A Categorial Syntax for Verbs of Perception[*]

Robin Clark and Gerhard Jäger

1. Introduction

"Categorial Grammar" is not a particular grammar formalism, let alone a theory of grammar, but rather a cover term for a family of quite diverse approaches to natural language syntax. This cover term is nonetheless useful, since all these theories share important characteristics. Besides the common foundation in the works of Ajdukiewicz 1935 and Bar-Hillel 1953 and the use of complex syntactic categories built up from atoms with slashes, they are based on two related premises that distinguish them from all other theories of grammar:

1. The locus of grammatical generalizations is the lexicon.

2. Constituent structure plays no role in grammatical theory.

This does not entail that Categorial Grammars deny the existence of constituent structure (in fact, Bar-Hillel's Basic Categorial Grammar and Lambek's 1961 non-associative grammar calculus assume a rigid binary branching structure). However, all Categorial Grammars assume that constituent structure cannot enter grammatical description.

The hypothesis that constituent structure is immaterial to grammatical descriptions contrasts sharply with the perspective found in the generative tradition. Generative grammar is largely grounded on relations like c-command, m-command and government which are based on tree geometry. Some theories have defined grammatical relations like *subject* and *object* entirely in terms of constituent structure; for example, the subject of a category X is that nominal which occurs the Spec(X).

Since such a strategy is not viable in Categorial Grammar (CG henceforth), researchers working in this tradition usually do without notions like "subject" etc. The bundle of properties that are associated with subjects are considered to be logically independent. So it seems that CG misses an important generalizations.

The paper tries to counter this objection by demonstrating that the configurational notion of "subject" in fact leads to analyses that are descriptively inadequate. This point will be made by a case study of naked infinitive (henceforth: NI) perception reports as in

(1) Jackie saw Oswald shoot Kennedy.

Several tests indicate that the accusative NP *Oswald* is the subject of the embedded VP *shoot Kennedy*. Under a configurational notion of "subject" this implies that the string *Oswald shoot Kennedy* forms a sentential constituent. On the other hand, there is firm evidence both from syntax and semantics that this string should not be considered a constituent. We will attempt to show that (1) in a categorial setting, some of the subject properties of the accusative NP in NI perception reports can be derived without recurse to constituent structure, and (2) that this frees the way to a fairly simple semantics of perception verbs that solves most puzzles from the literature in a straightforward way.

2. Subject properties and NI perception reports

Compare the following two sentences:

(2) a. John saw that Bill left.

b. John saw Bill leave.

Examples (2a) and (2b) both involve John's perception of something, although their entailments are rather different. Example (2a), epistemic perception, entails that John has perceived that Bill left and has understood that Bill left. In other words, John has understood the content of his perceptions; the following should be anomalous:

(3) John saw that Bill left but he didn't know it.

Notice that example (2a) does not entail that John actually saw the event of Bill's leaving. He could, in fact, have drawn the inference that Bill left through a fairly complex chain of deductions.

Example (2b) is quite different. In this case, John must actually have visually perceived the event of Bill's leaving although he may not have understood that that was what he saw. As Barwise 1981 observes, the following sentence:

(4) Nixon saw Mrs. Wood erase the tape

does not imply that Nixon understood what he was seeing. He may have thought that she was engaged in a peculiar calisthenic exercise, for example. Barwise argues that the tensed complement to a perception verb, as in (2a) involves "epistemic perception," that is, perception with some cogitation, while the naked infinitive complement, as in (2b), involves non-epistemic perception, that is raw perception without any additional non-perceptual cogitation.

Thus, I can see John embezzle money without seeing that John is embezzling money simply because I can perceive events in the world without understanding their import.

These intuitions have led to the standard analysis of the syntactic properties of verbs of perceptual report. Epistemic perception is a relation between an individual and a proposition. Assuming that the syntactic category CP corresponds to propositions, then, syntactically, this corresponds to a CP complement to the perception verb. What about non-epistemic perception? Barwise argued that the proper syntactic analysis was along the lines shown in (5):

(5) $[_{VP}$ see$[_{XP}$ John $[_{V^n}$ run$]]]$

where XP is some category distinct from CP. The representation in (5) is meant as a cover for a set of analyses that take the immediately post-verbal NP as forming a constituent with the naked infinitive. A direct consequence of this analysis is that the denotation of XP must be distinct from CP since (2a) and (2b) are not synonymous. Since CPs denote propositions, XP, whatever its category, must denote something other than a proposition; Barwise argues that XP must denote a scene, a visually perceived situation. This, in turn, lends support to his thesis that situations are a basic semantic category.

Syntactic facts *prima facie* support this syntactic analysis. To start with, non-thematic elements can occur in the postverbal position (as pointed out by Gee 1977:468):

(6) a. John saw it rain.

b. ?I've never seen there be so many complaints from students before.[1]

c. John saw the shit hit the fan.

Example (6a) shows that weather *it* can occur in the postverbal position. In (6b), presentational *there* occurs in this position and in (6c) an idiom chunk can occur in this position with its idiomatic interpretation. Note that weather *it*, presentational *there* and idiom chunks have the property that they are, in some sense, non-referential.

Consider, first, examples (6a) and (6b). According to classical Government-Binding theory, the only way that the postverbal NP could be a direct object of the verb is if the verb assigns it a thematic role; this is the content of the θ-*Criterion*. By "direct object" we mean, of course, that the NP is a sister to the verb in the parse tree. But if the position were associated with a thematic role, then non-referential elements would be excluded from that position. The only

---

[1]This is given as grammatical be Gee 1977.

syntactic position that is both non-thematic and associated with a grammatical function is the structural subject position and, therefore, non-referential elements are restricted to this position unless they are part of an idiom that includes the entire verb phrase as well. Turning to example (6c), we see that the postverbal NP receives its idiomatic interpretation. Therefore, it is non-referential and cannot be a sister to the main verb. This again shows that the postverbal NP in NI complement examples is a structural subject and not a direct object. The only way to satisfy this condition is if the postverbal NP forms a constituent with the naked infinitive.

Thus, NI constructions seem to class with so-called "Exceptional Case Marking" constructions, shown in (7), and small clause constructions[2], shown in (8), in allowing a non-thematic element to interceded between the verb and the embedded predicate.

(7)　a.　John believes Bill to have stolen the car.

　　b.　John believes it to be raining.

　　c.　John believes it to be obvious that Bill stole the car.

　　d.　John believes there to have been a riot in the park.

　　e.　John believes the shit to have hit the fan.

(8)　a.　John considers Bill a genius.

　　b.　John considers it obvious that Bill stole the car.

The crucial point here is that the presence of "a-thematic" material is diagnostic of the grammatical function *subject*; the grammaticality of the (b-e) examples in (7) show that the postverbal NP is a true subject of the following predicate and not the structural object of *believe*. We must, therefore, contrast the behavior of the postverbal NP in (7) with its behavior in an object control construction:

(9)　a.　John persuaded Bill to steal the car.

　　b.　*John persuaded it to rain.

　　c.　*John persuaded it to be obvious that Bill stole the car.

　　d.　*John persuaded there to be a riot in the park.

　　e.　*John persuaded the shit to hit the fan.

<hr>

[2] Because their predicates are not verbal, small clause constructions do not show the same range of non-thematic material in the following position. We take this as largely tangential to our main point.

The contrast between the ECM constructions in (7) and the control constructions in (9) present CG with an interesting problem. Generative grammar accounts for the contrast by associating subject properties with a particular piece of tree geometry, where, by subject property we mean things like:

(10) a. The subject is allowed to be non-thematic;

b. The subject is the "target" (or "landing site") of raising operations;

c. The subject is a "licensed" controller;

d. The subject is a "trigger" for certain agreement relations;

e. The presence of a subject defines local domains for binding.

f. Subjects are islands to extraction.

The list in (10) can, of course, be expanded and clarified. Our point is that in classical generative accounts all of the properties in (10) are unified under a particular geometric approach to grammatical relations; thus, establishing one of the properties in (10) is sufficient to establish constituent structure and endow to element in question with the full array of subject properties.

Grammatical subjects are traditionally treated as possible landing sites for raising processes like subject-to-subject raising (SSR) and passive. NI constructions admit *get* passives but *be* passives are more marked:

(11) a. John saw Bill get examined by a doctor.

b. *?John saw Bill be examined by a doctor.

Furthermore, the post-verbal position in NI constructions admits only a few cases of SSR:

(12) a. John saw Bill appear to unlock the safe.

b. ?John saw Bill seem to escape from the handcuffs.

c. *John saw Bill be likely to drink too much.

d. *John saw Bill tend to drive on the wrong side of the road.

The unacceptability of examples (12c) and (12d) are easily accounted for on the basis of semantic properties of the embedded predicate; it is difficult to imagine the exact visual manifestations of being likely to drink too much and tending to drive on the wrong side of the road, both properties being propensities that should be treated modally. Appearing and seeming, on the other hand, can involve deliberate deceptions that can be visually realized—stage magicians make this their stock in trade. Because of this intentionality, *appear* and,

to a lesser extent, *seem* may involve semantic relations between the "raised" subject and the predicate that are unavailable in the true raising constructions associated with *likely* and *tend*. Similarly, *get* passives may be preferred over *be* passives in NI constructions because of secondary semantic properties associated with the former but unavailable in the latter; compare, for example, the contrast between *get* and *be* in certain imperative constructions:

(13)   a. Don't get killed.

        b. *Don't be killed.

The contrast in (13) is probably attributable to differences in the aspectual properties associated with *get* and *be*. These differences may also account for the contrast between (11a) and (11b). In particular, *be* passives tend to have a more stative flavor than *get* passives, a fact which may limit their distribution in NI constructions. As was the case for the distribution of pleonastics, then, raising and passive provide only equivocal support for the subject status of the post-verbal NP in NI constructions; while the post-verbal NP does show some subject properties, other factors associated with the semantics of perceptual reports intervene.

A further subject property involves the distribution of anaphors and here the facts are much more straightforward. Putting aside formal details, let us suppose, following classical Government-Binding Theory, that subjects create a minimal domain for binding; that is, the presence of a structural subject on a constituent guarantees that a syntactic anaphor like *himself* or *each other* must be bound within that constituent while pronominals like *her* or *them* must be unbound in the same domain. It follows that if the postverbal NP in NI constructions is a structural subject, then it should create a minimal domain for binding. The following data are consistent with this view of binding domains:

(14)   a. John saw Mary touch herself.

        b. *John saw Mary touch himself.

        c. John saw Mary touch him.

        d. *John saw Mary touch her.

Examples (14a) and (14b) show that anaphors like *herself* must indeed find their antecedent within the domain defined by the postverbal NP; *Mary* in example (14a) is proximate to the anaphor *herself* and, so, is a legal antecedent for it; *John* in example (14b) is too distant to serve as a legal antecedent for *himself* since the NP *Mary* inscribes an opaque domain for binding due to its status as a subject. Similarly, examples (14c) and (14d) show that *John* can be a possible antecedent for the pronoun *him* because *Mary* defines the

minimal domain within which the pronoun must be free. Equally, the pronoun *her* in (14d) cannot be coreferential with *Mary* because the latter is within the pronoun's minimal domain; it cannot be coreferential with *John* because they disagree in gender.

Finally, we note that subjects tend to be islands to extraction:

(15)   a. *who$_i$ did friends of t_i visit Bill

        b. *which saint$_i$ does Fred consider stories about t_i utter fabrications

The ungrammaticality of the examples in (15) can be attributed to the fact that the *wh*-element is associated with a gap inside a subject, a tensed clause in (15a) and an small clause in (15b). The post-verbal NP in NI constructions is likewise an island for extraction:

(16) *who$_i$ did John see a friend of t_i steal a car

The analysis of the post-verbal NP in NI constructions as a true subject can immediately treat (16) as a violation of the islandhood of subjects.

As we have seen, the accusative NP in NI perception reports shows properties that are traditionally taken to be indicative of subjects. Under the configurational definition of "subject", it is thus inevitable to consider the string $[_{NP} NP_{acc} VP_{inf}]$ as a constituent.

In the next section we will collect a series of syntactic arguments that challenge this conclusion.

3. Other syntactic tests for constituency

Observations concerning coordination and anaphora also point towards a one-constituent analysis. This can be seen from examples (17a) and (17b).

(17)   a. John saw Mary enter and Bill leave

        b. John saw Mary enter, and Bill saw it too.

Akmajian 1977 points out that virtually all tests for constituency apart from coordination and pronominalization indicate that the complement of NI perception verbs do not form a constituent though. So they cannot appear in the postcopular position of pseudoclefts:

(18) *What we saw was Raquel Welch take a bath.

Neither can they be inserted into clefted positions:

(19) *It was Raquel Welch take a bath that we saw

They cannot be right node raised:

(20) *?We could hear, but we couldn't see, Raquel Welch take a bath.

Finally, they cannot undergo object deletion:

(21) *Raquel Welch take a bath is a breathtaking sight to see

Additional evidence against a one-constituent analysis comes from top-icalization in German. The underlying sentence structure in German is verb final. Main clauses display V-2, i.e. one constituent is obligatorily fronted, and the finite verb is placed immediately after this constituent. Thus if a string is a constituent, we expect that it can be topicalized. Let us apply this test to NI perception reports. The underlying word order can be seen in an embedded clause like (22):

(22) weil der Polizist jemanden fliehen gesehen hat.
since the policeman[nom] somebody[acc] escape[inf] seen[part] has.
since the policeman saw somebody escape.

The topicalization test indicates that $NP_{acc} + VP$ do not form a con-stituent, while the sequence "embedded VP+matrix Verb" do:
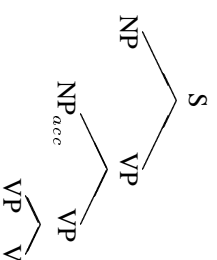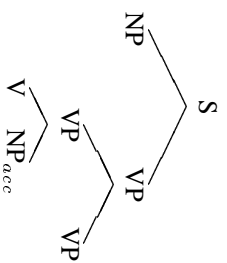
(23) a. $^{???}$Jemanden fliehen hat der Polizist gesehen.
Somebody escape has the policeman seen.

  b. Fliehen gesehen hat der Polizist jemanden.
Escape seen has the policeman somebody.

So the appropriate bracketing for (22) should be (24a) rather than (24b):

(24) a. weil der Polizist [jemanden [fliehen gesehen hat]]

  b. weil der Polizist [jemanden fliehen] gesehen hat.

All these observations indicate that the appropriate syntactic structure for NI perception reports should be like the trees in (25a) for English and Ger-man respectively rather than the structure in (5), neither of which involves an embedded small clause.

(25) a.

```
        S
       / \
     NP   VP
          / \
        VP   VP
             / \
            V  NP_acc
```

```
          S
         / \
       NP   VP
            / \
          NP_acc VP
                / \
               VP  V
```

4. Vlach's puzzle

All detailed studies of the semantics of perception verbs that we are aware of start with a one-constituent analysis (see for instance Barwise 1981, Higginbotham 1983, Vlach 1983, van der Does 1991). And even though the ontological background differs considerably, they agree on the following:

1. Perception is a relation between an agent and an abstract object (scene/situation, event, partial model etc.).

2. The $[NP_{acc} VP]$ constituent in NI perception reports denotes a set of situations (events ...).

3. *John sees NP VP* can be paraphrased as *John sees a situation (event ...) s, and $s \in \|[NP VP]\|$*.

Let us suppose, as seems reasonable, that active and passive sentences are supported by the same set of scenes and, so, denote the same proposition.[3] In particular, (26a) and (26b) are true paraphrases, differing only in their pragmatic contributions and that (26c) differs from the other two only in the contribution of *get* to the interpretation of the sentence:

(26) a. Oswald assassinated Kennedy.

b. Kennedy was assassinated by Oswald.

c. Kennedy got assassinated by Oswald.

Furthermore, let us follow the standard assumption that the semantic contribution of a passive sentence in an embedded context is exactly comparable, up to pragmatics, to the semantic contribution of an active sentence in the same context; thus, (27a) is a paraphrase of (27b):

---

[3] Vlach likely wouldn't agree with this. His view is discussed below.

(27) a. John saw that Oswald assassinated Kennedy.

  b. John saw that Kennedy was assassinated by Oswald.

Notice, in particular, that no privileged relationship holds between John and either Oswald or Kennedy in either (27a) and (27b).

  Compare this situation in (27) with the pair of sentences in (28), first observed by Vlach 1983:

(28) a. John saw Oswald shoot Kennedy.

  b. John saw Kennedy get shot by Oswald.

The behavior of the sentences in (28) is peculiar given the small clause analysis of NI complements, since (28a) and (28b) are not paraphrases of each other and differ by more

  than the contribution of *get* to (28b). In particular,

for (28a) to be true it must be the case the John saw Oswald exactly when the latter shot Kennedy; John need not have seen Kennedy at all. For (28b) to be true, on the other hand, John must have seen Kennedy at the moment that he got shot by Oswald; he need not have seen Oswald at all. Thus, while (28a) is not true of anyone, many people have shared John's visual experience in (28b).

  In brief, it would seem that the subject of the perception verb and the postverbal NP stand in some special relationship in non-epistemic perceptual reports, a relationship that is wholly absent in epistemic perceptual reports. To be more precise, we claim that the inference pattern in (29a) is valid, but the one in (29b) isn't.

(29) a. x saw y $VP \models x$ saw y

  b. x saw y [$_{VP}$ V z] $\models x$ saw z

  The invalidity of (29b) is demonstrated by (28). To substantiate the claim that (29a) is valid, let us consider three putative counterexamples. Suppose, first, that John is standing behind an opaque plastic screen, using magnets to move metal puppets on the other side of the screen. Suppose Mary observes the movement of the puppets, without seeing John. Can Mary use (30a) to report her perception?

(30) a. I saw John move the puppets.

Gee 1977 claims that she can. The judgments of those we have asked is that although (30a) is marginal in this context,

it can be so used just in case Mary is absolutely certain that no one else could be responsible for the movement of the puppets. In this case, seeing puppet movement is tantamount to direct perception of John.

Similarly, consider the case where Mary is separated from a forest by a large hill, so that she cannot see the forest (this example is also due to Gee 1977). Observing a huge billow of smoke rising over the hill, can Mary later use (31a) to report her experience?

(31) a. I saw the forest burn.

Again, the consensus of those we have asked is that (31a) is odd in the above context. We can sharpen the intuition by considering the following sentence:

(32) a. I saw the forest burn even though I didn't see the forest.

According to our intuitions, this sentence is contradictory, no matter what background knowledge we assume.

The same argumentation applies *ceteris paribus* to an argument from van der Does 1991, who in turn attributes it to Robin Cooper. He raises the question whether an entailment relation holds between (33a) and (33b):

(33) a. Daniel saw Lucia phone Henry.
      b. Daniel saw Lucia.

Van der Does (op. cit., p 245) discusses the following scenario: "Imagine Lucia, Henry and Daniel each sitting in separate rooms. There are phones which enable Lucia and Daniel and Henry to speak to each other, but only when Lucia phones Henry an oscilloscope in Daniel's room will show a patterns characteristic for Lucia's voice. Now suppose Daniel saw the patterns, can one report the fact by saying [(33a)]?" Van der Does claims that at least some subjects answered affirmatively, since "perceiving the pattern on the oscilloscope is perceiving a representation of Lucia, much as perceiving a video-recording of her would have been. And clearly in the latter sense [(33a)] might be used." We agree that (33a) might be true in such a situation, but so would (33b), and for the very same reason. In other words there might be some vagueness as to how direct direct perception should be, but the inference pattern is not affected by that.

To sum up the discussion so far, we observe a special semantic relationship between the matrix subject and the accusative NP in NI perception reports. In other words, $NP_{acc}VP$ don't form a semantic unit. This indicates that they don't form a syntactic unit either.

It should be mentioned that Vlach 1983, who was presumably the first to notice this special relationship, nevertheless uses a one-constituent analysis.

According to him, the difference in meaning between (28a) and (28b) is due to the fact that the denotation of *Kennedy shot by Oswald* consists of events that include Kennedy's location, while *Oswald shoot Kennedy* denotes a set of events that are locally connected to Oswald. Vlach doesn't give an explanation for this asymmetry, but apparently he assumes the event descriptions that include a subject denote events that are located at or around the location of the referent of the subject.

To test this assumption, consider (34)

(34) Jackie saw Oswald's assassination of Kennedy.

Despite the fact that the subject of the event description is *Oswald* and Jackie didn't see Oswald, the sentences is true. So the location of an event that is described by an event noun is not determined by the location of the referent of the subject. Events that are described by tensed sentences do not confirm a special status of the subject either.

(35)   a.   Oswald shot Kennedy.

  b.   ?That happened in the Texas Book Depository.

  c.   ?That happened in the Presidential Limousine.

  d.   That happened in Dallas.

It appears that an event that is described by a tensed clause has to include all participants, not just the referent of the subject. So we may conclude that our argumentation above is supported. The accusative NP and the embedded VP shouldn't be considered to be a semantic unit.

What about the arguments in favor of a one-constituent analysis? There were two that didn't rely on grammatical functions, coordination and anaphora. In the next section, we will demonstrate that the former argument is not conclusive; non-constituents may be conjoined. Anaphora isn't an conclusive argument either. It is generally hold nowadays that anaphora resolution operates on semantic entities rather than on syntactic constituents. We will argue below the meaning of (17b) does involve the event of Mary's entering, even though it does not correspond to any constituent. So we expect anaphoric reference to it to be possible.

5. The semantics of verbs of perception

As our starting point we take the semantics of verbs of perception as proposed by Higginbotham 1983. This decision is of little significance, other proposal

likes Barwise's or van der Does' could be modified in a similar fashion. Higginbotham assumes that verbs that can occur in the complement of NI perception reports have an event argument, and that the logical form of a sentence like (36a) is (36b).

(36) a. John saw Mary leave

b. $\exists e(\text{LEAVE}(\text{M}, e) \land \text{John sees } e)$

The variable e ranges over events here. The verb *see* that occurs with NI complements is thus semantically reduced to simple transitive *see*. As argued above, this semantics cannot be correct since then *seeing Oswald shoot Kennedy* would come down to seeing the whole event of Oswald's assassination of Kennedy, which in turn entails seeing Kennedy. The truth conditions of

(37) Jackie saw Oswald shoot Kennedy

are much weaker. To establish its truth, it is sufficient that Jackie saw that part of the complex assassination event the directly involved Oswald, i.e. his aiming and pulling the trigger. To accommodate this intuition, let us assume that for each participant x of an event e, there is a unique subevent e_x of e that has x as its only participant. We won't spell out this operation formally here, but the intuition should be clear enough. So the logical form of (37) should be

(38) $\exists e(\text{SHOOT}(\text{LHO}, \text{JFK}, e) \land \text{Jackie sees } e_{\text{LHO}})$

It goes without saying that such a logical form can only be derived compositionally if the accusative NP *Oswald* is an argument of the matrix verb. This in mind, we can give the following lexical semantics of *see*:

(39) $\lambda P x y. \exists e(P x e \land \text{SEE}(y, e_x))$

This is compatible with the following syntactic category of *see*, the Categorial counterpart to structure given in (25a).

(40) $(N \setminus S)/VPI/N$

We leave irrelevant morphosyntactic details open. In particular, we do not spell out the internal structure of NI VPs but abbreviate its category with VPI.

In the sequel we will show that this semantics of verbs of perception, paired with a categorial syntax, meets the main criteria that are discussed in the literature.

Verdicality

This is Barwise's name of the inference scheme

(41) John saw Mary leave $\models$ Mary left.

In an event based semantics, the logical form of *Mary left* is $\exists e : \text{LEAVE}(M, e)$. As under Higginbotham's original account, this follows from the premise by simple first order reasoning.

Extensionality

All elements of an NI perceptual report are transparent, i.e. they can be replaced by extensionally equivalent expressions *salva veritate*. Since no intensional operators are involved in our semantics of *see*, this is predicted.

Absence of scope ambiguities

Generally, all scope inducing items that might occur in one of the complements of *see* have matrix scope. Under our approach, this has nothing to do with the semantics of *see* but follows from its syntax. To start with, quantifiers in the accusative position always have matrix scope, for instance

(42) John saw Q leave $\equiv Qx$ are such that John saw x leave.

Since the quantifier occupies an argument position of the matrix verb, it must at least take scope over the matrix VP, no matter what particular approach to quantifier scope we adopt.

Coordination behaves similarly, i.e. the following two equivalences hold

(43)   a. John saw Mary swim and Bill walk $\equiv$ John saw Mary swim and John saw Bill walk

   b. John saw Mary swim or Bill walk $\equiv$ John saw Mary swim or John saw Bill walk

In any version of CG, conjunctions are considered as polymorphic items. Their category is $X \backslash X / X$, where X ranges over Boolean categories.[4] Roughly, a category is Boolean iff the corresponding semantic type ends up in t. So $S, N \backslash S, CN$ etc. are Boolean categories. The meaning of the coordination

---
[4] Steedman's 1996 syncategorematic treatment of conjunctions amounts to the same thing.

and is $\lambda Q P \bar{x}. P \bar{x} \wedge Q \bar{x}$. So in the first sentence in (43a), the substrings *Mary swim* and *Bill walk* have to be assigned a Boolean category each to make them conjoinable.

Under any version of CG, the accusative NP and the NI phrase cannot be combined directly to yield a Boolean category. Thus as in the case of quantifier scope, the absence of a narrow scope reading is expected. We have to answer the question how the wide scope reading is to be derived though.

Up to the present point, we remained neutral as to which version of CG is to be used. To handle this puzzle, we have to be more specific. Since *Mary* and *swim* do not form a functor-argument structure here, we need a certain degree of associativity to deal with this instance of non-constituent coordination. So the example can be handled in any version of Combinatory Categorial Grammar (CCG, cf. Ades and Steedman 1982) that contains the operation of function composition, and in any descendant of Lambek's 1958 associative CG.

As shown in Fig. 1,[5] the reading in question can be derived in CCG using only type lifting and backward function composition. We abbreviate $N \backslash S$ as VP for convenience. The predicate SEE_1 is shorthand for the meaning of *see* (cf. (39)). Since both combinators are theorems of the Lambek calculus, this is simultaneoulsy a Lambek derivation.

---

[5]The Combinatory branch of CG uses a format for complex categories and for derivation trees that differs somewhat from the Lambek trad ition. We chose a compromise here in using the backslash in Lambek's sense ("$A \backslash B$" takes an argument of type A and yields a value of type B) while choosing a CCG-style derivation tree.

$$\cfrac{\cfrac{\cfrac{\cfrac{Mary}{\mathrm{M}}\ lex}{\cfrac{N}{\lambda y.y\mathrm{M}}\ T}{((VP/VPI)/N)\setminus(VP/VPI)} \quad \cfrac{\cfrac{swim}{\mathrm{SWIM}}\ lex}{\cfrac{VPI}{\lambda z.z\mathrm{WALK}}\ T}{(VP/VPI)\setminus VP}}{\cfrac{((VP/VPI)/N)\setminus VP}{\lambda y.y\mathrm{M\ SWIM}}}<B \quad \cfrac{and}{\lambda QP\vec{x}.P\vec{x}\wedge Q\vec{x}}\ lex}{X\setminus X/X} \quad \cfrac{\cfrac{\cfrac{Bill}{\mathrm{B}}\ lex}{\cfrac{N}{\lambda y.y\mathrm{B}}\ T}{((VP/VPI)/N)\setminus(VP/VPI)} \quad \cfrac{\cfrac{walk}{\mathrm{WALK}}\ lex}{\cfrac{VPI}{\lambda z.z\mathrm{WALK}}\ T}{(VP/VPI)\setminus VP}}{\cfrac{((VP/VPI)/N)\setminus VP}{\lambda y.y\mathrm{B\ WALK}}}<B$$

$$\cfrac{((VP/VPI)/N)\setminus VP}{\lambda u,v.u\mathrm{M\ SWIM}v\wedge u\mathrm{B\ WALK}v}<$$

$$\cfrac{\cfrac{John}{\cfrac{N}{\mathrm{J}}}\ lex \quad \cfrac{\cfrac{\cfrac{saw}{VP/VPI/N}\ lex}{\mathrm{SEE_1}} \quad \cfrac{Mary\ swim\ and\ Bill\ walk}{\cfrac{((VP/VPI)/N)\setminus VP}{\lambda u,v.u\mathrm{M\ SWIM}v\wedge u\mathrm{B\ WALK}v}}}{\cfrac{VP}{\lambda v.\mathrm{SEE_1\ M\ SWIM}v\wedge\mathrm{SEE_1\ B\ WALK}v}}<}{\cfrac{S}{\mathrm{SEE_1\ M\ SWIM\ J}\wedge\mathrm{SEE_1\ B\ WALK\ J}}}<$$

Figure 1: Derivation of *John saw Mary swim and Bill walk*

Failure of logical equivalence

Although the complements of perception verbs can be combined by the classical propositional connectives, complements that are equivalent in classical logic cannot always be exchanged *salva veritate*. For instance, (44b) doesn't have a reading that is equivalent to (44a).

(44)   a.   Hegel saw Schelling sneeze.

   b.   Hegel saw ((Schelling sneeze and Hölderlin eat) or (Schelling sneeze and Hölderlin not eat)).

To handle this problem, it has to be remarked that even though the use of propositional connective in the "complement" of perception verbs looks suggestive, we assume a different treatment of conjunction and disjunction on the one hand, and of negation on the other hand. The former connectives always receive a wide scope interpretation, while negation is predicate negation. Therefore we do not expect patterns that look like classical validities to be sustained. Under this treatment (44b) is synonymous with

(45)   (Hegel saw Schelling sneeze and Hegel saw Hölderlin eat) or (Hegel saw Schelling sneeze and Hegel saw Hölderlin not eat)

By simple propositional reasoning we can infer *Hegel saw Schelling sneeze* from this. But furthermore we infer that there is an event *e* that involves Hölderlin and that is seen by Hegel. This does not follow from (44a), so (44a) and (44b) cannot be equivalent.

The puzzle of Russell's schoolchildren

Barwise gives a further desideratum for an adequate semantics of perception reports which is illustrated by the following inference scheme.

(46)   a.   Russell sees each boy touch at least one girl.

   b.   Russell didn't see any girl being touched by more than one boy.

   c.   ⊨ There are at least as many girls as boys.

As Vlach correctly observes, this inference scheme is not valid. Imagine a gameshow where 10 boys have to find a partner among 5 girls. The participants cannot see each other, but everybody sits in a booth with several phones each of which connects to exactly one participant of the opposite sex (so the boys have 5 phones and the girls 10 phones each). Each boy calls one girl, and it

happens that each girl receives exactly two phone calls. So each girl picks up two receivers simultaneously and holds them to her ears (one receiver per ear). The TV audience can see all 15 participants, but they can only see the left side of the girls. Russell was watching this silly show on TV. In this situation a) and b) are true, but c) isn't:

(47)   a. Russell saw each boy calling at least one girl.

        b. Russell didn't see any girl being called by more than one boy.

        c. There are at least as many girls as boys.

This is problematic for the theories of Barwise, Higginbotham, and van der Does, since they uniformly predict that if Russell sees a calling b, he also sees b getting a call by a. So if Russell would see two boys calling the same girl, he would see this girl getting a call from two boys. This makes the argument valid. Since we don't claim that Russell sees a girl called if he sees a boy call her, no such prediction is made.

To sum up this section, we tried to demonstrate that a fairly innocent modification of Higginbotham's proposal is sufficient to accommodate Vlach's puzzle while preserving its general advantages. Likely a similar adjustment could be made with other theories of the semantics of perceptual reports. If we insist on compositional interpretation, this adjustment excludes a one-constituent analysis of the syntax of perceptual reports though. This in turn forces us to adopt a syntactic theory that is able to handle non-constituent co-ordination, as most versions of CG do.

6. Conclusion

In this paper, we have presented some arguments for a reanalysis of NI complements to verbs of perceptual report. We have argued that the semantic analysis that takes the NI complement as a constituent denoting a scene or situation fails to provide a satisfying account of certain entailments; these semantic properties follow directly on our account, which does not treat the NI complements as a single semantic unit.

Furthermore, our account handles many of the syntactic properties associated with NI constructions; indeed, it seems to fare at least as well as the standard small clause account. We are, however, left with a residual problem: How can we account for the subjectlike properties of the post-verbal NP? On the standard account, the subject properties of this NP follow because these properties are correlated with tree geometry. We believe, however, that this approach to grammatical relations requires an undesirable loosening of the relationship between the syntax and the semantics.

So while subject properties should not be considered as evidence for a particular constituent structure, they require an explanation nevertheless. As far as the Binding facts are concerned, this might be fairly straightforward if we assume that Binding means linking of the anaphor to an superordinate argument place of the local verb (see for instance the proposals of Szabolcsi 1988 of Hepple 1990). Under this perspective, the domain of Binding is the local VP rather than the local clause. The other subject properties discussed above have to be left as an open problem, however.

One virtue of CGs is that they maintain a homomorphic relationship between syntax and semantic structures. While CGs have a pleasingly axiomatic structure that clarifies the relationship between natural language syntax and logic, they provide no obvious account of grammatical relations. We believe, that one task for the grammarian is to elucidate the role that grammatical relations play both in syntax and in semantics. We have not, however, given such a theory in this article, contenting ourselves with posing the problem as clearly as we could.

References

Ades, Anthony E. and Steedman, Mark J., 1982. On the order of words. *Linguistics and Philosophy* 4, 517–558.

Ajdukiewicz, Kazimierz, 1935. Die syntaktische Konnexität. *Studia Philosophica* 1, 1–27.

Akmajian, Adrian, 1977. The complement structure of perception verbs in an autonomous syntax framework. In *Formal Syntax*, eds. Peter W. Culicover, Thomas Wasow, and Adrian Akmajian, 427–460. New York: Academic Press.

Bar-Hillel, Yehoshua, 1953. A quasi-arithmetical notation for syntactic description. *Language* 29, 47–58.

Barwise, Jon, 1981. Scenes and other situations. *Journal of Philosphy* 78, 369–397.

Gee, James Paul, 1977. Comments on the paper by Akmajian. In *Formal Syntax*, eds. Peter W. Culicover, Thomas Wasow, and Adrian Akmajian, 461–482. Academic Press.

Hepple, Mark, 1990. The grammar and processing of order and dependency: A categorial approach. Ph.D. thesis, University of Edinburgh.

Higginbotham, James, 1983. The logic of perceptual reports: An extensional alternative to situation semantics. *The Journal of Philosophy* 80, 100–127.

Lambek, Joachim, 1958. The mathematics of sentence structure. *American Mathematical Monthly* 65, 154–170.

Lambek, Joachim, 1961. On the calculus of syntactic types. In *Structure of Language and Its Mathematical Aspects*, ed. Roman Jakobson. Providence, RI.

Steedman, Mark, 1996. *Surface Structure and Interpretation*. MIT Press.

Szabolcsi, Anna, 1988. *Bound Variables in Syntax. (Are there any?)*. Institute of Linguistics, Hungarian Academy of Science, Budapest.

van der Does, Jaap, 1991. A generalized quantifier logic for naked infinitives. *Linguistics and Philosophy* 14, 241–294.

Vlach, Frank, 1983. On situation semantics of perception. *Synthese* 54, 129–152.

Robin Clark
University of Pennsylvania at Philadelphia
Department of Linguistics
rclark@babel.ling.upenn.edu

Gerhard Jäger
Zentrum für Allgemeine Sprachwissenschaft
Berlin, Germany
jaeger@zas.gwz-berlin.de