Smart Gesture Selection with Word Embeddings Applied to NAO Robot

Mario Almagro-Cádiz[1], Víctor Fresno[1], and Félix de la Paz López[2](✉)

[1] Departamento de Lenguajes y Sistemas Informáticos,
Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain
{malmagro,vfresno}@lsi.uned.es
[2] Departamento de Inteligencia Artificial,
Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain
delapaz@dia.uned.es

Abstract. Nowadays, Human-Robot Interaction (HRI) field is growing by the day, a fact which is evidenced by the increasing number of existing projects as well as the application of increasingly advanced techniques from different areas of knowledge and multi-disciplinary approaches. In a future where technology automatically controls services such as health care, pedagogy or construction, social interfaces would be one of the necessary pillars of HRI field. In this context, gesture plays an important role in the transmission of information and is one of fundamental mechanisms relevant to human-robot interaction. This work proposes a new methodology for gestural annotation in free text through a semantic similarity analysis using distributed representations based on word embeddings. The intention with this is to endow NAO robot with an intelligent mechanism for gesture allocation.

Keywords: Word embeddings · Co-verbal gesture · HRI · NAO robot

1 Introduction

Over the last several decades, service automation scope has been the focus of most technological advances. Robotics presents itself as one of the future main pillars of society for guaranteeing the quality of life in all areas. Lines of robotic research are tackling many varied fields such as medicine, rehabilitation, cleaning, refuelling, agriculture, construction or teaching [38]. Among these researches, one of the most promising lines is therapeutic robotics [15], and specifically psycho-pedagogic interventions for people with cognitive difficulties, both in childhood and elderly [31].

Coexistence of robots and humans in the same space inevitably leads to the development of social interfaces for mutual interaction through natural language and physical expressions. Thus, gesture as a linguistic complement for improving communication and interaction's naturalness assume special significance in *HRI* field. With the future role of robotics in society, affinity for robotic prototypes take on a new significance, with degree of humanity being a decisive factor [21].

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2017, Part II, LNCS 10338, pp. 167–179, 2017.
DOI: 10.1007/978-3-319-59773-7_18

In the context of gestures in robotics, different gestural taxonomies have been defined [19,24], which can be grouped into two main typologies: gestures of interaction with the environment and co-verbal gestures [13,25]. Whereas deictic or object-manipulation gestures have been the most studied [11,30,36], a shortage of studies on rhetorical gestures for accompanying speech, or symbolic gestures related to meaning, doesn't necessarily imply less importance. Whilst important progress in integrating co-verbal gesture has been achieved in the digital environment with avatars, the synchronization of gestures with speech has been a largely unexplored area in robotics [33]. The idea behind this synchronization is the gesture association with specific and fixed keywords applying rule-based and statistical model-based approaches, so new methods are needed in order to extend synchronization coverage. Some systems are based on analysis of speech-based video fragments to develop measures for establishing rules [17]. It is also common to use systems requiring text annotated with gestures in order to generate *Hidden Markov Models* (*HMMs*), as Chiu et al. show [6].

Considering the lack and limitations of interfaces that integrate co-verbal gestures, this paper proposes a new methodology for a intelligent association between symbolic gestures and speech words. These symbolic gestures are used to extend or reiterate verbal meaning, so its use has an enormous impact on human-robot communication.

Given that the techniques in association of terms with gestures have traditionally been limited to the fact that animations are activated when detects certain keywords, this paper outlines an approach based on semantic similarities that intends to associate both in a more flexible way. In this manner a gesture will be activated from any word having a similar semantic meaning. This new methodology for integrating co-verbal gestures employs Natural Language Processing (*NLP*) techniques and semantic analysis through word embeddings. To do this, it applies semantic similarity measures between gestures and words by determining which gestural expression is more in line with verbal meaning.

2 Related Work

Due to the success of gesture recognition researches, there are a large number of studies on robots that can react to human gestures, as a result of which projects such as *ALBERT* [32] and *BIRON* [9] robots have emerged.

Murthy and Jadon make a review of gesture recognition systems based on hand movements [22], noting that most of researches employ *HMMs* to identify static positions.

Recently scientific community have focused on gesture synthesis and integration, which remain largely unexplored, particularly in the robotic scope. Even though most of synthesis works use predefined gestures [34], there exists some projects generating real-time gestural expression such as *Fritz* robot [2]. Traditionally projects in this area focused on deictic [11] or collaborative [30] gesture integration. Also, significant progress has been achieved on the other typologies; for example, *WE-4RII* robot has been implemented for emotional expressions [12].

As regards co-verbal gestures, nearly any type of work derives from rule-based or grammatical model-based approaches for activating them according to fixed words. Our proposal aims precisely to expand that set of already-defined words in accordance with a semantic similarity criterion. Co-verbal gesture generation has been a line of research more recurrent in virtual interfaces, with access to lexicons being the most common method. A number of more complex systems have emerged by using visual information such as *Greta* system [27] or *Max* agent [14].

Co-verbal gestures are semantically, temporarily and pragmatically synchronised with speech in an unconscious way according to McNeil [18]. Bergmann et al. bring together the main complexities of gestural synchronization in two issues, information distribution and packaging [3]. As distribution indicates how verbal concepts and gestural ideas provide different aspects of the meaning, packaging relates to amount of information contained in those aspects.

For the purpose of determining the information packaging, video fragments containing hand motions have been analysed [35]. Similarly, Levine et al. extract measures from gestural movements in various different videos to generate motions in real time [17]. However, measures are useful for quantifying emphasis and emotional degree but says nothing about semantic. Thereby different systems assume text entry annotated with types of gestures and parameters [10,14]. Neff et al. employ manually annotated semantic tags to create gestures through a probabilistic trained model [23]. Likewise, Endrass et al. use techniques aimed at gestural corpus [8]. Other systems apply dialogue managements to planning puntual gestures from communicative targets [37]. There are also teleoperated gestural systems and based on *Wizard of Oz* methods [7].

Some systems based on *NLP* have emerged like *REA* architecture, which lexicalizes gestures to manage them as words in a language generator [4], or *BEAT*, a system that suggests a more advanced approach for synchronising gestures and speech [5]. The last one applies a rule set to determine what types of gesture must be activated, selecting rhetorical gestures by default. Finally, Ng-Thow-Hing et al. proposes an improved system with the integration of all types of gestures implemented in a *Honda* robot [26]. For that, one Part-Of-Speech (*POS*) tagging process [39] and five grammatical models are used. These are attached to each type of gesture defined by McNeill to determine to which it belongs and hence what rule system based on the identification of keywords should be applied. Grammatical models consist of gestural lexicons created by the annotation of video conferences under the following assumptions: there are simpler gestures to model than others, those are usually associated with certain words, rhetorical gestures become accentuated in topic changes and a word can be combined with several gestures due to the context.

3 A New Methodology for Selecting NAO Robot's Gestures from Free Text

Nowadays, the most consolidated tools for co-verbal gesture integration are built on the idea of relating relevant words with gestures, so that they are activated when detects one of those speech words. To improve this basic approach, our proposal consists of a new speech pre-processing methodology based on the use of semantic similarities through word embeddings. The main idea is to increase the coverage of terms with which a gesture will be activated through that semantic search. This methodology analyses incoming speech and a series of gestural representations, and associate it with significant words depending on its semantic similarity.

In order to develop the proposal, it is necessary to prepare a gestural representation set, each consisting of an identifying tag and a set of related terms for constituting a semantic space representation. Those vector representations are used in quantifying the similarity of gestures with word meanings by comparing every term through similarity measures.

Once the gestural representation list is made up, this new methodology implies a speech segmentation into sentences first, and a text tokenization after. Assuming that some grammatical categories provide most of the meaning to every speech sentence, a *POS* tagging process is subsequently established to determine those grammatical categories and identify the most significant words in the message. From a pre-trained vector-based model of word embeddings, it is pretended to quantify the similarity between sentences and related terms to finally select gestures in accordance with a semantic similarity criteria. The entire process is shown in Fig. 1.

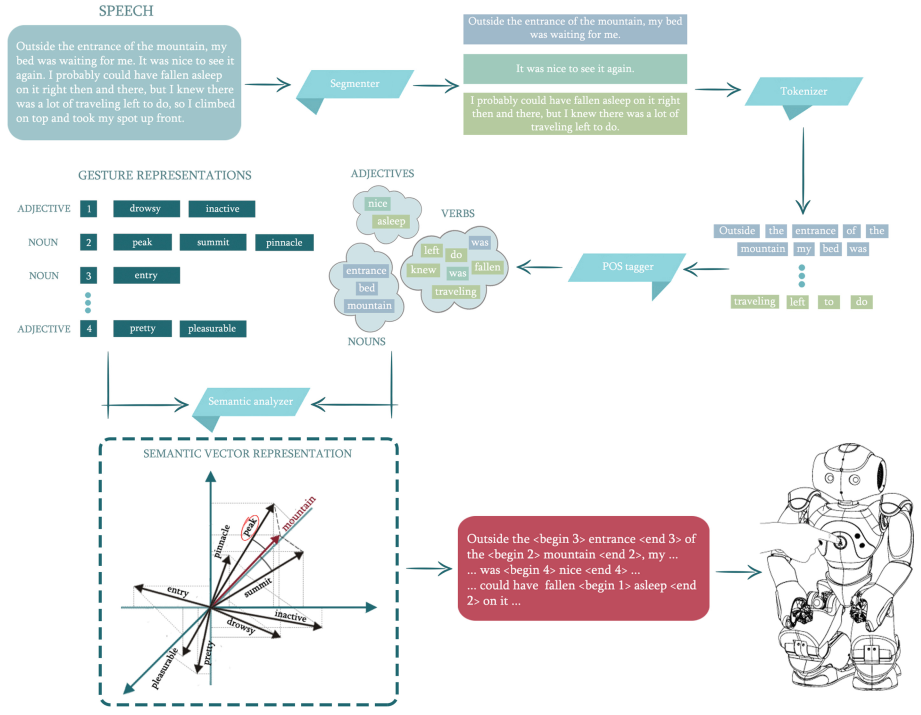
SPEECH
Outside the entrance of the mountain, my bed was waiting for me. It was nice to see it again. I probably could have fallen asleep on it right then and there, but I knew there was a lot of traveling left to do, so I climbed on top and took my spot up front.
Segmenter
Outside the entrance of the mountain, my bed was waiting for me.
It was nice to see it again.
I probably could have fallen asleep on it right then and there, but I knew there was a lot of traveling left to do.
Tokenizer
Outside the entrance of the mountain my bed was
traveling left to do
POS tagger
ADJECTIVES
nice
asleep
VERBS
was left do knew was fallen traveling
entrance bed mountain
NOUNS
GESTURE REPRESENTATIONS
ADJECTIVE 1 drowsy inactive
NOUN 2 peak summit pinnacle
NOUN 3 entry
ADJECTIVE 4 pretty pleasurable
Semantic analyzer
SEMANTIC VECTOR REPRESENTATION
pinnacle
peak
mountain
summit
entry
inactive
drowsy
pleasurable
pretty
Outside the <begin 3> entrance <end 3> of the <begin 2> mountain <end 2>, my ...
... was <begin 4> nice <end 4> ...
... could have fallen <begin 1> asleep <end 2> on it ...

Fig. 1. Proposed methodology for applying to systems such as *NAO* robot.

Word embeddings are word vector representations obtained from its contexts through transformations into reduced semantic spaces; for that, neuronal network training process and other mathematical techniques are applied to reduce the initial dimension of word representations into a reduced semantic space [16]. As each word is represented by a parameter vector, the similarity between two words results in a vector comparison; in this way, the semantic similarity between the gesture meaning, defined by its related terms, and the word meaninig can be established.

In order to evaluate the proposal, an experimentation has been carried out by determining under what conditions of semantic comparison and what association criteria optimize the gesture selection.

4 Experimental Design

Experimentation consists of the behavior analysis of word embeddings in a context of gesture association issues from its semantic spaces, as defined by related terms. Different semantic similarities have been studied applying two measures to 1200 words and 60 gesture representation with two different pre-trained word embeddings models: *word2vec*[1] model of Mikolov et al. [20], trained from *Google News*[2], and *GloVe*[3] model [29], an Stanford implementation generated from *Wikipedia*[4] and *Gigaword*[5] texts. Although both cases are partially based on a news repository resulting in a global scope, *Wikipedia* data strictly belongs to an academic scope; for this reason, *word2vec* model is estimated to offer a better word semantic representation in a general context.

In regarding to measures, euclidean distance and cosine similarity are employed. In the semantic space word embeddings are located, euclidean distance represents relations between concepts, whereas cosine similarity means semantic proximity between concepts. Therefore, a more efficient approach is a priori expected to be achieved with this last measure.

The 60 gesture representations employed in experimentation have been objectively generated from most common terms in language, provided by *Word frequency data*[6] corpus, with the purpose of assuring the coverage of used concepts. Words have been selected from the words related to these terms; for that, related vocabulary search pages have been used in addition to an subsequent review by an expert for choosing those with a similar meaning among the recommended words. Knowing what gesture belongs to each bag of words, semantic similarities between all terms and words have been calculated through the described measures.

[1] http://code.google.com/archive/p/word2vec.

[2] http://news.google.com.

[3] http://nlp.stanford.edu/projects/glove.

[4] http://wikipedia.org.

[5] http://catalog.ldc.upenn.edu/ldc2011t07.

[6] http://www.wordfrequency.info.

The comparative analysis of contextual constraints and assignment methods about those similarities intends to determine which method is most effective in associating gestures with words in accordance with similarity measures, allowing the presented methodology to be configured.

5 Experimental Results and Discussion

Experimentation has been planned in three phases by evaluating different features to be considered for the methodology at each stage.

First analysis covers the similarity of all semantic relations by comparing two alternative systems for assigning gestures to words: a system that enables multiple assignments against another one which limits assignments to an unique gesture.

Multiple assignment system establish a lower minimum membership threshold in such a way as to associate every terms above this value with the respective word. The location of this threshold will be determined by representing Precision and Recall for each measures and model, shown in Figs. 2 and 3. In these graphs

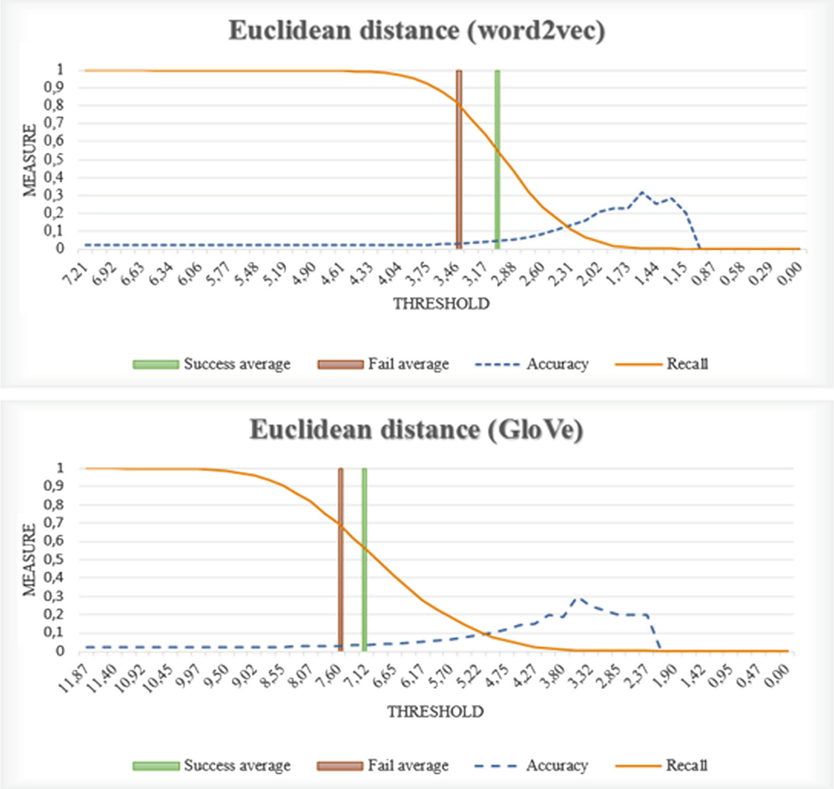

Fig. 2. Precision and Recall curves as a function of the threshold. Euclidean distance.

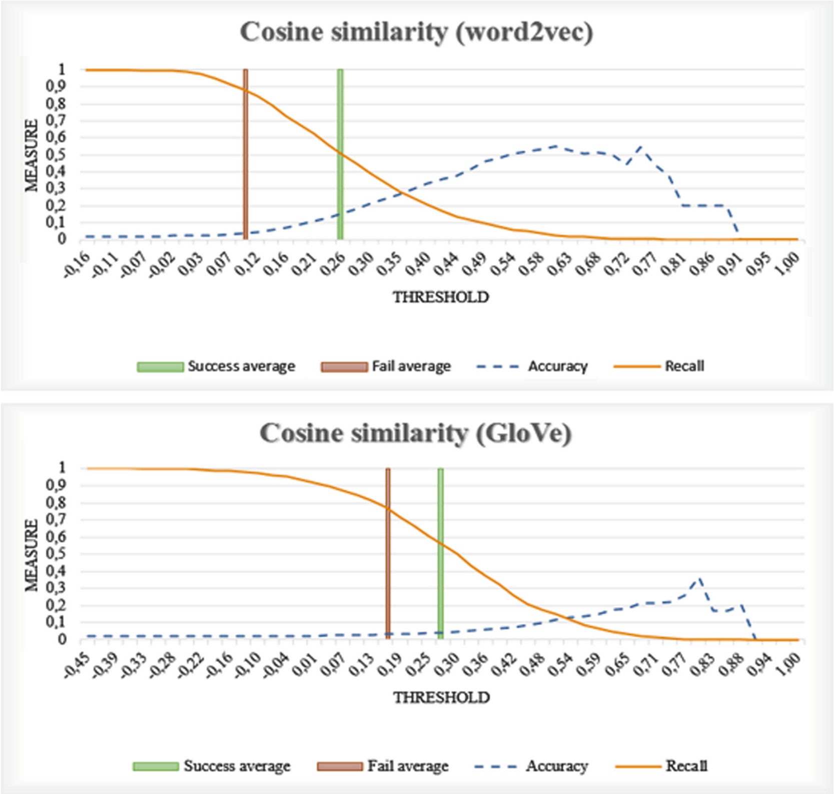

Fig. 3. Precision and Recall curves as a function of the threshold. Cosine similarity.

the average value of similarities assessed as correct as well as the average value of wrong similarities are drawn in order to stimate a range of possible locations. In this way, an intermediate threshold would exclude most of faulty relations and would allow a large number of right relations.

In this context, Precision curve represents successful gestures rate on the whole performed gestures, whereas Recall curve expresses successful gestures rate on the gestures which should have been activated. Whether a threshold is located in the right side, a high Precision and a near-zero Recall are obtained what means only a small number of gestures are performed, with half being successful. However, a threshold on the left side of graph results in a large number of executed gestures with a near-zero Precision. Therefore, the most interesting region is the middle area in which a Precision value no higher than 0.3 without compromising Recall is achieved.

Concerning unique assignment, system is limited to the closest gesture, linking simply each word to gesture containing the most similar term. Even though a minimum threshold is fixed, Precision is significantly increased in all values.

For measuring system performance, the minimum Precision value has been calculated from null threshold for each measure, shown in Table 1. Precision reaches 0.4 for cosine similarity with *word2vec* model.

Table 1. Unique assignment making no distinction between grammatical categories.

Similarity measure	Precision
Cosine similarity (word2vec)	**0.42**
Euclidean distance (word2vec)	0.31
Cosine similarity (GloVe)	0.29
Euclidean distance (GloVe)	0.21

Faced with the impossibility of locating a satisfactory threshold and reaching promising Precision, unique assignment method has been considered as the best option by establishing it during the whole experimentation.

Starting from the idea that the context surrounding a word depends on its grammatical category, one could think our system should consider grammatical information. For that reason, we have raised the need of applying a grammatical restriction for the semantic comparison with the goal of improving rates. Thus a second phase based on a category-by-category analysis is considered, in which each term is only evaluated against words in the same category.

The results shown in Table 2 support that division into categories by obtaining better Precision values when limits the comparison of similar context. The fact that cosine similarity measure get the highest rates for all categories confirms the initial suspicions about a better quantification of similarity based on proximity between context. High Precision over nouns is appreciated in these results divided by categories. This could be explained by the fact that nouns and verbs better reflect semantics in sentences than adjectives or adverbs; whilst adjectives qualify nouns, adverbs often describe circunstancial aspects of sentences.

On the other hand, the fact that not all grammatical categories appear with the same frequency in language must be also considered. For example, adverbs tends to be rarer and have fewer synonyms than nouns. One could then think that the poor results of the adverbs arise partly because until now we have forced experimentation to have the same related term proportion in each category;

Table 2. Unique assignment separated into grammatical categories.

Similarity measure	Precision				
	Global	Noun	Verb	Adjective	Adverb
Cosine similarity (word2vec)	**0.51**	**0.70**	**0.45**	**0.50**	**0.36**
Cosine similarity (GloVe)	0.44	0.64	0.38	0.48	0.27
Euclidean distance (word2vec)	0.43	0.55	0.41	0.47	0.28
Euclidean distance (GloVe)	0.36	0.47	0.32	0.37	0.26

this means that there is a possibility of considering non-existent similarities by demanding a large number of adverbs related to gestures. Thereby, a last phase has been taken into account by presenting a modification of data distribution in each grammatical category to analyse it. More specifically, less strongly relationships have been removed, giving rise to an decrease of adverbs and a lesser verb number. This has provided slightly greater Precision values, displayed in Table 3, which show an meaningful increase in adverb Precision but not sufficient to ignore the quantitative step between categories.

Given the preceding results, we can conclude that nouns should be prioritized against other categories when selecting a gesture from free text.

Table 3. Redistribution of data in each grammatical category.

Similarity measure	Precision				
	Global	Noun	Verb	Adjective	Adverb
Cosine similarity (word2vec)	**0.53**	**0.70**	**0.46**	**0.50**	**0.40**
Cosine similarity (GloVe)	0.48	0.64	0.39	**0.50**	0.32

6 Final Algorithm Proposal

Once the results have been analysed in every experiment phase, the following key conclusions can be drawn:

- A membership threshold for gestures is not feasible because of the overlap between successful and wrong similarities; association based on the closest gesture has better results.
- Cosine similarity measure is most successful at capturing semantic similarity with word embeddings. In turn, pre-trained embeddings by Mikolov have proved to be the most appropriate in a general context.
- Semantic comparison between words and terms into the same grammatical category is more effective due to the shared features of context.
- A better semantic capture by nouns is observed, what makes us consider an order of priority to narrow the search of related gestures to each category by analysing nouns first, followed by adjectives if no relationships are found, verbs and finally adverbs.

In view of all this, the final proposal for gestural assignment from free text in a NAO robot will should consider the comparison between a word meaning and each term meaning composing the semantic space of the gesture. Unique assignment based on the closest gestures will should be included, adding a minimum similarity threshold; in this way, the gesture containing the term with a greater degree of semantic similarity with respect to the speech word in the same grammatical category will be assigned, provided that exceeds minimum threshold for subsequent activation.

Our final proposed methodology is based on detecting the most significant words in speech through segmentation, tokenization and POS tagging processes.

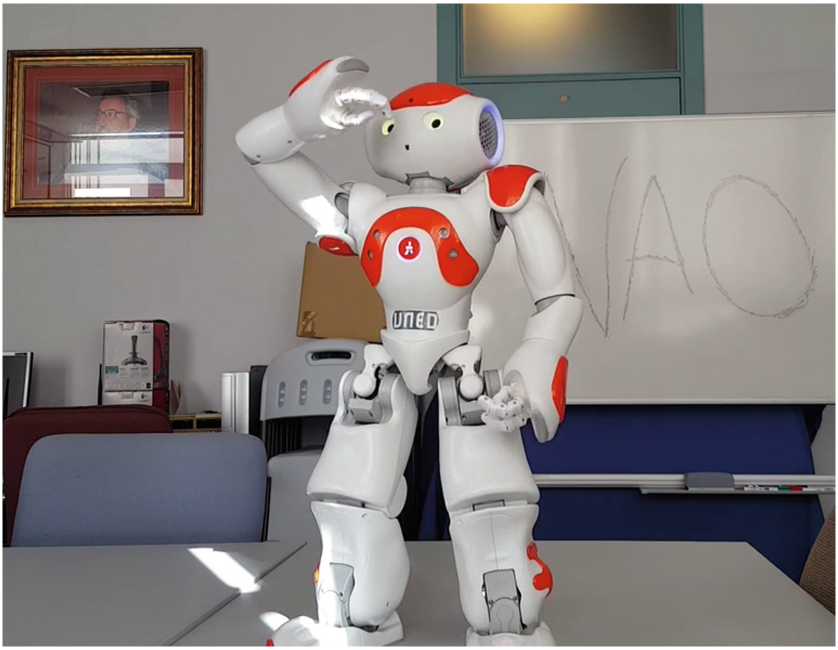

Fig. 4. Preview image of a video showing a *NAO* robot movements in accordance with the final algorithm proposal.

Starting with nouns, semantic similarity will should be analysed from the comparison of word embedding vectors with the goal of determining gesture assignment. For that, the word shall be compared to all terms belonging to the same category; if one or more gestures exceed a minimum cosine similarity value, the one most closest will be assigned.

To validate the methodology, we have implemented an algorithm including the *FreeLing* package [28] for segmentation and POS tagging processes and the pre-trained word embeddings by Mikolov. A story has been annotated with gestures of *Animations* library and applied to a Nao robot. Results have been recorded for a future distribution and can be visualized in a video (Fig. 4)[7].

7 Conclusions and Future Work

Robotics aims to draw a future marked by a high quality of life, encompassing both social and physical capabilities. Disposition of social interfaces for an interaction between people and machines more akin to interactions between human beings will be of central importance in this future. To this end the gesture integration into speech will be essential.

The importance of symbolic co-verbal gestures lies in the semantic transmission of the oral message. The suggested methodology is intended to integrate

[7] http://www.ia.uned.es/personal/delapaz/tfm_NAONLP_en.html.

them to provide a smart gestural annotation tool for gestural synchronization with language. In this sense, the use of naturalness improvement-oriented word embeddings involves a step forward compared to the techniques employed at the time in *HRI* scope.

Those semantic vector-based model also have a low computing cost after training, since that is estimated by calculating similarity through a simple vector operation such as cosine. Moreover, the proposed methodology is directly applicable to other languages whether the respective models are available.

Experimentation confirms that the inherent semantic in word embeddings contains information about the role of words in language, what penalizes the comparison between categories. Nouns are showing greater semantic characterization, partly due to an enormous amount of synonymous and a better conceptual reflection, whilst adverb meaning representation is considerably worse. Hence, the viability of word embeddings in the *HRI* field is confirmed through a tiered semantic approach from nouns to the rest of categories, leaving open other unexplored avenues for improving adverb captures.

The proposed methodology has been developed as first approximation of semantic-based gesture integration. Nevertheless, this proposal leaves open other lines of research and future improvements. Among these stand out the rule-based heuristic layer development to improve fluency, introduction of negation detectors, gestures memorisation for increasing variability or establishment of a confidence interval for toggling between gestures with similarity semantic values. Another confluent line of research would be *sentiment analysis* with which effusivity may be qualified based on the study of emotional issues in sentences through polarity classifiers. Finally, a point of view from a discourse analysis has arisen by using *Rhetorical Structure Theory* or *RST* [1]. It is intended to detect causal, contrast, justify, condition or concession relationships to activate animations for transition between concepts.

Acknowledgements. This research was partially supported by the Spanish Ministry of Science and Innovation (VoxPopuli Project, TIN2013-47090-C3-1-P).

References

1. Bateman, J., Delin, J.: Rhetorical structure theory. In: Encyclopedia of Language and Linguistics, 2nd edn. Elsevier, Oxford (2005)
2. Bennewitz, M., Faber, F., Joho, D., Behnke, S.: Fritz-a humanoid communication robot. In: The 16th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2007, pp. 1072–1077. IEEE (2007)
3. Bergmann, K., Kahl, S., Kopp, S.: Modeling the semantic coordination of speech and gesture under cognitive and linguistic constraints. In: Aylett, R., Krenn, B., Pelachaud, C., Shimodaira, H. (eds.) IVA 2013. LNCS, vol. 8108, pp. 203–216. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40415-3_18
4. Cassell, J., Bickmore, T., Campbell, L., Vilhjalmsson, H.: Human conversation as a system framework: designing embodied conversational agents. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.) Embodied Conversational Agents, pp. 29–63. MIT Press, Cambridge (2000)

5. Cassell, J., Vilhjálmsson, H.H., Bickmore, T.: BEAT: the behavior expression animation toolkit. In: Prendinger, H., Ishizuka, M. (eds.) Life-Like Characters, pp. 163–185. Springer, Heidelberg (2004)
6. Chiu, C.-C., Morency, L.-P., Marsella, S.: Predicting co-verbal gestures: a deep and temporal modeling approach. In: Brinkman, W.-P., Broekens, J., Heylen, D. (eds.) IVA 2015. LNCS, vol. 9238, pp. 152–166. Springer, Cham (2015). doi:10.1007/978-3-319-21996-7_17
7. Dahlbäck, N., Jönsson, A., Ahrenberg, L.: Wizard of Oz studies-why and how. Knowl.-Based Syst. **6**(4), 258–266 (1993)
8. Endrass, B., Damian, I., Huber, P., Rehm, M., André, E.: Generating culture-specific gestures for virtual agent dialogs. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS, vol. 6356, pp. 329–335. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15892-6_34
9. Haasch, A., Hohenner, S., Hüwel, S., Kleinehagenbrock, M., Lang, S., Toptsis, I., Fink, G.A., Fritsch, J., Wrede, B., Sagerer, G.: BIRON-the bielefeld robot companion. In: Proceedings of the International Workshop on Advances in Service Robotics, pp. 27–32. Stuttgart, Germany (2004)
10. Hartmann, B., Mancini, M., Pelachaud, C.: Implementing expressive gesture synthesis for embodied conversational agents. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) GW 2005. LNCS, vol. 3881, pp. 188–199. Springer, Heidelberg (2006). doi:10.1007/11678816_22
11. Hato, Y., Satake, S., Kanda, T., Imai, M., Hagita, N.: Pointing to space: modeling of deictic interaction referring to regions. In: 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 301–308. IEEE (2010)
12. Itoh, K., Miwa, H., Matsumoto, M., Zecca, M., Takanobu, H., Roccella, S., Carrozza, M.C., Dario, P., Takanishi, A.: Various emotional expressions with emotion expression humanoid robot we-4RII. In: First IEEE Technical Exhibition Based Conference on Robotics and Automation, TExCRA 2004, pp. 35–36. IEEE (2004)
13. Kendon, A.: Current issues in the study of gesture. Biol. Found. Gestures: Motor Semiot. Asp. **1**, 23–47 (1986)
14. Kopp, S., Wachsmuth, I.: Synthesizing multimodal utterances for conversational agents. Comput. Animat. Virtual Worlds **15**(1), 39–52 (2004)
15. Krebs, H.I., Hogan, N.: Therapeutic robotics: a technology push. Proc. IEEE **94**(9), 1727–1738 (2006)
16. Lebret, R., Legrand, J., Collobert, R.: Is deep learning really necessary for word embeddings? Technical report, Idiap (2013)
17. Levine, S., Theobalt, C., Koltun, V.: Real-time prosody-driven synthesis of body language. ACM Trans. Graph. (TOG) **28**, 172 (2009). ACM
18. McNeill, D.: Hand and Mind: What Gestures Reveal About Thought. University of Chicago Press, Chicago (1992)
19. McNeill, D., Levy, E.: Conceptual Representations in Language Activity and Gesture. ERIC Clearinghouse, Columbus (1980)
20. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
21. Minato, T., Shimada, M., Ishiguro, H., Itakura, S.: Development of an Android robot for studying human-robot interaction. In: Orchard, B., Yang, C., Ali, M. (eds.) IEA/AIE 2004. LNCS, vol. 3029, pp. 424–434. Springer, Heidelberg (2004). doi:10.1007/978-3-540-24677-0_44
22. Murthy, G., Jadon, R.: A review of vision based hand gestures recognition. Int. J. Inf. Technol. Knowl. Manag. **2**(2), 405–410 (2009)

23. Neff, M., Kipp, M., Albrecht, I., Seidel, H.P.: Gesture modeling and animation based on a probabilistic re-creation of speaker style. ACM Trans. Graph. (TOG) **27**(1), 5 (2008)
24. Nehaniv, C.L., Dautenhahn, K., Kubacki, J., Haegele, M., Parlitz, C., Alami, R.: A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction. In: IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2005, pp. 371–377. IEEE (2005)
25. Nespoulous, J.L., Lecours, A.R.: Gestures: nature and function. Biol. Found. Gestures: Motor Semiot. Asp., 49–62 (1986)
26. Ng-Thow-Hing, V., Luo, P., Okita, S.: Synchronized gesture and speech production for humanoid robots. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4617–4624. IEEE (2010)
27. Niewiadomski, R., Bevacqua, E., Mancini, M., Pelachaud, C.: Greta: an interactive expressive ECA system. In: Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems, vol. 2, pp. 1399–1400. International Foundation for Autonomous Agents and Multiagent Systems (2009)
28. Padró, L., Stanilovsky, E.: Freeling 3.0: towards wider multilinguality. In: LREC 2012 (2012)
29. Pennington, J., Socher, R., Manning, C.D.: Glove: global vectors for word representation. In: EMNLP, vol. 14, pp. 1532–1543 (2014)
30. Riek, L.D., Rabinowitch, T.C., Bremner, P., Pipe, A.G., Fraser, M., Robinson, P.: Cooperative gestures: effective signaling for humanoid robots. In: 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 61–68. IEEE (2010)
31. Robinson, H., MacDonald, B., Broadbent, E.: The role of healthcare robots for older people at home: a review. Int. J. Soc. Robot. **6**(4), 575–591 (2014)
32. Rogalla, O., Ehrenmann, M., Zollner, R., Becher, R., Dillmann, R.: Using gesture and speech control for commanding a robot assistant. In: Proceedings of the 11th IEEE International Workshop on Robot and Human Interactive Communication, pp. 454–459. IEEE (2002)
33. Salem, M., Kopp, S., Wachsmuth, I., Joublin, F.: Towards meaningful robot gesture. In: Ritter, H., Sagerer, G., Dillmann, R., Buss, M. (eds.) Human Centered Robot Systems, pp. 173–182. Springer, Heidelberg (2009)
34. Salem, M., Kopp, S., Wachsmuth, I., Joublin, F.: Towards an integrated model of speech and gesture production for multi-modal robot behavior. In: RO-MAN 2010, pp. 614–619. IEEE (2010)
35. Stone, M., DeCarlo, D., Oh, I., Rodriguez, C., Stere, A., Lees, A., Bregler, C.: Speaking with hands: creating animated conversational characters from recordings of human performance. ACM Trans. Graph. (TOG) **23**, 506–513 (2004). ACM
36. Sugiyama, O., Kanda, T., Imai, M., Ishiguro, H., Hagita, N.: Natural deictic communication with humanoid robots. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2007, pp. 1441–1448. IEEE (2007)
37. Tepper, P., Kopp, S., Cassell, J.: Content in context: generating language and iconic gesture without a gestionary. In: Proceedings of the Workshop on Balanced Perception and Action in ECAs at AAMAS, vol. 4, p. 8 (2004)
38. Ting, C.H., Yeo, W.H., King, Y.J., Chuah, Y.D., Lee, J.V., Khaw, W.B.: Humanoid robot: a review of the architecture, applications and future trend. Res. J. Appl. Sci. Eng. Technol. **7**, 1364–1369 (2014)
39. Voutilainen, A.: Part-of-speech tagging. In: The Oxford Handbook of Computational Linguistics, pp. 219–232 (2003)