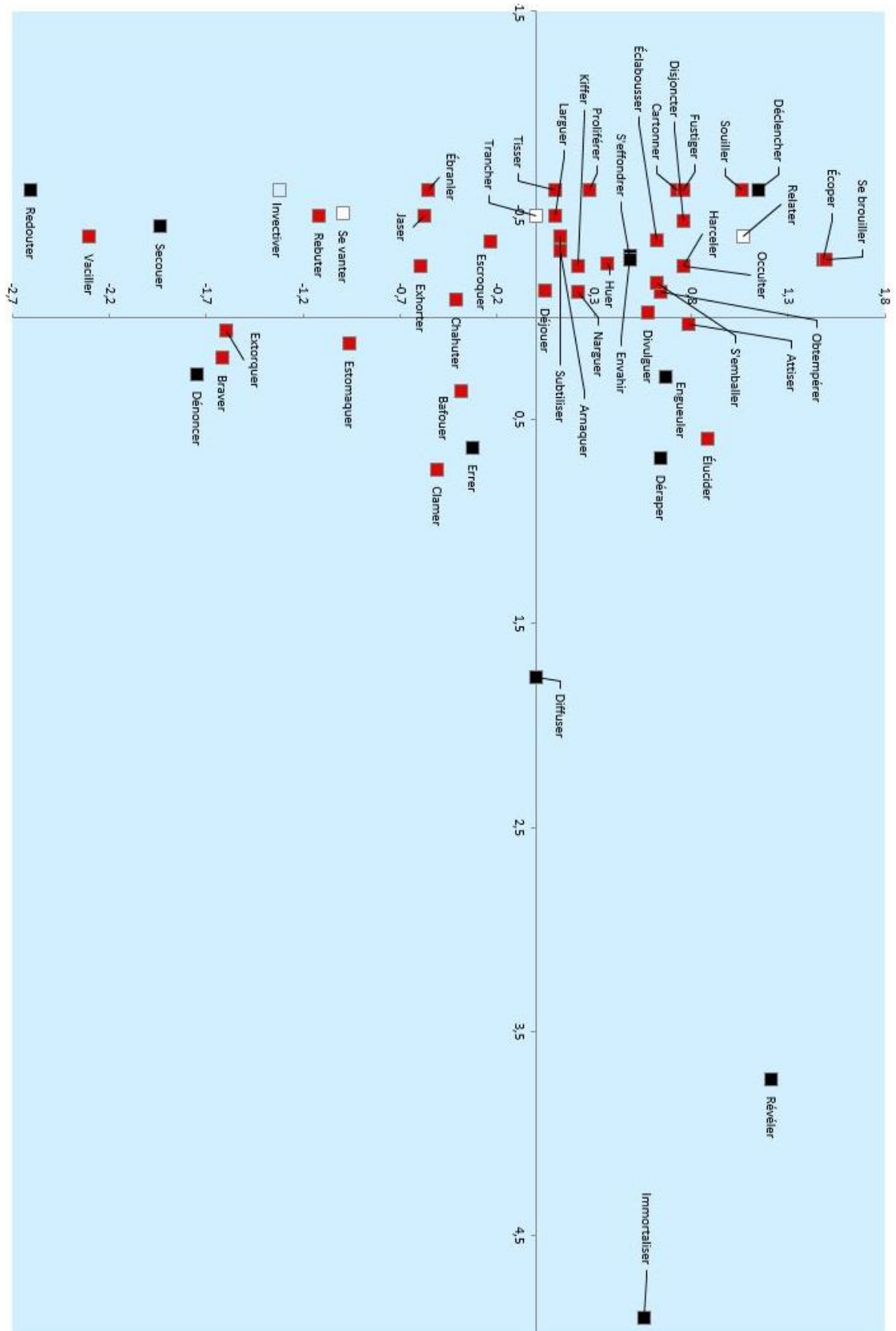


Supplement 1. Selection target words.

The scores on the meaning recognition test, the score on the inferencing test, and the score given by teachers were combined in order to select the target words. The score for the meaning recognition test and the inferencing test were standardized (z -transformed) and plotted on a cartesian grid (see figure below). For the teachers' ratings, a color code was given to each candidate target (red = not likely to be known, white = possibly known, black = likely to be known). Prime candidates would be red verbs in the upper left quadrant of the cartesian grid, that is, not likely to be known as rated by the teachers (red), most successful inferencing score (Y-axis), and lowest scores on the meaning recognition test (X-axis). Most targets (80 %) of the final selection of target words (20 items) met these three conditions. However, due to the concern of life-like authenticity of the news items (i.e., reflecting natural language use and the meaning of the targets had to be relevant to the content), some candidate targets that were considered to better fit the content of the articles (20 %) did only satisfy two of the three optimal conditions.



Candidate targets plotted on a cartesian grid. X-axis represents standardized scores on the meaning recognition test. Y-axis represents standardized scores on the meaning inferencing test. Red = not likely to be known, white = possibly known, black = likely to be known.

