First things first: cross-linguistic analyses of event apprehension

Apprehension is the rapid process during which we extract the 'gist' of a visual stimulus. People can retrieve information on an event's category, structure and roles in as little as 20-400ms (Dobel et al., 2007; Hafri et al., 2013; Castelhano & Henderson, 2007), when dealing with complex line drawings or real-world event photographs. It is however, notoriously difficult to isolate these initial visual processing stages from later processing, in particular when using verbal output ('name x or y') as the only measure.

Here, we investigate to what extent apprehension is guided by top-down factors, i.e., the *task demands* (nonverbal task, agent/action naming, event description) and the *language background* of the viewer, by analysing *first fixations* on event stimuli with only 300ms exposure time. Given the brevity of exposure, processing demands are high. Under increased cognitive load, people may resort to automatized visual processing routines that have been shown to be influenced by language-specific description patterns (Trueswell & Papafragou, 2010; Gerwien & Flecken, 2016).

We compared Dutch and Chinese (N=24 participants in each group), hypothesized to differ in the degree of *agent-orientedness* in causative event encoding: In contrast to Dutch, Chinese a) allows topic-drop (sentences with null subjects/agents), b) does not mark gender/number in pronouns, c) marks grammatical aspect (perfective LE, progressive ZAI), specifying action-contours, d) uses resultative verb compounds encoding action-results (e.g., to cut-become two pieces); these features should lead to general *lower* agent-saliency, compared to event encoding in Dutch.

We present crosslinguistic data from a novel eye-tracking paradigm in which participants were exposed to real-world event photographs of *causative actions* (agent engaged in action on object, e.g., woman opening a can, man peeling a banana) for 300ms only. Pictures appeared randomly in one of the four corners of the screen, ensuring that the location containing information on relevant event elements (agent, action/object) was unpredictable (Figure 1). Participants performed three tasks across blocks: 1) nonverbal task: some pictures were presented twice and participants were asked to respond to repetitions; 2) agent or action identification (between subjects): participants said aloud the name of the actor (e.g., Tom) or the action (e.g., cutting (paper)); 3) event description: participants constructed a full sentence answering "what happened in the picture?". Task order was pseudo-randomized across lists. We analyze the location of the very first fixation (on the agent or action/object) in isolation.

We hypothesize an interaction between language background and task demands: cross-linguistic differences in first fixations are not expected during agent/action naming; however, in event description, Chinese participants should be more likely to place their first fixation on the action/object, whereas Dutch speakers are more likely to fixate the agent as starting point for visual-linguistic processing (cf. Gerwien & Flecken, 2016, for German). An open question is, to what extent language-specific processing patterns are reflected in the nonverbal task.

We present our results (data collection is currently underway) and discuss them in the context of scene apprehension and cross-linguistic event encoding research.
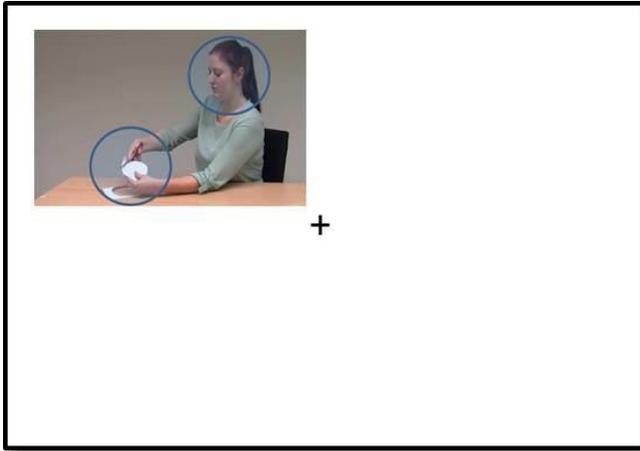
Figure 1: Stimuli appeared randomly in 1 of the 4 corners of a screen, after participants had fixated a centred fixation cross (circles indicate areas of interest, not visible to participants) (the fixation cross disappeared when a stimulus appeared).