Aktionsarten, speech and gesture

Raymond Becker [1], Alan Cienki [2], Austin Bennett [3],Christina Cudina [4], Camille Debras [5],
Zuzanna Fleischer [6], Michael Haaheim [7], Torsten Müller [8], Kashmiri Stec [9], Alessandra Zarcone [10].

[1] Cognitive Interaction Technology - Center of Excellence, Bielefeld University, Bielefeld, Germany
[2] Department of Language and Communication, Vrije Universiteit, Amsterdam, The Netherlands
[3] Independent Scholar, Cleveland, USA
[4] Department of German Linguistics, University of Bamberg, Bamberg, Germany.
[5] Département du Monde Anglophone, Université Sorbonne Nouvelle - Paris 3, Paris, France
[6] Department of English Language Acquisition, Adam Mickiewicz University, Poznań, Poland
[7] TELEM, Université Bordeaux 3 (Michel de Montaigne), Pessac, France
[8] English Department, Ruhr-University Bochum, Bochum, Germany
[9] Department of Communication Studies, University of Groningen, Groningen, the Netherlands
[10] Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Stuttgart, Germany

rbecker@techfak.uni-bielefeld.de, a.cienki@let.vu.nl, anb37@case.edu, cudinac@yahoo.com, camilledebras@yahoo.fr,
zfleischer@ifa.amu.edu.pl, mrhaaheim@u-bordeaux3.fr, torsten.mueller@ruhr-uni-bochum.de, k.k.m.stec@rug.nl,
zarconaa@ims.uni-stuttgart.de

Abstract

Two studies were conducted to investigate the relationship between the production and comprehension of Aktionsarten (verbally expressed event structure, also called *lexical aspect*) in speech and gesture. Study 1, a qualitative/observational study, demonstrates discrete gesturing styles for each of the three major Aktionsarten categories. Study 2, a comprehension study, shows sensitivity to multimodal representations of event structure. Participants were both more accurate and faster at verifying the verb when gesture and speech conveyed compatible event structure representation than when they did not. Together, these results demonstrate a coherent conceptualization of event structure which supports theories of thinking-for-speaking in relation to gesture.

Index Terms: gesture, speech, Aktionsart, event structure.

1. Introduction

Increasingly, attention is being given to how research on gesture with speech can provide ways of studying speakers' conceptualization of grammatical notions as they are speaking, that is: in the process of what may be called thinking-for-speaking-and-gesturing [1, 2]. For example, McNeill and Duncan [3], developing upon Vygotsky [4], argue that the process of expressing an idea when speaking consists of a dynamic interplay between the idea's imagistic nature and the linguistic category available for its expression in the given language, and some of this imagery becomes visible in gesture. The linguistic representation of event structure has been a topic of major interest in linguistics for decades and from various theoretical approaches. In this regard, some attention has been given to the reflection of grammatical aspect in speakers' gestures [5, 6, 7], with differences having been found between the type of grammatical aspect expressed in speech at any moment and the quality and duration of the co-speech gestures.

However a major way of analyzing the internal structure of events in linguistics is in terms of Aktionsart, dating back at least to Vendler [8]. His work provided a classification that traditionally serves as a fundamental starting point for research in this area, consisting of four types: states, activities, accomplishments, achievements (see Table 1). The following are examples of the four types:

1. He was hungry. (state)
2. He exercised at the gym. (activity)
3. He ate a pizza. (accomplishment)
4. He realized that he had no money. (achievement)

	Aktionsarten dimensions		
Aktionsarten types	Telicity	Durativity	Dynamicity
States	Atelic	Durative	Static
Activities	Atelic	Durative	Dynamic
Accomplishments	Telic	Durative	Dynamic
Achievements	Telic	Punctual	Dynamic

Table 1. The three dimensions of Aktionsarten coded for in experiment 1.

To date, few studies have been done on Aktionsarten in relation to spontaneous gesture with speech. The present research focuses on the conceptualization of the internal structure of events as evidenced by the production and comprehension of gestures accompanying speech which present particular categories of Aktionsarten. Specific research questions on the production side concern whether there are gestures that are specific to Aktionsart types, and if so, on what dimensions they vary; and whether there is a connection between durativity and/or telicity of action and spatial representation in gestures. If such differences are found, this raises questions for research on comprehension of speech with gesture in relation to Aktionsart, such as whether comprehension is affected by mismatches between Aktionsart expressed verbally and concomitantly in gesture.

Existing research on gesture in relation to Aktionsart categories has mainly focused on language production (particularly following the model of McNeill's work [9]). Conversely, much research on linguistic aspect within the paradigm of psycholinguistic experimentation has focused on language comprehension. Little to none of the latter research concerns the 'comprehension' of gesture, but focuses on issues such as the relations between language and picture comprehension (for grammatical aspect see [10 – 12]).

These previous studies on aspect use sentence-to-picture matching tasks, where participants hear or read a sentence and then are presented with a picture. The participants are next asked to judge whether the picture matches the sentence by pressing a button labeled 'Yes' or 'No' if it did not match. While this task can help researchers answer questions about event telicity (ongoing versus completed events), it is perhaps more difficult to study differences in event duration. A second limitation to the current methodology where event aspect has been studied is that while it can answer questions about how people integrate and comprehend the pictures of the events themselves along with the linguistic description, it is not always the case that people communicate about events that are present. So by studying the integration of a gesture with an event description, we can find out what properties of the event are visually depicted by the speaker when a picture of the event is unavailable. Further, we can investigate how the potential differences in the way that people gesture lead to easier integration of events, which are potentially cues provided by Aktionsarten categories.

We propose that studying the relation between gestural event depictions and event descriptions calls for integrating methods from linguistics, gesture studies, and cognitive psychology. Thus the studies performed by our research team take an interdisciplinary approach.

The research was conducted as part of a group project led by the first two authors in the workshop series Empirical Methods in Cognitive Linguistics [13], specifically in the first of two workshops in the fifth year that it has been held (thus: EMCL 5.1).

2. Study 1

In the first study we address the questions: Are there gestures that are specific to Aktionsart types? If so, on what dimensions do they vary? Is there a connection between durativity and/or telicity of action and spatial representation in gestures?

2.1 Method

We elicited and video-recorded semi-spontaneous narratives from participants at EMCL 5.1 and coded them for verb Aktionsart and for gesture types used with those verbs.

2.2 Participants

Five dyads (10 individuals) participated in this study. 2 individuals were native speakers of English; the others were highly proficient speakers who had completed advance coursework and graduate-level study in English. Since there were not enough native English speakers at hand in the research setting, we decided to open our pool of participants to non-native speakers of English, contending that any result would provide significant conclusions on the role of co-verbal information in the production and processing of meaning. One benefit of this situation was that the sample effectively allowed us to eliminate cultural or linguistic bias as a factor in the production of specific co-speech gestures. All provided their informed consent and participated without remuneration as part of the workshop.

2.3 Stimuli and Procedure

To elicit Aktionsarten-rich constructions in a natural way, pairs of participants were asked to interview each other using a semi-structured interview format, which we provided. Specifically, each participant was asked to tell in English about (1) a difficult situation they had experienced and (2) an unusual situation they had witnessed. The goal with asking for these different types of stories was to elicit a variety of verb Aktionsart types by involving different points of view: first-person narration in the first case and third-person perspective in the second. After this protocol was tested successfully as a pilot study on a pair of participants, the experiment was carried out with five other dyads. Of these, one pair was left out since one of the speakers produced no observable gestures for analysis. We ended up with four videotaped conversations of approximately 17 minutes each, amounting to a total corpus of 69 minutes and 2 seconds.

The narratives were transcribed and, in order to better define the scope of event structures analyzed, only verbs used in the past tense (simple past and past progressive) were coded for Aktionsart. Then gestures used with these verbs were coded, first for gesture strokes (as per Kendon [14: Ch. 7]) and then according to the manner of movement of each stroke. Members of the team worked in pairs to code the verbs and the accompanying gestures in independently assigned videos. The whole group eventually got together to decide on the more problematic cases. Coding for manner of movement began with more detailed characterizations, which were then grouped in patterns using a principled flexibility of the type employed in some metaphor research [15].

2.4 Results and Discussion

There were 33 Accomplishment, 19 Activity, and 28 Achievement verbs used in total. Coding was done on simple past-tense verbs expressing dynamic Aktionsarten in order to narrow down the scope of inquiry. In terms of gesture use, there was a salient pattern of gestures accompanying Achievement verbs, namely that they were characterized by having a punctual nature. No such punctual nature was observed in gestures with the verbs for Accomplishments or Activities. A secondary difference was noted in gestures with the Accomplishment and Achievement verbs, namely that the gesture stroke was completed on, or repeated on, the goal of the verb (the direct object or prepositional phrase), e.g., "cuz someone uh ... jumped *on the tracks* again". With Activity verbs, however, the stroke phase occurred contemporaneously with the verb, e.g., "went out the door an' *walking* ... along the street".

No Aktionsart had its own specific gesture (e.g., there is no "accomplishment gesture" per se). Rather, there was a tendency toward certain qualities: gestures with Achievement verbs tended to have a punctual ending, while those with Activity verbs had a tendency toward extended or repeated motion, and those with Accomplishment verbs did not exhibit a clear pattern -- often they lacked a gesture or there was a gesture hold on the verb. Additionally, gestures with Achievement verbs showed iterative action due to the broader linguistic context, use with plural direct objects (e.g., "grabbed all the blankets", "hid the knives").

Overall, two salient categories emerged from this study: a punctual nature in gestures accompanying Achievement verbs and the absence of that with Accomplishment and Activity verbs. We tested sensitivity to this distinction in a comprehension study, described below.

3. Study 2

Study 1 suggested a correlation between Achievement utterances and co-articulated gestures. Study 2 tests whether the comprehension of Achievement utterances is affected by mismatching gestures, i.e.: Is event comprehension affected by whether the gesture seen matches the event type in speech?

3.1 Method

Using excerpted video clips of the multimodal Aktionsart utterances obtained in Study 1, participants were asked to watch the clips and then complete a lexical decision task, the reaction time of which was recorded. This was followed by a test of participants' sensitivity to editing that had been done in half of the clips, as described below.

3.2 Participants

26 participants were recruited from the EMCL 5.1 pool, 3 from Stuttgart University, and 1 from Bielefeld University to participate in this study and provided their informed consent. Participants in Study 1 did not participate in Study 2. As in Study 1, participants were a mix of native (7) and highly proficient non-native (23) speakers of English.

3.3 Stimuli and Procedure

Stimuli materials were made by pairing audio and video clips from the 10 speakers in Study 1. We used 14 clips, all of which contained achievement utterances. For half of the 14 pairings (7 total) the video depicted either an activity or accomplishment (durative) gesture which mismatched with the spoken achievement verb utterance. The remaining clips (7 total) were matching audio and video clips (i.e., they were unedited from the original). Durative mismatching gestures were selected as plausible gestures for the verb utterances. Two counterbalanced lists were made such that in list 1, the 7 videos depicting durative gestures were paired with the first 7 achievement verb utterances, leaving items 8-14 unedited. In list 2, items 1-7 were unedited, and the last 7 achievement utterances were paired with the durative gestures. Playback order was randomized by participant in both lists. In both conditions, the video was edited so that it only showed the speaker from the neck down. This maintained the anonymity of our recorded speakers, and also prevented participants from noticing whether mouth and head movements, e.g., matched the audio track (see Figure 1).

Participants were tested in groups of 2 – 4 at a time. Each sat in front of a MacBook Pro (approximately 100 cm away) and wore headphones. Participants were asked to complete two tasks: verb recognition and edited-video recognition. For the verb recognition task participants saw an audio-video clip in each trial, each of which lasted approximately 2-3 seconds. Then a verb appeared on the screen in black 18pt Courier New font against a white background. This verb was either the same verb spoken in the clip (e.g., *walk*) or was a different verb (e.g., *paint*). The verb remained on the screen until the participant pressed the 'p' key marked with label 'Yes' or the 'q' key labeled 'No' to indicate that that verb was heard in the audio. Each participant completed 14 trials, which lasted about 2 minutes. Response times were recorded in milliseconds from the visual presentation of the verb until the participant made a response. For the edited-video recognition task, participants had to decide whether each clip was edited. In each trial one of the 14 clips that the participant had just seen appeared on the screen; the order was randomized by participant. This was followed by a 500 msec delay and then the question, *Does this video look edited?* appeared on the screen until the participant responded either 'Yes' or 'No'. Yes/No answers and response times were recorded as they were in the first task. Yes/No answers were converted into D-prime measures. D-primes are a measure of sensitivity to the editing, with scores closer to 4 mean participants are highly sensitive, and scores closer to 0 mean participants are not sensitive to the editing.

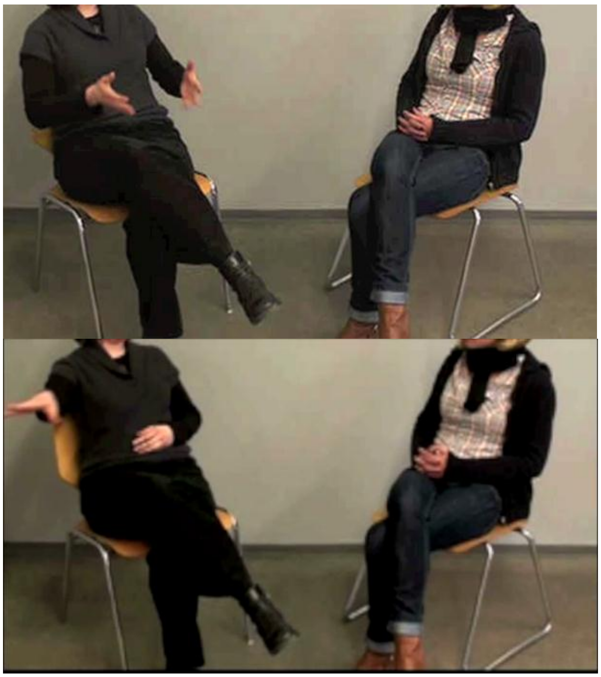

Figure 1: The top panel shows the gesture that was originally produced with an achievement verb phrase, 'it fell on the street' by the speaker on the left. The bottom panel is an example of an activity gesture video, produced by the speaker on the left, which was paired with an achievement verb phrase utterance.

3.4 Results and Discussion

Separate 2 X 2 within-subjects ANOVA were conducted on error rates and response times with audio-video match/mismatch and high/low sensitivity of editing as the two factors. Participants were grouped into high- and low-sensitivity by a median split of d-prime scores. Reaction times (RTs) greater than 10 seconds were excluded from analysis, and response times were then log-transformed.

The analysis of error rates yielded a significant effect of match/mismatch, $F(1,28) = 9.55$; $p < 0.01$ (see Table 1), but no effect for sensitivity $F(1,28) = 0.91$; $p = 0.35$, or match/mismatch by sensitivity, $F(1,28) = 2.00$; $p = 0.17$. The analysis of RTs yielded neither a significant effect for match/mismatch, $F(1,28) = 1.73$; $p = 0.20$, nor for sensitivity, $F(1,28) = 0.26$; $p = 0.61$. However, the interaction between match/mismatch and sensitivity was significant, $F(1,28) = 6.82$; $p = 0.01$. For low-sensitivity participants, the match condition showed shorter RTs compared to the mismatch condition, $t(14) = 2.34$, $p = 0.04$, but not for high-sensitivity participants, $t(14) = 0.99$, $p = 0.34$ (see Table 2).

	Match		Mismatch	
	M	*SD*	*M*	*SD*
Mean error rate	0.13	(0.20)	0.23	(0.17)
Mean RT (ms)	1480	(564)	1554	(593)

Table 2: Mean error rate and RT for Matched and Mismatched conditions

	Match		Mismatch	
	M	*SD*	*M*	*SD*
High sensitivity	1593	(557)	1515	(524)
Low sensitivity	1367	(566)	1594	(672)

Table 3: Mean RTs (ms) for Matched and Mismatched conditions for High sensitivity participants and Low sensitivity participants

Study 2 shows sensitivity to the mismatch of duration gestures with achievement verbs as evidenced by error rates among both groups, and RTs in the low-sensitivity group. In other words, if speech is accompanied by gesture, and both communicative streams convey compatible event structures, then comprehension is facilitated. Incompatible representation of the event structure, on the other hand, is perceptible and in fact inhibits comprehension.

An important factor that we considered and accounted for was how quickly and accurately a participant could decide on whether the video was edited or not, and in fact only participants who had difficulty doing so showed the match/mismatch effect for RTs. We called this factor sensitivity to editing, however, it could also be thought of in different ways. First, it could also mean that high-sensitivity participants were also sensitive to the mismatch between the event structure cues in gesture and speech, which allowed them to make better decisions about the video editing. So implicit versus explicit knowledge about event structure mismatches may be one possible explanation for the difference in RTs occurring in low-sensitivity and not high-sensitivity participants. This explanation is somewhat weak, because if it were the case that highly-sensitive participants are explicitly aware of mismatching speech and gesture, then we would predict that the effect might be as large as or larger for this group than the low-sensitivity group. An alternative to this implicit/explicit knowledge explanation could be as follows: instead of looking at participants as low-sensitivity, Wu, McQuire, and Coulson [15] suggest that they are *super-integrators*. These could be participants who take mismatching cues and make them work. So it is possible that some participants are good at making mismatching durative gestures work with achievement verbs, but it is then possible that this process is why it takes longer to process and leads to higher error rates.

One possible limitation with these findings is that we sampled from a group of high-expertise and widely diverse language backgrounds. However, it could be argued that we obtained these results even with speakers of English who have different first languages. Further, in the high- and low-sensitivity groups we had a balanced number of native- and non-native English speakers (4 native speakers in the high-sensitivity group and 3 native English speakers in the low-sensitivity group). One might predict different gesturing patterns by the various speakers based on their first language gestural patterns, particularly when it comes to motion events since, for some time, "L2 learners' gestures continue to align with L1-like units" [16: 113]. However, the fact that we did find consistencies among the speakers says something bigger about conceptualizing event structures and expressing them in terms of Aktionsart, perhaps even cross-linguistically. Further research would need to be conducted on cultural differences in gesture with respect to Aktionsart, and tested with methods such as those reported here (i.e., matching versus mismatching audio-visual clips and verb recognition) to see whether this is an effect that is exhibited in English speaking situations for native speakers of any language, or just specific linguistic and cultural groups. Nevertheless, the first step taken here is a starting point for what could potentially be a fruitful avenue of investigations in cross-linguistic and cross-cultural differences.

4. General discussion

In the field of cognitive linguistics, one of the fundamental tenets has been the idea that grammatical forms reflect different kinds of perspective-taking and ways of interpreting (including mentally visualizing) a scene [17, 18]. The findings in the present studies suggest that the discussion of Aktionsarten in linguistics in terms having to do with the "contour" of events is not a mere metaphor of the trade, but perhaps has more substantive underpinnings in terms of construing events in visuo-spatial and temporal terms, such as the punctual nature of events expressed with Achievement verbs. The gesture data suggests that Aktionsart is not merely a grammatical distinction, but that these categories have cognitive reality in terms of imagistic construal of events and the mental simulation of them, and the grounding of language in action [11, 19 - 21].

5. Conclusions

Data from co-speech gesture can reveal how speakers conceptualize event structure in spatio-motoric and durative-motoric terms. The use of co-speech gesture accompanying verbs displaying different Aktionsart categories shows key distinctions in the ways we express and comprehend events as we are thinking-for-speaking-and-gesturing. Our results suggest that speakers' gestures reflect features of Aktionsart categories which are important for event structure comprehension. This is consistent with the multimodal stance toward language taken by many in gesture studies, whereby gesture and speech are complementary aspects of the process of utterance [22], or indeed, that "gestures are part of language" [23: p. 4].

Finally, we believe that the complementary use of qualitative and quantitative forms of analysis, as in this research, helps provide a more comprehensive account of language use in context than can be obtained from only one or the other type of methodology.

6. Acknowledgements

The authors would like to thank the participants and the organization of EMCL 5.1, which was supported by the Freiburg Institute for Advanced Studies, the University of Freiburg, and the Deutsche Forschungsgemeinschaft. Many thanks also to Monica Gonzalez-Marquez for crucial assistance with the software for running Study 2 and in general for maintaining the EMCL series, and to Martin Hilpert for the local organization and support in Freiburg.

7. References

[1] Cienki, A. and Müller, C., "Verbal to gestural, and gestural to verbal, metaphoric expression", talk presented at the sixth Researching and Applying Metaphor conference, Leeds, UK, April 2006.

[2] Slobin, D., "Thinking for speaking", in Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society, 435-445, Berkeley Linguistics Society, 1987.

[3] McNeill, D. and Duncan, S., "Growth points in thinking-for-speaking", In D. McNeill [Ed], Language and Gesture, 141-161, Cambridge University Press, 2000.

[4] Vygotsky, L., Myshlenie i Rech' [Thinking and Speech]. Labirint, 1934/1996.

[5] Duncan, S. D., "Gesture, verb aspect, and the nature of iconic imagery in natural discourse", Gesture, 2(2): 183-206, 2002.

[6] McNeill, D., "Aspects of aspect", Gesture, 3(1): 1-17, 2003.

[7] Harrison, S., "Grammar, gesture, and cognition: The case of negation in English", PhD dissertation. Bordeaux, France: Université Michel de Montaigne, 2009.

[8] Vendler, Z., Linguistics in Philosophy. Cornell University Press, 1967.

[9] McNeill, D., Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press, 1992.

[10] Madden, C. J. and Therriault, D. J., "Verb aspect and perceptual simulations", The Quarterly Journal of Experimental Psychology, 62, 1294-1302, 2009.

[11] Madden, C. J. and Zwaan, R., "How does verb aspect constrain event representations?", Memory & Cognition, 31(5), 663-672, 2003.

[12] Yap, F. H., Chu, P. C. K., Yiu, E. S. M., Wong, S. F., Kwan, S. W. M., Matthews, S., Tan, L. H., Li. P., and Shirai, Y. "Aspectual asymmetries in the mental representation of events: Role of lexical and grammatical aspect", Memory & Cognition, 37, 587-595, 2009.

[13] Empirical Methods in Cognitive Linguistics 5.1. https://sites.google.com/site/emcl5freiburg/

[14] Kendon, A., Gesture: Visible Action as Utterance, Cambridge University Press, 2004.

[15] Wu, Y., McQuire, M., and Coulson, S., "Individual differences in comprehension of conversation", Supplement to the Journal of Cognitive Neuroscience. (Poster presented at Annual meeting of the Cognitive Neuroscience Society, San Francisco, CA, April, 2011).

[16] Cameron, L., "Confrontation or complementarity? Metaphor in language use and cognitive metaphor theory", Annual Review of Cognitive Linguistics, 5, 107-135, 2007.

[17] Gullberg, M., "Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon)", International Review of Applied Linguistics in Language Teaching, 44(2), 103-124, 2006.

[18] Langacker, R. W., Foundations of Cognitive Grammar. Volume 1: Theoretical Prerequisites. Stanford University Press, 1987.

[19] Langacker, R. W., Cognitive Grammar: A Basic Introduction. Oxford: Oxford University Press, 2008.

[20] Barsalou, L. W., "Perceptual symbol systems", Behavioral and Brain Sciences, 22:577-660, 1999.

[21] Glenberg, A. M. and Kaschak, M. P., "Grounding language in action", Psychonomic Bulletin & Review, 9:558-565, 2002.

[22] Bergen, B., "Experimental methods for simulation semantics", in M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, and M. Spivey [Eds], Methods in Cognitive Linguistics, 277-301, John Benjamins, 2007.

[23] Kendon, A., "Gesticulation and speech: Two aspects of the process of utterance", in M. R. Key [Ed], The Relation between Verbal and Nonverbal Communication, 207-227, Mouton, 1980.

[24] McNeill, D., Gesture and Thought, University of Chicago Press, 2005.