

Iconic Gestures with Spatial Semantics: A Case Study

Elizabeth Hinkelman¹

¹ Galactic Village Games, Inc., 110 Groton Rd., Westford MA 01886 USA
elizh@galactic-village.com

Abstract. The spontaneous gestures that accompany spoken language are particularly suited to conveying spatial information, yet their brevity, individuality, and lack of conventional linguistic structure impede their integration into NLU systems. The current work characterizes spontaneous size gestures in a manual task corpus, clarifying their form, discourse role and representation as a first step toward incorporating them into NLU systems.

Keywords: gesture, spatial language, knowledge representation.

1 Introduction

When gesture carries the primary load of communication, as in the major sign languages, it develops linguistic properties such as verb subcategorization [1] and lexicalization [2,3]. The spontaneous hand gestures that accompany speech, in contrast, do not show linguistic structure [4]. For this reason, computational research on spontaneous gesture has focused primarily on discourse functions, such as using long range video features to signal repair strategies [5] or shifts in topic [6]. Discrete-valued features extracted from gaze and body orientation have also been used for discourse functions such as signaling grounding. Much of this work emphasizes gesture production rather than recognition [7, 8, 9].

Yet the spontaneous hand gestures that accompany speech are increasingly recognized both as a cognitive aid to the gesturer, and an encoding of meaning [10, 11, 12]. Among the spontaneous gestures that accompany speech, *iconic* gestures are those which present “images of concrete entities and actions”[4]. Iconic gestures have in some cases (though not yet broadly) been shown to be effective in communicating spatial information between discourse participants [4, 11, 13].

The current work pursues the incorporation of spontaneous gesture into NLU systems: much groundwork must be laid. Amid the fluidity and abstractness of spontaneous gesture, we focus on concrete gestures with (relatively) straightforward spatial interpretations. We seek to answer the questions:

- What is the discourse purpose of the gestures?
- Do the gestures constitute intended communication?
- To what extent are they lexicalized?
- What are their semantics?
- How can they be related to the semantics of the co-occurring speech?

2 Corpus study

We collected a reference corpus for dialogue with intonation and gesture in a physical task context. The subjects were twelve pairs of University of Chicago undergraduate and graduate students, who were familiar with each other and had some cooking experience. They were recorded while performing a 30-45 minute cooking task (making chocolate truffles), using a single camera and lapel microphones. Some elements of the task include locating ingredients and equipment, dividing the labor, choosing flavorings, and activities such as measuring and washing up.

The resulting eight hours of videotape were examined for spatial gestures. These included pointing, displaying, miming of physical actions and manner[14], and size gestures. We selected the size gestures as a focus for possible NLU because they are the simplest and most imagistic of these groupings, and because they were relatively uniform in form.



All of the size gestures in our corpus stemmed from the recipe step: “Take a hunk of set ganache and roll into a walnut-sized ball between your palms.” An example can be seen in Illustration 1, where subject Chris reads the recipe step aloud, envisions the ball he will roll, and enlists Jason to confirm the ball size. In total he performs the gesture for about three seconds; Jason eventually turns his head to view it for about 800ms. We will refer to this example and similar gestures as 'the ball size gesture'.

2.1 results: ball size gesture use and discourse purpose

Of twelve pairs of subjects, two did not communicate about truffle size beyond reading the recipe. Ten discussed truffle size verbally; of these, three did not use gestures, and three used displays of ganache (dough). Four used size gestures: three ball size gestures and one caliper size gesture¹. Gestures were used in two main ways: to inform the partner of a desired size, or to request confirmation that a size was correct. In one case, multiple ball gestures were used to explain how an incorrect ball size leads to difficulties in baking. All gestures were used with co-occurring speech.

¹ A 'caliper gesture' shows the size of a small object using parallel thumb and forefinger .

2.2 Intended communication – ball size and display

We classify five of the seven gestures as intended communication, on the basis that: in three cases the gesturer used motion or location to attract visual attention; in two cases the gesturer made a verbal reference to the gesture (e.g. “like this?”), and in one case both were used. For the seventh gesture (the incorrect ball size explanation) we have no evidence that the gesture per se was intended communicatively. A further analysis of gaze and uptake in these cases is in progress. Although this is a very small sample, most of these gestures showed evidence of communicative intent.

2.3 Form constraints on the ball size gesture

We initially suspected that the ball size gesture was strongly lexicalized in comparison with spontaneous gesture generally. In all cases the thumb and forefinger circle to touch each other and embrace a notional ball, and are displayed as the focal side of the gesture. However, there is notable variation in other parameters. Either hand could be used, as in ASL. The position of the other three fingers is not conventionalized (where it might or might not be constrained in a sign language.) The location of the gesture relative to the gesturer is not as conventionalized as it would be in ASL. In the table, we refer to the gesturer as G and the observer as O.

The third column, the explanation of how two balls may melt into each other while baking, is more typical of spontaneous gesture in showing dynamic configurational elements with extended duration. The ball size gesture is not as conventionalized as an ASL gesture – nor can we say what lexicon it would belong to. More work is needed on this point. The ball size gesture contrasts with the caliper gesture in form.

Lexicalized?	Chris&Jason	Chris&Trish	Josh&Naomi
Hand	left	right	both
Handform	'OK'	'OK'	'OK', 'OK'
Fingers	splayed	curled	splayed, splayed
Orientation	O's visual plane	O's visual plane	Off G's vis plane
Location	At G's eye level	Near O's focus	Near G's chest
Path	static	static	Slowly together
Duration (ASL=250ms)	>3000ms (G) > 700ms (O)	260ms	1500ms

3 Representing Size

Finally we consider semantic representation. A size is a property of a physical object, generally represented as a value on a scale, where a scale is a partial ordering on a set of elements. The majority of verbal size descriptions followed the recipe text: 'the

size of a” small object, or simply mentioned a small object: walnut, half a walnut, meatball. The comparative “...smaller”, and (negated) intensifier “don't make it too big!” also occurred. The scale in this case seems to be based on the generics (types) of ball shaped food items, and the asserted relation is purely qualitative. Qualitative representations [15, 16] may prove extensible. Gesture's spatial medium, by contrast, is continuous rather than discrete; the underlying scale is tied to the visual or perhaps kinesic system. What representation could plausibly be generated by the visual system? Our preliminary work investigates low level features in the spirit of [17, 18].

Acknowledgments. This work was supported in part by NSF grant no. IRI-9109914. K-E. McCullough, C. Sidner and R. Jacobs provided valuable discussion.

References

1. Supalla, T.: Serial verbs of motion in American Sign Language. In S. Fischer (Ed.), *Theoretical Issues in Sign Language Research*. University of Chicago Press (1990)
2. Hoiting, N., Slobin, D.: From Gestures to Signs in the Acquisition of Sign Language. In Duncan, S. D., Cassell, J., Levy, E. T. (Eds.), *Gesture and the Dynamic Dimension of Language*, pp. 51 - 66. John Benjamins Publishing Company, Philadelphia (2007)
3. Goldin-Meadow, S. Gesture with Speech and Without It. In Duncan Cassell Levy, pp 31-50.
4. McNeill, D.(Ed.), *Language and Gesture*, pp.2-7. Cambridge Univ. Press, New York (2000).
5. Chen, L., Harper, M., Quek, F.: Gesture Patterns during Speech Repairs. In Proc. icmi, pp.155- Fourth IEEE International Conference on Multimodal Interfaces (ICMI'02), (2002)
6. Eisenstein, J., Barzilay, R., and Davis, R. 2008. Discourse topic and gestural form. In Cohn, A. (Ed.): *Proceedings of the 23rd NCAI*, pp. 836-841. AAAI Press (2008)
7. Cassell, J., Nakano, Y.I., Bickmore, T.W., Sidner, C., Rich, C.: Non-verbal cues for discourse structure, *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, p.114-123. Toulouse (2001)
8. Traum, D., Morency, L-P.: Integration of Visual Perception in Dialogue Understanding for Virtual Humans in Multi-Party Interaction. In Proc. AAMAS (in press). Toronto (2010)
9. Rich, C., Ponsler, B., Holroyd, A., Sidner, C.L.: Recognizing Engagement in Human-Robot Interaction. In: Proc. Human-robot Interaction. Osaka (2010)
10. McNeill, D.: *Gesture and Thought*. University of Chicago Press, Chicago (2005)
11. Tversky, B., Lozano, S. C.: Gestures aid both communicators and recipients. In K. Coventry, J. Bateman, T. Tenbrink (Eds.), *Spatial language and dialogue*. Oxford: Oxford University Press (forthcoming)
12. Goldin-Meadow, S. *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press (2003)
13. Beattie, G., Shovelton, H.: When Size Really Matters. *Gesture*, 6:1., pp. 63-84 (2006)
14. Hinkelman, E.: Spatiomotor Routines as Spontaneous Gestures. *Spatial Cognition* (2010)
15. Lovett, A., Forbus, K.: Shape is like Space: Modeling Shape Representation as a Set of Qualitative Spatial Relations. AAAI Spring Symposium Series, North America, Mar. 2010.
16. Bateman, J.A., Hois, J., Ross, R. J., Tenbrink, T. A Linguistic Ontology of Space for Natural Language Processing. In *Artificial Intelligence*, in press (2010)
17. Regier, T., Carlson, L.A.: Grounding Spatial Language in Perception: An Empirical and Computational Investigation. *Journal of Experimental Psychology*, Vol. 130, No. 2, pp 273-298 (2001)
18. Franconieri, S.L., Scimeca, J.M., Roth, J.C., Helseth, S.A.: Visual Spatial Relationship Representation as a sequence of attentional shifts. *Subm. J. Cognitive Science*.