Morphophonological and conceptual effects on Dutch subject-verb agreement

Inés Antón-Méndez

Utrecht University, The Netherlands

University of New England, Australia

Robert J. Hartsuiker

Ghent University, Belgium

Please address correspondence to:

Inés Antón-Méndez

Discipline of Psychology

School of Behavioural, Cognitive and Social Sciences

University of New England

NSW 2351

Australia

Tel. +61 0267733765

Email: iantonme@une.edu.au

Authors' note

Both authors contributed equally to the work presented here, the order in which they appear is not a reflection of their involvement. We thank Eva De Haan and Siene van Hoorick for their assistance, and several anonymous reviewers of a previous version of the paper for their useful comments and suggestions.

Abstract

Language production theories differ in their assumptions about the information flow between levels. Serial models hypothesize that different types of information, such as conceptual factors and morphophonological make up, would have an effect at different points during the implementation of agreement and would, therefore, not interact. Constraint-based models, on the other hand, entail an interplay of these two types of factors. Here, we present data from an experiment designed to test whether a conceptual factor (notional number) interacts with a morphophonological factor (determiner number ambiguity) resulting in an increased number of subject-verb agreement errors in Dutch. Analyses showed main effects of both factors but no interaction. We also carried out simulations of one specific serial model for production of subject-verb number agreement, the Marking and Morphing model (Eberhard, Cutting, & Bock, 2005). Our simulations in Dutch yielded an excellent fit between the model and these data (as well as other previously collected data). In conclusion, our results argue in favour of independent processing of these two types of information during agreement production and, more specifically, offer support for the Marking and Morphing model.

Key words: Language production, number agreement, morphophonology, distributivity.

Most languages possess one or another form of agreement (Ferguson & Barlow, 1988), meaning that one sentence constituent covaries in accordance with another sentence constituent with respect to a specific grammatical property within a given syntactic domain. Speakers, therefore, need a mechanism to ensure the application of this covariance where needed, irrespective of a sentence's complexity or distance between the agreeing terms. Like in other areas of psycholinguistics, researchers are divided as to how they consider this particular phenomenon to be implemented on-line. Some researchers argue for a deterministic procedure mostly blind to not directly relevant influences (e.g., Bock & Levelt, 1994), and some argue for a more probabilistic approach where numerous factors can leave a mark on the final product (Haskell & MacDonald, 2003; Vigliocco & Franck, 1999). The road to determining which account is correct has been meandering ever since Bock and Miller (1991) found subject-verb number agreement in English to be affected by the presence of a second noun or attractor in the sentence's subject phrase (e.g., "the key to the cabinets was old"). Initially, Bock and Eberhard (1993) found that only syntactic manipulations had an effect on agreement error elicitation while morphological and semantic factors in the form of pseudoplurals and collective nouns, on the other hand, had no effect. These results accorded nicely with a very strict serial model, in which agreement processes operate on abstract symbols, without regard to meaning and form. This strong claim, however, was relatively quickly disconfirmed by experimental results showing both meaning and form effects.

Evidence of semantic influences on agreement comes from three sources. First, and most importantly for our purposes, effects of notional number have been found in several languages (Eberhard, 1999; Vigliocco, Butterworth, & Garrett, 1996; Vigliocco, Butterworth, & Semenza, 1995; Vigliocco, Hartsuiker, Jarema, & Kolk, 1996) where subject phrases with a distributive meaning – that is, which conjure a plural concept even though the subject heads are syntactically singular (e.g., *the label on the bottles*) – result in more agreement errors than non-distributive phrases (e.g., *the road to the mountains*). Second, there are plausibility influences (Thornton & MacDonald, 2003); (see also Hupet, Fayol, & Schelstraete, 1998) to the effect that the likelihood of the attractor noun as a subject of the verb modulates the degree of attraction. For example, for a preamble such as *the album by the classical composers*, a verb like *praised* induces more agreement errors than a verb like *played*, which can only refer to *the album*. Third, there is an effect of semantic integration, described as "the degree to which phrases are tightly linked at the conceptual level" (Solomon & Pearlmutter, 2004), where phrases which are more tightly linked in memory induce more agreement errors, presumably because tightly linked concepts are more likely to be activated simultaneously.

With respect to morphophonological effects on agreement, there are two types of evidence. The most relevant for us, is the effect of number ambiguity of the subject head (Vigliocco, et al., 1995) or the determiner accompanying the subject head (Hartsuiker, Schriefers, Bock, & Kikstra, 2003) which is associated with an increase in the rate of agreement errors. For example, the Dutch definite determiner used with common-gender nouns (*de*) is ambiguous for number, as it is used with both the singular and plural versions of the noun. In contrast, the definite determiner for neuter-gender nouns (*het*) is unambiguous, because it can only be used in the singular (plural neuter-gender nouns take *de*). Hartsuiker et al. (2003) observed a much larger attraction effect in the ambiguous *de*-head noun condition than in the unambiguous *het*-head noun condition. Furthermore, case ambiguity of the attractor noun phrase (NP; Hartsuiker, Antón-Méndez, & Van Zee, 2001; Hartsuiker, et al., 2003; Nicol & Antón-Méndez, 2008) also increases the rate of agreement errors. For example, the Dutch pronoun *ze* is ambiguous between nominative and accusative case (it can mean *she, they*, or *them*), whereas the pronoun *hen* is unambiguously accusative (*them,* with a human antecedent). Hartsuiker et al. (2001) observed a much larger attraction effect in the ambiguous *ze*-direct object pronoun condition than in the unambiguous *hen*-direct object pronoun condition.

Models of language production in general and of agreement implementation in particular can easily account for the bulk of the evidence in favor of both notional and morphophonological effects on agreement by allowing back and forth information exchange between processing levels either by virtue of generalized interactivity (Haskell & MacDonald, 2003; Thornton & MacDonald, 2003; Vigliocco & Franck, 1999), or by appealing to feedback loops between the different levels (Hartsuiker, et al., 2003).

These effects, however, have also been accounted for within serial or partly serial models of agreement production like the Marking and Morphing model (Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001; Eberhard, Cutting, & Bock, 2005), which is a serial model, and the model of Franck, Vigliocco, Antón-Méndez et al. (Franck, Vigliocco, Antón-Méndez, Collina, & Frauenfelder, 2008), which is a partially serial model. Franck et al.'s model is able to account for both types of effects by virtue of local interactivity at the level of lexical selection, without postulating more generalized interactivity encompassing syntactic processing. The Marking and Morphing model explains the two types of effects within a fully serial framework, however, and it is worth describing how in some detail. According to this model, there are two distinct agreement operations: number marking and number morphing. Number marking takes place during the mapping from the message to the lexical-grammatical representation, resulting in the assignment of number to the subject phrase rather than just specifying the number of the different lexical entries selected. This explains why phrases such as *Mary and John* or *the majority of my students* correctly agree with a plural verb even though the subject heads are singular. Consequently, it also naturally accounts for message level influences on attraction errors such as the distributivity effect. During number morphing the subject phrase is built by assembling the constituent morphological units and, after any discrepancy between an eventual morpheme and the number marking on the subject phrase has been resolved, the right number feature is implemented on the agreeing verb. At this point, therefore, any morphological manipulation would be able to exert an influence on the eventual result of the agreement operation, which would explain the morphophonological effects.

When considered independently, semantic and formal effects are thus unproblematically explained by both interactive and serial models. What has not yet, so far as we know, been investigated is whether the two factors interact. The general agreement in the literature is that interactions between factors associated with different levels of processing support interactivity at the processing level (Stenberg, 1969), and an absence of such interactions, conversely, is consistent with seriality in processing. Although we will revisit this assumption in the Discussion, at this point we would like to follow the standard additive factor logic to its natural conclusions.

The additive factors logic implies that an interaction between the two factors considered here would be compatible with interactive models of agreement computation, but not with strictly serial ones. In particular, while Franck et al's model would also be able to account for an interaction between conceptual factors (distributivity) and morphophonological factors (number ambiguity) by virtue of local interactivity during lexical access, a purely serial model such as the Marking and Morphing model, would not.

We tested whether indeed such an interaction existed. We first carried out an experiment to elicit subject-verb agreement errors in Dutch, in which we manipulated the two types of factors. We then ran a simulation of the Marking and Morphing model as described by Eberhard et al. (2005). We chose to evaluate Eberhard et al.'s version of the Marking and Morphing model because it is the only formal model available in the domain of agreement production.

Experiment

Method

Participants. Thirty-six native speakers of Dutch, students from Utrecht University, participated in this study in exchange for a monetary reward. There were 31 females and 5 males. Their mean age was 21.8.

Materials. The stimuli consisted of sentence preambles comprising two nouns: the subject head plus a second noun in a prepositional modifier to the head (e.g., *the slogan on the posters*).

The variables manipulated were: (a) the conceptual number of the preamble (multiple token meaning vs. single token meaning), and (b) the number ambiguity of the head noun phrase's determiner (unambiguous vs. ambiguous determiners). All the experimental preambles consisted of a singular subject head and a plural attractor. The following are example sentences for each condition:

Multiple token, ambiguous – *De asbak op de tafels* (The ashtray on the tables)
Multiple token, unambiguous – *Het menu op de tafels* (The menu on the tables)
Single token, ambiguous – *De rivier bij de villas* (The river by the villas)
Single token, unambiguous – *Het meer bij de villas* (The lake by the villas)

The 40 experimental stimuli were selected from a larger set of 88 items by submitting them to pretests. First, 18 judges, who were all graduate students, post-docs, or faculty in psycholinguistics and were native speakers of Dutch, rated the items for distributivity on a 7-point scale. Two counterbalanced lists were used, so that across the lists each item occurred once in the number-unambiguous condition, and once in the number-ambiguous condition, and so that, within each list, there were equal number of number-unambiguous and number-ambiguous single and multiple token preambles. Twenty-four items were excluded: either because the scores tended towards the mean, or because there was considerable deviation in scores between the unambiguous and ambiguous version. Second, 18 Ghent university students rated the remaining 64 items for imageability (using an adapted version of Eberhard's 1999, imageability test), and 18 further students rated these items for plausibility. In the imageability test, 16 (subjectively-judged) high-imageable and 16 low-imageable items were included, so as to promote full use of the scale. For the same reason, the plausibility test contained 16 plausible and 16 implausible fillers. Lists were counterbalanced in the same way as in the first pretest.

We selected 40 items (20 multiple token and 20 single token) so that they matched on the mean ratings of distributivity, imageability, and plausibility (Table 1; the Appendix lists the items). ANOVAs with items as a random effect showed no effects of number ambiguity or conceptual number, and no interaction (all p's > 0.1) except, as is to be expected, a main effect of conceptual number (single token vs. multiple token) on rated distributivity, $F_2(1, 38) = 356.91$, $p < .001$.

Table 1: Mean ratings (M) and standard deviations with respect to distributivity, imageability and plausibility for multiple and single token items in both ambiguous determiner and unambiguous determiner conditions.

	Distributivity		Imageability		Plausibility	
	M	SD	M	SD	M	SD
Single token						
Ambiguous	1.46	.50	5.86	.77	6.30	.71
Unambiguous	1.62	.57	5.77	.76	6.30	.63

Multiple token						
Ambiguous	5.48	.88	5.83	1.30	6.42	.65
Unambiguous	5.48	.83	5.75	.97	6.53	.44

Note. N = 18 for all ratings.

For the main experiment, the 40 experimental items were combined with a total of 200 fillers. Of these, 120 had the same syntactic structure as the experimental items: 40 containing a plural head and a singular attractor; 40 containing a plural head and a plural attractor; and 40 containing a singular head and a singular attractor. The other 80 fillers had a simpler syntactic structure, containing only one noun (the head noun): 40 with a singular head and 40 with a plural head. In this way, half of the responses required a singular inflected verb and half a plural inflected verb. Furthermore, of all the preambles consisting of two nouns (including the experimental items), half had nouns with matching number and half had nouns with mismatching number. The experimental items were divided into two presentation lists each containing 10 experimental items per condition.

Procedure. Participants were tested individually in a sound proof booth. Sessions lasted approximately 20 minutes. The participants' task was to repeat the preambles presented visually on the screen and immediately complete them in order to make a full sentence. Their responses were digitally recorded on digital audiotape (DAT) using a Sennheiser microphone in a sound-proof booth.

The experiment started with 5 practice items. After the practice session, participants could proceed to the main experimental session or request more information if necessary. Trials consisted of the following sequence of events: blank screen for 1000 ms, presentation for 500 ms of a fixation point in the center of the screen where the preamble was to appear, blank screen for 250 ms, presentation of the preamble for 800 ms. After the preamble disappeared, a deadline bar appeared to encourage participants to be quick. The empty bar filled up in 1440 ms after which a tone was heard signalling the end of the trial. Presentation was self-paced: participants initiated each trial by pressing the space bar in the keyboard. This procedure is based on the one employed, for example, in Hartsuiker et al. (2001). The brevity of the preamble presentation ensures that participants have to rely on short-term memory in order to repeat the preamble (Hartsuiker, et al., 2003).

Scoring. Responses were transcribed using a computer sound program. Two trained research assistants, who were naïve about the hypotheses and who were native speakers of Dutch, independently transcribed each experimental response. They also transcribed all repetitions that deviated from correct. They then independently assigned each response to a specific category with respect to the quality of the repetition (e.g., correct, head noun number error), the completion, (e.g., correct, agreement error), and whether the verb was produced on time or too late (after the warning tone at the end of the trial). Subsequently, they worked together in identifying each occasion in which either their transcriptions or scores on any of these three dimensions differed (less than 5% of the trials) and resolved these discrepancies by discussion or repeated listening to that trial.

The second author (also a native speaker of Dutch) then checked all transcriptions, listening to all cases in which the research assistants differed regarding the classification as an agreement error. This led to a reclassification of two number repetition errors as agreement errors. He then reclassified the assistant's three scoring dimensions into the categories listed below. For incorporation in all these categories, with the exception of miscellaneous responses, we required that the verb was produced on time. Responses were scored as *Correct* if the participant repeated the preamble correctly and produced a correct completion. An

Agreement error was scored if the participant repeated the preamble correctly and produced a verb that disagreed in number with that of the subject head noun (i.e., a plural verb). Any other type of error (e.g., repetition errors, incomplete preambles, disfluencies, use of syntactic constructions with postverbal subjects (e.g., *de kleur van de bloemen vond hij niet zo mooi* – lit. *the color of the flowers found he not very pretty*), production of verb after the deadline signal) was considered a *Miscellaneous error*.

Results

The multiple-token item *de hoes/het scherm van de GSM's* (the cover/the screen of the cell phones) yielded a much larger percentage of miscellaneous responses (50%) than the other items (range: 0% - 36%), and was excluded from further analysis. Of the remaining 1404 experimental responses, there were 1115 (79.4%) correct responses, 45 (3.2%) agreement errors, and 244 (17.4%) miscellaneous errors (which were late responses in 96 (6.8%) cases). Table 2 provides an overview of the number of responses in the different scoring categories.

Table 2: Distribution of responses by scoring category.

Condition	Correct	Agr. error	Misc.
Single token			
Amb[a]	284	7	69
	78.9%	1.9%	19.1%
Unamb[a]	302	1	57
	83.9%	0.3%	15.8%
Multiple token			
Amb[b]	255	29	58

	74.6%	8.5%	17.0%
Unamb[b]	274	8	60
	80.1%	2.3%	17.5%

Note. Scoring categories: Correct; Agr. error (agreement error); Misc. (miscellaneous response). Conditions: Amb (number-ambiguous determiner); Unamb (number-unambiguous determiner). a. N = 360; b. N = 342.

Since it has been extensively argued that ANOVA is not an appropriate statistical test for analysis of categorical data and, furthermore, it can specifically lead to spurious interactions (see Jaeger, 2008, and references therein), we opted for a logit mixed effects analysis. Logit models in general are a type of linear regression on the natural logarithm of odds instead of probabilities, which avoids the problems of uneven effects for the same probability differences along the probability continuum – differences in probabilities closer to .5 matter less than the same differences at the extremes of the probability range. Logit models present other advantages with respect to ANOVA. First, they do not depend on the usually unmet assumption of homogeneity of variances, as ANOVA does. Second, for categorical dependent variables, logit models have been shown in simulations to possess greater statistical power (Dixon, 2008; Jaeger, 2008, p. 443). Logit mixed models constitute a form of logit models that includes the effects of random factors (subjects and items). Basically, the test finds the best fit of the data to a model including certain pre-specified predictors and assigns each predictor a coefficient or weight. Subsequently, the probability of the model being the real underlying model for the population is evaluated (quasi-log-likelihood). By testing different combinations of predictors and comparing whether the fit of the different models is significantly different, it is possible to arrive at a final model that achieves the best fit with the least number of predictors.

The data were submitted to a logit mixed effects analysis with: agreement errors as the dependent variable; distributivity, ambiguity, and their interaction as fixed factors; and subjects and items as crossed random factors.

The log-likelihood of the full model, i.e. the model's fit, was -167.1. The interaction parameter was associated with a coefficient of 0.55, and found to be non-significant (p = 0.67). Excluding the interaction from the full model resulted in a log-likelihood for the reduced model of -167.2. The difference in log-likelihood between the full and the reduced models was not statistically significant (X^2 (1) = 0.23, p = 0.63), which meant the interaction could be excluded without a significant loss of fit. However, from the resulting model, neither distributivity could be excluded (X^2 (1) = 12.70, p < 0.001), nor ambiguity (X^2 (1) = 19.494, p < 0.001). See table 3 for a description of the parameters.

Table 3. Summary of the fixed effects in the mixed logit model.

Fixed effects	Coeficient	SE	Wald Z	P
Intercept	-4.81	0.49	-9.74	<2^e-16***
Distributivity	1.76	0.51	3.45	<0.001***
Ambiguity	-1.57	0.42	-3.79	<0.001***

Note. Log likelihood -167.2

Discussion

In line with earlier studies (Eberhard, 1999; Hartsuiker & Barkhuysen, 2006; Hartsuiker, Kolk, & Huinck, 1999; Vigliocco, Butterworth, et al., 1996; Vigliocco, et al., 1995; Vigliocco, Hartsuiker, et al., 1996), we observed both an effect of notional number and, replicating results by Hartsuiker et al. (2003), of morphophonological number ambiguity of the subject head noun's determiner (*de* vs. *het*). Although the effect of number ambiguity of the determiner is descriptively larger in the multiple token condition than in the single token condition, the model's fit did not improve when a term representing the interaction of the two factors was included. Therefore, we can conclude that the interaction of these two factors did not play a significant role in the final pattern of results.

These results do not support interactive models of agreement implementation and are compatible with strictly serial models such as the Marking and Morphing model proposed by Bock and colleagues (Bock, Eberhard, & Cutting, 2004; Bock, et al., 2001). In their model, the conceptual properties of the preamble should only be able to influence the first component of the agreement operation, namely, marking. A subject phrase with a distributive reading would be liable to be marked as plural even though its head may be singular, resulting in more plural marked verbs. The morphological processing, on the other hand, takes place at a later stage during the implementation of subject-verb agreement, namely during morphing. Here the morphological realization of number on the sentential subject components would be checked against the number of the subject phrase, a compromise would be reached in case of discrepancies, and the resulting number would determine the number feature on the verb. Since the two types of manipulations affect different stages, no interaction between conceptual effects and morphophonological effects such as number ambiguity of the determiner is expected.

Marking and Morphing simulations

Because the results are compatible with general properties of the Marking and Morphing model, and this model has been described in enough detail to test specific patterns, we evaluated whether our data fit with the predictions of Eberhard et al.'s (2005) computational version of the model. In fact, Eberhard et al. describe several versions of the model. Since the present study varied number ambiguity of the determiner, we used the version of the model that includes a contribution of the determiner's number specification.

For a given sentence preamble, this model calculates a tendency towards plural (which the authors call a "SAP" or "singular and plural" value), according to (1):

$$S(r) = S(n) + w_1 \times (Sm_{N1} + (w_{\text{det}} \times Sm_{\text{det}})) + w_2 \times Sm_{N2}{}^{1} \quad (1)$$

in which $S(r)$ denotes the total specification of plural at the root of the subject noun phrase. This value is then copied to the verb phrase and determines the verb's number value, with higher values of $S(r)$ increasing the probability of a plural verb. $S(r)$ is determined by additive contributions from the marking process (the first operand) and morphing (the second and third operands). The marking process determines the value of $S(n)$ which denotes the phrase's notional number. This number is fixed at 1 for unambiguous multitudes, 0 for units, and -1 for cases of specific individuation (e.g., singling out by a quantifier like *one*). In the simulations reported in Eberhard et al., the value for ambiguous notional number, as is the case for distributive phrases, was a free parameter ranging between -1 and 1, with a best fit of 0.48.

The contribution of the morphing process is a weighted sum of the lexical number specification associated with each noun phrase. These number specifications are a function of weighted determiner's specifications plus the basic noun's specifications (e.g., 0 for a singular count noun, 1 for a plural count noun) multiplied by a parameter that captures the relative frequency of a noun's singular vs. plural form (i.e., a parameter determined from linguistic corpora). The weights w_j decrease as a function of structural distance from the root of the subject phrase. Thus, in a phrase with the syntactic structure *NP(NP1 + PP (P NP2))* (as in our experiments) the lexical specification of Noun 1 exerts a much stronger influence than that of Noun 2. See Eberhard et al. (2005) for a more detailed description of the model and justification of these assumptions. Table 4 lists the parameter values used in Eberhard et al.'s and our simulations.

Importantly, the model transforms the value of $S(r)$ into a predicted proportion of plural verbs according to the logistic transformation (2):

$$P(plural) = \frac{1}{1 + e^{-[S(r)+b]}} \qquad (2)$$

in which b refers to a "bias" parameter which predisposes the model towards singular (plural probability of zero) in the absence of evidence for plurality.

In Eberhard et al.'s (2005) simulations, b was a free parameter, and so were $S(n)$ when notional number was ambiguous (with the constraint that it ranged from -1 to 1), and the weights for the head noun (w_1) and local noun (w_2), with the constraint that $w_1 > w_2$.

To evaluate the fit of this model to the Dutch data (so far, the only cross-linguistic extension of the model had been to Spanish, Eberhard, et al., 2005), we contrasted the results of the simulation (see Table 4 for an overview of the parameters' values as determined by Eberhard et al. with the combined results of several experiments: the experiment reported here, Vigliocco et al.'s (1996) two experiments in Dutch, Hartsuiker and Barkhuysen's (2006) experiment (condition without extrinsic memory load), and Harsuiker et al (2003). Viglioco et al.'s (1996) Experiments 1 and 2 were designed to test the possible effect of distributivity in Dutch. Hartsuiker and Barkhuysen (2006) tested the effect of conceptual number in subject-verb agreement in the presence or absence of an extrinsic memory load. Here only the results of the no-memory-load condition are included, since the other condition would not be expected to fit the Marking and Morphing model as is. Finally, Hartsuiker et al.'s (2003)

Experiment 3 tested the effect of the determiner's morphophonological ambiguity as well as effects of number match with plural head nouns on subject-verb agreement in Dutch.

Error percentages obtained with each of these experiments for each condition and the resulting Dutch averages can be found in Table 5, which also shows the results of the model's simulation. To allow for comparison with the data and simulations of Eberhard et al., the results in Table 5 are expressed as proportions of plural verbs out of all corrects and agreement errors (thus disregarding any possible differences on miscellaneous errors).

Table 4. Parameter values in Marking and Morphing simulation of the data.

Condition	S(n)	w_1	SN_1	w_{det}	S_{det}	w_2	SN_2	b
SgPl(ST, de)	0	18.31	0	0.28	0	1.39	1.19	-3.42
SgPl(ST,het)	0	18.31	0	0.28	-1	1.39	1.19	-3.42
SgPl(MT,de)	0.48	18.31	0	0.28	0	1.39	1.19	-3.42
SgPl(MT,het)	0.48	18.31	0	0.28	-1	1.39	1.19	-3.42
SgSg(de)	0	18.31	0	0.28	0	1.39	0	-3.42
SgSg(het)	0	18.31	0	0.28	-1	1.39	0	-3.42
PlPl	1	18.31	1.19	0.28	0	1.39	1.19	-3.42
PlSg	1	18.31	1.19	0.28	0	1.39	0	-3.42

Note. Parameters were identical to the ones in Eberhard et al. (2005). Note that the values for SN1 an SN2 reported here are a combination of the parameters for lexical number specification (1 for plural count nouns, 0 for singular count nouns), multiplied by *contrastive* *frequency*, a measure that indicates the relative frequency of the singular over the plural form and is given by the ratio of Log10 (frequency of plural form + frequency of singular form) and Log 10(frequency of plural form). Contrastive frequency was 1.16 on average in Eberhard et al. It was 1.19 for both our single token and multiple token items.

SgSg = singular head N-singular local N, SgPl = singular head N-plural local N, PlSg = plural head N-singular local N, PlPl = plural head N-plural local N. ST = single token, MT = multiple token. De = ambiguous determiner *de*, het = unambiguous singular determiner *het*.

S_{det} is always 0 in the local NP (always ambiguous), thereby cancelling the relevant term in the equation.

Table 5. Results of several experiments in Dutch, averages for this language, and results of the simulation with the Marking and Morphing model.

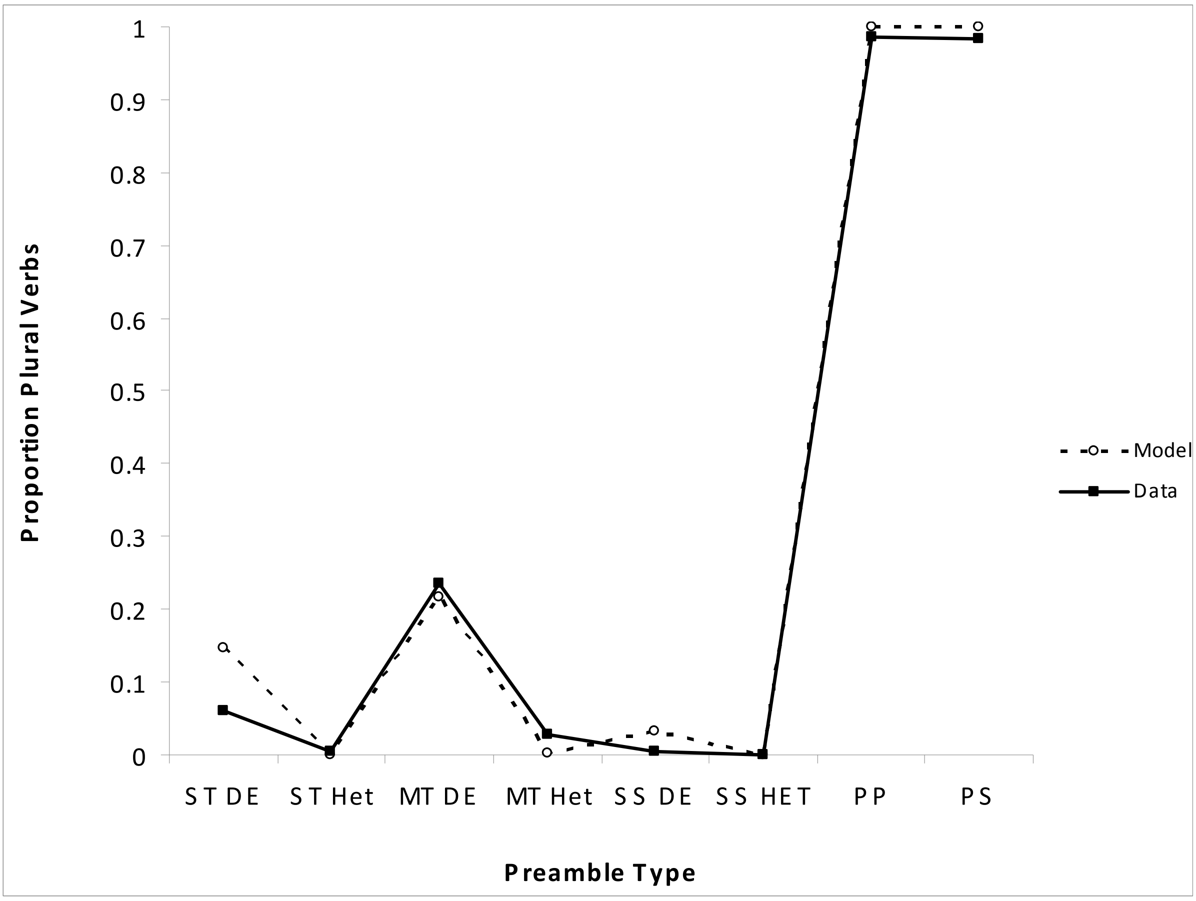
Condition	A-M&H	VHJK96;E1	VHJK96;E2	HB06;noload	HSBK03	Mean	Model
SgPl(ST,de)	2.4%	5.0%	3.7%	7.8%	11.4%	6.1%	14.6%
SgPl(ST,het)	0.3%				0.5%	0.4%	0.1%
SgPl(MT,de)	10.2%	36.9%	24.0%	23.3%		23.6%	21.7%
SgPl(MT, het)	2.8%					2.8%	0.2%
SgSg(de)		0.0%	0.7%	0.9%	0.0%	0.4%	3.2%
SgSg(het)					0.0%	0.0%	0.0%
PlPl					98.7%	98.7%	100%
PlSg					98.3%	98.3%	100%

Note1. A-M & H = Anton-Mendez & Hartsuiker (current study); VHJK= Vigliocco, Hartsuiker, Jarema & Kolk, 1996; HB06 = Hartsuiker & Barkhuysen, 2006; HSBK03 = Hartsuiker, Schriefers, Bock, & Kikstra, 2003.

Note2. SgSg-conditions not split out for ST-control vs. MT-control: in both control conditions, notional number is singular.

Note3. PlPl and PlSg conditions not split out for gender. In both conditions, the determiner ‘de’ is used in both the head and local noun.

Figure 1 displays the proportion of plural verbs for the combined data and Eberhard et al.’s Marking and Morphing model. Root mean square error was .0345 and the correlation between the model and the data was .997 for all conditions, and .901 when excluding the two conditions with a plural head (where performance is at ceiling).



RMSE (proportions) = 0.034

Figure 1. Fit between model and data. ST De = single token, ambiguous determiner; ST Het = single token, unambiguous determiner; MT De = multiple token, ambiguous determiner; MT Het = multiple token, unambigous determiner; SS De = singular head-singular local noun, ambiguous determiner; SS Het = singular head-singular local noun, unambiguous determiner; PP = plural head-plural local noun; PS = plural head-singular local noun.

Figure 1 clearly demonstrates a very high correspondence between model and data. These simulations provide proof that the Marking and Morphing model can in fact fit these data, with all of the parameters left unchanged from the original implementation.

The seemingly large proportion of agreement errors in the condition where both conceptual and morphophonological factors converged in facilitating a plural verb marking instead of singular is the result of the model mapping these variables via a non-linear transformation (as illustrated in Figure 2 which plots the transformation from Equation 2 with the bias parameter *b* set at -3.42 as specified in Eberhard et al, 2005). The figure shows that a given increase in *S(r)* when the starting value is relatively high results in a much larger increase in the predicted proportion of plural verbs than when the starting value is relatively low. This was the case in our simulations: the difference in *S(r)* between the unambiguous *het* and ambiguous *de* conditions was 5.1 in both the single token and multiple token conditions. But the single token unambiguous *het* condition had an *S(r)* of -3.5, whereas the multiple token ambiguous *het* condition had an *S(r)* of -2.9, so that the increase of 5.1 had a much larger effect on the proportion of plural verbs in the multiple token than the single token condition.

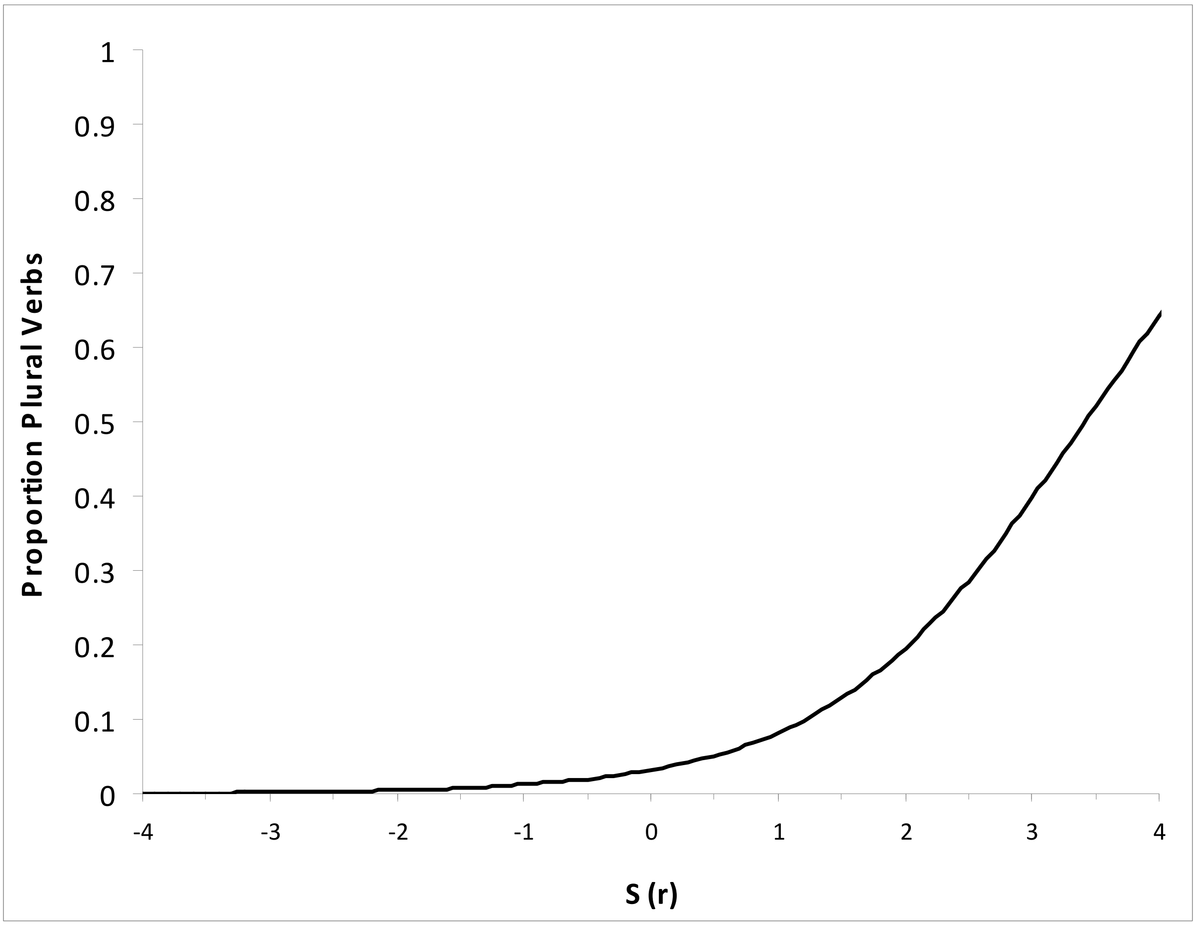


Figure 2. Function relating specification for plural (S(r)) to Proportion of plural verbs, with bias parameter b = - 3.42.

The model appears to be less apt at predicting the proportion of errors in the singular token conditions. While it is not clear why this condition shows a relatively larger discrepancy between model and data, it should be noted that we wished to evaluate the Marking and Morphing model without altering any of the original parameters and have therefore not optimized the parameter set.

General Discussion

This article reports two main findings. First, our experiment demonstrated that a meaning-related and a form-related variable do not interact in agreement production. Second, we find that the simulation of one particular model of agreement production which adopted a serial, non-interactive framework is capable of reproducing the pattern of results from the present experiment as well as results from previous experiments testing the effect of distributivity and morphophonological ambiguity on subject-verb agreement production.

Following the standard logic outlined in the introduction, the fact that meaning-related and form-related variables have additive effects on the proportions of agreement errors should be interpreted as evidence in favor of serial models of language production such as the Morphing and Marking model. That is, the absence of a significant interaction between factors that are postulated to affect different stages of language production can be interpreted as a reflection of the independence of the processing stages themselves.

Therefore these results offer considerable support for the Marking and Morphing model of agreement production. First, from the general principles of the model (i.e., serial processing of the two stages of marking and morphing), we derived the prediction that meaning-related and form-related variables should have additive effects on the proportions of agreement errors and this was clearly confirmed by the data. Second, simulations showed that a formal version of this model could capture the data without altering a single parameter (note that the only difference between the English and Dutch models concerned contrastive frequency, a variable determined from linguistic corpora – the difference, furthermore, was extremely small, namely 1.16 in English versus 1.19 in Dutch).

Nevertheless, we warn that these findings should be interpreted cautiously with respect to the debate regarding seriality vs. interactivity in language production. The common assumption that the presence of an interaction is evidence in support of interactive models while the absence of an interaction is supportive of serial models of agreement production is too simplistic. Although the *absence* of an interaction shows that, for the two variables in question, it is not necessary to postulate interactivity, it is logically possible that, on an underlying interactive architecture (conceptualized as the blurring of boundaries between levels of processing), two variables simultaneously affect the same processing stage without altering the effects each of them exerts on the behavioral outcome (as noted indeed by Sternberg, 1969, p. 282). Conversely, the *presence* of an interaction of variables affecting the proportions of speech errors, while most readily explained by an interactive architecture, must be interpreted with caution because of two reasons – one methodological and one theoretical. The methodological reason is that, in contrast to Reaction Time data for which Sternberg devised the additive-factors method, proportion data are bound between 0 and 1. This means that an increase in proportions close to 1, or a decrease in proportions close to 0 can only be very small. The Marking and Morphing model captures this by applying the logistic transformation (Equation 2, also Figure 2). But this implies a non-linearity in the mapping from a tendency toward plural (S[r] in the model) to proportions of plural verbs. If three sentence types A, B, and C differ in, say, notional number, resulting in S[r]'s of 1, 2, and 3, the increase in proportion of plurals is much smaller from A to B than from B to C. Similarly, if two variables exert additive effects on S[r], this may result in overadditivity (or underadditivity) in proportion of plurals (Figure 2) and thus in spurious interactions. Note that indeed, at least descriptively, both model and data here appear to show some overadditivity.

The theoretical reason is that, according to some proposals, patterns of speech errors are not only a function of the properties of the production system per se, but also of a biased monitoring system, that may detect and covertly correct some errors more often than others (e.g., Baars, Motley, & MacKay, 1975; Vigliocco & Hartsuiker, 2002). One example of this is the lexical bias effect in speech errors, by which phonological errors tend to form real words more often than chance would predict (e.g., Dell & Reich, 1981; Hartsuiker, Antón-Méndez, Roelstraete, & Costa, 2006). This phenomenon can be explained by feedback loops within the production system (Dell, 1986), but also by a monitoring system that uses lexicality as a criterion for correctness (Baars, et al., 1975) and so prevents more non-word errors from becoming overt than word errors.

In light of these considerations, the results of the experiment reported here cannot by themselves be taken as unequivocal evidence of a serial language production system. Nevertheless, the close fit between the combined response patterns of number agreement experiments in Dutch and that predicted by the implementation of the Marking and Morphing model does offer considerable support for this model of subject-verb agreement and, in so far as the Marking and Morphing theory and computational model are cast within a serial framework, the present data provide indirect support for seriality, at least in relation to this particular type of processing.

In general, it seems to us that, given the great difficulty in empirically distinguishing between interactive and serial models, one option for the field might be to ask fundamentally different questions about information flow and language production, replacing the general interactivity/seriality dichotomy with much more specific questions. For example, Roelofs (2004) asked *how* feedback can affect language production: Are there feedback connections within the production system, of are there rather connections between language comprehension and production, leading to indirect feedback? Similarly, Goldrick and colleagues (Goldrick, 2006; Goldrick, Costa, & Schiller, 2008) asked what the *limits* on cascading of activation are: Does it only involve adjacent levels, or can cascading extend further? In his review, Acuña-Fariña (2009) considers, among other things, whether specific properties of different languages could make agreement more or less likely to be influenced by one or another type of factor and more or less likely to allow "overrides", and whether the relative contribution of different factors may not depend on the domain of agreement. Finally, with respect to lexical bias, Hartsuiker and colleagues (Hartsuiker, 2006; Hartsuiker, Corley, & Martensen, 2005; Severens & Hartsuiker, 2009) and Nooteboom and Quené (2008) have begun to ask to which extent speech error patterns result from feedback, monitoring, or indeed a combination of the two and for which speech error patterns this is the case.

In conclusion, while our empirical results are compatible with serial computation of subject-verb agreement in language production and constitute evidence in favor of a process along the lines proposed by Bock and colleagues (Bock, et al., 2001; Bock, Nicol, & Cutting, 1999; Eberhard, et al., 2005), we do not want to extend the implications of independence of factors in relation to a specific part of the system as applying to the entirety of the system. The definitive proof for or against either framework is unlikely to be found soon and may never be found. Research that is inspired by the interactivity/seriality dichotomy but asks more specific questions is more likely to contribute to a more precise understanding of language production mechanisms themselves, and to a better understanding of how language production interacts with language comprehension and monitoring processes.

References

Acuña-Fariña, J. C. (2009). The linguistics and psycholinguistics of agreement: A tutorial overview. *Lingua, 119*, 389-424.

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14*, 382-391.

Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes, 8*(1), 57-99.

Bock, K., Eberhard, K. M., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language, 51*, 251-278.

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. [Journal article]. *Cognitive Psychology, 43*, 83-128.

Bock, K., & Levelt, W. J. M. (1994). Language production. Grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945-984). New York: Academic Press.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23*, 45-93.

Bock, K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language, 40*, 330-346.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20*, 611-629.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59*(4), 447-456. doi: 10.1016/j.jml.2007.11.004

Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject-verb agreement in English. *Journal of Memory and Language, 41*(4), 560-578.

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review, 112*(3), 531-559.

Ferguson, C. A., & Barlow, M. (1988). Introduction. In M. Barlow & C. A. Ferguson (Eds.), *Agreement in natural language: Approaches, theories, descriptions* (pp. 1-22). Stanford: CSLI, Stanford University.

Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U. H. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes, 23*(3), 329-374.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes, 21*(7/8), 817-855.

Goldrick, M., Costa, A., & Schiller, N. O. (2008). Situating language production within the matrix of human cognition: The state of the art in language production research. *Language and Cognitive Processes, 23*(4), 489-494.

Hartsuiker, R. J. (2006). Are speech error patterns affected by a monitoring bias? *Language and Cognitive Processes, 21*, 856-891.

Hartsuiker, R. J., Antón-Méndez, I., Roelstraete, B., & Costa, A. (2006). Spoonish Spanerisms: A lexical bias effect in Spanish. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32*(4), 949-953.

Hartsuiker, R. J., Antón-Méndez, I., & Van Zee, M. (2001). Object attraction in subject-verb agreement construction. *Journal of Memory and Language, 45*, 546-572.

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes, 21*(1/2/3), 181-204.

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al. (1975). *Journal of Memory and Language, 52*(1), 58-70.

Hartsuiker, R. J., Kolk, H. H. J., & Huinck, W. J. (1999). Agrammatic production of subject-verb agreement: The effect of conceptual number. *Brain and Language, 69*, 119-160.

Hartsuiker, R. J., Schriefers, H. J., Bock, K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject-verb agreement. [Journal Article]. *Memory and Cognition, 31*(8), 1316-1326.

Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject-verb agreement. *Journal of Memory and Language, 48*, 760-778.

Hupet, M., Fayol, M., & Schelstraete, M. (1998). Effect of semantic variables on the subject-verb agreement processes in writing. *British Journal of Psychology, 89*(1), 59-75.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434-446.

Nicol, J., & Antón-Méndez, I. (2008). The effect of case marking on subject-verb agreement errors in English. In W. D. Lewis, S. Karimi, H. Harley & S. O. Farrar (Eds.), *Time and again. Theoretical perspectives on formal linguistics. In honor of D. Terence Langendoen*. Amsterdam: John Benjamins.

Nooteboom, S., & Quené, H. (2008). Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language, 58*(3), 837-861.

Roelofs, A. (2004). Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review, 111*, 561-572.

Severens, E., & Hartsuiker, R. J. (2009). Is there a lexical bias effect in comprehension monitoring? *Language and Cognitive Processes, 24*, 910-927.

Solomon, E. S., & Pearlmutter, N. J. (2004). Semantinc integration and syntactic planning in language production. *Cognitive Psychology, 49*, 1-46.

Stenberg, S. (1969). The discovery of stages: Extension of Donders' methods. *Acta Psychologica, 30*, 276-315.

Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language, 48*, 740-759.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition, 61*(3), 261-298.

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language, 34*, 186-215.

Vigliocco, G., & Franck, J. (1999). When Sex and Syntax go hand in hand: Gender agreement in language production. *Journal of Memory and Language, 40*, 455-478.

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin, 128*(3), 442-472.

Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes, 11*(3), 407-421.

APPENDIX

Experimental sentence preambles

Multiple token items

Ambiguous determiner	Unambiguous determiner
De jas van de mannen	*Het pak van de mannen*
The coat of the men	The suit of the men
De som op de salarisstroken	*Het getal op de salarisstroken*
The sum on the payslips	The number on the payslips
De waarschuwing op de verpakkingen	*Het advies op de verpakkingen*
The warning on the packages	The advice on the packages
De asbak op de tafels	*Het menu op de tafels*
The ashtray on the tables	The menu on the tables
De hoes van de GSMs	*Het scherm van de GSMs*
The cover of the mobile phones	The screen of the mobile phones
De voorkant van de boeken	*Het voorblad van de boeken*
The front of the books	The frontpage of the books
De handtas van de dames	*Het juweel van de dames*
The handbag of the ladies	The juwel of the ladies
De hoed van de cowboys	*Het pistool van de cowboys*
The hat of the cowboys	The pistol of the cowboys
De deur van de huizen	*Het raam van de huizen*
The door of the houses	The window of the houses
De kleur van de bloemen	*Het blad van de bloemen*
The color of the flowers	The leaf of the flowers

De schurk in de films	*Het kreng in de films*
The vilain in the films	The bitch in the films
De jurk in de etalages	*Het kleed in de etalages*
The dress in the display window	The cloth in the display window
De slogan op de posters	*Het plaatje op de posters*
The slogan on the posters	The picture on the posters
De vervuiling van de rivieren	*Het afval in de rivieren*
The pollution of the rivers	The waste in the rivers
De lamp van de taxi's	*Het licht van de taxi's*
The lamp of the taxis	The light of the taxis
De sleuf van de geldautomaten	*Het toetsenbord van de geldautomaten*
The slot of the cash machines	The keypad of the cash machines
De dop op de flessen	*Het etiket op de flessen*
The top on the bottles	The label on the bottles
De helm voor de soldaten	*Het uniform voor de soldaten*
The helmet for the soldiers	The uniform for the soldiers
De muts voor de koks	*Het schort voor de koks*
The hat for the chefs	The apron for the chefs
De datum op de munten	*Het jaartal op de munten*
The date on the coins	The year on the coins

Single token

Ambiguous determiner	Unambiguous determiner
De kaart onder de kranten	*Het blad onder de kranten*
The card underneath the newspapers	The magazine underneath the newspapers

De rivier bij de villas	*Het meer bij de villas*
The river by the villas	The lake by the villas
De hond bij de schuren	*Het paard bij de schuren*
The dog by the barns	the horse by the barns
De drank met de vitamines	*Het sap met de vitamines*
The drink with the vitamins	The juice with the vitamins
De kerk bij de heuvels	*Het huis bij de heuvels*
The church near the hills	The house near the hills
De auto van de inbrekers	*Het busje van de inbrekers*
The car of the burglars	The van of the burglars
De muur bij de putten	*Het hek bij de putten*
The wall near the pits	The fence near the pits
De boot van de vissers	*Het schip van de vissers*
The boat of the fishermen	The ship of the fisherman
De vaas met de rozen	*Het boeket met de rozen*
The vase with the roses	The bouquet with the roses
De tuin met de zonnebloemen	*Het perk met de zonnebloemen*
The garden with the sunflowers	The bed with the sunflowers
De jongen met de krukken	*Het meisje met de krukken*
The boy with the crutches	The girl with the crutches
De herberg aan de wegen	*Het motel aan de wegen*
The inn by the roads	The motel by the roads
De weg in de bergen	*Het pad in de bergen*
The road in the mountains	The trail in the mountains
De shampoo met de bloemengeuren	*Het parfum met de bloemengeuren*

The shampoo with the flower scents	The perfume with the flowerscents
De muur met de decoraties	*Het tapijt met de decoraties*
The wall with the decorations	The carpet with the decorations
De stad met de fietsroutes	*Het dorp met de fietsroutes*
The city with the bike routes	The village with the bike routes
De speeltuin met de glijbanen	*Het zwembad met de glijbanen*
The playground with the slides	The swimming pool with the slides
De doos met de CD's	*Het pakket met de CD's*
The box with the CDs	The packages with the CDs
De geit bij de sloten	*Het schaap bij de sloten*
The goat by the ditches	The sheep by the ditches
De kooi met de gorilla's	*Het hok met de gorilla's*
The cage with the gorillas	The cage with the gorillas

Footnote

[1] In the case of the sentences considered here, the contribution from the second determiner is always 0, thus cancelling the term in the equation weighing its contribution.