of Verb Features in Pisa (2010). Moreover, we would like to thank Nathan Schneider, Lori Levin and Ben Lambert for their encouragement and feedback. This research was partially supported by the Greek Scholarships Foundation (IKY).
References


Framenet http://framenet.icsi.berkeley.edu/


Major Trends in Theoretical and Applied Linguistics