

Placing objects before subjects in Dutch

A comparison between word order variation in experimental and natural sentence production*

Geertje van Bergen

Radboud University Nijmegen

1. Introduction

Cross-linguistically, subjects tend to precede objects in transitive clauses. In the World Atlas of Language Structures (WALS, Haspelmath et al. 2005), 1017 out of 1228 languages have a basic word order in which the subject precedes the object, which comes down to more than 82%; a similar percentage is found by Hawkins (1994). At the same time, there is a cross-linguistic tendency for animate entities to precede inanimate ones in various syntactic structures, as shown in multiple psycholinguistic studies. For instance, English has a preference for active over passive structures, but this preference is less strong if the patient (which becomes the first argument in a passive sentence) is animate (Bock & Warren 1985, McDonald et al. 1993, Prat-Sala & Branigan 2000).

Since subjects prototypically outrank objects in animacy in transitive clauses (see Comrie 1989; see also corpus studies of Swedish (Dahl and Fraurud 1996), Norwegian (Øvrelid 2004) and Dutch (Van Tiel & Lamers 2007)), the ‘subject-first’ preference and the ‘animate-first’ preference generally lead to the same constituent order. However, these preferences are sometimes overridden by the selectional restrictions of the verb. Different verb types require different semantic roles. For instance, psych verbs require an experiencer argument or a recipient of a cognitive stimulus. In both (1) and (2), *John* is the experiencer and *the painting* is the theme or stimulus:

- (1) John *admired* the painting.
- (2) The painting *depressed* John.

Lamers (2001, 2005, 2007) and Lamers & De Hoop (2008) investigated how grammatical function, animacy and verb type interact with constituent order in the comprehension and production of Dutch transitive sentences by means of several

experiments. Lamers & De Hoop (2008) explain their findings in terms of bidirectional optimization, arguing that the speaker thinks of the hearer when uttering a sentence and the hearer thinks of the speaker when interpreting a sentence.

This paper shows that the notion of ‘speaker’ as used in the approach of Lamers & De Hoop (2008) is problematic. I will argue that participants in a production experiment cannot be compared to natural language producers, on the basis of a comparison between experimental data and naturally produced language, the latter taken from the *Corpus Gesproken Nederlands* (Corpus of Spoken Dutch, henceforth CGN). First, in Section 2, I describe the relevant experimental comprehension and production studies of Lamers (2001, 2005, 2007), Lamers et al. (2006) and Lamers & De Hoop (2008). Next, in Section 3, I describe the results of my corpus study. In Section 4 I focus on the similarities and differences between the corpus data and the experimental results. Finally, Section 5 concludes.

2. Experimental approaches to object fronting

Although the experiencer of a psych verb is always animate, it is not always the subject of the sentence. The verb *admire* in (1) is a *subject-experiencer* verb, which requires its subject to be animate; *depress* in (2), on the other hand, is an *object-experiencer* verb, which requires an animate object. In sentences with a subject-experiencer verb, both the subject-first and the animate-first preference can be fulfilled if the word order is Subject-before-Object: in (1), the animate argument *John* is in sentence-initial position. With an object-experiencer verb as in (2), however, the two preferences cannot be fulfilled at the same time. In Dutch, the subject-initial (inanimate-first) order is preferred, but object-initial (animate-first) sentences are also considered grammatical (e.g. Lamers 2001):

- | | | | | | |
|--------|--|--|-----------|--|----------------|
| | S | | V | | O |
| (3) a. | De vraag | | verraste | | de politicus. |
| | the question | | surprised | | the politician |
| | O | | V | | S |
| b. | De politicus | | verraste | | de vraag. |
| | the politician | | surprised | | the question |
| | ‘The question surprised the politician.’ | | | | |

Lamers (2001, 2005, 2007) conducted several rating studies to investigate the interplay of grammatical function, animacy and verb type in sentence comprehension. Participants judged the comprehensibility of embedded transitive sentences, containing a subject and an object that differed in word order (subject-initial vs. object-initial order) and in animacy (animate vs. inanimate subjects and objects).

Three types of experiencer verbs were used in the studies: subject-experiencer verbs requiring an animate subject, and causative and unaccusative psych verbs, both requiring an animate object. The latter two verb types differ in that causative psych verbs can passivize in Dutch, whereas unaccusative psych verbs cannot (see Lamers & De Hoop 2008 for discussion):

- (4) a. De politicus werd verrast door de vraag.
the politician was surprised by the question
'The politician was surprised by the question.'
- b. *De politicus werd bevallen door de vraag.
the politician was pleased by the question

Lamers (2001, 2007) found that the disfavored object-initial order was easier to comprehend for sentences with verbs that select an animate object (making the order animate — inanimate) than for sentences with verbs that select an animate subject (making the order inanimate — animate). Furthermore, she found a difference between the two types of psych verbs: object-initial sentences were easier to comprehend for unaccusative than for causative psych verbs.

Following up on these comprehension studies, an experiment was conducted to investigate the role of grammatical function, animacy and verb type in sentence production (Lamers et al. 2006, Lamers & De Hoop 2008). Using a similar procedure as in a production study by Ferreira (1994), participants were presented with two definite noun phrases (one animate and one inanimate) and a verb (subject-experiencer, causative psych or unaccusative psych) and they were asked to write down a sentence with these words. The produced sentence types were classified into four groups: subject-before-object (SO) active sentences, passive constructions, object-before-subject (OS) active sentences and 'other' constructions. A difference between the two types of psych verbs was found in the production data as well: unaccusative psych verbs were used in OS active sentences relatively often, whereas causative psych verbs hardly occurred in OS active sentences.

Lamers & De Hoop (2008) involve both the speaker and the hearer when explaining the differences between causative and unaccusative psych verbs in sentence comprehension and sentence production. They argue that, in order to fulfill the animate first-preference, the speaker will front the object of a psych verb only if the hearer can still arrive at the right interpretation, i.e., if subject and object are still distinguishable (cf. Gibson 1998, De Hoop & Lamers 2006, De Swart 2007, Bouma 2008). For both types of psych verbs, an OS active sentence fulfills the animate-first preference, but not the subject-first preference. For causative psych verbs, however, there is a possibility to fulfill the animate-first and the subject-first preference simultaneously by the use of a passive construction, as in (4a). This is impossible for unaccusative verbs, since a passive construction is not available, cf.

(4b). For unaccusative verbs the only way of fulfilling the animate-first preference is to use an object-initial active construction. Lamers & De Hoop (2008) argue that the impossibility of fulfilling both preferences at the same time leads to the production of more OS-sentences (the speaker thinking of the hearer) and to a greater ease of comprehensibility of OS-sentences (the hearer thinking of the speaker) for unaccusative psych verbs compared to causative psych verbs.

This bidirectional approach, that is, speakers thinking of hearers and vice versa, is a very interesting way of explaining the phenomenon of object fronting. Yet, the question is whether it is legitimate to use this general notion of ‘speaker’ when accounting for the behavior of participants in a production experiment. The way in which a sentence is produced by an experimental participant is totally different from that of a natural language user. A sentence produced by a natural language user is the result of the intention to convey a particular meaning to someone else. A sentence produced by a participant in the production experiment, on the other hand, is the result of combining three given constituents in a particular way. Participants of a production experiment are constrained in the way they express themselves by experimental conditions; natural language users have much more freedom and may therefore have different preferences in their construction choice.

The main goal of this article is to investigate whether these differences between natural and experimental sentence production are reflected in preferences for the linear order of subjects and objects. In the following section, I will discuss the corpus study I performed to investigate object fronting in naturally produced language.

3. Object fronting in natural language

For this study, I used material from the CGN. Different types of naturally produced speech are represented in this corpus, including spontaneous face-to-face and telephone conversations, interviews, debates, radio shows and read-aloud books. I extracted all sentences that contained one of the following verbs (which were also used in the experiments of Lamers (2001, 2005, 2007) and Lamers et al. (2006)):

- 8 subject-experiencer verbs: *begrijpen* ‘understand’, *missen* ‘miss’, *haten* ‘hate’, *respecteren* ‘respect’, *bewonderen* ‘admire’, *afkeuren* ‘reject’, *bekritisieren* ‘criticize’, *verafschuwen* ‘detest’
- 12 causative psych verbs: *verbazen* ‘surprise’, *verwonderen* ‘amaze’, *irriteren* ‘annoy’, *motiveren* ‘motivate’, *schaden* ‘harm’, *benadelen* ‘harm’, *beangstigen*

- ‘frighten’, *behagen* ‘please’, *choqueren* ‘shock’, *overdonderen* ‘overwhelm’, *imponeren* ‘impress’, *mishagen* ‘displease’
- 5 unaccusative verbs: *opvallen* ‘strike’, *bevallen* ‘please’, *invallen* ‘strike’, *te binnen schieten* ‘come to mind’, *aanstaan* ‘please’

Sentences in which the verb in question was used as adjective or adverb, e.g. *een bewonderende blik* ‘an admiring look’, were excluded from further analysis; complement clauses such as *ik begrijp [dat het nu te laat is]* ‘I understand [that it’s too late now]’ were omitted as well. This yielded a total of 2,541 sentences: 1,697 sentences containing an experiencer-theme verb (67%), 344 sentences with a causative psych verb (14%) and 500 sentences containing an unaccusative psych verb (20%). All sentences were classified into four construction types (cf. the categorization of Lamers & De Hoop (2008)): SO active sentences, passive sentences, OS active sentences and a group of ‘other’ sentence types. The frequency distribution of the corpus data is represented in Table 1, where the experimental results have been included for the sake of comparison.

Table 1. Frequency distribution of experimental data (exp) and corpus data (CGN) over four sentence categories per verb type (relative frequencies in parentheses).

Verb type		SO active		Passive		OS active		Other		Total	
		n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
subject-experiencer	exp	142	(79)	28	(16)	0	(0)	10	(6)	180	(100)
	CGN	1108	(65)	77	(5)	444	(26)	68	(4)	1697	(100)
causative psych	exp	108	(60)	49	(27)	4	(2)	20	(11)	180	(100)
	CGN	284	(83)	18	(5)	6	(2)	36	(10)	344	(100)
unaccusative psych	exp	110	(61)			44	(24)	26	(14)	180	(100)
	CGN	306	(61)			47	(9)	147	(29)	500	(100)

As shown in (1), SO active sentences are the most frequent construction types in the corpus data as well as in the experimental data, irrespective of verb type. Yet, there is a significant distributional difference between the natural language data and the experimental data within each verb type (subject-experiencer verbs, $X^2(3) = 87.9$, $p < .0001$; causative psych verbs, $X^2(3) = 52.8$, $p < .0001$; unaccusative psych verbs, $X^2(2) = 34.0$, $p < .0001$). First, the proportion of passive constructions is much lower in the corpus data than in the experimental results. Subject-experiencer verbs and causative psych verbs occur in passive constructions in only 5 per cent of the corpus data, as opposed to 16 and 27 per cent (in that order) of the experimental data. Second, the proportion of subject-experiencer verbs in OS active sentences is much larger in the corpus data than in the experimental data. Third, while unaccusative verbs occur in OS active sentences in almost a quarter

of the cases in the production experiment, the proportion of this construction type is much lower in the natural language data; a much higher number of 'other' construction types is used instead.

In the next section, I will describe the distributional differences between the experimental data and the corpus data in more detail and try to give an explanation for these differences.

4. Comparing natural and experimental sentence production

The results of both studies show the strength of the subject-first preference in Dutch: all three verb types are used most frequently in SO active sentences. In spite of this, the distributional patterns differ between the two studies for each of the verb types. Let us first consider the distributional differences within the category of unaccusative psych verbs. For this verb type, the percentage of OS sentences in the corpus data is much lower than in the experimental data, while the percentage of 'other' construction types is much higher. This difference can be explained if we take into account the different conditions under which the sentences were produced.

In natural language, speakers have the possibility to produce many more sentence types than the participants in the experiment. One possibility is to leave out one of the arguments of the verb:

- (5) Als je na 't jaar niet bevalt schoppen we je d'r
 if you after the year not please kick we you there
 zo weer uit.
 like.that again out
 'If we're not satisfied with you after a year, we can just kick you out.'
 (CGN: fn000285.194)

The fact that unaccusative verbs cannot be used in passive constructions does thus not necessarily lead to an increase of OS active sentences in natural language; it can also yield an increase in 'other' strategies, such as expressing only one instead of two arguments. This option was unavailable for the participants in the experiment, since they were explicitly instructed to use the two nouns that were given beforehand.

This brings us to the frequency differences in the use of passive constructions between the corpus data and the experimental data.¹ Although a passive sentence is a way to satisfy both the subject-first and the animate-first preference for causative psych verbs, it is not the only option. When not constrained by experimental conditions, speakers may use many other constructions, such as leaving one argument unexpressed, as in (5).

Discarding passive sentences and ‘other’ construction types from further analysis, we are still left with the big frequency difference between OS active sentences in the experimental and the corpus data for subject-experiencer verbs. In the corpus data, OS active sentences are used in over a quarter of the cases. This is totally unexpected in terms of fulfilling the subject-first and the animate-first preference. However, the difference is less striking if we consider once more the different circumstances under which the sentences were produced.

In their article on the incremental interpretation of subjects and objects, De Hoop & Lamers (2006) discuss five different cues to distinguish subjects from objects. In (6), these cues are represented on a scale to illustrate their relative importance for Dutch:

- (6) {CASE, AGREEMENT} >> SELECTION >> PRECEDENCE >> PROMINENCE

De Hoop & Lamers (2006) show that CASE and AGREEMENT are the most important cues to disambiguate between subjects and objects in Dutch. Even when other cues like PRECEDENCE (S precedes O) and/or PROMINENCE (S outranks O in animacy) would guide the hearer towards a different interpretation, the sentences will still be correctly interpreted by disambiguating information from agreement (7) or case (8):

- (7) Koning, keizer, admiraal, Popla kennen ze allemaal!
king emperor admiral Popla know they all
‘It doesn’t matter whether they’re king, emperor or admiral; they’re all familiar with Popla.’
(Bouma 2008: 13)
- (8) Haar heeft de gorilla gebeten.
her has the gorilla bitten
‘It was her the gorilla bit.’
(Lamers & De Hoop 2008)

Because only full noun phrases were used in the production experiment, the influence of case and agreement was controlled for. Yet, case and agreement information are not excluded in the corpus data. If the speaker takes the hearer’s perspective when uttering a sentence, s/he will change the order of subject and object only if there are enough other cues that lead to the right interpretation, i.e. if subject and object are still distinguishable (e.g. Gibson 1998, De Hoop & Lamers 2006, De Swart 2007, Bouma 2008). Case and agreement being the most important cues for argument disambiguation, more object fronting might be expected in the corpus data when either case or agreement distinguishes the subject from the object. In order to investigate this, every SO and OS sentence in the corpus (n = 2,195) was checked for distinguishability of subject and object on the basis of case and/or

agreement. The distribution of distinguishable and non-distinguishable arguments over the two word orders is given in Table 2.

Table 2. Frequencies of SO and OS active sentences with and without disambiguating case and/or agreement information

	Constituent order				Total	
	SO		OS			
	n	(%)	n	(%)	n	(%)
disambiguation by case/agreement	1439	(77)	422	(23)	1861	(100)
no disambiguation by case/agreement	259	(78)	75	(22)	334	(100)
Total	1698	(77)	497	(23)	2195	(100)

As can be seen, this expectation is not borne out. Although case and agreement distinguish subjects from objects in 1,861 out of 2,195 sentences (i.e. over 80%), the OS order is not more frequent in sentences with disambiguating case/agreement information than it is in sentences without such information ($X^2(1) < 1$, $p = .99$): irrespective of the case/agreement cue, the object precedes the subject in about a quarter of the sentences.

There is, however, another difference between the experimental data and the naturally produced sentences that must be considered here. In the experiment, every sentence always contained one animate and one inanimate argument, but this is not the case in the corpus data. For subject-experiencer verbs, the subject indeed has to be animate, but the object can be either animate or inanimate. For the two types of psych verbs the object is obligatorily animate, but the subject can be either animate or inanimate. Therefore, it was determined for every sentence without disambiguating case or agreement information whether there was a difference in animacy between the subject and the object. The results are given in Table 3.

Table 3. Frequencies of SO and OS sentences with and without a difference in animacy between the arguments.

	Constituent order				Total	
	SO		OS			
	n	(%)	n	(%)	n	(%)
animate — inanimate	212	(74)	73	(26)	285	(100)
animate — animate	47	(96)	2	(4)	49	(100)
Total	259	(78)	75	(22)	334	(100)

The frequency distribution in this table shows that the animacy of the arguments has an effect on the constituent order ($X^2(1) = 9.9, p = .0016$): the number of OS sentences for sentences with an animate and an inanimate argument is much higher than for sentences with two animate arguments (there were no sentences with two inanimate arguments, as all verb types require one animate argument at least). This can again be explained in terms of distinguishability. When there is a difference in animacy between the arguments, subjects can always be distinguished from objects by the selectional restrictions of the verb: the animate argument will be the subject of a subject-experiencer verb, and the object of a causative or unaccusative psych verb. We have seen in (6) that SELECTION outranks PRECEDENCE as a cue to distinguish subjects from objects in Dutch: when arguments are distinguishable on the basis of SELECTION, the hearer will arrive at the right interpretation regardless of the order of the constituents. With two animate arguments, however, SELECTION no longer disambiguates, so that the hearer can only rely on PRECEDENCE (S precedes O) as a cue to distinguish S from O. In that case, OS-sentences are virtually absent, which can be taken as an indication of the fact that the speaker takes the hearer into account when uttering a sentence.

5. Concluding remarks

As this paper has shown, participants in a language production experiment cannot be equated with natural language users, for a number of reasons. I compared the results of a production experiment on object fronting (Lamers et al. 2006) with natural language data from the CGN and discussed similarities and differences. Despite the different patterns that I found between experimental and natural language production, I argued that object fronting in naturally produced Dutch can still be explained in a bidirectional way, i.e., the speaker taking the hearer into account: the speaker only places the object before the subject if there are enough cues left for the hearer to distinguish the subject from the object.

It should be noted that the factors discussed in this article are not the only determinants of the linear order of subjects and objects.² The experimental data deviates from the natural data in more ways than case, agreement and animacy; for instance, all nouns were preceded by a definite article, and every noun phrase had the same length (measured in words). Naturally produced sentences vary in definiteness and length of the constituents. It has often been shown that definiteness and length have an effect on the linear order of two arguments: definite arguments tend to precede indefinite ones, and shorter constituents generally precede longer constituents (see for instance Bouma (2008) for subject and object fronting in Dutch, Bresnan et al. (2007) for the order of direct and indirect objects in English,

and Rosenbach (2002, 2005) for the order of genitive and noun in English). Furthermore, in naturally produced language there is contextual information available, and there is variation in the position of the verb relative to the arguments. These factors might also be involved in determining the linear order of subject and object. Further research is required to establish whether and how all of these factors contribute to the mutual order of subjects and objects in Dutch.

Notes

* I would like to thank two anonymous reviewers, Monique Lamers, Helen de Hoop and the members of the Optimal Communication group for useful comments. Thanks also go to Roeland van Hout for statistical advice. The Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged for financial support (grant 360-70-220, *Animacy*).

1. An anonymous reviewer points out the surprisingly high number of passive sentences in the experiment, even for subject-experiencer verbs, where passive constructions violate the animate-first preference. Participants will have had their reasons to start their sentences with the theme argument (for instance, because they considered the theme argument to be the most important part of the message), but the experimental data are inconclusive as to the underlying motives of the participants.

2. I am grateful to an anonymous reviewer for pointing this out.

References

- Bock, K. & R. Warren. 1985. "Conceptual accessibility and syntactic structure in sentence formulation". *Cognition* 21:1.47–67.
- Bouma, G. 2008. *Starting a sentence in Dutch: A corpus study of subject- and object-fronting*. PhD diss., Groningen (published by University of Groningen).
- Bresnan, J, A. Cueni, T. Nikitina & H. Baayen. 2007. "Predicting the dative alternation". *Cognitive foundations of interpretation* ed. by G. Bouma, I. Krämer & J. Zwarts, 69–94. Amsterdam: Koninklijke Academie voor Wetenschappen.
- Comrie, B. 1989. *Language universals and linguistic typology: Syntax and morphology*. Oxford: Blackwell.
- Dahl, Ö. & K. Fraurud. 1996. "Animacy in grammar and discourse." *Reference and referent accessibility* ed. by Th. Fretheim & J.K. Gundel, 47–64. Amsterdam: John Benjamins.
- Ferreira, F. 1994. "Choice of passive voice is affected by verb type and animacy". *Journal of Memory and Language* 33.715–736.
- Gibson, E. 1998. "Linguistic complexity: locality of syntactic dependencies". *Cognition* 68.1–76.
- Haspelmath, M., M. Dryer, D. Gil & B. Comrie, eds. 2005. *The World atlas of language structures* (book with interactive CD-ROM). Oxford: Oxford University Press.
- Hawkins, J.A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

- de Hoop, H & M.J.A. Lamers. 2006. "Incremental distinguishability of subject and object." *Case, valency, and transitivity*. (Studies in Language Companion Series) ed. by L. Kulikov, A. Malchukov & P. de Swart, 269–290. Amsterdam: John Benjamins.
- Lamers, M.J.A. & H. de Hoop. 2008. Animacy and object fronting in Dutch. Unpublished Ms..
- Lamers, M.J.A. 2007. "Verb type, animacy and definiteness in grammatical function disambiguation". *Linguistics in the Netherlands 2007* ed. by B. Los & M. van Koppen, 125–137. Amsterdam/Philadelphia: John Benjamins.
- Lamers, M.J.A., K. Hofmans, P. Hagoort & H. de Hoop. 2006. The influence of animacy and verb type on word order in perception and production. Poster presented at the Conference Architecture and Mechanisms for Language Processing (AMLAP), Nijmegen, August 30.
- Lamers, M.J.A. 2005. "The online resolution of subject–object ambiguities with and without case-marking in Dutch: evidence from event-related brain potentials." *Competition and variation in natural languages: the case for case*. (Perspectives on Cognitive Science series) ed. by M. Amberber and H. de Hoop, 251–293. Amsterdam: Elsevier.
- Lamers, M.J.A. 2001. *Sentence processing: using syntactic, semantic, and thematic information*. PhD diss., Groningen (published by University of Groningen).
- McDonald, J.L., K. Bock & M. Kelly. 1993. "Word and world order: semantic, phonological and metrical determinants of serial position". *Cognitive Psychology* 25.188–230.
- Øvrelid, L. 2004. "Disambiguation of syntactic functions in Norwegian: modeling variation in word order interpretations conditioned by animacy and definiteness." *Proceedings of the 20th Scandinavian Conference of Linguistics* ed. by F. Karlsson. Helsinki: University of Helsinki.
- Rosenbach, A. 2005. "Animacy versus weight as determinants of grammatical variation in English". *Language* 81.613–644.
- Rosenbach, A. 2002. *Genitive variation in English. Conceptual factors in synchronic and diachronic studies*. Berlin: Mouton de Gruyter.
- Prat-Sala, M. & H.P. Branigan. 2000. "Discourse constraints on syntactic processing in language production: a cross-linguistic study in English and Spanish". *Journal of Memory and Language* 42.168–182.
- de Swart, P. 2007. *Cross-linguistic variation in object marking*. PhD diss., Nijmegen (published by LOT, Utrecht).
- van Tiel, B. & M.J.A. Lamers. 2007. "Animacy in verschillende teksttypes". *Tabu* 36.19–38.

Author's address

Geertje van Bergen
Radboud University Nijmegen
P.O Box 9103
6500 HD Nijmegen
The Netherlands
g.vanbergen@let.ru.nl