**Appendix G: Description of the matched data set**

**Summary dataset (Appendix D “df-summary.csv”)**

matchid: identifies participants who were matched with each other

edu: number of years in full-time education

native: whether the participant is a native speaker of English

native\_lang: the participants native language(s)

lang\_subfamily: the subdivision within a language family to which the participant’s first language belongs

lang\_family: the language family of the participant’s native language

first\_exp: age at first exposure to English

eng\_edu: estimated number of hours of instruction in English

LoR: length of residence in an English-speaking country in months

AoA: age of arrival in an English-speaking country

variety: variety of English spoken

group: experimental group (native, classroom learner or immersion learner)

score\_sp: score on the spoken GJT

mean\_rt\_sp: mean reaction time on the spoken GJT

mean\_wr: score on the written GJT

mean\_rt\_wr: mean reaction time on the written GJT

score\_ps: score on the picture selection task

mean\_rt\_ps: mean reaction time on the picture selection task

**Long format (Appendix E “df-long.csv”, Appendix F “df-long.RDS”)**

participant: individual participant ID

response: "1" for correct response, "0" for incorrect response or time-out

task: task ("sGJT" for spoken GJT, "wGJT" for written

GJT, "PST" for picture selection task)

item: question item (1-80 for the two GJTs, P1-P80 for PST)

group: experimental group (native, classroom learner, immersion learner)

taskident: dummy variable for model with "1" for sGJT and wGJT, and "0" for PST

gramm: grammaticality of the question item ("G" for grammatical, "U" for

ungrammatical, "ps" for all PST items)