SESSION 1: Eye-tracking technology: Latest developments

Tobii or not Tobii?
Assessing the validity of eye-tracking data: Challenges and solutions

Coralie Vincent[1], Efstathia Soroli[2], Helen Engemann[3], Henriëtte Hendriks[4], & Maya Hickmann[1]

[1] *CNRS & University of Paris 8, France*
[2] *CNRS & University of Lille, France*
[3] *University of Mannheim, Germany*
[4] *University of Cambridge, UK*

Eye tracking (ET) methods become more and more popular in psycholinguistic research because they offer the possibility to record visual processing in real time, allowing for the study of the relation between cognition and language, two systems often considered independent (Pinker, 1994).

In order to evaluate the impact of specific language properties on online visual processing, we coupled a production task with an ET paradigm. A total of 473 native speakers of two typologically different languages (234 English and 239 French) within three age groups (142 seven-year-old children, 155 ten-year-old children, and 176 adults) were tested in a production task involving 36 dynamic motion scenes (videos), that first had to be visually explored and then verbally described (e.g., "a man running up a hill").

With respect to the ET data, which is the main focus of the present paper, and in order to properly compare the gaze patterns of the groups, a thorough validity check (pre-processing and quality assessment) was necessary. Indeed, validity is an issue that is almost never addressed in psycholinguistic research, even though an increasing number of researchers report it as one of the main sources of methodological bias (Holmqvist et al., 2011). Apart from the fact that a recording may include segments that are irrelevant for the analysis (e.g., eye blinks, off-screen fixations), it has been found that low quality data may misleadingly point to group differences in gaze behaviour, for instance between adults and children (Wass et al., 2014). More specifically, low precision due to incorrect gaze detection may "flatten out" the gaze distribution across different areas of interest (AOIs) or across groups, while low robustness (i.e., resulting from missing or fragmented data) can make *visit durations* seem shorter than they actually are, and thus bias interpretation of results.

The present paper compares results obtained with a turnkey solution (namely, Tobii Studio) to results obtained with in-house developed algorithms that: (a) carefully discard irrelevant parts of the recording; (b) exclude gaze initiation latencies; and (c) detect and compensate for spatial inaccuracies of the ET data. The findings show that turnkey solutions may be only relevant for some designs (i.e., more appropriate for static/picture material). However, design-adapted validity checks (pre-processing the recordings and quality assessment) as well as target-related compensations of inaccuracies, as proposed in this paper, are crucial and should be common practice for researchers who wish to compare gaze patterns or to evaluate group differences objectively. Challenges related to typical ET measures, such as *gaze proportions* to different dynamic AOI and *visit durations* are also discussed as they seem to be sensitive and subject to change due to validity-related factors.

References

- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. van de. (2011). *Eye tracking: A comprehensive guide to methods and measures* (1st ed.). New York, NY: Oxford University Press.
- Pinker, S. (1994). *The language instinct: How the mind creates language*. New York, NY: William Morrow & Company.
- Wass, S. V., Forssman, L., & Leppänen, J. (2014). Robustness and precision: How data quality may influence key dependent variables in infant eye-tracker analyses. *Infancy*, 19(5), 427-460.