Clarifying noun phrase semantics in HPSG

Matthew Purver
King's College London

Jonathan Ginzburg
King's College London

Proceedings of the 10th International Conference on
Head-Driven Phrase Structure Grammar

Michigan State University

Stefan Müller (Editor)

2003

Stanford, CA: CSLI Publications

pages 338–358

Purver, Matthew & Jonathan Ginzburg. 2003. Clarifying noun phrase semantics in HPSG. In Stefan Müller (ed.), *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar, Michigan State University*, 338–358. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2003.19.

Abstract

This paper examines reprise questions: questions which request clarification of the meaning intended by a speaker when uttering a word or phrase.[1] As such they can act as semantic probes, providing information about what meaning can be associated with word and phrase types. We present corpus evidence regarding the meaning of nouns and noun phrases, and argue that this evidence runs contrary to the usual treatments of semantics in HPSG, and to the traditional generalised quantifier view of NPs as sets of sets. Instead we outline an analysis of NPs as (possibly functional) sets of individuals.

1 Introduction

Reprise questions allow a conversational participant (CP) to request clarification of some property of an utterance (or part thereof). In this paper we are concerned specifically with those reprise questions which concern the *meaning* intended by a speaker when uttering a word or phrase. By virtue of this, they can provide information about what meaning can be associated with word and phrase types. This paper discusses the evidence provided by reprise questions regarding the semantics of common nouns (CNs) and quantified noun phrases (QNPs), and outlines some general implications for NP semantics, together with some implications for semantic representation and inheritance in HPSG.

Our central claim is that reprise questions show that CNs denote properties, and QNPs denote (possibly functional) individuals, or sets of individuals. This runs contrary to common HPSG approaches where semantic content is inherited from heads or amalgamated across daughters. It also does not fit with the representation as generalised quantifiers (GQs) commonly assumed by semanticists. Instead we develop a witness-set-based analysis which treats all QNPs in a coherent manner, and allows a suitable analysis of reprise questions. We then briefly discuss some issues which arise from this, such as anaphora, quantifier scope and the representation of non-monotone-increasing NPs.

1.1 Corpus Evidence

As reprise questions manifest themselves in distinctive ways (e.g. sequences of words repeated from the immediately preceding turn), they are relatively easy to find in a corpus, and it is usually clear which word or phrase they are intending to clarify. We could therefore use the British National Corpus (BNC) (see Burnard, 2000) and the search engine SCoRE (see Purver, 2001) to provide actual occurrences of reprise questions in dialogue. By examining the examples in their surrounding context (including the responses of other CPs) we could then construct possible (and impossible) paraphrases of the meaning of the questions,

[1]The authors are supported by the UK Engineering and Physical Sciences Research Council under grant GR/R04942/01. They would also like to thank the anonymous reviewers, Bill Ladusaw, Gerald Penn, John Beavers and Christian Ebert for useful discussion and comments.

and therefore the intended meaning of the original word or phrase. This method is necessarily subjective, but a similar exercise attributing meaning types to clarification questions in this way has been shown to have reasonable statistical reliability when the judgements of two independent markers were compared (see Purver et al., 2001).

1.2 HPSG Notation

Our analysis assumes the (Ginzburg and Sag, 2000) version of HPSG. In order to save space and improve readability, we will use some abbreviations throughout, as shown in table 1.

AVM	Abbreviation
$\begin{bmatrix} \textit{parameter} \\ \text{INDEX} \quad \text{x} \\ \text{RESTR} \quad \left\{ \begin{bmatrix} \text{INSTANCE} & \text{x} \\ \text{PROPERTY} & \text{P} \end{bmatrix} \right\} \end{bmatrix}$	$x : property(x, P)$
$\begin{bmatrix} \textit{proposition} \\ \text{SOA} \mid \text{NUCLEUS} \quad \begin{bmatrix} \textit{verb\_rel} \\ \text{ROLE\_1} \quad \text{x} \\ \text{ROLE\_2} \quad \text{y} \end{bmatrix} \end{bmatrix}$	$verb(x, y)$
$\begin{bmatrix} \textit{question} \\ \text{PARAMS} \quad \{\} \\ \text{PROP} \quad verb(x, y) \end{bmatrix}$	$?\{\}.verb(x, y)$
$\begin{bmatrix} \textit{question} \\ \text{PARAMS} \quad \{x : property(x, P)\} \\ \text{PROP} \quad verb(x, y) \end{bmatrix}$	$?\{x\}.verb(x, y)$ or $?\{x : property(x, P)\}.verb(x, y)$

Table 1: HPSG AVM Abbreviations

In the next section we give some background on the analysis of reprise questions, and on various views of NP semantics. The subsequent sections 3 and 4 discuss the content of reprise questions for CNs and QNPs together with a corresponding semantic analysis, and some further issues arising from this are discussed in section 5.

2 Background

2.1 Reprise Questions

Ginzburg and Cooper (2001, forthcoming) (hereafter G&C) provide an analysis of proper name (PN) reprise questions which treats them as questions concerning the semantic content of the PN (taken to be a referential index). In this way, a reprise such as that in example (1) can be taken to be paraphrasable as shown, where the two readings are distinct, but both concern the content of the PN *Bo*:

(1)
A: Did Bo leave?
B: **BO?**

$\rightsquigarrow$ *"Is it* $\textbf{BO}_i$ *that you are asking whether* i *left?"*
$\rightsquigarrow$ *"Who do you mean by 'Bo'?"*

They analyse this via a representation which expresses contextual dependence: contextually dependent phrases such as PNs denote parameters which are abstracted to a set which is the value of a new C-PARAMS feature. This allows the sign to be viewed as a λ-abstract, or a *meaning* in the Montagovian sense (a function from context to content). This is shown in AVM (2) for A's original utterance in example (1)[2]:

(2)
$$\begin{bmatrix} \text{C-PARAMS} & \{x : named(x, Bo), a : speaker(a), b : addressee(b)\} \\ \text{CONTENT} & [ask(a, b, ?\{\}.leave(x))] \end{bmatrix}$$

An equivalent λ-abstract expression (ignoring the parameters associated with speaker and addressee, as we will do from now on for readability's sake) would be:

(3) $\lambda\{x : named(x, Bo)\}.ask(a, b, ?\{\}.leave(x))$

The grounding process for an addressee now involves establishing the referents of these parameters in context, in order to obtain the fully specified intended content. It is failure do this that results in the formation of a clarification question with the purpose of querying the sub-utterance associated with a troublesome parameter.

Clausal vs. Constituent Readings They give two possible readings for elliptical questions like *"Bo?"*: a *clausal* question, used to check that the hearer has instantiated the parameter in the correct way (made the correct link to the context), which corresponds to the first yes/no-question paraphrase given in example (1) above, and a *constituent* question used when the hearer cannot instantiate the parameter at all, the second *wh*-question paraphrase.

While the clausal and constituent readings are distinct, they both involve *querying the semantic content* of the relevant sub-utterance, following an inability to find

[2]Note also that the semantic representation includes the conversational move type ask, following Ginzburg et al. (2003) – this is important in order to give the correct interpretation for *clausal* questions (see below).

a suitable referent for that content in the hearer's context. This allows us to use them to investigate what semantic content can be attributed to various word and phrase types.[3]

G&C's analysis applies only to PNs. It is clear that other word and phrase types can be reprised, but it is also likely that not all reprises involve querying a simple referential index. On the other hand, it seems uncontentious to propose that these questions must query the semantic content of the fragment being reprised (or at least some part of it), and we take this as our basic hypothesis when examining NPs in this paper. Note that we do mean directly conveyed semantic content: reprise questions do not appear to be able to query, say, implicatures or other pragmatically inferred material (see Ginzburg et al., 2003).

2.2 NP Semantics

Common Nouns The semantic content of CNs is traditionally viewed as being a property (of individuals). Montague (1974) expressed this as a λ-abstract, a function from individuals to truth values (e.g. $\lambda x.dog(x)$), and this view is essentially shared by most strands of formal semantics. Variations (especially in representation) certainly exist: in situation semantics this might be expressed as a λ-abstracted infon (Cooper, 1995), in DRT as a predicative DRS (Asher, 1993), but these approaches share the basic view that CNs are properties of individuals.

Quantificational vs. Referential In contrast, the semantic representation of QNPs has long been a subject of lively debate. Traditional views of NP semantics can broadly be described as falling into two camps: the quantificational and the referential. The quantificational view, typified by Russell (1905) and Montague (1974), holds that QNPs contribute quantificational terms to the semantic representation of a sentence. This is exemplified by Barwise and Cooper (1981)'s GQ representation, in which sentences containing QNPs are given representations as follows:

(4) *"every A"* $\mapsto every(A) \quad \text{where} \quad [\![every(A)]\!] = \{X | A \subseteq X\}$

(5) *"every A Bs"* $\mapsto every(A)(B) \quad \text{where } [\![every(A)(B)]\!] = B \in [\![every(A)]\!]$

On this view, QNPs therefore denote families of sets (sets of sets, here the set of those sets which contain A).

In contrast, the referential view (going back to Strawson (1950) and Donnellan (1966)) sees some NPs as directly referential; particularly definites, but sometimes

[3] As G&C point out, reprise questions may have other possible readings apart from the two described above. In particular, a *lexical* reading concerning phonology or orthography of the words used by the speaker seems to be available in many situations. While seemingly common, we are not concerned with such readings in this paper as they do not shed any light on semantics. When we refer to reprise questions hereafter, this should be taken as referring to semantic content readings only.

also others such as specific uses of indefinites (e.g. Fodor and Sag, 1982).

Strict adherents to the quantificational view take it also to hold for definite descriptions: definites are not considered to be directly referential in the same sense as PNs, but are seen as defined by existential quantification with a uniqueness constraint, with any apparently referential nature argued to follow from pragmatic principles rather than any true semantic reference (see Kripke, 1977; Ludlow and Segal, forthcoming).

Other approaches such as the dynamic theories of Heim (1982) and Kamp and Reyle (1993) might be said to fall somewhere in between the two camps, with definites having some kind of reference (although this may be to a contextual discourse referent rather than a real-world object). In most views, however, NPs with other quantifiers (*every, most* etc.) are seen as quantificational.

2.3 HPSG Approaches to Semantics

Inheritance-Based One common framework for representing and constructing semantics in HPSG is the unification/inheritance-based method typified by e.g. (Sag and Wasow, 1999; Ginzburg and Sag, 2000). By default, CONTENT is inherited by mothers directly from head daughters: for QNPs, where the CN is usually treated as the head[4], this leads to a representation where the content of the QNP is identified with that of the head CN. This content is usually taken to be a parameter with a referential index, although this may be quantified over depending on the nature of the determiner.

(6)
$$\begin{bmatrix} \textit{np} \\ \textsc{phon} & \langle \text{the, dog} \rangle \\ \textsc{cont} & \boxed{1} \\ \textsc{dtrs} & \left\langle \begin{bmatrix} \textit{det} \\ \textsc{phon} & \langle \text{the} \rangle \\ \textsc{cont} & [\textit{quantifier}] \end{bmatrix}, \begin{bmatrix} \textit{noun} \\ \textsc{phon} & \langle \text{dog} \rangle \\ \textsc{cont} & \boxed{1}[x : dog(x)] \end{bmatrix} \right\rangle \end{bmatrix}$$

Amalgamation-Based Another approach commonly used by wide-coverage grammars is Minimal Recursion Semantics (MRS, see Copestake et al., 1999). Here CONTENT is (by default) amalgamated across daughters rather than being inherited directly from the head. Content is represented as *elementary predications*, pieces of propositional information. As can be seen below, this results in a representation of NPs wherein the NP content contains all contributions of its daughters,

[4] Although there are alternative views: see (Beavers, this volume) for a discussion.

including but not limited to the CN:

(7)
$$\left[\begin{array}{l} \textit{np} \\ \textsc{phon} \quad \left\langle \text{the, dog} \right\rangle \\ \textsc{cont} \quad \left[\begin{array}{ll} \textsc{hook} \mid \textsc{index} & x \\ \textsc{rels} & \left\{ \boxed{2}\left[h0 : the(x, h1, h2)\right], \boxed{1}\left[h1 : dog(x)\right] \right\} \end{array}\right] \\ \textsc{dtrs} \quad \left\langle \left[\begin{array}{ll} \textit{det} & \\ \textsc{phon} & \left\langle \text{the} \right\rangle \\ \textsc{cont} & \left[\begin{array}{ll} \textsc{hook} \mid \textsc{index} & x \\ \textsc{rels} & \left\{ \boxed{2} \right\} \end{array}\right] \end{array}\right], \left[\begin{array}{ll} \textit{noun} & \\ \textsc{phon} & \left\langle \text{dog} \right\rangle \\ \textsc{cont} & \left[\begin{array}{ll} \textsc{hook} \mid \textsc{index} & x \\ \textsc{rels} & \left\{ \boxed{1} \right\} \end{array}\right] \end{array}\right] \right\rangle \end{array}\right]$$

In the next section we examine CN reprise questions, and show that their meaning seems entirely consistent with the traditional view of CNs as denoting properties, but somewhat at odds with the HPSG approaches shown above. In section 4 we then discuss QNP reprise questions, show that their meaning disposes one towards the referential view of QNP semantics, and propose an HPSG analysis which accounts for CNs and QNPs. Section 5 then discusses some issues raised by the view put forward in section 4.

3 Common Nouns

The traditional view of CNs leads us to expect CN reprise questions to be able to query the property expressed by the noun, and this property only.[5] The clausal and constituent readings may both still be available, but the property should always be the element under question:

Clausal: *"Is it the property P that you are asking/asserting X(P)?"*

Constituent: *"What is the property P which you intend to convey by the word N?"*

In contrast, it should not be possible for CN-only reprises to be interpreted as questions about e.g. individual referents.

3.1 Corpus Evidence

Indeed, this appears to be the case: all corpus examples of CN reprises found confirmed this expectation. Examples are given here together with what appear to

[5]Note that we are setting mass nouns and bare plurals aside for the present, although we plan to investigate them in the same way in future.

be possible and impossible paraphrases – see example (8)[6]:

(8)

Monica:	You pikey! Typical!
Andy:	**Pikey?**
Nick:	Pikey!
Andy:	What's pikey? What does pikey mean?
Monica:	I dunno. Crusty.
⇝	*"Are you saying I am a **pikey**?"*
⇝	*"What property do you mean by the word 'pikey'?"*
⇝	*#"Which pikey are you saying I am?"*

The same appears to be true when the CN forms part of an indefinite NP as in example (9)[7]:

(9)

Emma:	Got a comb anywhere?
Helena:	**Comb?**
Emma:	Even if it's one of those `<pause>` tremmy `[sic]` pretend combs you get with a Barbie doll, oh this'll do! `<pause>` Don't know what it is, but it'll do!
⇝	*"Is it a **comb** that you are asking if I've got?"*
⇝	*#"Which comb are you are asking if I've got?"*

And indeed even when the CN is part of a seemingly referential definite NP as in example (10)[8]:

(10)

Carol:	We'll get the turkey out of the oven.
Emma:	**Turkey?**
Carol:	Well it's `<pause>` it's `<pause>` er `<pause>` what's his name? Bernard Matthews' turkey roast.
Emma:	Oh it's looks horrible!
⇝	*"Are you saying the thing we'll get out is a **turkey**?"*
⇝	*"What concept/property do you mean by 'turkey'?"*
⇝	*#"Which turkey are you saying we'll get out?"*
⇝	*#"Is it this/that turkey you're saying we'll get out?"*

Note that paraphrases which concern an intended referent of the NP containing the CN (e.g. the *"Which X . . . "* paraphrases) do not appear to be available, even when the NP might appear to be referential (see example (10)).

3.2 Analysis

As expected, we therefore suppose that the semantic representation of a CN must consist of a property of individuals (which we shall refer to as a *predicate* to differentiate it from a property-of-properties). An analysis entirely parallel to that of section 2.1 is possible if predicates are regarded as possible cognitive / contextual

[6] BNC file KPR, sentences 218–225. For the benefit of non-UK English speakers, *crusty* is a noun here, usually derogatory, and perhaps best thought of as somewhere between *hippy* and *tramp*.

[7] BNC file KCE, sentences 1513–1516

[8] BNC file KBJ, sentences 131–135. It may help non-UK residents to know that a Bernard Matthews' Turkey Roast is a processed meat product: turkey-like, but not actually a turkey.

referents. The CONTENT of a CN can then be a parameter whose INDEX is a named predicate. This parameter is also made a member of C-PARAMS: the hearer must ground it (by finding the intended (predicate) referent given its name) or make it the subject of a clarification question in case this grounding process fails (e.g. in the case of unknown, ambiguous or just surprising words).

(11)
$$\begin{bmatrix} \text{PHON} & \langle \text{dog} \rangle \\ \text{CONTENT} & \boxed{1}\left[P : name(P, dog) \right] \\ \text{C-PARAMS} & \{ \boxed{1} \} \end{bmatrix}$$

Note however that this does not correspond to the standard HPSG approaches of section 2.3. In the inheritance-based approach, CN CONTENT is a parameter whose INDEX is an individual (to be inherited as the referent of a NP mother). Including this parameter in C-PARAMS, as shown in AVM (12), would not give the correct reading for a clarification question, as this individual would become the referent to be grounded and thus the subject of the question (which we have seen is impossible).

(12)
$$\begin{bmatrix} \text{CONTENT} & \boxed{1}\left[x : dog(x) \right] \\ \text{C-PARAMS} & \{ \boxed{1} \} \end{bmatrix}$$

Similarly in the MRS approach, CN content consists of an EP which again concerns the individual referent which will be quantified over by the mother NP, and making tbis content contextually available would allow reprise questions which concern this referent.

These problems could be solved by alternative analyses for both approaches whereby only *part* of the content (the predicate) is abstracted, but these would then beg the question of why only that part is abstracted and available for clarification. This would be especially problematic for the inheritance approach where CN and NP content are identical: as we will see below, the two do not give rise to the same reprise questions.

4 Noun Phrases

The quantificational and referential views of QNP semantics would seem to predict different meanings for QNP reprises, at least for those examples which the latter view holds to be directly referential: referential definites and perhaps specific indefinites.

4.1 Definite NPs

Taking a referential semantic viewpoint, we might therefore expect reprises of definite NPs to concern individual referents, and be paraphrasable as follows:

Clausal: *"Is it the individual X about which you are asking/asserting P(X)?"*
Constituent: *"Which individual X do you intend to refer to by the phrase NP?"*

From a quantificational viewpoint, a paraphrase concerning a set of properties or sets might instead be expected:
Clausal: *"Is it the set of properties that hold of X about which you are asking/asserting . . . ?"*
Constituent reading: *"Which set of properties do you intend to convey by the phrase NP?"*

Our corpus investigation included many types of definite NP: PNs, pronouns and demonstratives as well as definite descriptions. PNs have already been discussed in section 2.1 above – we examine the others here.

4.1.1 Referential Definites

All reprises of demonstratives and pronouns, and most reprises of definite descriptions (over half of the examples we found) appeared to be directly referential, with both clausal and constituent readings available (see examples (13)[9] and (14)[10]).

(13)

John:	Which way's North, do you know?
Sara:	That way.
John:	**That way?** Okay.
⇝	*"Are you telling me **that way there** is North?"*
⇝	*"By 'that way' do you mean that way there?"*

(14)

John:	They would be working on the kidnapper's instructions, the police?
Sid:	**The police?**
John:	Aye
Sid:	On
Unknowns:	<unclear>
Sid:	aye the, the senior detectives
⇝	*"Is it **the police** who you are saying would be working . . . ?"*
(⇝	*"Who do you mean by 'the police'?"*)

Reprises using PNs Interestingly, it appears possible to reprise these definites not only by echoing verbatim as in example (13), but also by reprising with a co-referring PN as in examples (15)[11] and (16)[12]. This gives further weight to the idea that these reprises are genuinely referential (PNs are generally held to be referential

[9] BNC file JP4, sentences 755–758
[10] BNC file KCS, sentences 661–665
[11] BNC file KCE, sentences 4190–4192
[12] BNC file KPY, sentences 1005–1008

even by those who hold to the quantificational view of definite NPs).

(15)

Joanne:	It's, how many times did he spew up the stairs?
Emma:	**Julian?** Couple of times.
⇝	*"Is it **Julian**$_i$ that you are asking how many times i spewed up the stairs?"*
⇝	*"By 'he' do you mean Julian?"*

(16)

Unknown:	And er they X-rayed me, and took a urine sample, took a blood sample. Er, the doctor
Unknown:	**Chorlton?**
Unknown:	Chorlton, mhm, he examined me, erm, he, he said now they were on about a slide <unclear> on my heart. Mhm, he couldn't find it.
⇝	*"By 'the doctor' do you mean Chorlton?"*

Two points are perhaps worth reinforcing: firstly, definite descriptions, pronouns, demonstratives and proper names all seem to make the same kind of referential reprise questions available; secondly, it seems very hard to interpret any of these examples as querying a family of sets rather than an individual referent.

We therefore suppose that the content of definite NPs must at least contain, and perhaps consist entirely of, the intended referent (or for plurals, set of referents), as shown in AVM (17). An analysis of these referent reprise questions would then be available exactly as for PNs in section 2.1 – an identifiable referent for the contextual parameter must be found in context as part of the grounding process.

(17)

$$\begin{bmatrix} \textsc{phon} & \langle \text{the, dog} \rangle \\ \textsc{content} & \boxed{1}\left[x : the\_dog(x) \right] \\ \textsc{c-params} & \left\{ \boxed{1} \right\} \end{bmatrix}$$

4.1.2 Functional Definites

Most other examples of definite description reprises did not seem to be querying an individual referent, but seemed better understood as querying a functional referent or its domain. These examples were mostly *attributive* uses (example (18)[13]): we also expect *de dicto* and *narrow scope* uses, among others, to behave in this way.

(18)

Eddie:	I want you <pause> to write the names of these notes up here.
Anon 1:	**The names?**
Eddie:	The names of them.
Anon 1:	Right.
⇝	*"What situation/notes should I interpret 'the names' relative to?"*
⇝	*"What are you intending 'the names' to refer to in that situation?"*
⇝	*#"Which actual names are you referring to by 'the names'?"*

Again, a reading concerning properties of properties or sets of sets does not

[13] BNC file KPB, sentences 418–421

seem plausible. We therefore suppose that such uses are best captured by an analysis as sketched in AVM (19), this being the functional equivalent of the version in AVM (17) above, with its constituent function and domain becoming the members of C-PARAMS:

(19)
$$\begin{bmatrix} \textsc{phon} & \left\langle \text{the, dog} \right\rangle \\ \textsc{content} & \left[f(s) : s \in D \wedge s \models the\_dog(f(s)) \right] \\ \textsc{c-params} & \left\{ \left[f \right], \left[D \right] \right\} \end{bmatrix}$$

Both function f and domain D of the argument s must therefore be found in context, and failure to do so licenses clarification questions concerning either function or domain, or both. Note that the idea of domain identification being required for definite interpretation has precedent (e.g. Poesio (1993)'s view of definite interpretation as anchoring a parameter corresponding to the resource situation), but that on our view this is not *all* that is required.

As shown above, we take the function expressed by attributive uses to be one from resource situations to individuals, following (Barwise and Perry, 1983). Other types such as narrow scope uses might be better accounted for as functional on wide-scoping individuals rather than situations.

4.1.3 Sub-Constituent Readings

The few remaining examples of definite NP reprises seemed to have a predicate reading, identical to that which would be obtained by reprising the CN alone. No intonational information is available in the BNC, but these readings appear to be those that are made more prominent by stressing the CN (see example (20)[14]).

(20)

Anon 1:	They'd carry the sack on their back?
George:	On the back, the bushel, yes
Anon 1:	**The bushel?**
George:	`<unclear>`
Anon 1:	`<unclear>`
George:	The corn.
$\rightsquigarrow$	*"What are you referring to by 'the bushel'?"*
$\rightsquigarrow$	*"What property do you mean by 'bushel'?"*
$\rightsquigarrow$	*"Is it the thing with the property **bushel** that you're saying ..."*

This does not seem to be restricted to definites: we will see the same readings for all other NPs we examined (see below). We will also see below that it is not restricted to the CN predicate – readings corresponding to the logical relation expressed by the determiner are also possible (again, the reader may find this easier to capture by imagining intonational stress on the determiner). In other words, the readings available for reprises of sub-constituents of the NP are still available when reprising the NP, especially when the relevant sub-constituent is stressed. We

[14] BNC file H5H, sentences 254–257

therefore suppose that this reading is in fact a focussed reprise of a daughter rather than the NP as a whole, and we will come back to this below.

4.2 Indefinite NPs

Again, a referential viewpoint might lead us to expect that reprises of indefinites should involve a referent; otherwise we expect a set of sets or property of properties.

4.2.1 Sub-Constituent Readings

However, if they do exist, such readings seem to be uncommon. All singular indefinite examples were most felicitous when read as CN sub-constituent readings (see example (21)[15]), as described in section 4.1.3 above. Note that the constituent reading, paraphrased in the examples below as *"What property do you mean by 'N'?"*, might also be paraphrased *"What is a N?"* – but that this should not be confused with a *referential* constituent reading *"Which N do you mean by 'a N'?"*.

(21)

Mum:	I've been treating it as a wart.
Vicky:	**A wart?**
Mum:	A corn and I've been putting corn plasters on it
⇝	*"Is it the property* ***wart***$_i$ *that you're saying you've been treating it as something with* i*?"*
⇝	*"What property do you mean by 'wart'?"*
⇝	*#"Which wart are you saying you've been treating it as?"*

For plural indefinites the same holds (example (22)[16]), although a reading querying the determiner rather than the predicate is also available:

(22)

Anon 1:	It had twenty rooms in it.
Anon 2:	**Twenty rooms?**
Anon 1:	Yes.
⇝	*"Is it* ***twenty***$_N$ *that you're saying it had N rooms?"*
⇝	*"Is it* ***rooms*** *that you're saying it had twenty of?"*
⇝	*#"Which twenty rooms are you saying are it had?"*

Note that again, the set-of-sets reading does not seem at all plausible.

4.2.2 Possible Referential Readings

However, while no clear examples were found in our corpus study, we feel that there *is* a possibility of referential questions with specific indefinites where the hearer realises that the speaker has a particular referent in mind, and intends the hearer to be able to identify it (what Ludlow and Segal (forthcoming) call *definite* indefinites). Some BNC examples, while probably most felicitous when read

[15] BNC file KE3, sentences 4679–4681

[16] BNC file K6U, sentences 1496–1498

as CN predicate queries, do seem to offer a possible referential paraphrase, e.g. example (23)[17]:

(23)

Stefan:	Everything work which is contemporary it is decided
Katherine:	Is one man?
Stefan:	No it is a woman
Katherine:	**A woman?**
Stefan:	A director who'll decide.
Katherine:	She's good?
Stefan:	Hm hm very good.
$\rightsquigarrow$	*"Is it a **woman** you are saying it is?"*
$\rightsquigarrow$	?*"Which woman are you saying it is?"*

If these readings are possible, an analysis of indefinites should allow for them to be constructed. Given this and the implausibility of a set-of-sets reading, we propose that as for definites, the content of indefinites should be an individual (or set of individuals). In ordinary uses this content must be existentially quantified at sentence/clause level (via STORE) – definite uses are distinguished simply by making the content a member of C-PARAMS (see the two versions in AVM (24)).

(24)

$$\begin{bmatrix} \textsc{phon} & \langle \text{a, dog} \rangle \\ \textsc{content} & \boxed{1}\left[x : dog(x)\right] \\ \textsc{store} & \left\{\boxed{1}\right\} \\ \textsc{c-params} & \{\} \end{bmatrix} \quad \begin{bmatrix} \textsc{phon} & \langle \text{a, dog} \rangle \\ \textsc{content} & \boxed{1}\left[x : dog(x)\right] \\ \textsc{store} & \{\} \\ \textsc{c-params} & \left\{\boxed{1}\right\} \end{bmatrix}$$

4.3 Other Quantified NPs

Reprises of QNPs with other quantifiers are very rare in the BNC[18], so we cannot claim strong results; but what examples we could find show similar behaviour to indefinites. Set-of-sets readings seem impossible; most examples seem best interpreted as concerning sub-constituents (either the CN predicate or the logical determiner relation); but referential interpretations seem possible too (see exam-

[17] BNC file KCV, sentences 3012–3018

[18] This is not surprising, as these NPs are relatively rare in the BNC to begin with: there are more than 50 times more sentences containing *"the N"* as there are containing *"every N"*, and *"most N"*, *"many N"* and *"few N"* are even rarer.

ple (25)[19]):

(25)

Richard:	No I'll commute every day
Anon 6:	**Every day?**
Richard:	as if, er Saturday and Sunday
Anon 6:	And all holidays?
$\rightsquigarrow$	*"Is it* ***days***$_N$ *that you are saying you'll commute every N?"*
$\rightsquigarrow$	*"Is it* ***every***$_D$ *that you are saying you'll commute on D days?"*
$\rightsquigarrow$	*"Which days do you really mean by 'every day'?"*

We should perhaps not be surprised by referential readings with universal quantifiers: universals are sometimes considered as definites (see e.g. Abbott, 2001). But although other quantifiers were too rare in the BNC to provide evidence, we can imagine examples in which referential readings seem plausible, especially when using co-referring PNs in the reprise:

(26)

A:	Most people came to the party.
B:	**Most people?**
A:	Well, me, Brenda and Carmen.
$\rightsquigarrow$	*"Who do you mean by 'most people'?"*

Given this possibility, we propose to analyse these QNPs like indefinites: as existentially quantified sets of individuals, which are not contributed to C-PARAMS under normal circumstances. Referential uses are obtained simply by adding the content to C-PARAMS.

4.4 HPSG Analysis

QNPs as Witness Sets The evidence therefore leads us towards a representation whereby all QNPs denote sets of individuals, while CNs denote predicates. Referential NPs (including definites and referential uses of indefinites) are those where the set must be identified in context; for non-referential NPs, the set must be existentially quantified.

Such an existentially quantified set representation is justified for all monotone-increasing (MON↑) quantifiers if we take the sets as Barwise and Cooper (1981)'s *witness sets*: they show that a verbal predicate belonging to a GQ $D(A)$ is equivalent to the predicate holding of a witness set, where this is a set w which is both a subset of A and a member of $D(A)$. For an indefinite *a dog*, w can be any nonempty set of dogs; for the universal *every dog*, w is the set of all dogs; for *most dogs*, w is a set containing more than half of all dogs, and so on.

CONTENT **Specification** Note that under this analysis, NPs do not inherit their content directly from either daughter, or amalgamate it across daughters (the two common HPSG approaches): the referential set reprise reading is available when

[19]BNC file KSV, sentences 257–260

reprising NPs, but not when reprising daughters. Instead of using a general inheritance or amalgamation principle, we must therefore posit a type *qnp* for all QNPs which specifies how the semantic representation is built:

(27)
$$\begin{bmatrix} \textit{qnp} & \\ \textsc{content} & \left[w : w = Q'(P) \right] \\ \textsc{dtrs} & \left\langle \begin{bmatrix} \textit{det} & \\ \textsc{content} & Q' \end{bmatrix}, \begin{bmatrix} \textit{nominal} & \\ \textsc{content} & P \end{bmatrix} \right\rangle \end{bmatrix}$$

Here we are representing the CN as a predicate P and the determiner as a logical relation Q' between predicate and witness set. In Barwise and Cooper (1981)'s terms, this can be related to the standard GQ representation $Q(P)$ as follows:

(28) $w = Q'(P) \quad \leftrightarrow \quad w \subseteq P \wedge w \in Q(P)$

Note that the constraint expressed above is still monotonic (no semantic information is dropped in construction of the mother) and compositional (the content of the mother is obtained purely by functional application of daughter contents). But note also that by this nature it does not fit with the approaches we are used to in HPSG: content is not simply inherited nor amalgamated.

Existential Quantification and STORE Quantification uses the familiar lexically-based storage and retrieval method of (Ginzburg and Sag, 2000): existentially quantified elements are added to STORE, inherited via heads and retrieved into QUANTS. As only existential quantification is being used, the members of QUANTS can simply be parameters rather than quantifiers, and their order is not important. QUANTS can therefore be a set rather than a list, no longer requiring the `order` operator of Ginzburg and Sag (2000). The members of the QUANTS set are taken to be *simultaneously* quantified over, following Cooper (1993)'s definition of simultaneous quantification for STDRT.

Our version of the STORE Amalgamation Constraint therefore appears as in AVM (29):

(29)
$$\begin{bmatrix} \textit{word} & \\ \textsc{content} & \left[\textsc{quants} \quad \boxed{2} \right] \\ \textsc{store} & \{\boxed{1} \cup \ldots \cup \boxed{n}\} - \boxed{2} \\ \textsc{arg-st} & \left\langle \left[\textsc{store} \quad \boxed{1} \right], \ldots, \left[\textsc{store} \quad \boxed{n} \right] \right\rangle \end{bmatrix}$$

C-PARAMS **Amalgamation** We have seen that reprising a QNP mother can sometimes give a reading which queries only a focussed sub-constituent daughter; but reprising a daughter cannot query the content of the mother (or indeed its sisters, although we have not shown evidence for this here). Therefore the C-PARAMS value of NPs must include the amalgamated values of its daughters so that they can

form the subject of the query[20], but this cannot be inherited directly from any one of them. C-PARAMS must therefore be amalgamated by mothers directly across daughters (rather than via lexical heads and inheritance as assumed by G&C). We can express this as a default constraint:

(30)
$$\begin{bmatrix} \textit{phrase} & \\ \text{C-PARAMS} & \boxed{1} \cup \ldots \cup \boxed{n} \\ \text{DTRS} & \left\langle \left[\text{C-PARAMS} \quad \boxed{1}\right], \ldots, \left[\text{C-PARAMS} \quad \boxed{n}\right] \right\rangle \end{bmatrix}$$

However, definite NPs must override this default, as they also introduce a new parameter (their own content). Indefinites hold to the default, but we must ensure that their content is instead existentially quantified.

Definiteness Principle So indefinites contribute their content to STORE, while definites contribute it to C-PARAMS. We can therefore state a general Definiteness Principle: the content of a NP must be a member of either C-PARAMS or STORE. For words, this is simply expressed:

(31)
$$\begin{bmatrix} \textit{word} & \\ \text{CONTENT} & \boxed{1} \\ \text{STORE} & \boxed{2} \\ \text{C-PARAMS} & \left\{\boxed{1}\right\}- \boxed{2} \end{bmatrix}$$

For phrases, we must combine with STORE inheritance and C-PARAMS amalgamation (replacing AVM (30)):

(32)
$$\begin{bmatrix} \textit{phrase} & \\ \text{CONTENT} & \boxed{1} \\ \text{STORE} & \boxed{2} \cup \boxed{3} \\ \text{C-PARAMS} & (\left\{\boxed{1}\right\}- \boxed{2}) \cup \boxed{4} \cup \ldots \cup \boxed{n} \\ \text{HEAD-DTR} & \left[\text{STORE} \quad \boxed{3}\right] \\ \text{DTRS} & \left\langle \left[\text{C-PARAMS} \quad \boxed{4}\right], \ldots, \left[\text{C-PARAMS} \quad \boxed{n}\right] \right\rangle \end{bmatrix}$$

Definites and other referential words/phrases[21] can therefore be specified as having empty STORE values, forcing their content to be a member of C-PARAMS. Indefinites can be specified as contributing to STORE, and thus can make no contribution to C-PARAMS.

4.5 Summary

This section has shown that reprises of definite NPs query a (possibly functional) referent, and surmised that this may also be true for referential uses of other QNPs.

[20] We analyse this sub-constituent focussing using Engdahl and Vallduví (1996)'s HPSG treatment of information structure, but space precludes a full exposition here.

[21] On our account, this includes CNs, which are referential to a predicate.

Non-referential uses seem to query sub-constituents: questions about GQs or sets of sets are not plausible.

We have therefore proposed a semantic representation of NPs as witness sets rather than GQs, and shown how to express quantification and the alternation between definiteness and indefiniteness. The next section briefly examines some further implications of this representation.

5 Further Issues

Determiners The analysis of section 4.4 assumed that determiners denoted logical relations between predicates and witness sets. Determiner-only reprises should therefore query such relations, but they are rare in the BNC: the only suitable examples found involved numerals (see example (33)[22]). For these examples, the query appears to concern the cardinality of the witness set, which does fit quite nicely with the idea of determiners as denoting set relations.

(33)

Marsha:	yeah that's it, this, she's got three rottweiler's now and
Sarah:	**three?**
Marsha:	yeah, one died so only got three now `<laugh>`
$\rightsquigarrow$	*"Is it* ***three***$_N$ *you are saying she's got* N *rottweilers?"*

For other determiners, we have to rely on our intuition, and on those QNP reprise examples mentioned in section 4 above in which the determiner appears to be stressed, e.g. example (25) above, for which we gave a determiner paraphrase which again seems to query a relation. Of course, we hesitate to make any strong claims based on this limited evidence, but we can say that the determiner reprises we have seen provide no counter-evidence to the analysis of section 4.4.

Anaphora Intersentential anaphora has already been briefly discussed – pronouns appear to behave like referential definites in that their referents must be identified in context, and can be clarified otherwise. However, accounting for *intra*sentential anaphora requires a further step. If pronouns (and anaphoric definites) refer to existentially quantified elements within the same sentence, they can no longer have a C-PARAM associated with them: they do not refer to an element in the external context.

We therefore propose that elements of C-PARAMS can be removed if they can be identified with an element of QUANTS – i.e. a binding mechanism similar in concept to Poesio (1993)'s *parameter anchoring* and van der Sandt (1992)'s *presupposition binding*. This is implemented via a new feature B(OUND)-PARAMS: referential parameters can be members of either C-PARAMS or B-PARAMS, but membership of B-PARAMS is limited to those parameters which can be identified with members of QUANTS). This means we must update our definiteness principle

[22] BNC file KP2, sentences 295–297

to allow B-PARAMS membership:

(34)
$$\begin{bmatrix} \textit{word} & \\ \text{CONTENT} & \boxed{1} \\ \text{STORE} & \boxed{2} \\ \text{C-PARAMS} & \boxed{3} \\ \text{B-PARAMS} & \left\{\boxed{1}\right\} - \boxed{2} - \boxed{3} \end{bmatrix}$$

while B-PARAMS discharge is expressed through a similar mechanism to quantifier retrieval:

(35)
$$\begin{bmatrix} \textit{word} & \\ \text{CONTENT} & \left[\text{QUANTS} \quad \boxed{Q}\right] \\ \text{B-PARAMS} & \{\boxed{1_b} \cup \ldots \cup \boxed{n_b}\} - subset(\boxed{Q}) \\ \text{ARG-ST} & \left\langle \left[\text{B-PARAMS} \quad \boxed{1_b}\right], \ldots, \left[\text{B-PARAMS} \quad \boxed{n_b}\right] \right\rangle \end{bmatrix}$$

We ensure that all members of B-PARAMS are thus discharged by specifying top-level sentences (in our grammar, signs of type *root-cl*) as having empty B-PARAMS.

Quantifier Scope The functional representation of section 4.1.2 allows relative scope to be expressed by regarding narrow-scoping NPs as functional on other wider-scoping sets: the alternative readings of *"every dog_d likes a cat_c"* are produced by the alternative views of *a cat* being a simple existentially quantified individual c, or one that is functionally dependent on the set of dogs $f(d)$ via an existentially quantified function f.[23] This follows simply from the anaphora mechanism described above: the narrow-scope reading is produced by identifying the *domain* of the functional cat with the existentially quantified set of dogs via B-PARAMS, while the function is existentially quantified via STORE.

Monotone Decreasing Quantifiers A simple witness set representation cannot be sufficient for non-MON↑ quantifiers: the sentence *"few men work"* does not only convey the fact that working holds of some set w containing few men, but also that it does not hold of any men not in w.

One solution might be to appeal to pragmatics: Hobbs (1996) solves the problem by pragmatically strengthening the sentence meaning to the assertion that w is the *maximal* set of working men. Another would of course be to regard the content of QNPs as GQs rather than witness sets, but then we cannot explain why sets-of-sets reprise readings seem impossible. A third, which we favour, is to view non-MON↑ QNPs as denoting pairs of *reference set* (the men who work) and *complement set* (the men who don't). We would then expect reprises to be able to query both sets; again, as examples of non-MON↑ QNP reprises are rare, we are not sure

[23] This is similar to the choice function approach to scope (see e.g. Reinhart, 1997).

yet whether this is the case, but imagined examples are encouraging. Kibble (1997) gives the following example of complement set anaphora:

(36) | BBC News: Not all of the journalists agreed, among them the BBC's John Simpson.

where *them* is construed to refer to those who did *not* agree. An imagined reprise version seems possible to construe as querying the complement set:

(37)
A: Not all of the journalists agreed.
B: **Not all of them?**
A: John Simpson was pretty combative. Paxman didn't like it much either.

$\rightsquigarrow$ *"Who do you mean **didn't** agree?"*

More data is needed, but if plausible this might allow a neat way to explain complement set anaphora in general.

6 Conclusions

In this paper we have introduced the use of reprise questions as probes in order to investigate the semantic content of words and phrases, and examined the evidence provided thereby as regards the content of CNs and NPs. This has led us to a view of CNs as denoting predicates, and all MON↑ QNPs as denoting witness sets, with the difference between definite and indefinite uses expressed by contextual identification via C-PARAMS vs. existential quantification via STORE. We have shown how this can take into account relative scope and anaphora, and suggested a solution for non-MON↑ quantifiers via a representation as pairs of sets.

Along the way, we have seen that inheritance/amalgamation approaches common in HPSG do not fit with the evidence. This is not intended as a criticism of these approaches, which serve their intended purpose of building high-level sentence semantics extremely well: it is only once we start to look at this low level, at the semantics that individual words and phrases can have on their own, that we need to revise our thinking.

References

B. Abbott. Definiteness and indefiniteness. In L. Horn and G. Ward, editors, *Handbook of Pragmatics*. Blackwell, 2001. To appear.

N. Asher. *Reference to Abstract Objects in Discourse*. Kluwer Academic Publishers, 1993.

J. Barwise and R. Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219, 1981.

J. Barwise and J. Perry. *Situations and Attitudes*. MIT Press, 1983.

L. Burnard. *Reference Guide for the British National Corpus (World Edition)*. Oxford University Computing Services, 2000.

R. Cooper. Towards a general semantic framework. In R. Cooper, editor, *Integrating Semantic Theories*. ILLC/Department of Philosophy, University of Amsterdam, 1993.

R. Cooper. The role of situations in generalized quantifiers. In S. Lappin, editor, *The Handbook of Contemporary Semantic Theory*. Blackwell, 1995.
A. Copestake, D. Flickinger, I. A. Sag, and C. Pollard. Minimal recursion semantics: An introduction. Draft, 1999.
K. Donnellan. Reference and definite descriptions. *Philosophical Review*, 77:281–304, 1966.
E. Engdahl and E. Vallduví. Information packaging in HPSG. In C. Grover and E. Vallduví, editors, *Studies in HPSG*, pages 1–31. University of Edinburgh, 1996.
J. Fodor and I. A. Sag. Referential and quantificational indefinites. *Linguistics and Philosophy*, 5:355–398, 1982.
J. Ginzburg and R. Cooper. Resolving ellipsis in clarification. In *Proceedings of the 39th Meeting of the ACL*, pages 236–243, 2001.
J. Ginzburg and R. Cooper. Clarification, ellipsis, and the nature of contextual updates. *Linguistics and Philosophy*, forthcoming.
J. Ginzburg and I. A. Sag. *Interrogative Investigations: the Form, Meaning and Use of English Interrogatives*. CSLI Publications, 2000.
J. Ginzburg, I. A. Sag, and M. Purver. Integrating conversational move types in the grammar of conversation. In P. Kühnlein, H. Rieser, and H. Zeevat, editors, *Perspectives on Dialogue in the New Millennium*, pages 25–42. John Benjamins, 2003.
I. Heim. *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts at Amherst, 1982.
J. Hobbs. Monotone decreasing quantifiers. In K. van Deemter and S. Peters, editors, *Semantic Ambiguity and Underspecification*, pages 55–76. CSLI Publications, 1996.
H. Kamp and U. Reyle. *From Discourse To Logic*. Kluwer Academic Publishers, 1993.
R. Kibble. Complement anaphora and dynamic binding. In *Proceedings of the 7th annual conference on Semantics and Linguistic Theory (SALT VII)*, pages 346–368, 1997.
S. Kripke. Speaker's reference and semantic reference. In *Perspectives in the Philosophy of Language*, pages 6–27. University of Minnesota Press, 1977.
P. Ludlow and G. Segal. On a unitary semantical analysis for definite and indefinite descriptions. In A. Bezuidenhout and M. Reimer, editors, *Descriptions and Beyond: An Interdisciplinary Collection of Essays on Definite and Indefinite Descriptions*. Oxford University Press, forthcoming.
R. Montague. The proper treatment of quantification in ordinary English. In R. Thomason, editor, *Formal Philosophy: Selected Papers of Richard Montague*, pages 247–270. Yale University Press, 1974.
M. Poesio. A situation-theoretic formalization of definite description interpretation in plan elaboration dialogues. In P. Aczel, D. Israel, Y. Katagiri, and S. Peters, editors, *Situation Theory and its Applications*, volume 3, pages 339–374. CSLI Publications, 1993.
M. Purver. SCoRE: A tool for searching the BNC. Technical Report TR-01-07, Department of Computer Science, King's College London, October 2001.
M. Purver, J. Ginzburg, and P. Healey. On the means for clarification in dialogue. In *Proceedings of the 2nd ACL SIGdial Workshop on Discourse and Dialogue*, 2001.
T. Reinhart. Quantifier scope: How labour is divided between QR and choice functions. *Linguistics and Philosophy*, 20:335–397, 1997.
B. Russell. On denoting. *Mind*, 14:479–493, 1905.
I. Sag and T. Wasow. *Syntactic Theory: A Formal Introduction*. CSLI Publications, 1999.
P. Strawson. On referring. *Mind*, 59:320–344, 1950.
R. van der Sandt. Presupposition projection as anaphora resolution. *Journal of Semantics*, 9:333–377, 1992.