Distinct brain signatures of content and structure violation during action observation

L. Maffongelli, E. Bartoli, D. Sammler, S. Kölsch, C. Campus, E. Olivier, L. Fadiga, A. D’Ausilio

* Istituto Italiano di Tecnologia, Genova, Italy
\[b\] Max Planck Institute, Leipzig, Germany
\[c\] Institute of Neuroscience, Université Catholique de Louvain, Belgium
\[d\] Freie Universität Berlin, Germany
\[e\] Università di Ferrara, Ferrara, Italy

**Abstract**

Sentences, musical phrases and goal-directed actions are composed of elements that are linked by specific rules to form meaningful outcomes. In goal-directed actions including a non-canonical element or scrambling the order of the elements alters the action’s content and structure, respectively. In the present study we investigated event-related potentials of the electroencephalographic (EEG) activity recorded during observation of both alterations of the action content (obtained by violating the semantic components of an action, e.g. making coffee with cola) and alterations of the action structure (obtained by inverting the order of two temporally adjacent pictures of sequences depicting daily life actions) interfering with the normal flow of the motor acts that compose an action. Action content alterations elicited a bilateral posterior distributed EEG negativity, peaking at around 400 ms after stimulus onset similar to the ERPs evoked by semantic violations in language studies. Alteration of the action structure elicited an early left anterior negativity followed by a late left anterior positivity, which closely resembles the ERP pattern found in language syntax violation studies. Our results suggest a functional dissociation between the processing of action content and structure, reminiscent of a similar dissociation found in the language or music domains. Importantly, this study provides further support to the hypothesis that some basic mechanisms, such as the rule-based structuring of sequential events, are shared between different cognitive domains.

**1. Introduction**

Understanding how the nervous system deals with the organization of motor elements into a meaningful motor plan is a central problem in the cognitive neuroscience of action. Complex human actions are composed of simple motor constituents, which might be organized according to local or hierarchical dependencies. The fundamental notion of dependency, by which movements can be arranged into a specific sequence to achieve a given goal, has profound implications for motor control (Grafton and Hamilton, 2007).

The existence of higher order planning to deploy a certain sequence of events is supported by coarticulation phenomena. Coarticulation, defined as the kinematic blending between simpler organizational units, would imply the existence of a superordinate representational layer to impose a logical and temporal order to single motor elements (i.e., goal). Coarticulation between adjacent elements could emerge at different levels of complexity. In a simple reach-to-grasp action all the effectors (arm, hand, digits) move toward the object in a very structured way. In fact, finger kinematics reflect size, position and shape of the given object, prior to the actual grasping of the object (Jeannerod, 1984). Moving up in complexity, a reach-to-grasp action can be linked to other actions to allow more difficult behaviors, thus showing a movement/action planning hierarchy. Here, the same grasping action shows subtle electromyographic (EMG) differences when it is embedded in different complex actions like reaching and grasping for food with the goal of eating or placing food away (Cattaneo et al., 2007). This evidence reveals coarticulation-like effects at higher levels of complexity.
Motor-related activity during action observation presents similar phenomena of coarticulation. Monkeys inferior parietal lobule neurons, coding for a simple grasping action, show a modulation of activity when this movement is part of different actions (e.g., grasp to eat/grasp to place). This is true for both action execution and observation (Fogassi et al., 2005), suggesting that the same neural circuitry involved in action execution has the capability to support the discrimination of subtle kinematic dependencies between adjacent action units.

Everyday actions are typically complex and thus formed by relatively long sequences of subunits (motor acts). Each of them will be slightly different according to the local goal and context (Cattaneo et al., 2007; Lacoboni et al., 2005). As a result, complex actions are built according to a very limited number of correct sequences leading to the same goal. Thus, motor knowledge about the correct sequencing of actions might be exploited in action sequences leading to the same goal. Generally, action knowledge about the correct sequencing of actions might be exploited in action execution and observation by constraining the anticipation of the goal of an observed action. If this is the case, violations occurring in the normal flow of an observed action should elicit specific brain responses that could give information about the mechanisms supporting others’ action processing.

Generally, action violation paradigms consist of the manipulation of action content rather than action structure. In fact, two kinds of content violations have been devised, such as the manipulation of tool use knowledge or tool canonical function. In the first case, the violation entails the wrong use of a tool in a given action (i.e., to accomplish the action of “teeth brushing”; the toothbrush has to be grasped at the handle side and not at the brush side). In the second case, the violation consists of using the wrong or non-canonical tool in a given action (i.e., grasping a knife to brush teeth). Both violations do not permit goal achievement and elicit a large frontal negative deflection in the electroencephalogram (EEG) (Amoruso et al., 2013; Balconi and Caldiroli, 2011).

On the other hand, it might be possible to formulate another kind of action violation, by manipulating the structure of the observed action. According to the structural dependencies existing between action units (e.g., reach to grasp) to form complex actions (e.g., make coffee), a certain sequence will be disrupted at one specific point if the order of two adjacent elements is inverted.

We thus designed an event-related potential (ERP) study using an action-observation paradigm in which sequences of pictures representing complex familiar goal-directed actions were presented. Each sequence picture depicted an action requiring an agent interacting with different objects in the correct temporal order. The experimental material was manipulated by introducing a “Content violation” or a “Structure violation.” We hypothesized that action structure and action content violations would elicit different ERPs patterns. We further predicted that the dissociation would match the one often observed in other domains, such as language (Kutas and Hillyard, 1980; Lau et al., 2008) and music (Koelsch, 2011; Steinbeis and Koelsch, 2008). Content violations in language (i.e., semantic violations) usually elicit a negative ERP peaking around 400 ms after stimulus onset (for reviews see, e.g., Kutas and Federmeier, 2011; Amoruso et al., 2013). In contrast, a structure violation should elicit an earlier anterior negativity (with latencies of around 180–300 ms). When task-relevant, such early negative ERPs are followed by a later posterior positivity (with a latency of around 500–800 ms) in both language (Friederici et al., 1993; Steinhauer and Drury, 2012) and music (Koelsch et al., 2005; Patel et al., 1998).

2. Methods and materials

2.1. Participants

Twenty participants took part in the electroencephalographic (EEG) study (mean age: 26 years; SD: 2.9; 11 females) after giving informed consent, in accordance with the Declaration of Helsinki. The protocol was approved by the local ethical committee ASL-3 (“Azienda Sanitaria Locale”, Local Health Unit, Genoa). All the subjects included in the study had normal or corrected-to-normal vision and were right-handed (Oldfield, 1971). Two participants were excluded due to excessive EEG artifacts, and two for low task performance.

2.2. Materials

We used sequences of static pictures depicting 20 different complex actions. Each action sequence was composed of 8 color pictures (resolution 1920 × 1280 pixels). Every sequence showed a goal-directed action requiring an agent interacting with different objects in a specific temporal order (the list of actions is provided in Appendix A). The first picture (scenario frame) displayed all the objects needed to execute the forthcoming action. The subsequent 7 pictures captured critical agent-object interactions, important for the understanding of every single step of the specific action (i.e., coffee making action: scenario frame: a table with a coffee maker, a cup, coffee powder and a water carafe; picture 2: the actress picks up the coffee machine; picture 3: she opens the coffee machine; picture 4: she fills the machine with water; picture 5: she fills the machine with coffee powder; picture 6: she closes the coffee machine; picture 7: she puts the coffee machine on the stove; picture 8: she serves coffee).

Actions were manipulated in two ways, by introducing either a Content violation or a Structure violation. In the Content violation condition the actor used an incongruent object in one of the pictures (i.e., to put cola instead of water in the coffee machine) while in the Structure violation condition we inverted the temporal order of two adjacent pictures in the sequence (i.e., the sequence order was 1, 2, 3, 5, 4, 6, 7, 8; see Fig. 1).

Each of the 20 sequences was repeated 3 times (counterbalanced order across sequences, leading to 60 trials) in each of the 3 experimental conditions (Structure violation condition, Content violation condition and Control condition, in which no manipulation was introduced), resulting in a total of 180 trials.

2.3. Procedure and trial structure

The participants were seated comfortably in a darkened room, in front of a 17-in. computer screen (distance 70 cm). They were asked to avoid movements, and to blink, if necessary, between trials. The experimental session was divided into 4 blocks, each containing 45 random trials from all conditions. Participants took short breaks between blocks and could also rest during the presentation of videos (2 seconds long) showing images of landscapes.

An internet-based experiment tested whether the experimental material (pictures) was conveying the action content and goal. Order and timing of picture presentation was the same as in the EEG experiment. After the scenario picture, the participants had to write down the action that was going to be executed (Task 1). At the end of the sequence subjects had to decide (forced choice) whether the action contained a violation of some kind (Task 2). All the 20 actions were presented in the 3 conditions (Control, Structure and Content), for a total of 60 trials. Fifty-seven subjects (mean age = 28.8 years, 27 of which were males) completed the experiment. Subjects were always recognized the action following the Scenario (Task 1). Accuracy in Task 2 was 83% (SD = 0.37%). In the control condition was 91% (SD = 0.28%), 82% (SD = 0.37%) in the Content and in the Structure condition 75% (SD = 0.43%).
(every 20 trials). The experiment lasted about 2.5 h.

Specifically, each trial began with the presentation of a fixation point (1 s). Participants were exposed to 8 pictures, and were instructed to attend to all of them. The first picture of each sequence represented the scenario picture (3.5 s) and immediately after its display the subjects were asked to describe the type of action (Task 1). They had to answer verbally to this task (i.e. “making coffee”) and the experimenter took note of that. After the participant’s response, the experimenter pressed a button to show the remaining 7 pictures. These pictures were displayed for 1 s each and were separated from the successive one by an inter-stimulus interval of 1 s during which a fixation point was presented (Fig. 1). At the end of the sequence, an action-unrelated question (i.e. “was the background in the pictures white?”) appeared on the screen to assess the subject’s attention (Task 2). The subjects’ reaction times were recorded with a two-buttons response pad. The assignment of the left and right buttons to “yes” and “no” responses was counterbalanced across participants. The inter-trial interval was 3.5 s. The experiment was controlled with Psychtoolbox functions (Brainard, 1997), running in Matlab (Mathworks, Inc.).

2.4. EEG Recording and pre-processing

EEG activity was acquired with an electrode cap (32 active Ag/AgCl electrodes, arranged according to the international 10–20 system) using Brain Vision Recorder software (Brain Products, München, Germany). The data were recorded with a Brain Amp MR+ amplifier (Brain Products, above) applying a sampling rate of 500 Hz, a 10 s time constant and a high cutoff frequency of 250 Hz, referenced to Afz. Electrode impedance was kept below 5 KΩ. Data were re-referenced offline to the algebraic mean of TP9 and TP10 (Light et al., 2010), filtered offline with a 0.1–45 Hz band-pass filter and then down-sampled to 250 Hz. Artifacts were removed through visually inspected Independent Component Analysis (ICA) implemented in EEGLAB (Delorme and Makeig, 2004), considering time, topographic and spectral distribution of the components. Data were segmented in epochs from −200 to 1000 ms relative to the onset of the critical picture in the sequence, using a common baseline (from −200 ms to 0 ms).

2.5. Data analysis

Subjects always recognized the action following the scenario picture (Task 1). Response times (RTs) to the attentional question (Task 2) were calculated from the offset of the last picture within a time window of 2 s. RTs were evaluated using a repeated measures analysis of variance (rm-ANOVA) considering SEQUENCE TYPE as the within-subject factor (Control, Structure, Content).

ERPs were obtained for each of the sequence types (Control, Content, Structure) by averaging corresponding epochs, and compared in EEGLAB with an rm-ANOVA. The choice of statistical time windows for both types of violations was based on previous studies on action, music and language processing (i.e. Balconi and Caldironi, 2011; Koelsch et al., 2005; Friederici et al., 1993). Therefore, in each condition, we used the same two time windows to investigate the hypothesized early (100–250 ms) and late effects (300–750 ms) of violations. Regions of interest (ROIs) were defined as follows: left-anterior (F3, F7, FC1, FC5), right-anterior (F4, F8, FC2, FC6), left-posterior (CP1, CP5, P3, P7) and right-posterior (CP2, CP6, P4, P8).

Mean amplitude values in each ROI, time window and condition were then exported and statistically compared with R (R Development Core Team, 2008). Two separate analyses (rm-ANOVAs) were performed depending on the violation type (Content versus Control and Structure versus Control). The two rm-ANOVAs included the factors CONDITION (Violation, Control), ROI (Left-Anterior, Right-Anterior, Left-Posterior, Right-Posterior) and TIME (Early, Late), with the average potential as dependent variable.

When sphericity assumptions were violated, based on Mauchly’s test, we report Huynh-Feldt-corrected p-Values. Post-hoc analyses were performed by means of paired t-tests applying false discovery rate (FDR) correction for multiple comparisons. The level of significance was set at p < 0.05. In order to explore the possible modulation of effects due to the repeated presentation of picture sequences we performed an additional ANOVA on a subset of trials, including only the first presentation of each picture.
The complete results of all ANOVAs (full dataset and first presentation) are reported in Table 1 for the Content condition and in Table 2 for the Structure condition.

### Table 1

<table>
<thead>
<tr>
<th>Effect</th>
<th>df1</th>
<th>df2</th>
<th>All trials</th>
<th>1st presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>p</td>
<td>F</td>
<td>p</td>
</tr>
<tr>
<td>Condition</td>
<td>1</td>
<td>15</td>
<td>2.25</td>
<td>ns</td>
</tr>
<tr>
<td>ROI</td>
<td>3</td>
<td>45</td>
<td>5.59</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>15</td>
<td>1.18</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Condition × ROI</td>
<td>3</td>
<td>45</td>
<td>0.11</td>
<td>ns</td>
</tr>
<tr>
<td>Condition × time</td>
<td>1</td>
<td>15</td>
<td>2.75</td>
<td>ns</td>
</tr>
<tr>
<td>ROI × time</td>
<td>3</td>
<td>45</td>
<td>17.9</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Condition × ROI × time</td>
<td>3</td>
<td>45</td>
<td>3.82</td>
<td>&lt; 0.05</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Effect</th>
<th>df1</th>
<th>df2</th>
<th>All trials</th>
<th>1st presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>p</td>
<td>F</td>
<td>p</td>
</tr>
<tr>
<td>Condition</td>
<td>1</td>
<td>15</td>
<td>0.44</td>
<td>ns</td>
</tr>
<tr>
<td>ROI</td>
<td>3</td>
<td>45</td>
<td>5.32</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>15</td>
<td>7.67</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Condition × ROI</td>
<td>3</td>
<td>45</td>
<td>0.64</td>
<td>ns</td>
</tr>
<tr>
<td>Condition × time</td>
<td>1</td>
<td>15</td>
<td>7.12</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>ROI × time</td>
<td>3</td>
<td>45</td>
<td>13.7</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Condition × ROI × time</td>
<td>3</td>
<td>45</td>
<td>9.99</td>
<td>&lt; 0.05</td>
</tr>
</tbody>
</table>

3. Results

3.1. Behavioral data

The rm-ANOVA on RTs in Task 2 did not reveal any effect of Condition ($F(2,30)=3.06; \ p=0.07$; Control: Mean = 1.76 s; SD = 0.30 s; Content: Mean = 1.84 s; SD = 0.38 s; Structure: Mean = 1.79 s; SD = 0.30 s). Accuracy was very high (Control = 96%, Content = 90%, Structure = 92%), suggesting that the subjects were engaged in the observation of the experimental material.

3.2. Neurophysiological data

3.2.1. Effect of Content violation

The Content condition showed a bilateral posterior negativity in the 300–750 ms time window post stimulus onset (Fig. 2 and S1).²

Post-hoc analyses run on the triple interaction CONDITION × ROI × TIME revealed significantly different values for the Content violation compared to the Control condition in the Late time window in both the left-posterior ROI ($t(15)=-3.18, \ p<.05$) and the right posterior ROI ($t(15)=-2.80, \ p<.05$) (Figs. 2 and 4). Regarding the main effect of the factor TIME, the late time window showed more negative values compared to the early time window (mean difference = –1.3 μV; SD = 2.1 μV).

3.2.2. Effect of structure violation

The Structure condition showed two different ERP components, an early left anterior (100–250 ms) negative deflection followed by a left anterior positive potential starting around 300 ms post stimulus onset and lasting until 750 ms (Fig. 3 and S1).

Post-hoc analyses run on the triple interaction CONDITION × ROI × TIME, showed significantly different values for the Structure violation with respect to the Control condition in the left-anterior ROI in both the Early time window ($t(15)=-2.49; \ p<.05$) and in the Late time window ($t(15)=6.18; \ p<.05$; Figs. 3 and 4). Thus, this result points out an enhanced involvement of the left-anterior ROIs compared to the left-posterior ROIs. Post-hoc analysis run on the two-way interaction CONDITION × TIME, revealed that in the early time window the Structure violation evoked more negative amplitudes than the Control condition ($t(63)=2.33, \ p<.05$); in the late time window the Structure violation led to more positive values compared to the Control condition ($t(63)=-2.54, \ p<.05$). Post-hoc analysis run on the two-way interaction ROI × TIME, showed that in the early time window anterior ROIs showed more positive values compared to the posterior ROIs, in both left ($t(31)=5.09, \ p<.05$) and right hemisphere ($t(31)=4.83, \ p<.05$). Regarding the main effect of the factor TIME, the early time window displayed more negative values than the late time window (mean difference = –0.9 μV; SD = 2.0 μV).

4. Discussion

This study investigated the neural markers of complex action violation processing during action observation. With the introduction of violations affecting either the content or the structure of actions, we found typical ERP patterns reminiscent of those previously described for semantic versus syntactic violations in the language and music domains (Friederici, 2011; Koelsch, 2011), namely a negativity between 300 and 750 ms for the Content violation and an earlier negativity (100–250 ms) followed by a positivity (300–750 ms) for the Structure violation.

4.1. Content violation

The Content violation elicited a larger negative deflection in the 300–750 ms time window, associated to the use of a non-canonical tool to fulfill one of the action sub-goals. Recent studies report a similar N400 potential when the so-called “world-knowledge” about typical human actions is violated (Proverbio and Riva, 2009), or when tools are used in a non-canonical manner (Balconi and Caldiroli, 2011). As a consequence, it has been proposed that the N400 may not be a purely linguistic marker of semantic processing, but it may rather extend to other cognitive domains, including action observation (Amoruso et al. 2013). Our results corroborate this interpretation, with some intriguing differences.

In the action domain, as far as the topography is concerned, the N400 deflection was previously found to be frontally distributed, in response to incongruous conditions (Balconi and Caldiroli, 2011; Balconi and Vitaloni, 2014). Since the linguistic N400 generally shows a central-parietal distribution, a semantic violation in the action domain is believed to cause a frontal shift (West and Holcomb, 2002). Our data show a bilateral parietal scalp distribution for the N400. An important point to emphasize is that all the aforementioned studies introduced the violation in the last frame (i.e. end of the action sequence), whereas our violation was in the middle of the sequence. In this way we manipulated only one inner sub-goal of the whole action sequence instead of the final action goal. However, further investigations are needed to demonstrate whether the frontal N400 reflects the neural processing.

² When analyzing only trials in which action sequences were presented for the first time, thus eliminating repetitions effects, we found that the interaction detected for the overall data remained significant (all $p<.05$), indicating that the results of our main analysis are robust and not merely due to repeated stimuli presentation.
of the main action goal as opposed to sub-goals.

Nevertheless, the pattern found for N400 suggests that this is a neural marker of a temporally quite specific process, which probably monitors the online unfolding of actions and extracts locally relevant information about tool use, as opposed to a process that extracts the global meaning of actions. Noteworthy, other evidence could support a more posterior negativity. For instance, studies on apraxia (Kalénine et al., 2010) and fMRI (Gallivan et al., 2013) indicate that various aspects of tool knowledge are associated with the left posterior middle temporal gyrus. It is noteworthy to mention that the same region, bilaterally, has been identified as one of the generators of the N400 component (Lau et al., 2008).

Concerning the lateralization aspects, studies in the gesture processing domain also reported a centro-parietal negativity for the comparison of meaningful versus meaningless hand postures (Gunter and Bach, 2004). A similar non-lateralized negative deflection was observed when comparing pictures that are semantically incongruous to a previously presented object name (Hamm et al., 2002), or when a tool is used in an inappropriate way (Bach et al., 2009). On the contrary, the preparation of meaningful vs. meaningless actions embedded in a semantic categorization task preceding the required motor act, elicited a right lateralized N400 (Van Elk et al., 2008).

More generally, our results contribute to the growing body of evidence suggesting an amodal processing of content information. Indeed, the N400 has been found for semantic violations in lexical decision processes (Kutas and Hillyard 1980; Kutas and Federmeier 2000), violation in the musical domain (Daltrozzo and Schön, 2009; Goerlich et al., 2011; Koelsch, 2011) as well as in visual scene processing (Võ and Wolfe, 2013). All these studies would primarily corroborate the conception of a qualitative similarity between language and music content processing which extends to other domains such as action processing.

4.2. Structure violation

The action structure condition elicited an early negativity (100–250 ms) followed by a positivity (300–750 ms). The effects reported in the current study are suggestive of the (early) left anterior negativity (ELAN) that appears to be correlated with the initial syntactic processing of local dependencies in language, whereas a late centro-parietal positivity (P600) is associated with the difficulty of syntactic integration and explained in terms of reanalysis of the entire structure (i.e., Hahne and Friederici, 1999). Likewise, in the music domain, a similar ERP pattern was shown in response to structural irregularities (Koelsch, 2011; Patel et al., 1998). However, early negativities elicited by music-structural irregularities, often reveal a right anterior scalp distribution (ERAN) (Koelsch, 2011).

We observed that an early negativity followed by a late positivity effect is only present in the Structure condition. This pattern seems to indicate the specificity of such violation as opposed to the Content condition. More precisely, the early negativity may reflect the monitoring of local structural dependencies between coarticulated adjacent elements. The following positivity might instead represent the later reanalysis of the whole sequence. Indeed, actions consist of simple motor constituents that can be organized in a sequence to reach a desired goal (Grafton and Hamilton, 2007). However, the number of sequences that can be performed to reach a specific goal is limited. Thus, each goal would be characterized by temporal rules like those of the coffee-making example: first grasp for the coffee machine and then fill it with water. These rules constitute the motor knowledge by which actions are formed. We postulate that this biphasic pattern (ELAN-positivity) mirrors the fast and local detection of an incongruent action unit followed by a later re-analysis of the entire sequence, respectively. Regarding the topography of these responses, our results prove that both the early negativity and the late positivity were observed in the left-anterior ROIs. The finding of an early negativity in anterior regions is in agreement with language and music-related ELAN and ERAN, respectively. Concerning the late positivity, we found an anterior distribution, which is at odds with that of language and music, usually showing a centro-parietal localization (Hahne and Friederici, 1999; Patel et al., 1998). This anterior shift may be interpreted in two manners. The first is based on the fact that action stimuli could call for action-specific processes located in anterior brain regions. This assumption is in accordance with neuroimaging studies showing that action
observation elicits bilateral frontal activations (Aziz-Zadeh et al., 2006) with some degree of prominence of the left hemisphere (Gazzola and Keysers, 2009). Furthermore, such a pattern of lateralization is consistent with previous evidence highlighting how alternated finger movements, as opposed to repeated ones, induce extra-activity on the left hemisphere (Dirnberger et al., 2006).
Otherwise, the anterior shift can be explained by specific characteristics of our stimuli. In fact, it is worth mentioning that a late positivity with a frontal distribution has also been observed in language tasks (Kaan and Swaab, 2003). In particular, the frontally distributed late positivity is probably related to ambiguity resolution and/or to an increase in discourse level complexity. In this sense, the anterior shift may be explained by the use of long sequences of sub-actions concatenated to each other for the achievement of a complex goal. Nevertheless, future studies are needed to elucidate the nature and modulation of the late positivity evoked by structural action violation.

4.3. Communalities between language, music and action

Emerging evidence indicates that from an evolutionary point of view, gestural development in infants takes place before the appearance of communication skills. Indeed, the relationship between gestures and language during the development of communication abilities in children is very tight (Kraljević et al., 2014). Similar assumptions, along with recent findings in great apes (Bard et al., 2014), has led to theories suggesting that spoken languages derive from sign languages (Caselli et al., 2012).

One aspect of these theories is the quest for common primitives shared by the domains of language and action. A current area of debate on this issue is whether “action syntax” exists and, if so, what its characteristics are (Moro, 2014; Pulvermüller, 2014). The notion of syntax was further applied also to human actions in studies exploring the cognitive processes (Allen et al., 2010; Greenfield, 1991) and their neural correlates (Farag et al., 2010; van Schie, et al., 2006); Allen et al. (2010) demonstrated that purposive action sequences are processed in terms of a means-ends parse, which is a formal/abstract specification of how actions are linked together in achieving a goal. Means-ends parses may relate to the syntactic frames that bear the abstract structure of a sequence. In the same year, Farag et al. (2010) showed the contribution of the frontal cortex to clustering events according to the hierarchical as opposed to linear-sequential organization of complex familiar activities. The novelty of our findings is that violations of action Structure, elicit electrophysiological patterns resembling those elicited in language and music studies, corroborating the hypothesis that language, music and action share some structural analogies (Fadiga et al. 2009).

Furthermore, additional theoretical support to the action-language analogy is given by the typological generalization regarding the majority of human languages. Effectively, the basic word order is either SOV (subject-object-verb) or SVO (subject-verb-object). Such ordering arrangement is explained by the prototypicality of transitive action scenarios in which an animate agent acts on an inanimate patient (i.e. the entity upon whom an action is carried out) in order to induce a change of state. Indeed, actions, like verbs in language, show a similar argument-structure, which connects agents and objects. Generally, in a transitive scenario there are two principles (based on semantics) that cannot be “violated”. These are the “subject salience” and the “verb–object–contiguity”. The first principle refers to the fact that a subject usually precedes objects because the agent is at the head of the causal chain affecting the patient; the second principle concerns the fact that verbs and object are usually adjacent because it is the agent’s action to change the state of the patient (Comrie, 1989; Greenberg, 1963). Following this proposal, we suggest that our action sequences were transitive in nature and here we implemented a violation of the transitive relation between specific pictures (sequence disruption), thus interrupting the essential relation between agent and patient of the specific scenario.

Studies focused on the neural correlates of these processes, proposed that the prevalent word order pattern reflects the way of linearizing and nesting the core conceptual components of action in Broca’s area (Kemmerer, 2012). In support of such neural localization, lesions in Broca’s area have been associated with deficits in gesture recognition (Pazzaglia et al., 2008) and impaired capability to reorganize the correct order of pictures representing separate steps of human action sequences (Fazio et al., 2009). Neuroimaging research revealed that Broca’s area, and its right homologue, may control selection and nesting of action units, integrated in hierarchical behavioral plans, regardless of their temporal structure (Koechlin and Jubault, 2006). Finally, further proof for the contribution of Broca’s area (and its right homologue) to syntactic processing comes from studies localizing the sources of the (early) negativities for language and music syntactic violations (Maess et al., 2001). Taken together, these results seem to indicate that Broca’s area may form a node of a neural circuit responsible for processing supra-modal hierarchical structures (Fitch and Martins, 2014).

Acknowledgments

This work was supported by European Community grant POETICON + + (STREP Project ICT-288382) and by Italian MIUR (PRIN). We would also like to acknowledge Laura Taverna for helping us in the preparation of the experimental stimuli.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuropsychologia.2015.05.020

Appendix A

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Control condition</th>
<th>Structure violation</th>
<th>Content violation</th>
</tr>
</thead>
<tbody>
<tr>
<td>To wash the hair</td>
<td>To wet the hair</td>
<td>To apply shampoo in the hair</td>
<td>To wet the hair</td>
</tr>
<tr>
<td>To hang up a picture</td>
<td>To apply shampoo in the hair</td>
<td>To apply gel in the hair</td>
<td>To take a cup</td>
</tr>
<tr>
<td>To send a letter</td>
<td>To take the hammer</td>
<td>To hammer with the hammer</td>
<td>To hammer with a cup</td>
</tr>
<tr>
<td></td>
<td></td>
<td>To hammer with the hammer</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>To affix the stamp on the envelope</td>
<td>To insert the letter in the envelope</td>
</tr>
</tbody>
</table>
Short description of the stimuli (i.e. the action used in all conditions). For the experimental conditions, the reported sentences refer only to the critical frames of the sequence that were manipulated.

Experimental trial timeline. (A) Sequence of events in the Control trial for one action (to prepare a coffee). After the presentation of the first picture (Scenario picture), participants were asked to report which action was going to be executed (Task 1). At the end of the entire sequence, an action-irrelevant attentional question was presented (Task 2). (B) The sequence critical pictures in the Structure (red frames)
and Content (green frames) condition, as opposed to the control condition (black frames).

Grand-average event-related potentials (ERP) of the Control condition (black line) compared to the Content violation (red line) measured in the selected regions of interest (ROIs). The vertical dashed line represents the onset and offset of the selected time windows. Light gray boxes represent non-significant time windows; dark grey boxes refer to significant time windows. Please refer to Fig. 4 for a schematic description of the selected scalp ROIs.

Grand-average event-related potentials (ERP) of the Control condition (black line) compared to the Structure violation (red line) measured in the selected regions of interest (ROIs). The vertical dotted line represents the onset and offset of the selected time windows. Light gray boxes represent non-significant time windows; dark grey boxes refer to significant time windows. Please refer to Fig. 4 for a schematic description of the selected scalp ROIs.

(A) The left side shows the topographical maps depicting the distribution of the effect in the Control condition and in the Structure condition over both time windows. The right side shows maps of the Control condition and the Content violation condition.

(B) Bar graphs describing the ERP amplitudes of violations (Structure violation in red, Content violation in green and the Control condition in black). Results are reported only for the ROIs showing significant effects for the early (100–250 ms) and late (300–750 ms) time windows. Significant effects are highlighted by an asterisk. On the right side, the scalp channel configuration is shown, together with the ROIs in which effects are reported, for the Structure violation (red) and the Content violation (green).

References


