Edinburgh Research Explorer

Colliding Complementizers in Dutch: Another Syntactic OCP **Effect**

Citation for published version:

Ackema, P 2001, 'Colliding Complementizers in Dutch: Another Syntactic OCP Effect' Linguistic Inquiry, vol. 32, no. 4, pp. 717-727. DOI: 10.1162/002438901753373069

Digital Object Identifier (DOI):

10.1162/002438901753373069

Link:

Link to publication record in Edinburgh Research Explorer

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Linguistic Inquiry

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Squibs and Discussion

Colliding Complementizers in Dutch: Another Syntactic OCP Effect Peter Ackema Utrecht University

1 Introduction

According to the phonological Obligatory Contour Principle (OCP), it is undesirable to have adjacent segments with identical feature specification. In morphology and syntax, too, sequences of identical morphemes or heads sometimes produce unacceptable results (see George's (1980) Stuttering Prohibition, Menn and MacWhinney's (1984) Repeated Morph Constraint). Grimshaw (1997:170) adopts a general constraint along the lines of (1) as a morphosyntactic counterpart to the OCP:

(1) Sequences of identical functional heads are ill formed.

Menn and MacWhinney (1984) mention three strategies grammars employ to avoid violations of (1): omission (one of the offending elements is left out), avoidance (another construction is used), and suppletion. In some cases suppletion takes the form of replacing one of the offending heads by another head of the same class. This head has a different form (hence, (1) is satisfied), but a feature specification that does not match the morphosyntactic context. In other words, a violation of (1) can be avoided by using a "wrong" functional head. Most such cases in Menn and MacWhinney's overview involve clitic clusters. Discussing these, Grimshaw (1997) argues that the phenomenon can be explained as the result of the interaction of several hierarchically ordered constraints. In this squib I argue that the phenomenon of "replacement by a wrong head" also occurs in syntax proper—specifically, in cases where two identical complementizers find themselves adjacent in Dutch (section 3). I further argue that this phenomenon can be explained in parallel fashion to Grimshaw's account of clitic replacement and that a closer look at the possibilities with respect to this kind of replacement sheds additional light on the precise nature of the morphosyntactic OCP (section 4). First, however, in

I would like to thank Hans den Besten, Eric Hoekstra, Fred Weerman, and two anonymous reviewers for very useful comments on an earlier version of this squib.

section 2 I briefly sketch Grimshaw's analysis of clitic replacement in Romance.

2 Clitic Substitution

Some Italian clitics happen to have the same form despite having different feature specifications. One such pair is the impersonal subject clitic si and the reflexive clitic si. If these two are combined in one clause, the result is not the expected si si sequence. Instead, one of the clitics is replaced by ci, the clitic for first person plural arguments. It seems likely that a condition like (1) is responsible for the impossibility of si si; however, the assumption that there is a universal principle ruling out such sequences appears to be untenable, given that in the dialect of Conegliano si si is perfectly acceptable. Grimshaw (1997) argues that (1) can nevertheless be adopted as a universal principle, claiming that in cases in which it conflicts with another constraint languages differ as to which constraint must be satisfied and which can be violated, as in Optimality Theory (Prince and Smolensky 1993). Now, it is clear that in cases of clitic substitution (1) indeed conflicts with another principle—namely, the principle that when the syntactic context requires an element with a particular feature specification to be present, this element should indeed appear.

All clitics are listed in the lexicon with a particular feature specification. For instance, one si has features appropriate for impersonal subjects, which Grimshaw argues to be third person plural; one si is reflexive; and ci is specified as first person plural. Now, the computational system generates structures that contain clitic positions with certain features. Which clitic(s) is (are) the appropriate one(s) to use in the structure under consideration is then determined by assessing how well the various clitics satisfy a set of hierarchically ranked spellout constraints. (In other words, clitics are seen as the spell-out of feature bundles in syntax, rather than providing syntax with these features themselves; see also Miller and Sag 1997, Monachesi 1999, and others.) In simple cases with only one clitic, the best spell-out will be a clitic whose feature specification fully expresses and does not contradict the one provided by morphosyntax (if the lexicon of the language contains such a clitic). In particular, every feature in the morphosyntactic structure should be properly expressed by a clitic specified for that feature. This is expressed by PARSE constraints, the general format of which is given in (2). Conversely, a clitic specified for a feature that is not present in the morphosyntactic structure should not be used. This is expressed by FILL constraints, the general format of which is given in (3).

(2) Parse(Feature)

A feature in the input must appear in the output.

(3) Fill(Feature)

A feature in the output must appear in the input.

Considering the $si \ si \rightarrow ci \ si$ switch again, suppose that the si that is replaced by ci is the impersonal subject clitic. (A similar analysis can be given if reflexive si is replaced.) Since the impersonal subject is specified as third person plural, and the lexical specification of ci is first person plural, both Parse(3rd person) and Fill(1st person) are violated if ci is used as the spell-out for the impersonal subject. That this nevertheless occurs in the standard language indicates that (1) is more important than these faithfulness constraints here. Finally, it must be ensured that some clitic appears, even if it is the wrong one, instead of one of the si's being deleted. For this reason Grimshaw introduces a general Parse(Clitic) constraint, which states that every feature bundle must have some spell-out. The analysis is then as shown in (4).

(4)	Input: impersonal reflexive	Parse (Clitic)	*XX	Parse (3rd pers)	FILL (1st pers)
	si si		*!		
	r ci si			*	*
	si	*!		*	

In the Conegliano dialect PARSE(3rd person) is more important than *XX (the OCP), so here a *si si* sequence results, rather than *si* replacement.

3 Complementizer Substitution in Dutch

So, a sequence of two identical functional heads is sometimes avoided by replacing one of them with a functional head of the same type whose form differs but whose feature specification is inappropriate for the morphosyntactic context. In this section I argue that the same phenomenon occurs with complementizers in Dutch. In this language the form of the coordinating complementizer 'or' happens to be the same as that of the interrogative subordinating complementizer 'whether', namely, of. In some dialects an embedded interrogative clause can start with the sequence of dat 'whether that'. In standard Dutch this is normally impossible; nevertheless, of dat sequences do occur in specific contexts in standard Dutch as well. Consider for instance (5), from De Rooij 1978:147.

(5) Ik vraag hem of hij zich verslapen heeft of dat hij I ask him whether he REFL overslept has or that he ziek is. ill is According to De Rooij, (5) is only possible when one particular constituent occurs after the first of 'whether': namely, an embedded interrogative clause consisting of two conjoined clauses (I ask him, 'Have you overslept or are you ill?''). De Rooij further claims that (5) does not have a reading such that there are two conjoined main clauses, each with its own embedded interrogative clause, to which conjunction reduction has applied (either I ask him whether he overslept or I ask him whether he is ill). The assumption that an of dat sequence signals one embedded interrogative clause that consists of a conjunction is also apparent from Hoekstra's (1992) analysis of examples like this. Hoekstra assumes that dat is deleted when it is adjacent to subordinating of, but not when it is adjacent to coordinating of, so (5) receives the analysis in (6) (where — indicates a deleted dat).

(6) ik vraag hem [of [[— hij zich verslapen heeft] of [dat hij ziek is]]]

If Hoekstra's assumption is correct, the occurrence of *dat* in the second clause is not really surprising—this clause is simply conjoined with another *dat*-clause, whose *dat* is deleted. Crucially, however, it can be shown that the *of dat* sequence can also occur when the second clause is a dependent interrogative by itself and hence should be introduced by *of* 'whether'. Imagine the following context: I am sitting absentmindedly in a train. Suddenly I notice that another traveler is pointing at the seat next to me, and I realize that she has just asked me something about the availability of that seat. Because my thoughts were elsewhere, I didn't hear whether she asked, 'Is that seat free?' or 'Is that seat occupied?', and so I don't know whether to answer yes or no. Hence, I ask her:

(7) Vroeg je nou of die plaats bezet is of dat asked you now whether that seat occupied is or that hij vrij is? he free is 'Did you ask whether that seat is occupied or whether it is free?'

Here we are dealing with two embedded interrogatives (i.e., two conjoined main clauses plus conjunction reduction), because what I want to know is which of the two options she asked a question about. (The interpretation of (7) is 'Did you ask A or did you ask B?', not 'Did you ask, "A or B''?'. The latter interpretation is not impossible, but it is not intended in the given context.) Hence, this example must be analyzed as follows:

(8) vroeg je nou [of [die plaats bezet is]] of vroeg je nou [dat [hij vrij is]]

This means that the right conjunct really consists of an embedded interrogative clause that is introduced by dat 'that' rather than by of

'whether'. Retaining of 'whether' here results in a sentence that is significantly worse than (7). 2

(9) ?*Vroeg je nou of die plaats bezet is of of hij vrij is?

That (1) is responsible for the occurrence of *dat* in (7) becomes even clearer when it is compared with (10). Here conjunction reduction has not applied, so the subordinating complementizer in the second conjunct is not adjacent to the coordinator. The expected interrogative *of* appears; replacing it with *dat* is absolutely impossible now. Also, if the coordinator in a construction like (7) is not *of* but for instance *en* 'and', substitution is impossible as well; see (11).

- (10) Vroeg je of die plaats bezet is of vroeg je asked you whether that place occupied is or asked you of/*dat hij vrij is?
 whether/that he free is
- (11) Ze vroeg of die plaats vrij was en of/*dat she asked whether that seat free was and whether/that het raam open mocht.

 the window open might 'She asked whether that seat was free and whether she might open the window.'

4 Discussion

4.1 Analysis of Complementizer Substitution in Dutch

A matrix verb like vragen 'ask' selects an interrogative complement clause, that is, a complement clause headed by a [+Q] C-position. Now, on a par with Grimshaw's (1997) view of clitics, let us suppose that it is not so much that the insertion of a particular complementizer supplies the relevant CP with a [+Q] feature, but that the computa-

- (i) Ik weet niet of ie zijn stuk al af heeft of I know not whether he his piece already finished has or *of/dat ie lui is geweest.
 - *whether/that he lazy is been

'I do not know whether he has finished his piece already or *whether/ that he has been lazy.'

It seems evident that this example must be interpreted such that I do not know one thing (namely, which of the two situations holds), not that I either do not know the first thing or do not know the second thing. This example must hence be analyzed on a par with (6), so it still fits the pattern expected by De Rooij and Hoekstra, in contrast to (7)–(8).

² De Rooij (1978) appears to accept this kind of example. All speakers I asked (about ten) agree with the relative judgments of (7) and (9), however: (9) is judged either as simply unacceptable or at least as significantly worse than (7).

¹ Den Besten (1983:90) also assumes of-substitution by dat in (i).

tional system only provides structures in which the complement CP to interrogative verbs is headed by a [+Q] C-position. Spell-out principles must then determine which complementizer is the optimal one for this position. Since of 'whether' is lexically specified as being [+Q], it will normally be the best spell-out for a [+Q] C-position, since none of the faithfulness constraints in (12)-(14) (cf. the constraints in section 2) are violated in that case (α is '+' or '-').

- (12) PARSE([Q])If $[\alpha Q]$ appears in the input, $[\alpha Q]$ must appear in the output.
- If $[\alpha Q]$ appears in the output, $[\alpha Q]$ must appear in the input.
- (14) PARSE(C)A C-position must have some spell-out.

In cases like (7) the OCP in (1) comes into the picture. If (1) is more important than (12)–(13), the C-position is spelled out by a "wrong" complementizer, provided that (14) also outranks (12)–(13). But why is it spelled out by dat 'that'? Suppose, as a first option, that dat is lexically specified as [-Q]. In that case the $of \rightarrow dat$ switch is indeed explained under the indicated ranking; see (15).

(15)	Input: coordinating C - [+Q] subordinating C	Parse(C)	*XX	Fill([Q])	Parse([Q])
	of of([+Q])		*!		
	r of dat([−Q])			*([-Q])	*([+Q])
	of –	*!			*([+Q])

However, this option has several disadvantages. First, it does not explain why it is dat that replaces of in this context, rather than any other complementizer—for instance, als 'if'. Second, it does not explain why we do not find the reverse of the $of \rightarrow dat$ switch in Dutch. A dat dat complementizer sequence is quite awkward, as illustrated by the well-known oddity of uttering a that-clause in the subject position of another that-clause, which holds in Dutch just as in English; see (16a). However, this cannot be remedied by replacing one dat by of; see (16b). Instead, speakers adopt one of the other strategies mentioned by Menn and MacWhinney (1984), namely, avoidance: as shown in (16c), speakers use a structure in which the offending thatclause is extraposed.

- (16) a. ?*Dat dat Jan rookt Piet verontrust is duidelijk. that that Jan smokes Piet bothers is obvious 'It is obvious that it bothers Piet that Jan smokes.' b. (*Dat of/*Of dat) Jan rookt Piet verontrust is duidelijk.

Dat het Piet verontrust dat Jan rookt is duidelijk.
 that it Piet bothers that Jan smokes is obvious

Such problems can be solved by an alternative assumption, namely, that dat is in fact not specified for the [Q] feature, or for any other feature at all, and that in this it contrasts with all other complementizers. If this assumption is correct, dat never violates any FILL constraint. It follows that dat replaces of in OCP contexts, since the structure with dat is more acceptable under any constraint ranking than structures with any other complementizer acting as substitute. That dat is not replaced by of can now also be made to follow, namely, by ranking Fill(Q) above the OCP. That ranking was not possible under the assumption that dat is specified as [-Q], since the of \rightarrow dat switch would be prohibited then as well. If dat is unspecified, this prohibition no longer holds. Instead, the crucial ranking for the $of \rightarrow$ dat switch is now just OCP \gg PARSE([Q]). However, a dat \rightarrow of switch is prohibited when $Fill(Q) \gg OCP$, because the [+Q] specification of of triggers a violation of Fill([Q]) in a syntactically non-[+Q] context. Compare (17a) (of \rightarrow dat in OCP contexts) and (17b) (no $dat \rightarrow of$ in OCP contexts).

(17) a.	Input: C C[+Q]	Parse(C)	Fill([Q])	*XX	Parse([Q])
	of of([+Q])			*!	
	r of dat([0Q])				*([+Q])
	of –	*!			*([+Q])

b.	Input: C C[−Q]	Parse(C)	Fill([Q])	*XX	Parse([Q])
	r dat dat([0Q])			*	*([-Q])
	dat of([+Q])		*!([+Q])		*([-Q])
	dat –	*!			*([-Q])

Note that the best option for a *dat dat* sequence thus is to retain the sequence. As noted, the OCP violation this incurs still makes such examples rather bad, although not all speakers agree on this and some find them quite acceptable (Hans den Besten, personal communication). In this case most speakers prefer using another structure as in (16c).

4.2 On the Nature of the Morphosyntactic OCP: Content or Form

So far I have been assuming that *identical* in (1) refers to forms—in the syntactic Dutch case discussed above, the form *of* 'whether'. However, as pointed out by Kornfilt (1984), morphosyntactic OCP effects can also occur when two elements with comparable *content* are adja-

cent. Kornfilt discusses a case from Turkish in which two nominal agreement morphemes cannot be adjacent, despite having different forms. Similarly, Grimshaw (1997) discusses a case in which clitic substitution takes place even though the clitics involved do not have identical shapes, namely, the case of "spurious se" in Spanish. Here, when a third person accusative clitic (lo in isolation) and a third person dative clitic (le in isolation) are combined, the dative is replaced by se. As Grimshaw suggests, possibly the shared third person feature is enough to violate the OCP. Yet another case is discussed by Mohanan (1994), who shows that in Hindi OCP effects are triggered when two elements with the same Case features are adjacent.

That content rather than form can be relevant for (1) is also shown in the case of the Dutch $of \rightarrow dat$ complementizer switch. In colloquial registers Dutch allows doubly filled Comp violations in embedded interrogatives. In some dialects an $of \rightarrow dat$ switch takes place in this context (see Hoekstra 1992).

(18) a. Standard Dutch

Ze vroeg wie het boek gelezen had. she asked who the book read had 'She asked who had read the book.'

- b. Colloquial variant of (18a)

 Ze vroeg wie of het boek gelezen had.
- c. Regional variant of (18b)

 Ze vroeg wie dat het boek gelezen had.

Hoekstra argues that the replacement of of by dat in this environment pleads against analyzing cases like (5) in terms of substitution, since the same rule would then apply in two different contexts. However, an analysis parallel to the one provided for (7) can be given for (18c). This is because the switch in (18c) can be assumed to be triggered by the OCP (so an analysis similar to (17a) applies), the only difference being that this is a case of shared featural content rather than shared form: both the wh-word and of are interrogative, so they share a [+Q] feature, which dat lacks.

Given these two possible interpretations of the OCP, one might wonder if there is a connection between them. I think it can be argued that the "content" interpretation is basic and that the "form" interpretation only comes into play in some cases as an additional factor. This is because in those cases where identical form seems to be the relevant factor, it appears the two elements in question must still share some featural content in order for the OCP effect to be triggered. In particular, it appears they must at least be of the same (functional) category. For example, Monachesi (1999) points out that a *si si* sequence (see section 2) is perfectly acceptable in Italian if the first *si* is the complementizer 'if' and the second is a clitic. Something similar can be observed for the Dutch case discussed above. As noted, a sequence of *dat dat* complementizers is deviant for many speakers; recall (16a). But a *dat dat* sequence in which the second *dat* is not a complementizer

but a determiner is perfectly fine (*Ik geloof dat dat meisje Marie heet* 'I believe that that girl is called Marie'). Yet another phenomenon indicating that phonologically identical elements do not trigger a syntactic OCP effect if they do not share at least some (categorial) features is that the effect disappears if one of the two offending elements is used metalinguistically, as a name for itself; see (19) (see also Battus 1981:113–114). When used as a name, the element in question arguably has been converted to a noun, since in all respects it now behaves like the head of a nominal phrase (see Kruisinga 1932). This means it no longer shares the functional categorial features with its nonconverted counterpart (whereas, obviously, it is still identical in form).

(19) Hij vroeg of *of* een voegwoord is. he asked whether *whether* a complementizer is 'He asked whether *whether* is a complementizer.'

We may conclude that the featural interpretation of the morphosyntactic OCP is basic. Moreover, a qualification can be added to the effect that the two feature-sharing elements involved are prohibited from occurring in adjacent positions only if they *also* have the same form.

(20) OCP *[f]
$$\alpha$$
 [f] if α = 0 (if [f] is spelled out by the same form twice)

The OCP can hold either in its unqualified form or with the parenthe-sized qualification. Whether or not the qualification holds in a particular context in a particular language does not seem predictable. In general, cases where the OCP holds in its unqualified form are likely to involve features that not many elements share; whereas if more general features (e.g., category features) are involved, it is more likely that the OCP only holds with the qualification—otherwise, all occurrences of two categorially identical elements would be ruled out, in all probability an overly severe filter on possible outputs.

5 Conclusion

Data concerning complementizer substitution in Dutch show that in syntax, as in morphology, a condition that bans adjacent identical heads can have the effect that one head is replaced by a counterpart that has the wrong feature specification for the syntactic context. This can be explained relatively straightforwardly in a model in which only nodes and their associated features are present in syntax—overt lexemes are not. Principles of selection in (morpho)syntax then can concern selection only of particular features, not of some particular form (see Ackema and Neeleman 2000 for more discussion). Syntactic selectional principles hence are not violated in the case of complementizer substitution discussed here. Besides the (morpho)syntax-internal principles there must then be mapping principles that state

how the (morpho)syntactic features must be spelled out (see, e.g., Sproat 1985, Halle and Marantz 1993). It is with these principles that the OCP in (20) interacts (i.e., (20) is a condition on outputs, not on inputs), and it is this interaction that sometimes gives rise to the phenomenon that one feature is spelled out by a form lexically specified to spell out a different feature.

References

- Ackema, Peter, and Ad Neeleman. 2000. M-selection and phrasal affixation. In *UCL working papers in linguistics* 12, 307–342. Department of Phonetics and Linguistics, University College London.
- Battus. 1981. Opperlandse taal- en letterkunde. Amsterdam: Querido. Besten, Hans den. 1983. On the interaction of root transformations and lexical deletive rules. In On the formal syntax of the Westgermania, ed. Werner Abraham, 47–131. Amsterdam: John Benjamins.
- George, Leland. 1980. Analogical generalizations of natural language syntax. Doctoral dissertation, MIT, Cambridge, Mass.
- Grimshaw, Jane. 1997. The best clitic: Constraint conflict in morphosyntax. In *Elements of grammar*, ed. Liliane Haegeman, 169–196. Dordrecht: Kluwer.
- Halle, Morris, and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In *The view from Building 20: Essays* in linguistics in honor of Sylvain Bromberger, ed. Kenneth Hale and Samuel Jay Keyser, 111–176. Cambridge, Mass.: MIT Press.
- Hoekstra, Eric. 1992. *Of* en *dat* nader bekeken. *De Nieuwe Taalgids* 85:441–445.
- Kornfilt, Jaklin. 1984. The Stuttering Prohibition and morpheme deletion in Turkish. In *Proceedings of the Turkish Linguistics Conference*, ed. Ayhan Aksu Koç and Eser Erguvanli Taylan, 59–83. Istanbul: Bogaziçi University Publications.
- Kruisinga, Etsko. 1932. A handbook of present-day English. Vol. 2. 5th ed. Utrecht: Noordhoff.
- Menn, Lise, and Brian MacWhinney. 1984. The Repeated Morph Constraint: Toward an explanation. *Language* 60:519–541.
- Miller, Philip, and Ivan Sag. 1997. French clitic movement without clitics or movement. *Natural Language & Linguistic Theory* 15:573–639.
- Mohanan, Tara. 1994. Case OCP: A constraint on word order in Hindi. In *Theoretical perspectives on word order in South Asian languages*, ed. Miriam Butt, Tracy Holloway King, and Gillian Ramchand, 185–216. Stanford, Calif.: CSLI Publications.
- Monachesi, Paola. 1999. A lexical approach to Italian cliticization. Stanford, Calif.: CSLI Publications.
- Prince, Alan, and Paul Smolensky. 1993. Optimality Theory. Ms.,

Rutgers University, New Brunswick, N.J., and University of Colorado, Boulder.

Rooij, Jaap de. 1978. Het ene *of* is het andere niet. *De Nieuwe Taalgids* 71:146–148.

Sproat, Richard. 1985. On deriving the lexicon. Doctoral dissertation, MIT, Cambridge, Mass.

Some Problems with a Lowering Account of Scrambling Jason C. Johnston University of Sydney Iksan Park University of Sydney Bošković and Takahashi (1998) (B&T) propose that scrambled arguments are base-generated in their surface positions (adjoined to IP), but may undergo covert lowering to their canonical positions in order to be assigned a θ-role at LF, in accordance with Last Resort. They argue that since no independent principle of grammar requires it, such movement need not leave a trace, and therefore the Proper Binding Condition does not come into play.

They cite examples, principally from Japanese, showing that the appropriate LF configuration to account for various binding and scope phenomena is that in which the scrambled element appears in its canonical position (and nowhere else). For instance, in (1) (B&T 1998:354) the scrambled dative QP *daremo-ni* 'everyone' cannot take scope over the nominative QP *dareka-ga* 'someone' despite its surface position.

= for some x, x a person, x thinks that for every y, y a person. Mary met y

 \neq for every y, y a person, there is some x, x a person, such that x thinks that Mary met y

Similarly, in (2) (B&T 1998:355) the anaphor *otagai-no* 'each other' cannot be bound by the scrambled argument *Mary to Pam-ni* 'Mary and Pam'.²

(2) *[Mary to Pam]_i-ni [otagai_i-no hahaoya]-ga
Mary and Pam-dat each.other-gen mother-nom
[John-ga e_i atta to] omotteiru.
John-nom met that think
'Mary and Pam, each other's mothers think that John met.'

We are grateful to the following people for encouragement, useful comments, and suggestions: Hee-Don Ahn, Sung-Ho Ahn, Daeho Chung, Jeong-Shik Lee, Hongkeun Park, Jane Simpson, and two anonymous *LI* reviewers.

¹ They also propose an alternative mechanism for clause-internally scrambled arguments, but we do not deal with this here.

² An account of this based on the A/Ā distinction is of course available, but B&T wish to minimize or eliminate this distinction. In any case, as they point out, it remains possible that the scrambled position in multiple-subject languages like Japanese is in fact a specifier of IP.