SupMat_1: Annotation development process

A pilot study was conducted to determine which variables could be reliably annotated. In brief, 500 instances (250 spoken, 250 written) from the BNC were annotated twice by the same annotator, a native speaker of British English with an MA in Linguistics. Variables included were those 10 that are typically mentioned in relation to article usage; these variables are listed in the table below.

The effect of these variables on article choice was measured by establishing the cross-tabulated frequency association between the article counts and variable counts. First, we retained those variables which returned a significant effect on article choice, as per the chi-squared test (column 4). Next, the training corpus was annotated a second time for those variables which passed the test, by the same annotator; the Cohen's Kappa test was employed to assess intra-rater reliability (agreement determined at Cohen Kappa rating >0.80). 5 variables were found to have a significant effect on article choice ($p<0.05$), i.e., Count, Number, Elaboration, Specific Referent and Hearer Knowledge. The data was annotated again for these 5 variables; the Cohen-Kappa ratings given in column 5. Because the Cohen Kappa rating was >0.80, all 5 variables were retained for annotation of the full sample.

Name	**Variable**	**Values**	**Significance**	**Cohen-Kappa**
Count	Does the article precede a count noun?	Yes, no	>0.0001***	0.94
Number	Does the article precede a noun which is singular, plural, or non-count?	Singular, plural, no	>0.0001***	0.87
Proper_Noun	Does the article precede a proper or common noun?	Proper, common	0.09193	-
Concrete	Does the article precede a noun which is abstract or concrete?	Abstract, concrete	0.9075	-
Elaboration	Is the noun elaborated on and, if so, where?	Before, after, both, neither	>0.0001***	0.95
Sentence_Initial/ Turn_Initial	Is the article sentence/turn initial?	Yes, no	0.1103	-
Phrase_Initial	Is the article phrase initial?	Yes, no	0.6484	-
Phrase	Is the article part of a phrase (eg idiom)?	Yes, no	0.4248	-

Specific_Referent (SR)	Does the article precede a noun which has a specific referent?	Yes, no	<0.0001***	0.90
Hearer_Knowledge (HK)	Does the article precede a noun of which the hearer is aware	Yes, no	>0.0001***	0.89