Extraction with deep anaphora? The role of VP external orphans

Enrico Boone and Anikó Lipták (Leiden University Centre for Linguistics)

1. The issue Since Hankamer and Sag (1976), anaphora is classified in two basic types: syntactically atomic deep anaphora and fully articulated surface anaphora containing syntactic structure. The two types are predicted to show differences in extraction possibilities, which has become a standard diagnostic in identifying the two types. Extraction data, however, are often contradictory: Danish *det*, argued to be a surface VP anaphor (VPA) by Houser et al (2007), is compatible with A- but not with Ā-extraction, similarly to its surface anaphoric Norwegian cognate *det* (Bentzen et al 2013).

In this talk, we present a case study of Dutch VP anaphora and show that extraction data are an unreliable diagnostic for the presence of syntactic structure. While Dutch *dat* is a deep anaphor, it allows for A-extraction and some instances of Ā-extraction. We show that this behavior is fully compatible with the atomic nature of *dat* once one recognizes that the extractable constituents (both in the A- and Ā-domains) must be capable to syntactically integrate into the structure. Moreover, we show that the extractable constituent must be semantically integrated via a well-formed λ-expression as well (along the lines of Mikkelsen et al 2012).

2. *Dat* VPA is a deep anaphor In Dutch, an anaphoric VP can be overtly realized by the demonstrative pronoun *dat* (or the neuter personal pronoun *het*). According to standardly used diagnostics for deep vs. surface anaphora (originating from Hankamer and Sag 1976, see also Houser et al 2007, Mikkelsen et al 2013), *dat* is without a doubt a deep anaphor: it can have non-linguistic antecedents and it allows for voice mismatches (e.g. passive / active) between the pronominalized VP and its correlate in the antecedent (cf.1).

(1) Het vuilnis moest buiten worden gezet, en dus heeft Jan **dat** gedaan.
the litter must outside BE.INF placed and PRT has J. that done

Further supporting the assumption that *dat* is a deep anaphor, is its syntactic distribution, which patterns fully with that of a DP in that: (**i**) *dat* can only occur in syntactic positions that are reserved for nominal constituents (including a scrambling and topicalization position); and (**ii**) *dat* can only co-occur with verbs that c-select DP arguments, see e.g. (2/3):.

(2) Jan doet / ziet / kan / *laat / * dwingt / * begint { een truckje / **dat** }.
J. does / sees / can / let / forces / begins a trick that

(3) Jan begint liedjes te zingen. * Klaas begint **dat** ook.
J. begins songs *te* sing.INF K. begins that too

3. Extraction with *dat* VPA Given that *dat* is a deep anaphor, the behavior of *dat* VPA with respect to 'extraction' is at first sight puzzling: it allows for A-extraction with passives and unaccusatives (cf. 4), disallows Ā-extraction with DPs (5a), and for 8 out of 10 informants, it allows Ā-extraction with some PP arguments (5b).

(4) a. Piet werd gearresteerd en Marie werd **dat** ook. (*A-extraction*)
P was arrested and M. was that too

b. Dit vliegtuig is geland en dat vliegtuig is **dat** ook.
this plane is landed and that plane is that too

(5) a. Jan leende Marie wel een boek. * Wie deed hij **dat** niet? (*Ā-extraction*)
J. lent M. AFF a boek who did he *dat* not?

b. Jan leende wel een boek aan Marie. Aan wie deed hij **dat** niet?
J. lent AFF a book P M. P who did he *dat* not?

The possibility of Ā-extraction for the speakers who accept (5b) furthermore co-exists with that of base generating the PP argument next to *dat* (cf. 6b), something that is only allowed with PP but not with non-subject DPs (cf. 6b). Such a PP associate of *dat* forms a single VP constituent with *doe dat* (which can be evidenced by e.g. VP topicalization, not illustrated here):

(6) a. ✓ Hij deed **dat** ook aan Susan. b. * Hij deed **dat** Susan ook.
he did that too P Susan he did that Susan too

Base generation moreover is possible only for argument PP, (subject) DP, but not for AP or CP associates:

(7) * Jan heeft gezegd [$_{CP}$ dat hij zal komen] en Piet heeft **dat** [$_{CP}$ dat hij niet komt].
Jan has said dat he shall come and Piet has that dat he not comes

The PP-category restriction on associates of *dat* make the Dutch facts look a lot like so-called *orphan* PPs that appear next to the VP anaphors *do it/do the same* in English (Mikkelsen et al 2012):

(8) You have [jilted <u>two fiancés</u>] and I expect you would [$_{VP}$ [$_{VP}$ *do the same*] [$_{orphan}$ to me]]

However, whereas in English such orphan PPs are restricted to PPs headed by a specific set of prepositions (*to/for/with*), PP associates of *dat* in Dutch are not so restricted: prepositions selected by the antecedent predicate can also be orphans (cf. 6a).

4. The account As noted, *dat* is a DP proform selected by a DP-selecting verb like *doen*, auxiliaries or modals: *[$_{VP}$ doen **dat**]*. Orphan associates that occur next to *dat* must be able to be syntactically integrated in the extended structure above this VP. Importantly, DPs cannot be introduced into this extended structure as adjuncts, as they would fail to receive case if integrated this way. This is precisely what rules out (6b): the only head available for assigning case on *Susan* is v^0. Crucially, v^0 already assigns accusative case to *dat*, as illustrated in (10).

(9) * [$_{TP}$ Hij [$_{vP}$ t_{hij} [$_{vP}$ $v^0$$_{\text{~~acc~~}}$ [$_{VP}$ deed [$_{DP}$ **dat**$_{\text{~~acc~~}}$]] [$_{DP}$ Susan$_{acc}$]]]]] (cf. 6b)

Subject DPs, even in the unaccusative and passive examples in (4), successfully integrate into the structure, due to the fact that they can start out in Spec,*v*P (instead of inside the VP, cf. Neeleman & Weerman 1999) and subsequently get assigned case by T^0.

PP arguments (but not CPs or APs) can successfully combine with the *doe dat* VP syntactically, as (**i**) they are allowed to be merged outside their selecting predicate in line with works like Barbiers (1995), Helmantel (2002), Koster (2000) and (**ii**) they do not need to receive case. Our account explains the extraction pattern in (4-6): all possible extractees are introduced VP-external (i.e. external to *dat*) and thus never actually extract *out of* the anaphor.

5. Semantic integration of orphan associates The semantic composition of orphan associates with their internal argument slot proceeds as proposed by Mikkelsen et al (2012): the presence of the associate forces λ-abstraction over the antecedent of *dat*. The resulting lambda function can then apply to the argument (i.e. the orphan) via functional application:

(10) a. [λx.jilt x] (me) English (8) *from Mikkelsen et al 2012*
b. [λx.leende x een boek] (Susan) Dutch (6a)

Importantly, this successfully rules out cases in which the orphan associate is a predicative function itself, like a secondary predicate, as in (11). This shows that next to syntactic integration (as described above), semantic integration must also be successful.

(11) * Ik beschouw Jan als intelligent, en jij doet **dat** als stom.
I consider Jan as intelligent and you do that as stupid

Overall, an important result of our theory is that extraction cannot be taken as first hand evidence for structure: what looks like an extracted constituent might originate outside the anaphor.

Selected references **Bentzen**, **Merchant** & **Svenonius** 2013 Deep properties of surface pronouns: Pronominal predicate anaphors in Norwegian and German • **Houser**, **Mikkelsen** & **Toosarvandani** 2007 Verb phrase pronominalization in Danish: Deep or surface anaphora? *WECOL*. • Koster 2000 Extraposition as parallel construal. Ms. • **Mikkelsen, Hardt & Ørnses** 2012 Orphans hosted by VP anaphora *WCCFL* 29. • **Neeleman & Weerman** 1999 *Flexible syntax: a theory of case and arguments*. Kluwer.