Janusz Badio
University of Łódź
Łódź, Poland

Looking for events in a video of mundane activity

ABSTRACT

The present work focuses on the unit of event, its psychological reality, and linguistic coding. The category of event has been defined as a segment of time with a beginning and end (Zacks & Tversky 2001). During this time something happens. Two related studies are presented based on Badio (2014). The first of them is a non-verbal study of video unitization in the search of events and their psychological reality, as well as ways in which people mark them off the stream of continuous activity. Perception of change of variable parameters of action has a decisive role in this process. The second study investigates verbalisations of the same video to conclude that the main independent variable, Polish-native and English-foreign verbalisations are similar with regard to the number of coded events, but a little different as far as some processing parameters are concerned. The paper presents the design, analysis of inferential statistics of the experiments, and provides suggestions for improvement. It also focuses on a few other variables related to this topic that can be used in future research.

Key words

event, construal, coding, segmentation, retelling, English, Polish, processing cost, corpus

1. Introduction

Use of language, typically for some communicative purpose, places a demand on the language producer to assume a certain point of view and perspective. Selection of the salient participants and features is performed to effect a conceptualisation. This process, often referred to as construal (cf. Langacker 2008), involves multiple aspects of attention (cf. Badio 2014) and subsequent or almost simultaneous recruiting of language form to code and instruct about the details of such a conceptualization. Construal (here understood to be mental, conceptual) and coding (the term restricted to selection and use of language form) naturally pertain to *events*, *activities*, *action*, *scenes*, and *situations*.

Though related, the above terms differ in some important details. Events are defined as things that happen (Lewandowska-Tomaszczyk 2011), so they are contrasted with actions and activities, which are understood as typically human and intentional. Scenes and situations invoke events and activities together with the accompanying context apart from the main participants. As an example, the difference between event and scene/situation of *riding a bike* is that the focus implied by the event of riding is narrowed to the main participants, i.e. *rider* and *bike*, whereas the term situation of *riding a bike* tends to involve a larger scope of attention to time and other parameters of context. The term *scene* or *state* is preferentially used instead of the term event to designate a static relation, i.e. when the configuration of participants does not change over the portion of time that it comprises (cf. Vendler 1957; Bach 1986: 6; Radden & Dirven 2007: 171-332). In sum, the terms *event*, *scene* and *situation* are not used consistently in linguistics literature.

The present paper applies the term *event* to refer to dynamic as well as static scenes, both those that involve human activity, and those that do not. Events have been defined in variable ways, but this work accepts the definition provided by Zacks & Tversky (2001), who argue that they involve a segment of time at a given location with a beginning and end, during which a certain entity interacts with other entities. This view shows striking similarities to the so-called FIT (Feature Integration Theory) model of human attention offered by Treismann & Gelade (1980). During the first pre-attentive stage, an object is automatically broken down into its features, and subsequently during the second stage of processing, the features are recombined by focused attention that functions as *glue* uniting the two streams of processing. In a similar fashion, Barsalou (1999) puts forward the argument that during language comprehension modality-specific brain regions responsible for vision, hearing, smell, and kinaesthesia become active. This procedure is facilitated, or even made possible at all, because of an immediate attentional decomposition of a perceptual gestalt into component, modality-specific features.

Once an object is constituted at a certain location in space/time, its evolution in real and processing time also needs to be monitored and registered. People have a way of knowing when exactly one event finishes and another begins. Language labels, mainly verbs but also nominals, are used to instruct about them.

The present paper has two related goals. First, the plan and results of the non-linguistic task of video segmentation are described to consider the psychological reality of event category, followed by a description of the design and results of the related verbal task. The latter sub-study leads to the creation of a corpus of recordings in Polish (native) and English (foreign) languages.

For lack of space, the paper will not discuss details of event classifications, and the related topic of event partonomies (event parts) and taxonomies (kinds of events). These topics have been discussed elsewhere (Bach 1986; Badio 2014; Bohnemeyer & Pederson 2011; Radden & Dirven 2007; Vendler 1957; Tversky 1990). The views of cognitive linguistics have been presented by Langacker (1991), Talmy (2000) and Croft (1998) to name only a few works.

2. *VIDEO SEGMENTATION: BACKGROUND AND QUESTIONS*

The main research question this study attempted to answer concerned the psychological reality of the event category. In particular, it focused on how people segment ongoing activity and hence how they categorise it. Such a focus and methodology was motivated by earlier studies into event segmentation by Newtson (1973), Newtson & Enquist (1976), who had used this methodology for the first time, and Zacks, Tversky & Iyer (2001), who later used the procedure of unitization to investigate whether people activate background schemata during event segmentation.

Such a formulation of the general question in the latter work required an operational definition of the event category as a segment of video-recorded activity. The segment was marked off by a viewer who pressed a computer space bar each time s/he felt a meaningful part had commenced and/or terminated. One group of viewers was instructed to mark off segments only when necessary, whereas the other group's instruction was to mark off a new segment as frequently as possible. The former group marked off coarse grained segments, but the other group marked off finer segments. This is illustrated in Table 1.

The careful reader will have already noticed that though there are more so-called finer breakpoints than coarse breakpoints, they co-align at the 4th, 8th and 10th seconds. This tendency was revealed at a frequency that was statistically significant, which led to the conclusion that the participants must have used both bottom-up and top-down processes to perform the task.

Seconds	1	2	3	4	5	6	7	8	9	10
Course Breakpoints	1			1			1			1
Fine breakpoints		1		1		1	1	1		1

Table 1. Example division of video into 1 s. bins

To paraphrase, the perception of action on the video, which is a bottom-up process, was also guided by the top-down activation of action schemata, resulting in the high, statistically significant alignment of the fine and coarse-grained units.

The present study, however, did not divide the participants (*N*=45) into fine-grain and coarse-grain coders. Instead, they all received the same, only one instruction to mark off meaningful segments by inserting breakpoint during the video presentation (2 minutes). The guiding question was how the video recording would be segmented, whether the segments would be similar among participants, and what strategies would be used during the unitization task.

2.1. The Task, Participants, Materials and Procedures

The first of the questions presented above, regarding ways in which the participants would divide the video material, required the creation of a recording and use of actors. As signalled in the title, the study chose to focus on mundane activity, whose events do not constitute a narrative in Labov's (2001) sense for lack of what was termed *most reportable event*, i.e. an event that is worth presenting because it is surprising, or unusual. As a result, the causal structure of the event sequence on the prepared video was not as clear as is typically expected from an elicited story. This decision was made on purpose so as to minimize the activation of background schemata during the performance of the task. The following script was prepared for the unprofessional actresses to perform later:

C1: K is sitting in her room. Leaves the room.
C2: Goes downstairs /camera follows/
C3: Puts on shoes,
C4: Takes a ball,
C5: opens door and leaves,
C6: Goes downstairs into the garden
C7: Walks along the lawn
C8/9: A reading /continuous recording without cuts/
C9: K signals /invites her to play/,
C10 A stands up

C11: They play (throw ball to each other) – 10 sec.
C12: Mistake and the ball lands in bushes
C13: A goes to fetch the ball, enters the bush /camera follows her inside the bush/
after 10 s. camera focus goes back to the previous location (scene)
C16: Alex sitting and reading.

The sequence of such events does not constitute anything worth retelling and presenting to another person for fear of being treated with a shrug of the shoulders, to say the least. The script consists of sixteen moves, each of which can be interpreted as a single event, or can be broken down into smaller subevents. Nevertheless, the video recording presented them as a continuous flow of activity with events merging into one another.

The video recording was performed after the actresses had understood and practised their roles. The 2 minute video was subsequently used to ask a group (N=45) of MA students at the Institute of English of the University of Łódź, Poland to watch the video and insert breakpoints between meaningful wholes. Technically, a media player called GOM was used, as it has the functionality of capturing images in a convenient way by pressing one button. When this button is pressed (clicked on, in fact), GOM Player saves the captured image to the designated folder and its name is the number of seconds and milliseconds that have elapsed from the beginning of the recording. One after another, after receiving some practice using this computer application, the participants were asked to take a seat at the laptop and perform the task of pressing the button whenever they thought a meaningful whole began and finished. The next section deals with data and results.

2.2. *Video segmentation: results*

The segmentation decisions were recorded in the names the application assigned to the captured images. The continuous video was discretised into 144 one-second slots, and coding involved using number "1" horizontally in an Excel spreadsheet for a given participant whenever s/he had decided to insert a breakpoint during that second. This is illustrated in Table 2.
In general, it turned out that some slots were marked off considerably more frequently than others. For example, this tends to be true of slots 4, 5, 6. This is the time when the girl on the video gets up from the chair she has been sitting on behind her desk and additionally turns around to walk across the room towards the exit door to the hall on the first floor of the house.
Another example, slots 9 and 10 on this portion of the whole table, refer to the moment of her beginning to walk down the first flight of stairs. The same procedure was applied to the whole sequence.

participants/ seconds	1	2	3	4	5	6	7	8	9	10
AK		1	1	1	1	1	2			1
BA		1			1					1
KI					1		1			
KK		1	1		1	1				1
KS						1				1
MJ		1		1	1	1			1	
NL					1		1		1	
PS					1					1
sums in columns		4	2	2	7	4	4		2	5

Table 2. Breakpoints inserted in the video

The moments of the maxima of change of motion, direction of motion, character, goal, change of activity, change of body part, and change of object were used by the viewers to click the button and insert a breakpoint between what they considered a meaningful part, and what the present study terms *event*. The following descriptive statistics were obtained in the analysis of the segmentation task:

N	45
Mean	18.97
SEM	1.61
Median	19
SD	10.83
Skewness:	1.03
Range:	46
Min.:	4
Max.:	50
Confidence interval (95.0%)	3.25

Some interesting points to observe are as follows. First, the values of the median and the mean are very similar, which induces a tentative conclusion that the frequency distribution of the dataset is normal, i.e. the average

number of 17-18 events were the most common tendency and all values are evenly distributed about the mean. However, the values of dispersion show considerable variability. The number of actual breakpoints inserted by the viewers, i.e. the range, spreads from 1-4 to 37-40, whereas the *SD* (*standard deviation*) amounts to *SD*=10.83, with the whole frequency distribution skewing to the left, i.e. towards the smaller values, between approximately 5-8 to 21-24. The intervals larger than 33-36 appear to be outliers and were not considered in the analysis. They are larger than plus two standard deviations (to the right of the mean). This is presented in Figure 1.

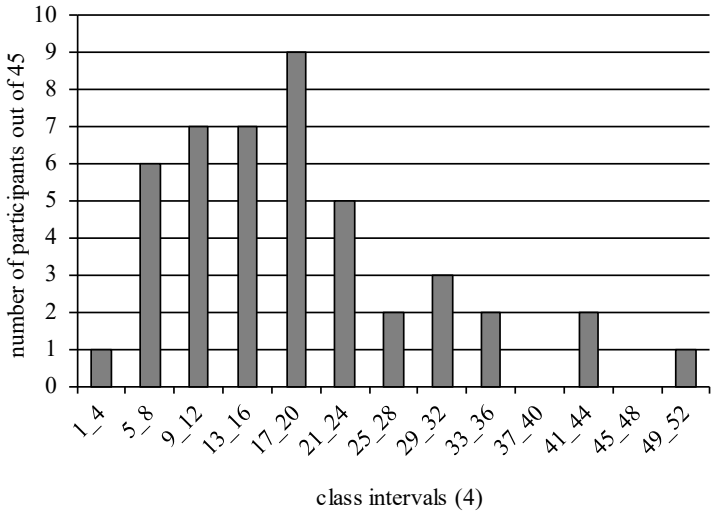

Figure 1. Frequency distribution of the breakpoints in the video segmentation task (Badio 2014: 151)

After each participant had finished this first task, s/he was asked to comment on how s/he performed the task, and what motivated the actual decisions. The answers were then grouped around the following (paraphrased) categories:

a) It was hard to predict what would happen next and segment.
b) It was hard to follow the girl's movements.
c) Sometimes I had a problem deciding when exactly one activity finished and another began.
d) I followed (observed, focused on) the girl or an abject.
e) I did not click the capture every time I should have. Sometimes it was too late.
f) The segmentation task would have been easier if I could have done it again.
g) At the beginning I marked off every movement (e.g. turning of a page at the beginning of the video recording), not treating a sequence of movements as belonging to one activity (event). Then, as the movie continued, I changed the strategy and began to subsume physical movements into a larger category.

Comments (a, b) clearly indicate that the participants were not able to predict the next scene. This design feature was deliberate, and helped prevent the viewers from activating background schemata that would guide their coding decisions. Point (c) signals that activity is indeed fluid, and continuous, unlike discretised verbalisations of the same activity. Point (e) is an important observation that could help improve the design of this experiment. One must remember that the viewers watched the video only once, so sometimes they hesitated, deciding whether to click the button immediately or a half a second later, which resulted in hesitations, false starts, and even omissions. Perhaps allowing the viewers to work at their own pace, or/and view the same material again, would be an essential improvement of the task. Otherwise, the participants acted as if it was their reflexes and fast responses that were being tested. On the positive side, not having enough time to rethink their categorising decisions, they had no choice but to focus attention on the developing action, rather than their schemas regarding the observed stimulus. But the final comment (g) is an important one as it points to an important tendency to begin with the strategy to be guided by perceptual input followed by an increasing reliance on the action schemas activated during the task. Figure 2 below presents the results of the video segmentation task.

Especially, but not only, simultaneous change of variable parameters of activity signals a new event. The event category was studied without linguistic coding, and the participants had no problem performing the task, let alone understanding the instructions, which confirms the prediction that the category of *event* is psychologically real, not solely tied to use of language.

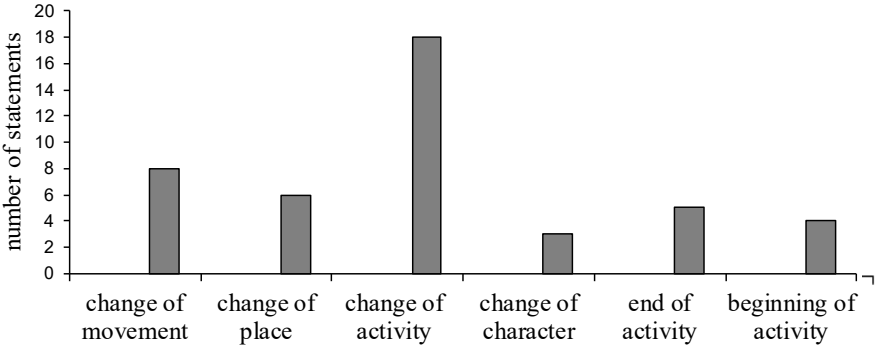

Figure 2. Motivations in video segmentation tasks (Badio 2014: 152)

3. *Video presentation and verbal coding in English-foreign and Polish-native languages: introduction, questions, hypotheses*

The first study of video segmentation described above was followed by the study of spoken, verbal reports of the same video recording in English-foreign and Polish-native languages. The following general questions provided the motivation for the study. Will the non-verbal video segmentation and verbal tasks differ as regards the sheer number of inserted breakpoints? Additionally, we were interested in some possible differences between the way native and non-native speakers of Polish and English (accordingly) would code the input video sequence of events.

The general questions, formulated in this way, were then interpreted for the purposes of the task, which required each participant first to watch the video and then retell its content in either English or Polish. The students who took part in this part of the experiment were second year MA students in the Department of English, all at the upper-intermediate to advanced level of English. The instructions were carefully presented to them individually, and they were not informed which language they would have to use. It was hypothesised that there would be fewer events in the verbalisations than the breakpoints in the video segmentation task. Additionally, the second hypothesis predicted that the Polish-native language retellings would contain more events than the English-foreign language retellings, due to the general common-sense agreement that it is more difficult to use a foreign than a native form of language. The interest in this question is arguably feasible as the present author considers the processing of a foreign language to be a legitimate object of study (cf. Hulstijn 2007: 197).

3.1. *Verbal retellings: materials, procedures, corpus, and data collection format*

Each participant sat in front of a 15-inch laptop, and s/he was explained the two step procedure, to watch and then retell the video sequence in either the English-foreign or Polish-native languages. The retellings were recorded with the use of the Microsoft Windows Vista voice recording function and application. The instruction to retell in English or Polish was provided after the watching of the video. After the recordings were collected (26 English-foreign, 14-Polish-native), there was a need to code the data and store it in a convenient way in order to analyse it appropriately later.

3.1.1. Verbal retellings: coding and corpus

The coding of the recorded data required a format that would allow convenient search, grouping, and filtering of the interesting parameters of the verbal recordings. First, they were transcribed using the system proposed by Chafe (1994, 2003). The main unit in this system of speech transcription is the so-called *intonation unit*, defined with the use of a set of perceptual, auditory and syntactic parameters. They usually appear together: falling pitch at the end of the intonation unit, a pause that signals a speaker is planning what to say next, the full syntactic phrase (for example: NP, VP, PP, Clause). The interested reader is also referred to Badio (2004) and Stelma & Cameron (2007). As the coding proceeded, the intonation units were fed into the format of a dBase of the Microsoft Access with two tables named SPEAKERS and INTONATION UNITS, related in one to many, i.e. one speaker, many intonation units. The intonation units table had the following parameters describing each unit: ID, the actual unit, event (part of event) the unit coded, and language (English or Polish). The following lines from the dBase illustrate the case.

ID_INT UNIT	ID_SPEAKER	THE INT UNIT	event	LANGUAGE
54	3	so I saw a girl,	GIRL SIT	ENGLISH AS L2
55	3	who was sitting in her . y . bedroom,	GIRL SIT	ENGLISH AS L2
56	3	..y in front of eeeee… her computer,	GIRL SIT	ENGLISH AS L2

Table 3. One sentence, but three intonation units coding one event GIRL SIT

In this case, lines 54-56 in the corpus code a sentence *I saw a girl who was sitting in her bedroom in front of the computer.* This sentence has been coded as three consecutive intonation units, each providing different information concerning the event, arbitrarily called GIRL SIT. Line (54) informs that the viewer noticed it, so it is about her/his perception, whereas (55) constitutes the core of the scene/event, and (56) includes additional information about the circumstances of the event.

The above-described coding procedure started from the first recording, and when a new event was coded by someone, it was added to the list. As a result, the following list of event types appeared in the data of spoken English and Polish retellings of the original video. Their numbers were subsequently compared between the Polish and English data.

The descriptive statistics indicated that the use of Polish-native language might have caused a narrowly larger number of events compared to the corpus of English-foreign recordings. However, the inferential test comparing the two groups of recordings ($N = 14$, $df = 21$) rendered the t value statistically insignificant, i.e. $t = 1.03 < t\text{-stat} = 1.80$; $p > .05$, and the null hypothesis, that the choice of language (Eng. vs. Pl.) does not tend to be related to the av-

erage number of events coded in these two different experimental conditions had to be accepted. Interestingly, it turned out that the preferred strategy to code events is to stretch their content into more than one intonation unit (see example on the previous page), ($N = 14$, $df = 21$), the obtained value of t was: $t = 3.35 > t\text{-stat} = 2.07$ in the two-tailed, non-directional test.

Table 4. Events found in the video verbalisations, both English and Polish

Events arranged by time	Polish events	English events
GIRL SIT	10	15
GIRL WORK	0	4
GIRL CLOSE LAPTOP	3	6
GIRL STAND UP	1	1
GIRL LEAVE ROOM	2	4
GIRL WALK DOWNSTAIRS	8	3
GIRL OPEN WINDOW	1	5
GIRL LOOK OUT WINDOW	7	2
GIRL CONTINUE WALK DOWNSTAIRS	8	5
GIRL PUT ON SHOES	6	5
GIRL PUT ON HER LEFT SHOE	0	1
GIRL PUT ON HER RIGHT SHOE	0	1
GIRL TAKE BALL	13	4
GIRL OPEN DOOR	0	1
GIRL GO OUTSIDE HOUSE	8	7
GIRL WALK ROUND HOUSE	0	5
GIRL OTHER SIT GARDEN	3	1
GIRL ASK GIRL OTHER PLAY BALL	10	5
GIRL OTHER STAND UP	0	1
GIRLS PLAY BALL	8	9
GIRLS NO CATCH BALL	8	9
GIRLS STOP PLAY	2	1
GIRL LOOK OTHER GIRL	1	0
GIRLS LOOK FOR BALL	4	5
GIRLS FIND BALL	9	0
GIRLS DO NOT FIND BALL	0	1
GIRLS GET BALL OUT	0	2
GIRL OTHER SIT GARDEN – LAST SCENE	2	0
GIRL RETURN HOME	1	0

3.1.2. Verbal retellings of the video: some qualitative comments

Some qualitative analysis followed the above-presented quantitative studies. The analysis compared events coded in the Polish and English corpora. First, the foreign language speakers tended to spend more time in order to present the context at the beginning of their retellings, which is evidenced in a considerably larger number of people who decided to code the event called GIRL SIT. This strategy may have appeared in response to a greater processing load, and need for planning in the case of English as a foreign language. Second, the native speakers of Polish seemed to code more events that create the causal structure of the video event sequence, whereas their English-foreign counterparts preferred to hold on to the first available scene, or clue to carry on with the verbalisation. Third, all participants presented the event in the chronological order. Fourth, all of them also tended to divide events into manageable components and presented them by using at least two intonation units. Fifth, the presented events and intonation units that coded them frequently tended to code interactions between the characters and objects. Sixth, both Polish and English samples contain evidence as to how demanding, and non-obvious it is to try to retell the content of even a simple (two characters, 2 minute long) video presentation. Last, both Polish and English samples contain examples of coding information that was absent in the original recording. One must remember that the instruction clearly asked them only to retell, and not to add anything that did not take place.

Summary and conclusions

The previous section finishes the presentation of the two studies, non-verbal video segmentation and verbal retelling of the video. The search for events demonstrated the psychological reality of event category, and uncovered some reasons why the participants decided to mark off a new event. These were change of activity, beginning/end of activity, change of character, or location. Without direct instruction to use either fine-grained or coarse-grained coding, the viewers nevertheless coded in variable ways, but they tended to be guided more by perceptual information at the beginning. This strategy, however, was gradually replaced by predicting what would happen next on the basis of activated background schemas for action.

The verbalisation task required the decision to select a coding format for the recordings of speech and a manageable and convenient format for storing the data. The analysis of a created corpus, a database of transcriptions of the recordings, revealed that there was no statistically significant difference in the number of coded events in the Polish-native and English-foreign transcriptions. This conclusion further augments the belief that the event

category is psychologically real, apart from a reflection that the task itself turned out to be easy enough to perform regardless of language. However, the additional qualitative analysis also pointed out that despite the similarities, there were also some differences in how the speakers behaved in Polish-native and English-foreign conditions. In general the use of the native language generated higher processing effort as evidenced by the need to have more time to plan a verbalisation, and by the tendency to focus on the available clues in order to maintain fluency at the cost of building the causal structure of the retelling. The two groups, however, behaved similarly as regards the amount of pauses, and the inclusion of events that were absent in the original input, which indicates how creative our retellings of past experience are regardless of the clear instruction to be objective.

Finally, a few comments are warranted regarding the possible pitfalls of the presented studies and related analyses, together with ideas for improvement and future research. The non-verbal study of video segmentation did not allow the second or even third viewing and careful, exact insertion of the breakpoints where the participants felt a meaningful part began and ended. Instead, they only viewed the two-minute video once and had to click the button, which was probably too demanding and required quick responses. As a result, some subjects may have been late, or did not click but later felt they should have, as the post-task reports revealed. It could be advantageous to change the task so that the participants have no time limit, during which they insert the breakpoints. Also, one should perhaps usefully introduce an independent variable, e.g. type of input, with one condition involving a familiar and unfamiliar activity.

As for the verbalisations, a follow-up study could use English-native and English-non-native forms of language, with language controlled for. As it stands now, the focus is on the difference between type of language (native vs. non-native). The comparison of Polish and English, though arguably legitimate, may be less interesting than keeping language constant. Certainly, one reason for the actual choice was a limitation and no access to sufficient numbers of native speakers of English. Last, the relationships between intonation units and events deserves further consideration. They show the workings of our attention. For example, mental scanning is used in the case of presenting an event in the format of a few intonation units, whereas coding it as one intonation unit enlarges the scope of attention, but involves processing fewer details (summary mental scanning).

References

Bach, Emmon 1986: The algebra of events. *Linguistics and Philosophy* 9, 5–16.

Badio, Janusz 2014: *Construal and Linguistic Coding of Narrative Events*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Badio, Janusz 2004: *Conceptual and Phonetic Chunking Procedures of English Speech*. Unpublished PhD dissertation. University of Łódź: Łódź, Poland.

Barsalou, Lawrence 1999: Perceptual symbol systems. *Behavioral and Brain Sciences* 22, 577–609.

Bohnemeyer, Jürgen, Eric Pederson (eds.) 2011: *Event Representation in Language and Cognition*. Cambridge: Cambridge University Press.

Chafe, William 1994: Discourse, *Consciousness and Time*. Chicago: The University of Chicago Press.

Chafe, William 2003: Prosodic and functional units of language. In: Jane A. Edwards, Martin D. Lampert (eds.) 2003: *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale, NJ: Lawrence Erlbaum Associates, 33–43.

Croft, William 1998: The structure of events and the structure of language. In: Michael Tomasello (ed.) 1998: *The New Psychology of Language, Cognitive and Functional Approaches to Language Structure*. London: Lawrence Erlbaum Associates Publishers, 67–93.

Hulstijn, Jan, H. 2007: Fundamental issues in second language acquisition. *EUROSLA Yearbook* 7.1, 191–203.

Labov, William 2001: Uncovering the event structure of narrative. *Georgetown University Round Table*. Georgetown: Georgetown University Press. URL: www.ling.upenn.edu/~wlabov/uesn.pdf. ED: 9 April 2004.

Langacker, Ronald 2008: *Cognitive Grammar. A Basic Introduction*. Oxford: Oxford University Press.

Lewandowska-Tomaszczyk, Barbara 2011: Events as they are. In: Piotr Stalmaszczyk (ed.) 2011: *Turning Points in the Philosophy of Language and Linguistics*. Frankfurt am Main: Peter Lang, 35–63.

Newtson, Darren 1973: Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology* 28.1, 28–38.

Newtson, Darren, Gretchen Enquist 1976: The perceptual organization of ongoing behaviour. *Journal of Experimental and Social Psychology* 12, 336–348.

Radden, Günter, René Dirven 2007: *Cognitive English Grammar – Series: Cognitive Linguistics in Practice* 2. Amsterdam: John Benjamins Publishing Company.

Stelma, Juurd H., Lynne J. Cameron 2007: Intonation units in spoken interaction. *Text and Talk* 27.3, 361–393.

Talmy, Leonard 2000: *Towards a Cognitive Semantics*, Vol. 1 and 2. Cambridge, Mass: The MIT Press.

Treisman, Anne, Garry Gelade 1980: A feature-integration theory of attention. *Cognitive Psychology* 12, 97–136.

Tversky, Barbara 1990: When taxonomies and partonomies meet. In: Savas L.Tsohatzidis (ed.) 1990: *Meanings and Prototypes: Studies in Linguistic Categorisation*. London: Routlege, 334–344.

Vendler, Zeno 1957: Verbs and Times. *The Philosophical Review* 66.2, 143–160.

Zacks, Jeffrey M., Barbara Tversky 2001: Event Structure in Perception and Conception. *Psychological Bulletin* 127, 3–21.

Zacks, Jeff, Barbara Tversky, Gowri Iyer 2001: Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General* 130.1, 29–58.

Streszczenie

Wydzielanie zdarzeń z nagrania wideo przedstawiającego szereg codziennych czynności

Użycie języka naturalnie wiąże się z przyjęciem pewnej perspektywy i punktu widzenia. W językoznawstwie kognitywnym mówi się, że użytkownik języka *konstruuje* scenę lub zdarzenie, wyodrębniając przy tym ich wybrane podstruktury. Badio (2014) rozróżnia *konstruowanie* od językowego *kodowania*: pierwszy termin odpowiada *konceptualizacji*, czyli poziomowi myśli, a drugi utożsamiony jest z procesem doboru środków językowych. Termin *zdarzenie* w artykule definiowany jest za Zacksem & Tversky (2001) jako segment czasu w określonej przestrzeni. Uwaga jest istotnym procesem, który umożliwia reprezentację wiedzy i użycie języka. Jej różne cechy (np. selektywność, przerzutność, podzielność) są kluczowe w procesach konstruowania i językowego kodowania. Artykuł ma dwa cele: przedstawia eksperyment dzielenia sekwencji wideo na znaczące części oraz związaną z tym zadaniem werbalizację ustną tej samej sekwencji wideo w języku polskim, ojczystym, i angielskim, obcym.

Główne pytanie badawcze związane z celem pierwszym, czyli segmentacją materiału nagranego na wideo, dotyczyło sposobu segmentacji tegoż wideo. Nagranie było krótkie, trwało około dwóch minut, i przedstawiało zwykłe domowe czynności, a całość – choć ciągła – nie układała się w spójną historię. Interesujące było, jakie części wydzielą badani, studenci filologii angielskiej UŁ. W zadaniu dzielenia nagrania wideo wykorzystany został program GOM; studenci naciskali odpowiedni przycisk w programie, a kopia obrazu w momencie naciśnięcia była zapisywana w odpowiednim folderze. Nazwa każdego kolejnego obrazka była dokładnym czasem naciśnięcia przycisku, czyli decyzji segmentacji. Badani naciskali przycisk, aby zaznaczyć w ten sposób granice znaczących części. Granice te, po przeprowadzonej analizie, okazały się zbieżne ze zmianą: kierunku ruchu, postaci, celu, przedmiotu, części ciała. Respondenci tym chętniej zaznaczali granice, im więcej takich parametrów czynności zmieniało się jednocześnie. Wyniki powyższej analizy potwierdziły również krótkie wypowiedzi pisemne po wykonaniu segmentacji. Dominującą strategią wykorzystywaną w zadaniu segmentacji było wychwytywanie jakichkolwiek zmian uznawanych za sygnały końca i początku jakiejś całości, a następnie stopniowe włączanie wyższych struktur poznawczych do oceny oglądanego materiału.

Badani zwrócili też uwagę na to, że trudno było przewidzieć, co będzie dalej, zdecydować, gdzie kończy się jeden segment aktywności, a zaczyna drugi. Ponadto niektórzy przyznali, że nie zawsze zdążyli wcisnąć przycisk. Ta ostatnia uwaga jest niezwykle wartościowa i prowadzi do wniosku, iż może lepszym sposobem przeprowadzenia badania byłoby pozwolenie respondentom na kilkukrotne przejrzenie materiału, nieograniczone czasem jego trwania. Z drugiej strony, zauważa się, że tak zaprojektowany eksperyment zmuszał badanych nieco bardziej do skupienia uwagi na dynamicznym procesie percepcji i kategoryzacji, bez czasu na dłuższą konsultację ze schematami pamięciowymi i pojęciowymi zdarzeń i aktywności, które i tak były aktywowane.

Zadanie drugie polegało na opowiedzeniu z pamięci obejrzanego materiału nagrania wideo w języku polskim jako ojczystym lub angielskim jako obcym. Celem eksperymentu było sprawdze-

nie, czy kodowanie językowe doprowadzi do wyodrębnienia tej samej lub podobnej liczby zdarzeń w porównaniu do zadania pierwszego, w którym badani musieli podzielić obejrzane nagranie, nie posługując się językiem. Ponadto nie było pewności, czy użycie języka polskiego, ojczystego, i angielskiego jako obcego da wyniki podobne co do ilości i jakości kodowanych zdarzeń. Po nagraniu wypowiedzi respondentów dokonano ich zapisu (transkrypcji) według metodologii wyznaczonej w pracach Chafe'a (1994) oraz Stelmy & Cameron (2007). Jednostką mowy jest tu tzw. jednostka intonacji skorelowana z jednym centrum uwagi. Do zapisu transkrypcji oraz późniejszej analizy zbudowano prostą relacyjną bazę danych z dwiema tabelami: MÓWCY oraz JEDNOSTKI INTONACJI powiązanymi w ten sposób, że danemu mówcy odpowiadało wiele jednostek intonacyjnych. Tabela MÓWCY miała m.in. następujące ważne kolumny: jednostka intonacji, typ zdarzenia. Typ zdarzenia pochodził z arbitralnej listy przygotowanej przed badaniem. Do typu zdarzenia, nazwanego arbitralnie, np. GIRL SIT, kwalifikowano następnie odpowiednie jednostki intonacji, na przykład:

GIRL SIT:

so I saw a girl,	no i zobaczyłem dziewczynę,
who was sitting in her .y . bedroom,	która siedziała w swojej sypialni,
..y in front eeeeeee .. her computer	yy.. przed .. eeee. Komputerem,

Oczywiście, ten sam typ zdarzenia kodowany był przez różnych uczestników badania odmiennie. Porównanie ilości zakodowanych zdarzeń w językach polskim i angielskim doprowadziło do przyjęcia hipotezy zerowej, tj. że ilość zdarzeń nie zależała od wyboru języka ($t=1.03$; $p>.05$). W ankiecie wypełnionej po badaniu uczestnicy napisali, że nie mieli trudności w kodowaniu (choć popełniali błędy językowe) i nie unikali werbalizacji niejasnych sytuacji. Wykazano ponadto statystycznie istotną różnicę w ilości zakodowanych zdarzeń w porównaniu do ilości jednostek intonacji. Zdarzenie zwykle kodowano przy pomocy przynajmniej dwóch jednostek intonacji. Użycie języka obcego wiązało się z tendencją do dłuższego przedstawienia kontekstu na początku odpowiedzi w porównaniu do użycia języka polskiego ojczystego. Ponadto przy użyciu języka ojczystego, polskiego, zakodowano więcej zdarzeń, które budowały strukturę przyczynowo-skutkową w przedstawionej na wideo sekwencji w porównaniu do kodowania przy pomocy języka angielskiego. Wszyscy uczestnicy tej części eksperymentu odtworzyli zdarzenia w sposób chronologiczny, opisując interakcje pomiędzy uczestnikami a przedmiotami. Analiza danych pochodzących od wszystkich uczestników badania wykazała również, jak trudny to proces, bez względu na wybór języka. Chociaż instrukcje mówiły jasno, że należy opowiedzieć sekwencję zdarzeń przedstawioną na nagraniu, okazało się, że pamięć i powiązane z nią użycie języka bywają twórcze. Wiele szczegółów nagrania wideo zostało bowiem zniekształconych w opisie językowym.

Jak już wspomniano, być może interesująca byłaby zmiana pewnych warunków eksperymentu. Po pierwsze, jak już sygnalizowano wcześniej, prawdopodobnie lepszą opcją w zadaniu segmentacji byłoby nieograniczanie czasu i przekazanie respondentom instrukcji, aby w celu wyodrębnienia znaczących części obejrzeli nagranie kilka razy. Po drugie, porównanie języka angielskiego jako obcego z angielskim jako językiem ojczystym byłoby chyba bardziej naturalne. Ponadto nie do końca jasny, a wart bardziej szczegółowych badań jest związek pomiędzy poziomem zdarzenia a poziomem jednostki intonacyjnej.

Bibliografia

Badio, Janusz 2014: *Construal and Linguistic Coding of Narrative Events*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Chafe, William 1994: *Discourse, Consciousness and Time*. Chicago: The University of Chicago Press.

Stelma, Juurd H., Lynne J. Cameron 2007: Intonation units in spoken interaction. *Text and Talk* 27.3, 361–393.

Zacks, Jeffrey M., Barbara Tversky 2001: Event Structure in Perception and Conception. *Psychological Bulletin* 127, 3–21.