**Table S1**

*Stimuli characteristics for target words*

| | Cognates ($n$ = 40) | | Noncognates ($n$ = 40) | |
|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* |
| Length | 5.98 | 2.02 | 5.38 | 1.66 |
| No. of syllables | 1.90 | 0.84 | 1.68 | 0.80 |
| SUBTLEX-US (log10) | 3.26 | 0.68 | 3.24 | 0.72 |
| Orthographic neighbors | 4.39 | 4.84 | 3.78 | 4.25 |
| Phonological neighbors | 8.92 | 10.94 | 8.50 | 8.37 |
| Normalized Levenshtein distance | 0.75 | 0.24 | 0.16 | 0.15 |

## Appendix S1

### Self-paced reading task (Experiment 2): Filler sentences

The fillers comprised 90 implausible and 30 plausible sentences of different syntactic types that did not resemble the target sentences, such as temporal clauses. Implausibility of the filler sentences was defined as either a very unlikely event (e.g., someone watering the TV) or an impossible event (e.g., a dancing tomato). The plausible-implausible ratio was unbalanced because of the participants' possible different sentence interpretation strategies for the experimental sentences. A target-like English syntax strategy or an agent-first strategy (Bever, 1970; VanPatten, 2015) would mean that 40 sentences would appear plausible and 40 implausible (i.e., a 50% plausible-ratio). A plausibility strategy (assigning semantic roles according to world knowledge rather than syntax; Ferreira, 2003) would mean 80 sentences would appear plausible and 0 implausible (i.e., a 100% plausible-ratio). Hence, a 75% implausible-ratio was used for the filler sentences so that even participants who applied a plausibility strategy for the experimental sentences would encounter a reasonable number of implausible sentences and would thus not be pushed to abandon their strategy due to an insecurity that may arise from choosing one response option considerably more often than the other.

**Appendix S2**

**Self-paced reading task (Experiment 2): Plausibility norming study**

Twenty-four advanced students of English from the University of Groningen (mean age = 22.4, *SD* = 2.55; mean self-rated English proficiency = 9.24/10, *SD* = 0.85) who did not take part in the main experiment were instructed to assess the plausibility of experimental and filler sentences on a scale from 1 (*highly implausible*) to 5 (*highly plausible*). The 80 reversible experimental sentences were presented as main clauses in both plausible and implausible versions across two lists, so that each participant rated 40 plausible and 40 implausible experimental sentences. Each participant saw each item in only one of two possible versions. Additionally, each list contained all 120 filler sentences, amounting to each participant rating a total of 200 sentences. The results of the norming study are presented in Table S2. A paired-sample *t*-test yielded a significant difference of plausibility, with the plausible experimental sentences being rated as far more plausible than the implausible ones, $t(23) = 75.05$, $p < .001$.

**Table S2**

*Mean and standard deviation (in parentheses) of the plausibility norming study*

| Sentence type | Example | Plausibility norming[a] |
|---|---|---|
| Experimental | | |
| Plausible | *A woman buys a dress in the store.* | 4.77 (.23) |
| Implausible | *A dress buys a woman in the store.* | 1.11 (.17) |
| Filler | | |
| Plausible | *An athlete runs a marathon.* | 4.82 (.26) |
| Implausible | *A cow practices the piano.* | 1.89 (.42) |

[a]on a scale from 1 (*highly implausible*) to 5 (*highly plausible*)