1. Introduction

Speas & Tenny (2003) (henceforth S&T) have put forward a forceful and thought-provoking argument that the relation between syntax and pragmatics can be modeled more insightfully if syntax takes over a larger share of the burden. In particular, S&T suggest that well-established tools from formal syntax rather than philosophico-conceptual a priori considerations or logico-semantic discourse representations provide the correct constraints for grammatical encodings and their empirical consequences. The domains of application are grammaticalized illocutionary force, i.e. sentence mood, and point of view related phenomena. Combining recent work on functional projections in the C-domain (Cinque 1999; Rizzi 1997) and on the mapping between lexicon and argument structure (Hale and Keyser 2002), S&T hypothesize that grammatically relevant properties for the mentioned domains are encoded in or deducible from the following syntactic configurations.

(1) represents the formal encoding of sentence mood in terms of a layered "speech act phrase" (SA(*)P) configurationally defining the "P(ragmatic)-roles" SPEAKER (S-P), UTTERANCE CONTENT (UC-P), and HEARER (H-P).  

(2) represents the formal basis for point of view phenomena in terms of a layered "sentience phrase" (Sen(*)P) configurationally defining the P-roles SEAT OF KNOWLEDGE (SK-P) and EVIDENCE (E-P).

\[ \text{(1) represents the formal encoding of sentence mood in terms of a layered "speech act phrase" (SA(*)P) configurationally defining the "P(ragmatic)-roles" SPEAKER (S-P), UTTERANCE CONTENT (UC-P), and HEARER (H-P).} \]

\[ \text{(2) represents the formal basis for point of view phenomena in terms of a layered "sentience phrase" (Sen(*)P) configurationally defining the P-roles SEAT OF KNOWLEDGE (SK-P) and EVIDENCE (E-P).} \]

---

1 There is an intentional analogy to the configurational definition of Θ-roles AGENT (A-Θ), THEME (TH-Θ), and GOAL (G-Θ) in approaches like Hale & Keyser (2002).

2 With some hesitation, S&T take SenP to be a unification of the more familiar "Evaluation Phrase" (EvalP = SenP) and "Evidential Phrase" (EvidP = Sen*P), for which one may consult Cinque (1999).
specifier of SA*P and thus, acquiring the P-role UC-P, encodes the utterance content. Sen*P further embeds CP, which stands as placeholder for the remainder of the sentence.

Crucially, S&T claim that (1) forms the basis for explaining why grammars usually encode no more than four basic sentence moods, namely, declarative, interrogative, imperative, and subjunctive. Varying the finiteness of YP_{UC-P} in (1), S&T define declarative [+finite] and subjunctive [-finite] mood. The other two types are transformationally derived from (1) by an application of a counterpart of dative shift (cf. Larson 1988), as shown in (3).

\[ \text{(3)} \]

Again, the finiteness of YP_{UC-P} makes a difference, defining an interrogative if [+finite] and an imperative if [-finite] is chosen. Independent formal assumptions ensure that this exhausts the possibilities. Likewise (1) and (2) together define the maximal range of P-roles, guaranteeing that no more than three "sentient" P-roles, i.e. S-P, H-P, and SK-P can be grammatically active. S&T then go on to provide empirical evidence for their system, which they claim is predictable in terms of formal notions like c-command, control via binding, and locality.

The aim of our discussion is to show that S&T's approach is inadequate in substantial respects. In particular we will show in section 2 that S&T's inventory of sentence moods is a product of arbitrary stipulation rather than formal syntactic deduction. In section 3 we will further demonstrate that S&T's empirical arguments in favor of their system cannot be upheld in the light of both technical problems and empirical counterevidence. We take this as an interesting and important result in so far as it seriously challenges the kind of syntactic reductionism S&T are pursuing. It is our impression that more satisfactory revisions of S&T's system will only be forthcoming if one allows for substantial and explicit assumptions from an outside theory of language and discourse to interact with syntax.

2. Deriving the Inventory of Sentence Moods and the Size of SAP and SenP
S&T take "inspiration" for postulating structures (1)/(2) from Hale & Keyser (2002) (henceforth H&K), who set out to configurationally define the bounds on possible lexical projections and thus indirectly define the notion "possible word." However, (1) and (2) correspond to only one out of four admissible basic configurations allowed by H&K, namely,

---

3 Sadock & Zwicky (1985) find only three major ("frequent") types, excluding S&T's "subjunctive." We follow S&T's terminology, noting that a proper distinction between formal sentence types and "functional" sentence moods is called for (cf. for example Grewendorf and Zaefferer 1991). It is also interesting to note that Grewendorf (2002:66ff) relabels Rizzi's influential "ForceP", the counterpart of "SAP", as "TypeP". This shift toward a more neutral formal characterization of sentences is extremely useful, especially so when it comes to subordination.

4 The standard use of (root) "subjunctives" is to express wishes (S&T:318). Thus "optative" may be an alternative label for this sentence mood. S&T note the tendency of subjunctives to occur in embedded structures.

5 For a critique of H&K's approach, see Fodor & Lepore (1999).
the maximal one. Structures (4a)-(4c), although illustrated in S&T (p.320), are discarded as candidates without discussion.

(4) a. X° b. XP c. XP

X° ZP YP X'

In order to motivate such an omission, it would, as a first step, have to be argued why languages could not dispose of four different basic SAP-domains. Otherwise, (4a)-(4c) could be taken to define between three and five additional basic sentence moods, bringing the total inventory up to between seven and nine, contrary to common assumptions. As a second step, it would have to be argued why (1) is the correct choice for SAP rather than any of (4a)-(4c). It is unclear whether it would suffice to claim that SA₁*₂ is supposed to encode an abstract predicate of communication, there being a whole range of different such predicates with different argument structures. Such a claim would not be based on purely formal grounds anyway. Likewise, recourse to the theory of functional projections would at most rule out (4a), if we take it as given that functional heads require at least one argument. Clearly, the obvious point that (1) meets the "virtual conceptual necessities" of any standard theory of communication (cf. Shannon and Weaver 1949) by containing representations of a sender (XP₁₁), a receiver (ZP₁₂), and a message (YP₁₃), would lie outside formal syntax, which according to S&T provides a superior explanatory perspective nevertheless.

Sideling this fundamental question, S&T instead touch on the issue of checking resources and recursivity as imposing further restrictions. It is conceded right away that "[i]n most of Hale and Keyser's work, it is not clear how recursion is to be ruled out" (p.339,fn.5). In fact, in dealing with double object configurations, H&K (p.163) appeal to a "recursive (b)-type structure embedded in an (a)-type structure," where (b)-type corresponds to (4c). This shows that no principled ban on recursion can be derived from H&K's theory. Neither does such a ban seem to be part of current approaches to functional projections, where e.g. Rizzi (1997), driven by empirical considerations, makes use of * (the Kleene star operator) to allow for an unbounded number of TopPs in the C-domain.

S&T's more substantial attempt at limiting the size of SAP/SenP involves the idea that "the head can move only once" (p.319), which itself would follow from the "stipulation [...] that each head can only have one feature to check (or, all features must be checked in the same position)" (p.339,fn.7). Of course, the one-move-constraint has a trivial reading once pied piping is disregarded. Thus, classical V°-to-I°-to-C°-movement by adjunction involves only one instance of V°-movement, given that [I° V° I° ]-to-C° counts as I°-movement that pied pipes V°. But, of course, ruling out pied piping in general does not seem to be desirable. Thus, imposing an one-feature-per-head-constraint on checking may be a more promising strategy. Assume SA*° has an attractable feature, ∧f, and SA° an attracting one, ∨f. If these cancel after SA*°-to-SA° in (1), the size of the structure is fixed. However, the projections in (1)/(2) have to connect to the sentence they are heading. In particular, given that Sen*P is modeled after Cinque's (1999) "Evidential Phrase," which can be realized by overt

---

6 We assume that finiteness of at least ZP, i.e. the complement of X° could be the source for variation as before. Sadock & Zwicky's (1985) subtypes and minor types might potentially be mapped onto such additional structures. See also Truckenbrodt (2004), who argues for a "hearer-less" subtype of interrogatives in German.

7 For examples, see McCawley (1977) and Ballmer & Brennenstuhl (1981). From this perspective it would actually be plausible to argue for an extended SAP-structure introducing an additional "aboutness argument": "speaker communicates to hearer about x that . . . " Of course, this function is often coded sentence-internally in terms of TopP.

8 We disregard the possibility of allowing multiple specifiers as argued for by Chomsky (1995).
morphology in some languages, Sen° would have to check such features against (a category enclosing) V° eventually. Thus, in addition to a °g, canceling against some °g of Sen°, Sen° needs some °h in order to license evidential morphology in combination with a substantial bearer. Consequently, there is no deeper (formal) reason why SA° could not have some °i in addition to its °j, i.e. no reason why SAP could not be extended further up, licensing additional (more complex) sentence moods. The theoretical point about these considerations is that generative syntax in general and head movement in particular are formally designed to enable recursion. Thus, local bits of non-recursive syntax have to be motivated by independent principles.

Another arbitrary feature of S&T's system is the exclusive appeal to dative-shift, i.e. 3-2 promotion, as generating additional sentence moods. Passive- or middle-like 2-1 promotions, which can be found in H&K's toolkit, are not considered. Certainly, S&T could not claim that 3-2 diatheses are less marked and therefore more suitable for defining unmarked sentence moods. Instead, 2-1 promotions would seem to lead to unwelcome empirical predictions having to do with binding and locality. The crucial point is that a multiplication of possible sentence moods by application of 2-1 promotions is ruled out by mere stipulation. The underlying formal syntax does not of itself warrant such an exclusion.

3. Explaining the Behavior of Indexicals and Logophoric Reflexive(s) in Terms of C-Command and Locality

Let us move on to the independent empirical evidence S&T put forward in favor of their syntax-based approach. The main claim is that c-command and locality are the proper tools for treating (the interpretation of) among other things indexicals, logophors and logophoric reflexives. The aim of our discussion is to demonstrate that S&T's system leads into a tangle of puzzles and inconsistencies. We take this as a serious challenge, our intuition being that the problems are less a matter of purely technical failure but of a misconception about the division of labor between core syntax and more peripheral systems.

One purported advantage of bringing locality of binding into the picture has to do with Slave (Athapaskan) 1st and 2nd person pronouns in complements to "direct discourse verbs" (Rice 1986). The highly intricate facts warrant the following generalization (S&T:325).

(5) a. Slave first person pronominal is bound by the most local SPEAKER

b. Slave second person pronominal is bound by the most local HEARER

S&T follow Tsoulas & Kural (1999) in taking indexical pronouns to be operator-bound variables. They further assume that bindings are computed over structures involving representations of the P-roles S-P and H-P as well as their Θ-theoretic counterparts, all of which function as variable-binders. Thus, as shown schematically in (6) (S&T:325), the interpretation of you changes if there is an overt c-commanding expression bearing H-Θ in the sentence. (Coreference is indicated by cosuperscription.)

---

9 It is an open question to what extent S&T's approach is intended to capture Cinque's (1999) idea that the functional projections in question should host semantically compatible adverbs in their specifiers.
10 Note that Ross (1970:224) modeled the speech act domain as a full-fledged (performativistic) superordinate clause, i.e. as a structure of category S. Translated into current views about clause structure one would minimally expect IP or (AgrP and) TP to embed SAP. In such a system, SA° would clearly need an additional °i in order to be licensed by I°. Use of TP would further evoke the difficult issue of how to interpret SA° temporally.
11 See section 3. As can be inferred from our discussion there, making YP the most prominent argument in SAP would lead to numerous cases of binding/control failure and thus, potential uninterpretability.
12 Trying to determine the exact nature of this structural level would take our discussion too far afield.
While (5) makes correct predictions for (6), it clearly fails to rule out cases like (7), i.e. cases where the bound indexical is outside of the reported clause.\textsuperscript{13}

(7) a. *\([\text{SAP} ^i \text{XP}_{S-P} ^j \text{ZP}_{H-P} [\text{CP} ^k \text{Simon}_{s-\Theta} \text{told} ^n \text{him}_{h-\Theta} [\text{CP} ^n \text{you not visit} ^k \text{me} ] ] ]\]

b. *\([\text{SAP} ^i \text{XP}_{S-P} ^j \text{ZP}_{H-P} [\text{CP} ^k \text{Simon}_{s-\Theta} \text{said} [\text{CP} ^j \text{you not visit} ^k \text{me} ] ] ]\]

What S&T's approach misses is that it is the fact that reports are sufficiently like speech acts which allows shifting of indexicals. (7) would be correctly ruled out right away if an embedded SAP-layer were employed for the complement of "direct discourse verbs." The S-P and H-P roles of this SAP would have to be sensitive to higher S-\(\Theta\) and H-\(\Theta\) if present, but the binding of 1st and 2nd person indexicals would have to be the exclusive job of (bearers of) S-P and H-P. Yet, S&T (p.338) explicitly opt against embedding SAP in an attempt at steering clear of the infinite regress dilemma lurking behind the traditional performative analysis.\textsuperscript{14}

Another technical deficiency concerns c-command, which is a prerequisite for syntactic operator-variable binding. A swift glance at (1) reveals that ZP H-P does not c-command the remainder of the sentence, i.e. YP UC-P, in declaratives and subjunctives. Thus, indexical \textit{you} should go unbinding and therefore be uninterpretable in structures bearing these sentence moods, contrary to fact of course. S&T's representations in (6) are misleading in this respect. Now, heeding the operator-like nature of the binders one could, as a remedy, postulate an additional QR-step adjoining ZP H-P to a sufficiently high projection. This, however, jeopardizes what motivated the distinction between (1) and (3), since, according to S&T, declaratives and interrogatives differ essentially according to whether XP S-P or ZP H-P are coindexed with XP SK-P in Spec,SenP (p.334f). S&T "take this coindexing to be a sort of control, which requires that the controller c-command the controlee" (p.335). Now, it may be reasonable to assume that such a kind of "sk-control" requires the controller to be in an A-position, while QR targets A-bar-positions and thus leads to A-bar-binding. This would sufficiently distinguish sk-control from "indexical binding." We then have to interpret the P-role positions in (1)-(3) as A-positions, something made plausible by the analogy to H&K's lexical projections.

The QR-solution just sketched has further disadvantages though in that it leads to a weak crossover configuration and should be ruled out independently.\textsuperscript{15} A properly QR-ed ZP H-P would bind its trace, i.e. a variable, in the complement position of SA*P and could thus not bind another (indexical) variable inside YP at the same time.

Likewise, sk-control, which, as we have seen, S&T take to presuppose c-command, leads into substantial difficulties. Consider the licensing of a logophoric \textit{SELF}-anaphor in (8) (cf. Ross 1970:228).

(8) \([\text{SAP} ^i \text{XP}_{S-P} [\text{SenP} ^i \text{XP}_{SK-P} [\text{CP} \text{This paper was written by Ann and} ^i \text{myself} ] ] ]\]

Coindexation of the (bearer of the role) "seat of knowledge" with XP S-P ("speaker") is S&T's encoding of the logophoric role \textit{source}, taken to be a primitive in Sells (1987). Given that

\textsuperscript{13} Although we do not have the necessary data from Slave, we conjecture that the bindings and interpretations in question are not well-formed in this (or any other, except, perhaps, a sign) language.

\textsuperscript{14} In this respect S&T diverges from Speas (2004). For an early critique of the performative analysis, as developed by Ross (1970), see Grewendorf (1972).

\textsuperscript{15} See Heim & Kratzer (1998:chapter 10) for ample discussion.
XP_{S,P} is the highest specifier of the sentence, c-command is guaranteed. However, it has repeatedly been pointed out in the literature (e.g. Reinhart and Reuland 1991:317; Sells 1987:451) that antecedents need not c-command logophors. Take for example the licensing of Japanese *zibun* in (9) from Sells (1987:454).

(9) Taroo wa ^1^Takasi kara [ Yosiko ga ^1^zibun o niku
dei ru to ] kiita

*Taroo TOP Takasi from Yosiko SUBJ self OBJ be_hating COMP heard*

"Taroo heard from ^1^Takasi that Yosiko hated ^1^him"

Here, *Takasi* is the source of information. However, irrespective of whether this would provide that DP with something like SK-Θ to be coindexed with and control *zibun* directly or with S-Θ to be coindexed with and control an intermediate XP_{SK,P} in (some) Spec,SenP, *Takasi* is not in a position to c-command its controlee given that is is inside a PP, [pp Takasi kara ]. Again, a purely formal syntactic perspective seems to be misguided.16

Apart from this technical issue, there are substantial empirical problems too. Thus, consider the contrast in (10) (S&T:335).

(10) a. Honestly, Mary knew the victim

b. Honestly, who knew the victim?

In a declarative, *honestly* is speaker-oriented, while it is hearer-oriented in an interrogative. According to S&T this follows if the interpretation of *honestly* is dependent on SK-control, which itself underlies a closeness principle. XP_{S,P} is the closest controller of XP_{SK,P} in declaratives, (1), trivially, while ZP_{H,P} is closest in interrogatives, (3).17 However, in imperatives, which are also defined by configuration (3), speaker-orientation of speech-act adverbials is permitted, contrary to S&T’s predictions, i.e. in violation of the closeness principle. This is shown in (11).18

(11) Seriously, don't waste your time with polemics!

Also, the evaluator responsible for the interpretation of adverbs like *unfortunately* in (12) is claimed to vary with sentence mood (S&T:335).

(12) a. Mary unfortunately knew the victim

b. Who unfortunately knew the victim?

16 We are well aware of the possibility to reanalyze PPs in ways that they become "transparent" for c-command by DP (cf. Cinque 2004). The point is that such costly reanalysis, even if it can be made theoretically watertight without voiding c-command as a formal tool, is superfluous once one allows an interpretational notion of source of information into the system as has been done by Sells (1987). The same issue of c-command failure arises (e.g. in Icelandic) for genitive subjects of DP ("John's opinion") and probably combinations like ("We gathered from John's report that . . .").

17 S&T systematically confuse the terms "closest" and "highest" in their crucial section 3. This may have to do with their earlier assumption that English indexicals, as opposed to their Slave counterparts, are bound by the highest potential binder (p.326). We have shown that indexical binding and SK-control have to be kept apart.

18 Other examples, like those in (i), are admittedly slightly marked.

(i) Honestly / Frankly, leave me alone!

It is clear, however, that the speech act adverbials must be construed as speaker-oriented not hearer-oriented. Thanks to Chris Potts for providing native speaker judgments. The argument can be reproduced for German.
Yet, (13) indicates that speaker-orientation must be an option in interrogatives too. Again this violates the closeness principle.

(13) a. Why did John unfortunately leave? # Something I personally find extremely fortunate

b. Who knew the victim that unfortunately died?

If *unfortunately* in (13a) were hearer-oriented, the continuation made by one and the same speaker should not sound contradictory. Equally, the acceptability of (14) would seem to come as a surprise, given that ZPₜₜ₋₂ as a closer controller should block speaker-orientation of *myself* in interrogatives.¹⁹

(14) Why shouldn't this paper be written by Ann and myself?

Again, our suspicion is confirmed that purely formal syntactic mechanisms, locality in the case at hand, do not provide any systematic account of the facts.

### 4. Conclusion

In this paper, we have argued against S&T's thought-provoking attempt at reducing an explanation of the inventory of sentence moods and the licensing of indexicals, logophoric reflexive)s and other point of view related phenomena to formal syntax. In particular, we showed that S&T's method for keeping the inventory of (major) sentence moods small boils down to arbitrary stipulation rather than formal syntactic deduction. This involves (i) arbitrary in-/exclusion of deep structures underlying SA(ₜ₋₁)P, (ii) arbitrary in-/exclusion of diatheses allowed to operate on SA(ₜ₋₁)P, and (iii) arbitrary constraints on recursivity and feature checking (resources).

We also showed the inadequacy of S&T's employment of formal syntactic tools (c-command/binding/control/closeness) in dealing with empirical phenomena. Thus, c-command is not systematically required for a proper account of logophoric reflexives. Also, requiring c-command for an operator-binding account of indexical pronouns leads to a dilemma. Either 2nd person pronouns like *you* go unbound in declaratives and subjunctives due to c-command failure. Or, if one allows an additional movement step for the binding constituents another dilemma arises. If the movement step is an instance of A-movement, S&T's sentence mood distinctions threaten to break down. If it is an instance of A-bar-movement, a violation of the weak crossover constraint follows. Equally, closeness does not play any systematic role in deriving the facts contrary to S&T's assumptions. Thus, (i) accounting for Slave (Athapaskan) indexical shift in the complement of "direct discourse verbs" is in need of additional localization, which, however, conflicts with S&T's ban on embedding SAP. (ii) The behavior of speech-act adverbs in imperatives and the behavior of evaluative adverbs and logophoric reflexives in interrogatives is not dependent on the closest binder, i.e. it is not hearer-oriented.

As already indicated, we take the deficiencies of S&T's system as an interesting and important result in so far as they seriously challenge the kind of syntactic reductionism S&T are pursuing. It is our impression that more satisfactory revisions of S&T's system will only be forthcoming if one allows for substantial and explicit assumptions from an outside theory of language and discourse to interact with syntax.

---

¹⁹ Thanks to Philippa Cook, Laura Downing, and Benjamin Shaer for native speaker judgments. Note that S&T could not assume that (14) is fine because the actual speech act involved may be more of a suggestion than of an information-seeking question. Their approach is explicitly designed to capture abstract sentence moods instead of full-fledged speech act representations (p.317).
References:


