# Motion Capture of Modern Greek Verbs: Measuring aspects and relations among actions

Panagiotis Kouris*†, Marietta Sionti*, Chrysovalantis Korfitis*, Stella Markantonatou*
**Institute for Language and Speech Processing, Athena RIC*, Greece
†*School of Electrical & Computer Engineering, National Technical University of Athens*, Greece
pkouris@islab.ntua.gr, {marietsionti, ckorfitis}@gmail.com, marks@athenarc.gr

***Abstract*—In the context of human motion analysis and human-centered computational sensing, this work presents a methodology for the investigation of the relations among actions of a set of (Modern Greek) motion verbs. The actions denoted by these verbs fall in the motion categories of pushing, pulling, hitting, and beating. Motion data were collected with motion capture technology, and measures of correlation and distance are used to identify existing relations among actions. Finally, hierarchical clustering analysis was applied to identify groups of actions. The results are in line with a semantic categorization of the corresponding verbs. The overall experimental procedure and data analysis indicate that the employed methodology could be useful in promising applications of motion recognition or motion clustering, aiming at the identification of related captured actions.**

***Index Terms*—MoCap, Motion verbs, Human motion analysis, Hierarchical clustering analysis**

## I. Introduction

Recently, human motion analysis and event segmentation have attracted the interest of computer vision engineers, computational linguists, and psychologists [1]–[7]: engineers attempt to match language with videos in order to assist activity recognition while linguists and psychologists anticipate that the visual data will offer crucial information to linguistic analysis and embodiment. The driving force for this extensive engagement is useful and promising applications such as security surveillance environments, healthcare systems, human performance analysis in sports activities, human-computer interfaces, educational systems.

We report on work aiming at the identification of relations among the actions denoted by a set of motion verbs of Modern Greek, namely actions of pushing, pulling, hitting, and beating. We used motion capture technology to collect a dataset of motion data for the actions denoted by 20 motion verbs. After data preprocessing, we use two approaches to specify the relations among the motions denoted by the selected 20 verbs; correlation and distance measures. Both these metrics return analogous results and validate the existence of related actions. Next, hierarchical clustering based on correlation and hierarchical clustering based on distance are applied to identify consistent groups of motions. Since the presented methodology can identify clusters of actions (denoted by semantically related verbs), it could be useful to applications that associate new observations (i.e., captured motions) with relevant groups of actions and contribute to the identification of actions drawing on segmented or partial data.

The novelty of the presented framework lies in the methodology for motion analysis of Modern Greek verbs of particular conceptual categories, as this constitutes a study that has not been discussed in the relevant literature. In this context, this work utilizes motion capture technology and methodology of data analysis to group motions or distinguish among them, giving advance to the field of human-centered computational sensing.

The rest of this paper is organized as follows: Section II offers a brief overview of the related work. In Section III are discussed the studied motion verbs, the task of motion capturing, and the preprocessing of the data. Section IV describes in detail the analysis process up to clustering, while Section V discusses the obtained results. In Section VI, some final remarks are made, and future work directions are discussed.

## II. Related work

In the context of human motion analysis, high level applications of human action recognition require low level technology such as feature extraction and representation, mainly based on single person activity recognition [1], [5]. These features may vary depending on the approach. In the case of the generic model recovery, a 3D model is required, while the motion-based model exclusively utilizes motion characteristics. The appearance–based model is a 2D shape model derived from images and videos [4]. In this line of research, the majority of researchers who attempt to detect activities work on three dominant motion features, namely space, time and frequency [5]. Motion verbs, in particular, have received considerable attention in studies that try to ground language in sensorimotor data because their

This research is co-financed by Greece and the European Union (European Social Fund-ESF) through the Operational Programme “Human Resources Development, Education and Lifelong Learning 2014-2020” in the context of the project “The sensorimotor basis of causality and aspect and their representation with the verbs of pushing, pulling, hitting and beating of Modern Greek” (MIS 5006565).

denotation is grounded in action rather than in mental or emotional situations (e.g., for visual data see in [2], for kinematic data see in [3], [8], [9]).

## III. Motion capture data

### A. Modern Greek Motion Verbs

This research focuses on capturing and analyzing actions of pushing, pulling, hitting and beating that are denoted by a set of Modern Greek verbs (see Table I).

In particular, we chose these verbs because we have already studied motion verbs such as *walk*, *jump*, and *march* that only involve one entity, namely the moving one [8]. In this study, we turned to movements that involved (a) more than one participant, that is, an actor and an entity affected by the action, respectively realized as a subject and an object in Modern Greek, and (b) the hands of the actor. Furthermore, we decided to work with actions that require a forceful movement of the actor rather than a slight one. To this end, we chose the aforementioned four general categories, namely "hitting", "pushing", "pulling", and "beating". We used the acclaimed Modern Greek lexicon *Onomastikon* [10] that is conceptually organized to find verb predicates in the particular semantic fields. Of them, we chose those verbs that could be easier represented in our experimental setup.

More specifically, in *Onomastikon*, lemmas of the same Part of Speech are organized in sets of near synonyms; the sets, in turn, are organized in groups representing a "concept" (e.g., the concept of beating). The verbs of pushing, pulling, hitting and beating (e.g., σπρώχνω 'push', κλωτσώ 'kick') have been preferred for this study because they denote a "countable" motion of the human being that performs the corresponding action and is normally denoted by the subject of the active verb. Any physical body affected by the action is typically denoted with the direct object of the active verb. Furthermore, these verbs denote actions that are close to walking-like actions (e.g., walk, run), which have been a popular research topic in Modern Greek [3], [8].

The English translation of the employed verbs (Table I) is used as the label of verb or action in the rest of the paper.

### B. Motion capturing

Twenty predicates of pushing, pulling, hitting and beating (Table I) were selected from *Onomastikon* [10]. The selected verbs fulfilled requirements imposed by the nature of sensorimotor experiments and lab limitations. More specifically:

- The verb is used in spoken Modern Greek with a literal meaning (some of these verbs belong to older versions of Greek and are most often used with a metaphorical meaning on within fixed expressions).
- Each action is performed by one human.

Table I
Verbs of pushing, pulling, hitting and beating of Modern Greek

| N | Verb | English translation |
|---|---|---|
| 1 | σκουντώ (skoundo) | Prod |
| 2 | ωθώ (otho) | Push |
| 3 | σπρώχνω (sprochno) | Shove |
| 4 | στριμώχνω (strimoxno) | Cram/Squeeze |
| 5 | κλωτσώ (klotso) | Kick |
| 6 | αποκρούω (apokrouo) | Puch back |
| 7 | ποδοπατώ (podopato) | Trample |
| 8 | τσαλαπατώ (tsalapato) | Trample (clumsier) |
| 9 | χτυπώ έμψυχο (xtipo empsixo) | Hit (animate) |
| 10 | χτυπώ πόρτα (xtipo porta) | Hit (door) |
| 11 | σφυροκοπώ (sfirokopo) | Hammer |
| 12 | κρούω (krouo) | Knock/Ring |
| 13 | γρονθοκοπώ (gronthokopo) | Punch |
| 14 | δέρνω (derno) | Beat |
| 15 | καρπαζώνω (karpazono) | Slap (on cervix) |
| 16 | χαστουκίζω (chastoukizo) | Slap (on cheek) |
| 17 | κοπανώ (kopano) | Pound |
| 18 | τσουγκρίζω (tsougrizo) | Clink |
| 19 | μαχαιρώνω (maxerono) | Stab |
| 20 | μαστιγώνω (mastigono) | Whip |

- Reciprocal or medio-passive verbs (e.g., χτυπιέμαι 'to beat myself') were excluded.

**Equipment:** We used a full body Synertial IGS-C420 system[1] containing 18 inertial motion trackers. Figure 1 illustrates the Synertial mocap suit and Figure 2 depicts the employed skeleton with 18 bones of motion capturing. Each sensor module comprises 3D gyroscopes, 3D accelerometers and 3D magnetometers. The inertial motion trackers give absolute orientation values that are used to transform the 3D linear accelerations to global coordinates, which in turn give the translation of the body segments [11]. The mocap software is the animate v1.3.

Figure 1. Synertial mocap suit

**Participants:** All participants (N=12 participants) were native Greek speakers and were encouraged to act intuitively. The age range of the participants is 20 - 60 years. It should be noted that the collected captured data

[1]Synertial mocap suit: https://www.synertial.com/

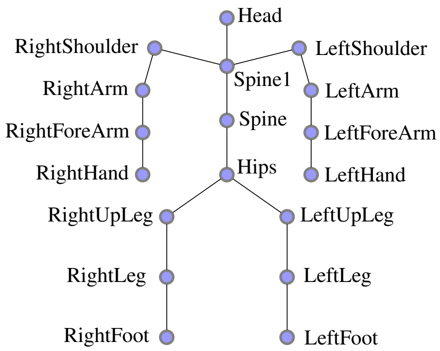


Figure 2. Skeleton of motion capturing

require a significant amount of post-processing effort of high computational and time cost. This is a deterrent factor for evolving a large number of participants in such experiments. For example, two prominent motion capture databases are the CMU motion capture database of the Carnegie Mellon University[2], which contains data from 14 participants each one performing approximately 10 - 12 actions, and the HDM05 database of the Hochschule der Medien in Stuttgart[3], which contains data about 70 motions performed by 5 actors.

**Action Performance:** The experimenter uttered each verb and the participant performed the denoted action. All the performed verbs are transitive. For standardization reasons, we used a minimal number of objects, the same for all the experiments.These objects were: a dummy doll (verbs of table I: *Prod, Push, Shove, Cram/Squeeze, Hit animate, Punch, Beat, Slap on cervix, Slap on cheek, Stab, Whip*), one or more balls (*Kick, Puch back*), paper (*Trample, Trample clumsier*), a door (*Hit door*), a hammer (*Hit animate, Hammer*), a small bell (*Knock/Ring*), a glass (*Clink*), a plastic knife (*Stab*), a rope (*Whip*), and a table (*Pound*).

### C. Data preprocessing

Motion capturing produced a BioVision Hierarchical (BVH) file [11] for each action; these raw data have to be transformed in order to allow for measurements. We followed three approaches of transforming motion capture data:

i. *Rotation data*: The relevant Euler angles of the x, y, and z-axis for each bone of the skeleton, as it is a typical practice in related work [6], [12], [13].

ii. *Positive rotation data*: similar to [12], this approach assumes that each $angle$ of the rotation data is transformed to a positive angle $p\_angle$ around an equilibrium angle $e$, as described in the Equation 1 (e.g., in *positive rotation data*, the movement of the head to the right is assumed to be equivalent to the movement to the left). In the experimental procedure, for each rotation, we assume that the equilibrium angle is equal to zero ($e = 0$).

$$p\_angle = \begin{cases} e - angle, & \text{if } angle < e. \\ angle - e, & \text{otherwise.} \end{cases} \tag{1}$$

iii. *Position data*: The BVH data were transformed into positions of spatial coordinates. In this transformation, for each bone of the skeleton, we transform its relevant Euler rotation angles (i.e., angles of x, y, and z-axis of each bone) to absolute coordinates. The absolute position of each bone is computed with respect to the absolute coordinates of the root bone, which is captured by the equipment, and the overall hierarchical structure of the skeleton, as it is described in [11].

To extract features from the data [7], [13], we used the point estimations of the minimum, maximum, mean and standard deviation of each movement or rotation (i.e., the point estimations of time-series for each bone of x, y, and z-axis). These point estimations are the “features” of each bone movement.

Finally, the data were standardized (i.e., $mean = 0$ and $standard\ deviation = 1$) because our research focuses on clustering analysis that presupposes data standardization in order to improve clustering performance [14], [15].

### D. Data overview

Preprocessing and feature extraction result in a dataset, one per each of the three data approaches described in Section III-C, consisting of 240 usage examples (12 participants x 20 verbs). Each usage example comprises 217 features (e.i., 1 for time + 18 bones x 3 for x, y, z-axis x 4 for the minimum, maximum, mean and standard deviation of each bone movement).

## IV. Investigating the data of motion verbs

Aiming at the identification of clusters of verbs that denote similar actions, this research employs clustering analysis based on the correlation and the distance among them.

### A. Correlation of actions

We first measured the correlation among the actions denoted by the 20 verbs by specifying the degree of association among motions. To measure correlation, we have experimented with two methods; Pearson’s correlation and Spearman’s rank correlation coefficient [16], [17]. Spearman’s correlation is most appropriate for our dataset of non linear data because it examines the association between variables in terms of monotony. The Spearman’s correlation metric is applied to each one of the three data approaches; *rotation*, *positive rotation*, and *position data*. The obtained correlation matrices represent the degree of association among the 20 used motion verbs. In particular, the correlation $c_{ij}$ between

[2]CMU graphics lab mocap database: http://mocap.cs.cmu.edu/

[3]Mocap database HDM05: www.mpi-inf.mpg.de/resources/HDM05

the motions denoted by a verb $i$ and a verb $j$, which are represented by the corresponding vectors $v_i$ and $v_j$, is given by the Equation 2.

$$c_{ij} = \begin{cases} 1.0, & \text{if } i = j. \\ correlation(v_i, v_j), & \text{otherwise.} \end{cases} \quad (2)$$

### B. Distance of actions

To compute the distance among the actions denoted by the verbs, we considered the following metrics: Euclidean distance, Cosine distance, and Earth mover's distance [18], [19]. Since these metrics returned similar results, especially the cosine and the Euclidean distance, we use the cosine distance as a representative measure of our dataset.

Cosine distance was applied to the three data approaches; *rotation*, *positive rotation*, and *position data* yielding the corresponding distance matrices. More specifically, the distance $d_{ij}$ between the motion $v_i$ denoted by a verb $i$ and the motion $v_j$ denoted by a verb $j$ is given by the Equation 3.

$$d_{ij} = \begin{cases} 0.0, & \text{if } i = j. \\ distance(v_i, v_j), & \text{otherwise.} \end{cases} \quad (3)$$

### C. Clustering of actions

Hierarchical cluster analysis [20] is used to form clusters of similar actions (denoted by the studied verbs). The method of agglomerative clustering [21] is employed; with this bottom-up approach, each observation starts in its own cluster and then the clusters are successively merged until all of them have been merged into a single one.

We utilize the methodology of hierarchical clustering because it enables us to examine and study the clusters in a hierarchical order, moving from small groups to large ones.

Since the method of agglomerative clustering requires a distance matrix, we use two approaches;

i. *Distance-based on correlation*: The correlation matrix (which is described above in Section IV-A) is converted to distance matrix by transforming each element of the correlation matrix $c_{ij}$ to a distance one $dc_{ij}$ as it is described in Equation 4.

$$dc_{ij} = 1.0 - c_{ij} \quad (4)$$

ii. *Cosine distance*: The distance matrix, as mentioned above in Section IV-B, is used in agglomerative clustering.

For the linkage criterion of hierarchical clustering, which determines the distance between groups of observations, we used the method of average linkage, where the distance between two clusters is defined as the average distance between the observations in the clusters. This method of average linkage is preferred because it produces more compact clusters than other approaches in the case of our dataset.

## V. Results and discussion

### A. Correlation and distance

Figure 3 shows the correlation among the actions denoted by the verbs for the *rotation data*, while Figure 4 illustrates the cosine distances among the actions. The dark areas in these figures correspond to high correlation or short distance, while the lightboxes represent pairs of verbs with low correlation or big distance, respectively. It can be seen that Spearman's correlation and cosine distance lead to comparable results verifying the associations that exist in the data.

In particular, the actions that show very strong pairwise correlation are intuitively consistent with the usage and the semantics of these verbs in Modern Greek (e.g., the motion of the verbs *Prod, Push* and *Shove*, which have a similar meaning, are strongly correlated). This may be an indication that the actions of the semantically correlated verbs are also correlated. We draw the same conclusions about distance.

For both *positive rotation* and *position data*, the obtained results show very strong correlations. For all actions in pairwise, Spearman's correlation value is greater than 0.5 for *positive rotation data* and greater than 0.6 for *position data*. Thus, all pairs of actions are strongly correlated, especially in the case of *position data*. Therefore, both versions of *positive rotation* and *position data* fail to sufficiently discriminate among actions. These strong correlations may be due to the fact that the *positive rotation data* ignores the movement direction of each part of the body. Also, in the case of *position data*, we have a very strong correlation for each pair of actions because the whole body moves in the same direction for each studied action, in terms of spatial coordinates. Since the *positive rotation* and *position data* show very strong correlations, and they do not include any significant information for the relation or the discrimination of the actions, they are not taken into account. In contrast to *positive rotation* and *position data*, *rotation data* capture sufficiently the discrimination and the relation between actions; the matrices of correlation and distance are shown in Figures 3 and 4, respectively.

### B. Hierarchical clustering

Figure 5 illustrates the dendrogram of hierarchical clustering that is based on the *correlation matrix*, while Figure 6 shows the dendrogram of clustering that is based on *cosine distance*. Considering the small groups with no more than five actions, both clustering approaches produce the same groups. Clusters with cardinality larger than 5 differ slightly, especially in the case of actions {*4.Cram/Squeeze*} and {*9.Hit (animate), 10.Hit (door)*}, which are grouped in different clusters by the two approaches.

The fact that the two approaches of clustering provide the same groups of actions in the case of small groups

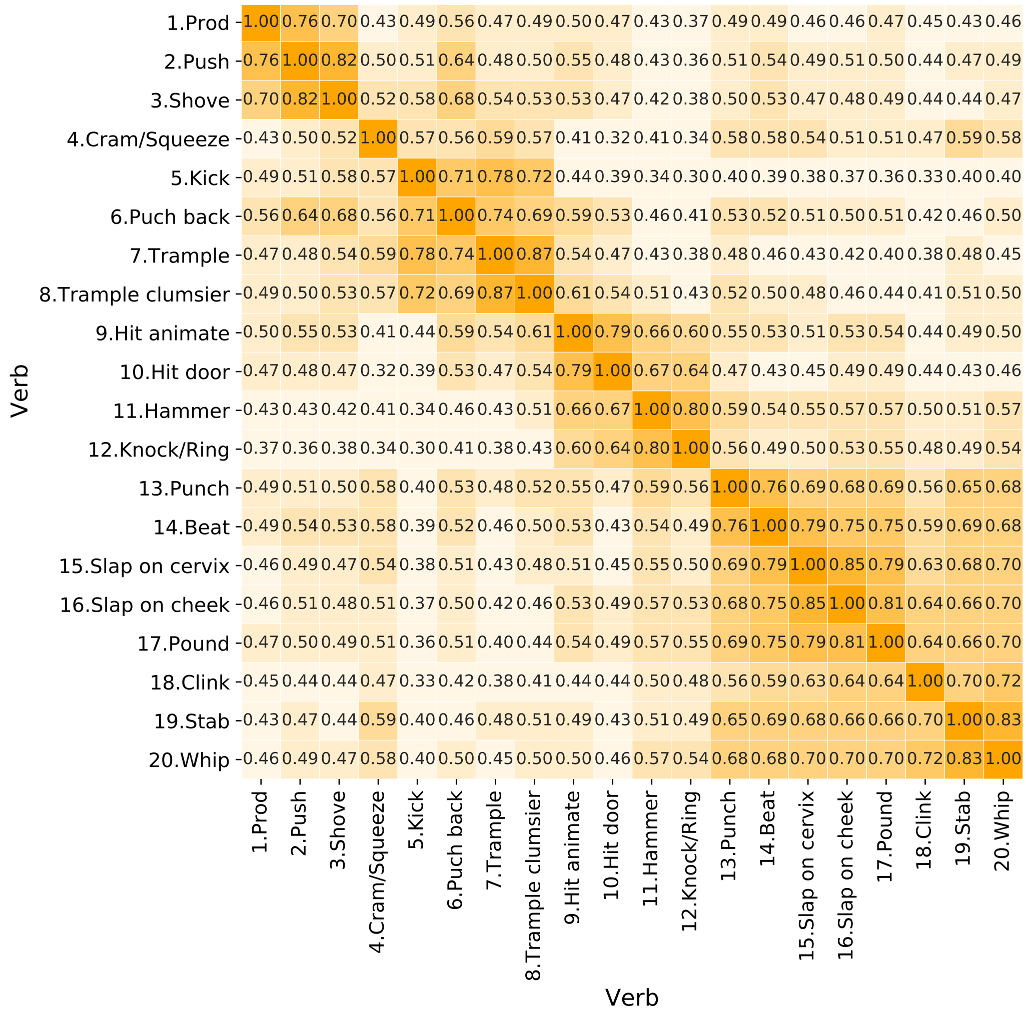


Figure 3. Spearman's rank correlation coefficient matrix ($p-value < 0.001$ for each pair of actions).

(i.e., no more than five actions per group) confirms that these clusters are valid and compact; their respective actions have similar characteristics. If we use the intersection of two sets of clusters of the corresponding two clustering approaches (*clustering based on correlation* and *clustering based on distance*) as a criterion of determining the optimal clusters $C$, we designate that the optimal clusters would be those that are specified as follows:

$$C = clustering(correlation) \cap clustering(distance) = \left\{ \begin{array}{c} \{1.Prod,\ 2.Push,\ 3.Shove\}, \\ \{4.Cram/Squeeze\}, \\ \{5.Kick,\ 6.Puch\ back,\ 7.Trample, \\ 8.Trample\ clumsier\}, \\ \{9.Hit\ animate,\ 10.Hit\ door\}, \\ \{11.Hammer,\ 12.Knock/Ring\}, \\ \{13.Punch,\ 14.Beat,\ 15.Slap\ on\ cervix, \\ 16.Slap\ on\ cheek,\ 17.Pound\}, \\ \{18.Clink,\ 19.Stab,\ 20.Whip\} \end{array} \right\}$$

With respect to the *rotation data* and the produced clusters of the two clustering approaches.

Therefore, the optimal number of clusters in our dataset is equal to 7, where the first group {*1.Prod, 2.Push, 3.Shove*} contains actions of pushing while action {*4.Cram/Squeeze*} is grouped on its own. Also, the actions of the set {*5.Kick, 6.Puch back, 7.Trample, 8.Trample clumsier*} are grouped in the same cluster as three of them are performed with the feet. Moreover, actions {*9.Hit animate, 10.Hit door*} correspond to the same verb of hitting and beating, while the group {*11.Hammer, 12.Knock/Ring*} contains close motions. Furthermore, the set {*13.Punch, 14.Beat, 15.Slap on cervix, 16.Slap on cheek, 17.Pound*} includes actions of beating, and the final group {*18.Clink, 19.Stab, 20.Whip*} contains actions that are performed with the hand.

The proposed unsupervised clustering analysis of *rotation data* provides a valid and effective grouping of actions, which supports linguistic explanation from the point of embodiment. Tight clusters {*5.Kick, 6.Puch back, 7.Trample, 8.Trample clumsier*}, {*1.Prod, 2.Push, 3.Shove*} and {*11.Hammer, 12.Knock/Ring*} correspond to manner verbs, verbs that encode the manner of the action and provide the path related information with

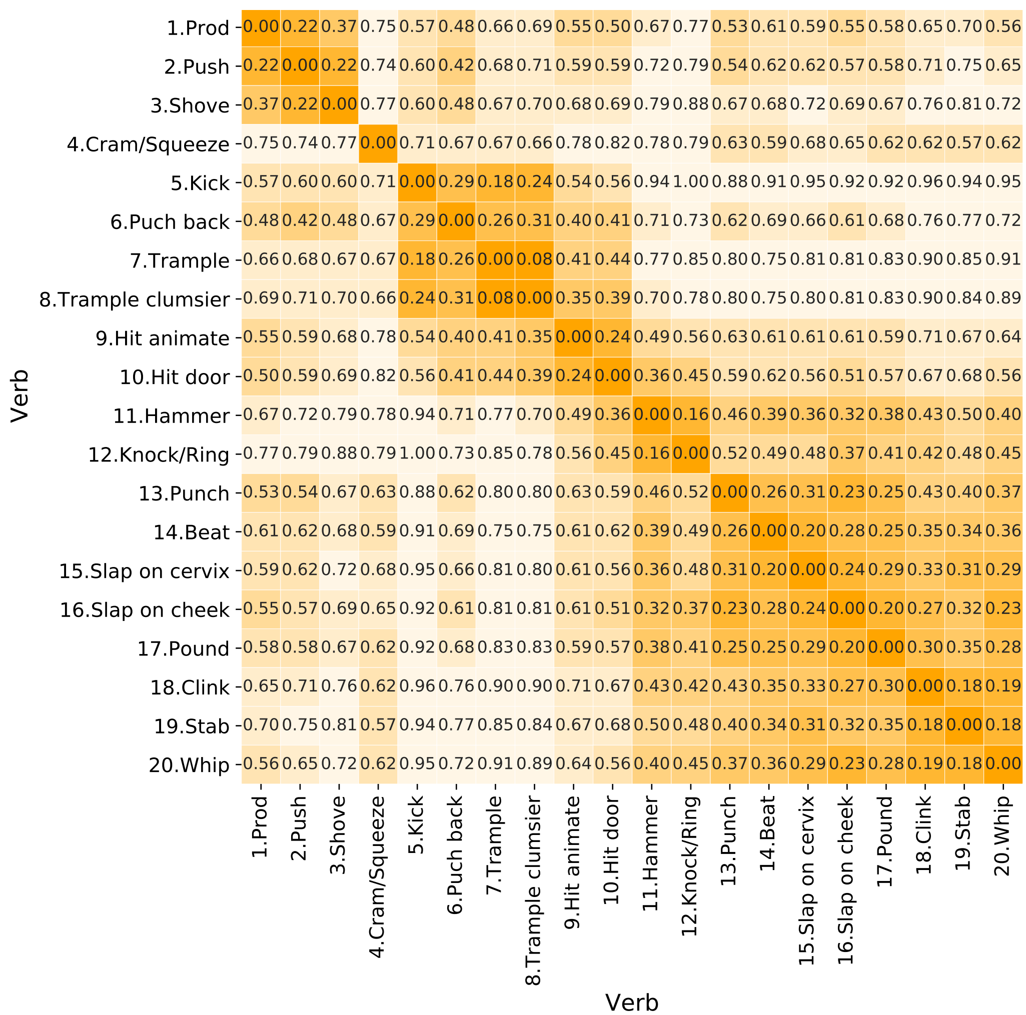


Figure 4. Cosine distance matrix

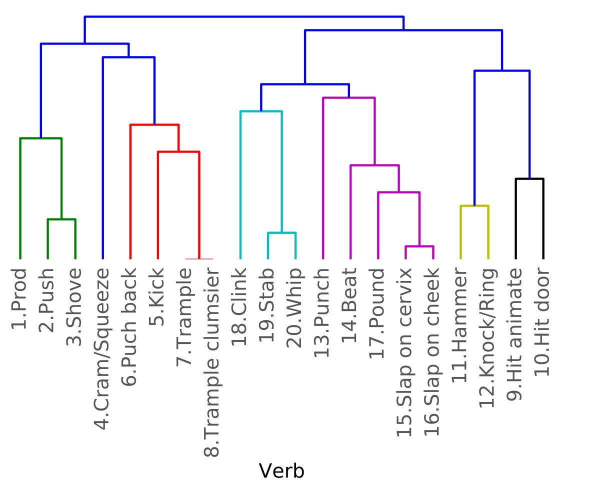


Figure 5. Hierarchical clustering based on Spearman's rank correlation coefficient.

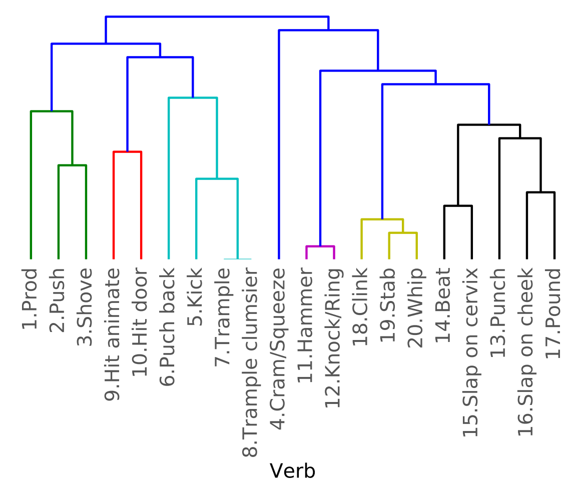


Figure 6. Hierarchical clustering based on cosine distance.

adjuncts (e.g., I fly to America). Given the diglossia phenomena of Greek, there are many pairs of near synonyms with one colloquial and one formal member, the latter related to older or ancient versions of Greek.

Such is the pair {*2.Push, 3.Shove*}. The actions could be equally performed with the hands or the feet, the shoulders or the back. These are transitive verbs that denote the intention of the agent but they do not entail that the patient moves eventually.

Moreover, the members of the pair {*7.Trample, 8.Trample clumsier*} differ in style, the second predicate being colloquial and probably more emphatic than the first one. Verb {*5.Kick*} is not a (near) synonym but shares with the pair {*7.Trample, 8.Trample clumsier*} the property of performing the action with legs. As opposed to the previous pair, these predicates are transitive verbs that focus on the way the agent moves. Finally, {*11.Hammer, 12.Knock/Ring*} comprises verb predicates that are hardly used literally in Modern Greek while, especially the first one, is widely used metaphorically (e.g., bombing, the effect of hard rain). In their literal sense, they both denote actions performed with the hands and describe the way the action is performed.

Although this research is monolingual, it can be extended to other languages with verb predicates with the same semantics as the studied Modern Greek ones.

## VI. Conclusion and Future Work

We have used twenty Modern Greek verbs of pushing, pulling, hitting, and beating to capture and analyze the actions denoted by them and investigate the relations among these actions. The analysis assumes tree versions of the data; *rotation*, *positive rotation*, and *position data*. Of them, the last two fail to highlight the relation and discrimination among the motions. To investigate the relations among the actions, we measured the correlation and the distance among the actions drawing on rotation data. The obtained results demonstrate sets of actions with similar characteristics. Therefore, the methodology of hierarchical clustering adopted in this research leads to intuitively consistent groups that are in line with the semantics of the verbs denoting the studied actions.

The employed clustering analysis could be useful to practical applications that identify captured motions. The same methodology could be extended to classify motions or textual data to predefined classes. This extension, which constitutes our plan for future work, may aim at the investigation of the correlation between the motions denoted by verb predicates and the linguistic properties of these predicates (such as their argument structure or their aspectual properties). Additionally, to measure this association between actions and linguistics, an appropriate methodology should be developed, leading to reliable results. This research would be useful in the fields of human motion analysis or human-centered computational sensing.

## References


[1] X. Yu, C. L. Teo, Y. Yang, C. Fermüller, and Y. Aloimonos, "Action attribute detection from sports videos with contextual constraints." in *BMVC*, 2013.

[2] K. Pastra and Y. Aloimonos, "The minimalist grammar of action," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1585, pp. 103–117, 2012.

[3] M. Sionti, T. Schack, and Y. Aloimonos, "The language of motion mocap ontology," in *Advances in Computer Vision*, K. Arai and S. Kapoor, Eds. Cham: Springer International Publishing, 2019, pp. 710–723.

[4] P. C. Ribeiro, J. Santos-Victor, and P. Lisboa, "Human activity recognition from video: modeling, feature selection and classification architecture," in *Proceedings of International Workshop on Human Activity Recognition and Modelling*. Citeseer, 2005, pp. 61–78.

[5] S.-R. Ke, H. L. U. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, and K.-H. Choi, "A review on video-based human activity recognition," *computers*, vol. 2, no. 2, pp. 88–131, 2013.

[6] E. Protopapadakis, A. Voulodimos, A. Doulamis, S. Camarinopoulos, N. Doulamis, and G. Miaoulis, "Dance pose identification from motion capture data: a comparison of classifiers," *Technologies*, vol. 6, no. 1, p. 31, 2018.

[7] S. Wawrzyniak and W. Niemiro, "Clustering approach to the problem of human activity recognition using motion data," in *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2015, pp. 411–416.

[8] M. Sionti, L. Claudino, Y. Aloimonos, C. P. Rose, and S. Markantonatou, "Semantic clusters combined with kinematics: The case of english and modern greek motion verbs," in *Major Trends in Theoretical and Applied Linguistics 1*. Sciendo Migration, 2014, pp. 495–510.

[9] M. Sionti, T. Schack, and Y. Aloimonos, "An embodied tutoring system for literal vs. metaphorical concepts," *Frontiers in psychology*, vol. 9, 2018.

[10] T. Vostantzoglou, *Antilexicon i Onomastikon tis Neas Ellinikis Glossis*. Athens, 1962.

[11] M. Meredith, S. Maddock *et al.*, "Motion capture file formats explained," *Department of Computer Science, University of Sheffield*, vol. 211, pp. 241–244, 2001.

[12] A. Kleinsmith and N. Bianchi-Berthouze, "Form as a cue in the automatic recognition of non-acted affective body expressions," in *International Conference on Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 155–164.

[13] A. Ball, D. Rye, F. Ramos, and M. Velonaki, "Unsupervised clustering of people from 'skeleton'data," in *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2012, pp. 225–226.

[14] P. Trebuňa, J. Halčinová, M. Fil'o, and J. Markovič, "The importance of normalization and standardization in the process of clustering," in *2014 IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE, 2014, pp. 381–385.

[15] G. W. Milligan and M. C. Cooper, "A study of standardization of variables in cluster analysis," *Journal of classification*, vol. 5, no. 2, pp. 181–204, 1988.

[16] J. H. Zar, "Spearman rank correlation," *Encyclopedia of Biostatistics*, vol. 7, 2005.

[17] A. Rebekić, Z. Lončarić, S. Petrović, and S. Marić, "Pearson's or spearman's correlation coefficient-which one to use?" *Poljoprivreda (Osijek)*, vol. 21, no. 2, pp. 47–54, 2015.

[18] A. S. Shirkhorshidi, S. Aghabozorgi, and T. Y. Wah, "A comparison study on similarity and dissimilarity measures in clustering continuous data," *PloS one*, vol. 10, no. 12, p. e0144059, 2015.

[19] A. Andoni, P. Indyk, and R. Krauthgamer, "Earth mover distance over high-dimensional spaces," in *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2008, pp. 343–352.

[20] D. Xu and Y. Tian, "A comprehensive survey of clustering algorithms," *Annals of Data Science*, vol. 2, no. 2, pp. 165–193, 2015.

[21] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, 2012.