

Incremental processing of telicity in Italian children

Jasmijn E. Bosch, Mathilde Chailleux & Francesca Foppolo*

Abstract. A sentence like ‘Lyn has peeled the apple’ triggers a *telicity inference* that the event is telic and to a *culmination inference* that the event has reached its telos and has stopped. This results in the final interpretation of the sentence that Lyn has completely peeled the apple. We present an eye-tracking study to test children’s ability to predict the upcoming noun (e.g., *the apple*) during the incremental processing of sentences like ‘Show me in which picture Lyn has peeled the...’ in which the predicate is telic and the verb appears in the perfective form. By means of the Visual World Paradigm, our aim was to investigate children’s ability to use the lexical semantics and aspectual morphology of verbs during language processing and comprehension. To test if children can predict the target (e.g., a completely peeled apple) by exploiting the lexical-semantic meaning of the verb, we contrasted the picture of the target with the picture of an object that cannot be peeled; to test if they can predict on the basis of the verb’s perfective morphology, we compared the target with the picture of a half-peeled apple. Our results show that Italian children anticipate the upcoming noun in both cases, providing evidence that they can incrementally exploit the morphosyntactic cue on the verb (perfective aspect) to derive the culmination inference that the telos is reached, and the event is completed. We also show that the integration of aspect requires some additional time compared to the integration of basic lexical semantics of the verb.

Keywords. aspect, semantics, anticipation, eye-tracking, visual world paradigm

1. Introduction. It is a well-known fact that, during sentence processing, we can anticipate upcoming nouns on the basis of the lexical semantics of prenominal verbs. In their pioneer study, Altmann & Kamide (1999) showed that participants anticipated the word ‘cake’ when hearing the verb ‘eat’, which was revealed by increased looks towards the picture of a cake in a visual scenario in which the cake was the only edible object. Other studies have shown that we can also integrate morphosyntactic information, such as tense expressed on the verb, to anticipate the target referent. In another study employing the Visual World Paradigm, Altmann & Kamide (2007) showed that, when participants heard a sentence like ‘The man *will drink...*’, they looked more at a full glass of beer (in which its content is yet to be drunk), while they looked more at the empty glass of wine when hearing the sentence ‘The man *has drunk...*’. This suggests that, as listeners, we quickly and incrementally integrate the inference conveyed by the verb’s morphology that an event of drinking has taken place and was completed, resulting in an empty glass.

In fact, there is more than lexical information and tense in a verb. Verbs denote events, and events are complex phenomena. They stretch along the time dimension and can be classified as *punctive* or *durative*. For example, events such as ‘blowing out a candle’, which happen immediately, are called punctive and events like ‘peeling an apple’, which take some time, are called durative. Independently of their duration, predicates can also have a *telos* (or a culmination point); this property, however, is not (or not always) intrinsic to the lexical properties of the predicate itself,

*This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska Curie grant agreement No 765556.

The affiliation of all authors is the University of Milano-Bicocca. Contact emails: jasmijn.bosch@unimib.it, mathilde.chailleux@gmail.com, francesca.foppolo@unimib.it.

but it pertains to the combination of a certain predicate (e.g., ‘peeling’) with a certain complement. This is called the *aktionsart*. Considering this property of events, predicates can be classified as *atelic* or *telic*. Under this distinction, predicates like, for example, ‘running’ or ‘peeling apples’ are classified as atelic, since they do not have a clear ending or culmination point. On the other hand, predicates like ‘running the Boston Marathon’ or ‘peeling an apple’ are classified as telic, due to the fact that there is a clear point in which the event culminates (namely, when the finish line in the marathon is reached or when the apple is completely peeled).

There is another layer to add to the time dimension of the processing of events: in many languages, verbs carry tense and aspectual morphology. While *tense* information deals with the collocation of the event in time along a linear dimension of past-present-future, *aspect* deals with the status of the completion of the event. With respect to this distinction, a verb like ‘is peeling’ carries the imperfective aspectual information, signaling the fact that the event of peeling has started at some point in the past, but is still ongoing. In the case of telic predicates like ‘is peeling the apple’, this means that the action of peeling is still happening, and the *telos* has not been reached yet. When combined with perfective aspect, instead (i.e., ‘has peeled the apple’), the information conveyed is that the event of peeling has stopped, and the *telos* has been reached.

Thus, in a simple sentence like ‘Lyn has peeled the apple’, two different layers of information are carried by the verb and are integrated online during sentence comprehension. The first one is the layer of the *aktionsart*, which carries the information that the event is durative and telic (since the durative predicate is combined with a definite noun phrase, ‘the apple’). The second layer consists of the aspectual morphology realized on the verb, which carries the information about the status of completion of the event. Together, these two layers of information point to a *telicity inference* that the event is telic and to a *culmination inference* that the event has reached its *telos* and has stopped. This results in the final interpretation of the sentence that Lyn has completely peeled the apple.

The nature and interplay of these inferences in the derivation of the meaning of the sentence are debated in theoretical semantics, and they go beyond the purposes of this paper. From a psycholinguistic perspective, the question that can be tested experimentally is whether (and when) this interpretation is carried out incrementally during sentence processing.

Previous works have focused on the time course of the telicity inference. For example, Proctor et al. (2004) used a self-paced reading study to examine the speed and accuracy with which readers draw telicity inferences during on-line language comprehension. Participants read sentences containing either a consumption verb (e.g., ‘consume’) or an observation verb (e.g., ‘monitor’) followed by either a mass or a count object (e.g., ‘ice water’ vs. ‘ice cube’), which could trigger an atelic or telic interpretation of the predicate. Each sentence ended with an adverbial phrase that was either consistent (e.g., ‘in 8 minutes’) or inconsistent (e.g., ‘for 8 minutes’) with a telic verb. Proctor and collaborators report no slowdown in the final adverbial region in sentences like ‘Leslie consumed Polar Purity’s ice cube for eight minutes’, while a slowdown was observed in the same sentence when the adverbial modifier ‘in eight minutes’ was used. They interpret this result as evidence that the inference that an event is telic seems not to be computed until there is evidence that it is needed and that, when it is computed, this inference is associated with a computational cost (see also, Pickering et al. 2006, Townsend 2012).

Few works have used the visual world paradigm to test the incremental processing of aspect. Among these, Zhou, Crain & Zhan (2014) tested Mandarin-speaking adults and children aged 3 to 5 with sentences in which the verb carried a perfective aspectual morpheme (*le*) or a durative aspectual morpheme (*zhe*) in a scenario contrasting a completed vs. an uncompleted event. The results show that both the adults and the children (of all age groups) looked more at the completed

event (e.g., a woman who has finished planting a flower) when hearing the perfective morpheme, while they looked more to the ongoing event (e.g., a woman in the process of planting a flower) when hearing the durative morpheme. This effect occurred immediately after the onset of the aspectual morpheme (appearing at the end of the verb), showing that even young children are able to use the temporal information encoded in aspectual morphemes as rapidly as adults to facilitate event recognition.

Summing up, we know that we process lexical and tense features of verbs incrementally, and that we make use of such cues to identify objects that are compatible with the lexical semantics of the verb and to separate completed from yet-to-be events (as shown in Altmann and Kamide's studies mentioned above). We also know that we make use of aspectual cues to separate ongoing vs. completed events, as shown by Zhou and colleagues. Previous reading studies, instead, suggest that we do not immediately commit to the fact that the event is telic and has a culmination point, even when we encounter a durative predicate followed by a definite noun phrase.

However, some questions still remain unanswered. Suppose that we already know, from the visual context, that we are facing a telic event such as peeling an apple, and we are shown different degrees of completion of the same peeling-an-apple event (e.g., a half peeled apple and a completely peeled apple). In such a case, it is unclear when listeners commit to the fact that the telos is reached, and consequently, when they start looking at the completely peeled apple while hearing a sentence like 'Show me where Lyn has peeled the...'. This question has been addressed experimentally by Foppolo, Greco, Panzeri and Carminati (2016), who tested Italian adults with a Visual World Paradigm. Results show that participants fixated on the completed event while hearing the verb, providing evidence for an incremental and rapid integration of aspectual cues during sentence processing. Building on this study, we tested incremental processing of verbs in Italian-speaking children, by contrasting the integration of lexical-semantic information and aspectual cues during sentence processing.

2. Methods.

2.1 PARTICIPANTS. We tested 35 monolingual Italian children between the ages of 8 and 10. The data of six children had to be excluded because of poor calibration of the eye-tracker. The final sample consisted of 29 children (11 boys and 18 girls), with a mean age of 9 years and 4 months ($SD = 11$ months). Participants were recruited at a primary school in the urban area of Milan. Prior to testing, all parents signed a consent form that was approved by the ethics committee of the University of Milano-Bicocca.

2.2 MATERIALS. We created a visual world eye-tracking experiment, in which we tested to what extent children were able to rely on aspectual and lexical information during online sentence processing. The experiment was implemented in E-prime 3 (Psychology Software Tools, Pittsburgh, PA). Participants saw a visual scenario with two pictures depicting completed or ongoing events. The pictures were colored photographs focusing on the hands of a person who was involved in an action or who had just completed an action using one or more objects. No faces were shown. We used partly the same pictures as Foppolo et al. (2016), although additional stimuli were created for the purpose of this study. The two images (577 x 408) were shown on the left and right side of a grey screen (1920 x 1080), with a clear space in between them.

Regarding the auditory stimuli, participants listened to transitive sentences in Italian with a verb in the *passato prossimo*, in which the auxiliary *ha* was combined with a past participle to trigger the completion interpretation. For example, participants heard *Guarda in quale foto ha colorato la stella* ('Look in which picture (he/she) colored the star'). Three different experimental conditions were

created: two in which the target could be anticipated while processing the verb and one in which no anticipation was possible (these are labelled Early and Late respectively). In the Early Lexical condition, the correct picture could be selected on the basis of the lexical meaning of the verb (e.g., an object that can be colored, like a drawing, vs. an object that cannot, like a Lego tower, cf. Figure 1, A-B). In the Early Aspect condition, the pictures displayed two actions at a different state of completion (e.g., a half-colored star vs. a fully colored star, cf. Figure 1, A-C), so that the correct picture could be selected based on the aspectual information morphologically expressed by the verb that should trigger the inference that the telos is reached. In the Late condition, the target picture could only be selected upon hearing the direct object; the predicate could apply to both objects, so the sentence remained ambiguous until the final noun (e.g., a fully colored star vs. a fully colored leaf, cf. Figure 1, A-D). An overview of the experimental conditions is provided in Figure 1.

The audios were recorded by a female native speaker of Italian, and manipulated using Praat (Boersma, 2001), so that the introduction of the sentence (*Guarda in quale foto ha* ‘Look in which picture he/she has’) was always the same. The auxiliary *ha* always started 2400 ms after the start of the trial, and the mean onset of the direct object (i.e., the article preceding the final noun) was at 3668 ms. The mean length of the experimental sentences was 4890 ms.

We used a Latin square design with three lists of 21 items (seven per condition). Participants were assigned to one of the three lists. Items were presented in a randomized order.

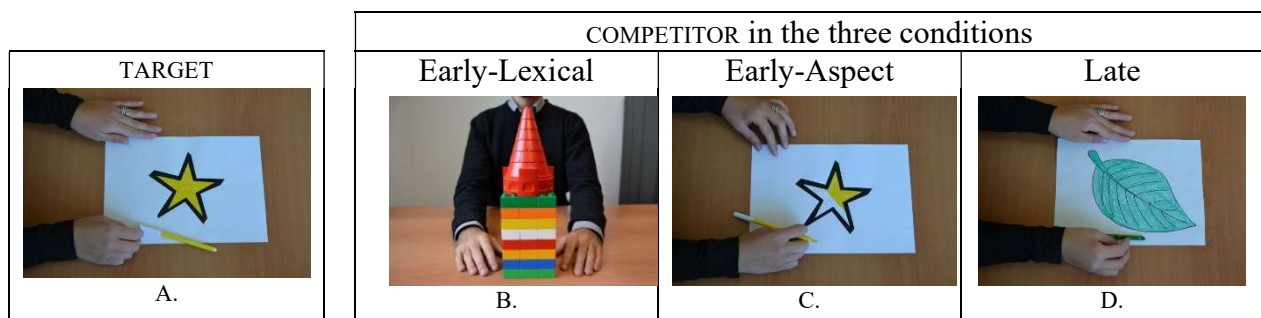


Figure 1: Target (A) and competitor images in each of the three conditions.

2.3 PROCEDURE. Children were tested individually in a quiet room within the school, using a portable Tobii Pro X3-120 eye-tracker which captured their gaze at 120 Hz. Participants were seated between 60 and 70 cm from the display. Calibration took place after a short familiarization phase, consisting of one example and three practice items.

During the experimental phase, participants listened to sentences through headphones while their eye movements were recorded. At the end of each sentence, a question mark appeared on the screen, and children could give their offline response by clicking on the mouse to select the correct picture. After that, a fixation cross appeared, ensuring that children were looking at the center of the screen before moving on to the next trial. Participants did not receive any feedback about their performance during the experiment.

2.4 ANALYSIS. We performed a track loss analysis on the eye-tracking data during the experimental sentences. Trials which had more than 35% data loss were removed from the analysis. As a result, 49 trials were removed. Moreover, we excluded trials with inaccurate offline responses, which left us with a total of 540 remaining trials for the analysis of the eye gaze data.

We used the *eyetrackingR* (Dink & Ferguson, 2015) and *ggplot2* (Wickham, 2016) packages in *R* (R Core Team, 2019) to visualize the eye gaze pattern. The data were then analyzed with a

generalized linear mixed effect models in *R*, using the *glmer* function of the *lme4* package (Bates, Maechler, Bolker & Walker, 2015).

Sentences were divided in three time windows: the introduction (*Guarda in quale foto* ‘Look in which picture’), the verb (e.g., *ha sbucciato* ‘has peeled’), and the noun phrase (e.g. *la mela* ‘the apple’). The boundaries of the time windows were shifted by 200 ms, to take into account the time that is required to plan and execute a saccadic eye movement (Altmann, 2011).

The statistical analysis tested whether the likelihood of looking at the target (versus competitor) during the noun phrase depended on the experimental condition (Early Aspect versus Early Lexical versus Late). When comparing Early Aspect against the baseline condition, Late was coded as -0.5 and Early Aspect was coded as $+0.5$. When comparing the two Early conditions, Early Aspect was coded as -0.5 and Early Lexical was coded as $+0.5$. The model also included random intercepts for Item and Subject.

3. Results. The analysis of the offline responses showed a high overall accuracy on the task; 96.7% of the trials were answered correctly. In the analysis of the eye-tracking data, we only focused on accurate trials only. The time course pattern of the eye gaze data is shown in Figure 2.

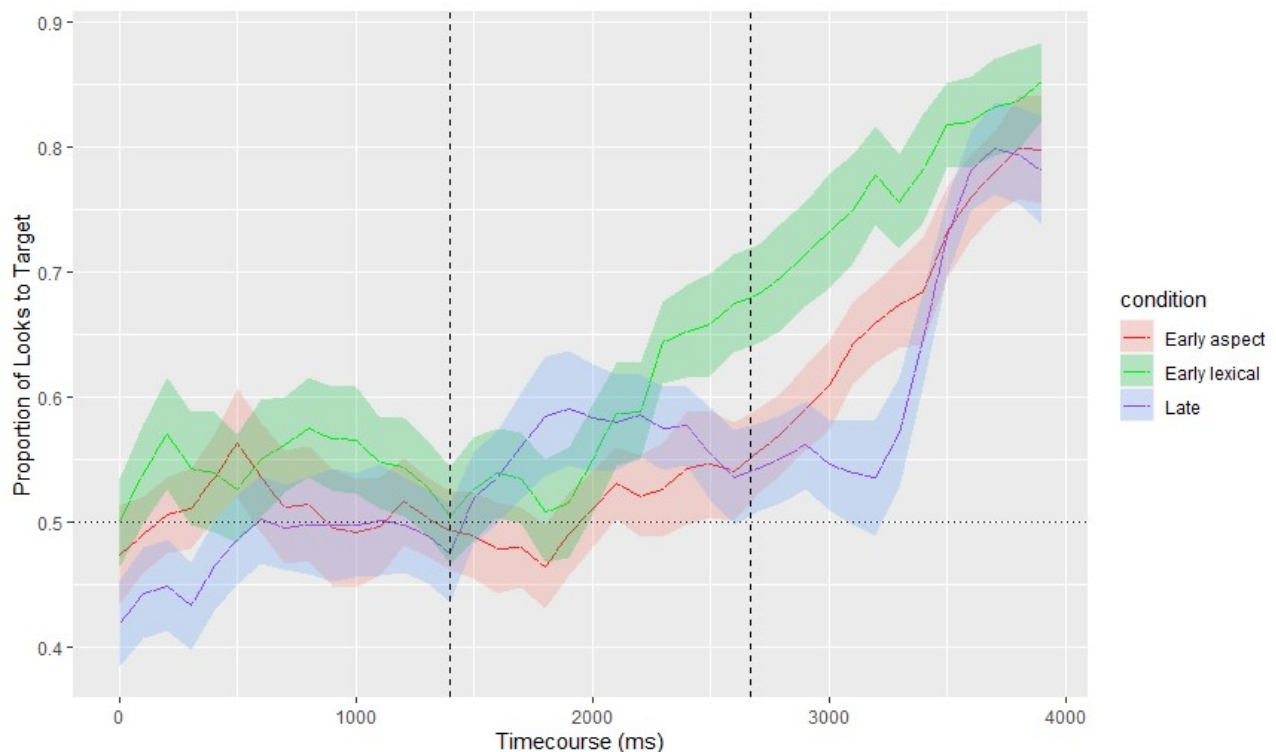


Figure 2: Time course of the proportions of looks toward the target (versus competitor) in the three conditions. The first vertical line indicates the onset of the auxiliary verb; the second vertical line indicates the average onset of the article preceding the final noun. The dotted horizontal line represents chance performance.

As can be seen from this plot, in the Early Lexical condition, participants started directing their gaze toward the target picture during the second time window, which suggests that they immediately integrated the lexical meaning of the verb while processing the sentence. In contrast, in the Early

Aspect condition and the Late condition, we only observe a shift toward the target picture during the noun phrase. Nevertheless, participants appear to be faster in the Early Aspect condition than in the Late condition, suggesting that participants were able to make rapid use of the aspectual information on the verb during online sentence processing.

In the statistical analysis we focused on the odds of looking at the target during the direct object time window in the three conditions. The summary of the model output is provided in Table 1.

| Fixed factor | Est. odds ratio | 95% CI | <i>p</i> |
|---|-----------------|--------------|----------|
| Condition (Early Aspect vs Early Lexical) | 2.06 | 1.94 .. 2.18 | <.001 |
| Condition (Early Aspect vs Late) | 1.83 | 1.74 .. 1.94 | <.001 |

Table 1: Output of the Generalized linear mixed model defined as:
Looks to target (yes or no) ~ condition + (1 | item) + (1 | subject)

These results show that participants were significantly more likely to look at the target in the Early Lexical condition than in the Early Aspect condition, but they were also significantly more likely to look at the target in the Early Aspect condition than in the Late condition. This confirms the observation that children in this study were sensitive to grammatical aspect, although they were significantly faster when they could rely on simple lexical semantics.

4. Discussion. In this paper we presented a Visual World eye-tracking experiment conducted with Italian children on the incremental processing of telic predicates. Previous results with Mandarin speaking children and adults show that listeners can distinguish between perfective and imperfective morphemes, providing evidence of a rapid integration of aspectual cues during sentence processing. Additionally, previous results with Italian adults on similar materials show rapid integration of the completion inference during sentence processing.

Our aim was to extend previous research to address two experimental questions: (1) Is the culmination inference derived incrementally by children? (2) If yes, at which stage of the derivation is it computed? To address question (1) we designed an experiment in which listeners could anticipate the target picture by exploiting linguistic cues on the verb; to address question (2) we contrasted two types of anticipatory cues: one related to verb’s lexical semantics and one related to aspectual (perfective) morphemes on the verb.

We discuss three main findings. First, our results show that Italian children can anticipate upcoming nouns on the basis of the lexical semantics of verbs, as already observed in classic studies with adults.

Second, children were faster to shift their gaze to the target in the Early-Aspect than in the Late (control) condition. Thus, children relied on the aspectual cue on the verb, since semantics alone did not provide enough information to disentangle the two events depicted in the scenario (recall that in this condition the event was the same, shown at different degrees of completion). This finding provides evidence that children can exploit a morphosyntactic cue denoting perfective aspect to derive the culmination inference that the telos has been reached and the event has been completed, and that they do so incrementally.

Third, we found earlier anticipatory eye-movements to the target in the Early-Lexical compared to the Early-Aspect condition. This result suggests that lexical-semantic information provides a faster cue than aspectual information, and that the culmination inference derived in the Aspect condition requires some additional time compared to the integration of basic lexical semantics of the verb. This may reflect an additional cost of the derivation of the culmination inference, which

might take more time for children, despite the fact of being derived incrementally. One possible explanation for this delay relates to the process of integrating visual and linguistic cues during sentence processing, which might be particularly challenging in the Early-Aspect condition. In this case, two steps are required for the identification of the target: (i) the identification of the event, which is triggered by the lexical semantics of the verb, and (ii) the identification of the degree of completion of the event, which is triggered by morphosyntax (i.e., the combination of auxiliary and past participle). While only step (i) is required in the Early-Lexical condition, step (ii) is also necessary to anticipate the target in the Early-Aspect condition. We speculate that this additional step might explain why, at least in children, the effect of anticipation shows up later in a condition that requires the additional integration of aspectual information, in comparison to a condition in which relying on lexical semantics suffices. Future research should investigate this issue further.

References

- Altmann, Gerry T. & Kamide, Yuki. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3). 247-264.
- Altmann, Gerry T. & Kamide, Yuki. 2007. The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57(4). 502-518.
- Altmann, Gerry T. 2011. The mediation of eye movements by spoken language. In *The Oxford handbook of eye movements*.
- Bates, Douglas, Maechler, Martin, Bolker, Ben & Walker, Steve. 2015. Fitting Linear Mixed Effects Models Using lme4. *Journal of Statistical Software*, 67(1). 1-48.
- Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5:9/10. 341-345.
- Dink, Jacob W. & Ferguson, Brock. 2015. eyetrackingR: An R library for eye-tracking data analysis. *Online: www.eyetracking-r.com*.
- Foppolo, Francesca, Panzeri, Francesca, Greco, Ciro, & Carminati, Maria N. 2016. The incremental processing of accomplishment predicates. Talk at the Workshop Events in Language & Cognition, 29th Annual CUNY Conference on Human Sentence Processing. University of Florida, Gainesville, Florida.
- Pickering, Martin J., McElree, Brian, Frisson, Steven, Chen, Lillian & Traxler, Matthew J. 2006. Underspecification and aspectual coercion. *Discourse Processes*, 42(2). 131-155.
- Proctor, Andrea S., Dickey, Michael W. & Rips, Lance J. (2004). The time-course and cost of telicity inferences. In *Proceedings of the Cognitive Science Society*, 26 (26). 1107-1112.
- Psychology Software Tools, Inc. [E-Prime 3.0]. 2020.
- R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Townsend, David J. 2013. Aspectual coercion in eye movements. *Journal of psycholinguistic research*, 42(3). 281-306.
- Zhou, Peng, Crain, Steven, & Zhan, Likan. 2014. Grammatical aspect and event recognition in children's online sentence comprehension. *Cognition*, 133(1). 262-276.