More Complement Selection, and Agreement

Doug Arnold
Language & Linguistics
doug@essex.ac.uk

1 Review

We have three rules (roughly):

(1) XP → Spec, Head (Head-Spec Rule)
(2) XP → Head, $C_1 \ldots, C_n$ (Head-Comps Rule)
(3) XP → Head, Adjunct (Head-Adjunct)

With extra constraints like the following:

- In Head-Subj construction, the Spec must match the single element of the head's SUBJ list.
- In a Head-Comps construction, the C_i must match the elements of the head's COMPS list.
- The POS value on the mother must be the same as the head's POS value.

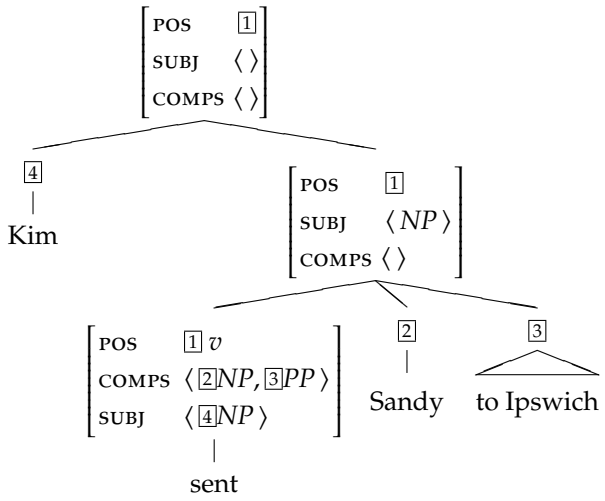
That is, essentially, but slightly more precisely (ignoring POS):

(4) $\left[\textsc{spr}\ \langle\ \rangle\right] \rightarrow \boxed{1},\quad \mathbf{H}\left[\textsc{spr}\ \left\langle \boxed{1} \right\rangle\right]$ (Head-Spec Rule)

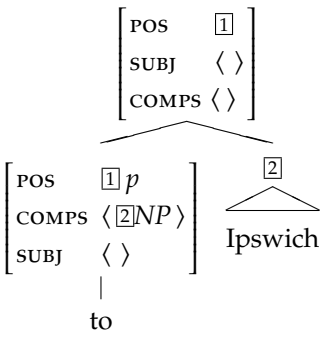
(5) $\left[\textsc{comps}\ \langle\ \rangle\right] \rightarrow \mathbf{H}\left[\textsc{comps}\ \left\langle \boxed{1}, \ldots, \boxed{n} \right\rangle\right],\quad \boxed{1}, \ldots \boxed{n}$ (Head-Comps Rule)

(6) XP → **H** $\boxed{1}$, $\left[\textsc{mod}\ \left\langle \boxed{1} \right\rangle\right]$ (Head-Mod Rule)

(7)



(8)



Notice that when we write a 'tag' in several places, it indicates the same structure, so it makes no difference where we write the actual value. ('Token identity')

Today, we will extend this approach to handle a variety of phenomena:

- more kinds of complement selection
- agreement

2 More Complement Selection

- Finite verbs require their subjects to be nominative, prepositions require their complements to be accusative:
  (9) She/*Her walked behind *she/her.
- Some verbs select PPs with a particular preposition, or particular kinds of S complement:
  (10) Sam relies on Sandy (*at/*under/*near Sandy),
  (11) It *bothered* Sam [*(that) Kim left].

In dealing with some of these phenomena, it will be useful to assume that POS is just one of several features that is shared between a head and its mother.

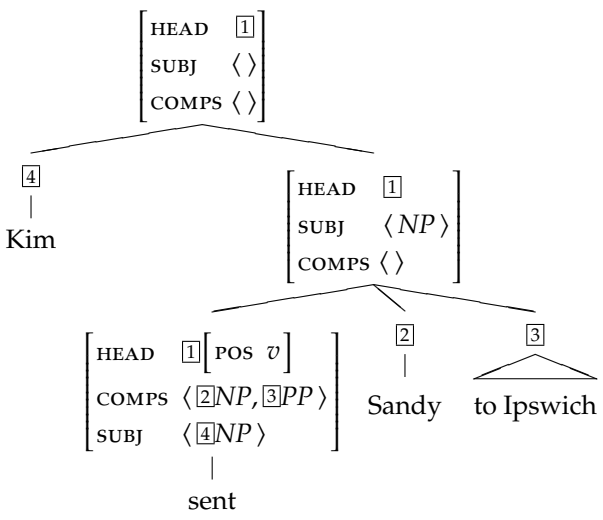
(12) $\begin{bmatrix} \textsc{head} & \boxed{2} \\ \textsc{spr} & \langle\,\rangle \end{bmatrix} \rightarrow \boxed{1}, \quad \mathbf{H}\begin{bmatrix} \textsc{head} & \boxed{2} \\ \textsc{spr} & \left\langle \boxed{1} \right\rangle \end{bmatrix}$ (Head-Spec Rule)

(13) $\begin{bmatrix} \textsc{head} & \boxed{5} \\ \textsc{comps} & \langle\,\rangle \end{bmatrix} \rightarrow \mathbf{H}\begin{bmatrix} \textsc{head} & \boxed{5} \\ \textsc{comps} & \left\langle \boxed{1}, \ldots, \boxed{n} \right\rangle \end{bmatrix}, \quad \boxed{1}, \ldots \boxed{n}$ (Head-Comps Rule)

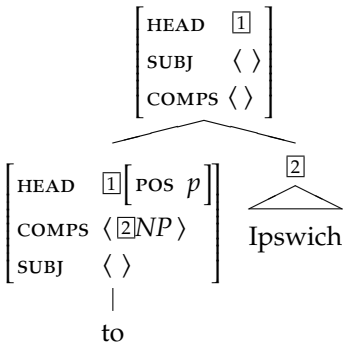
(14) $\left[\textsc{head}\ \boxed{2}\right] \rightarrow \mathbf{H}\left[\textsc{head}\ \boxed{2}\right]\boxed{1}, \quad \left[\textsc{mod}\ \left\langle \boxed{1} \right\rangle\right]$ (Head-Mod Rule)

We now get structures like (15) and (16):

(15)



(16)



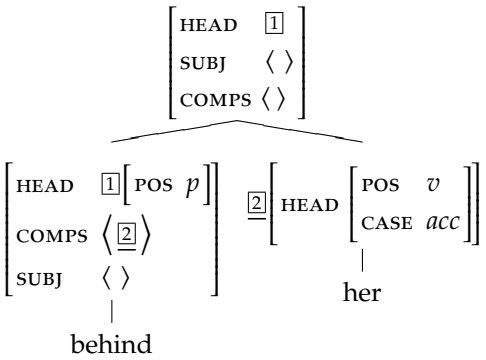
Now suppose we introduce a head feature CASE, with values *nom*, and *acc*.

If we assume *she* is specified as HEAD|CASE *nom*, and *her* is specified as HEAD|CASE *acc*, then we will allow (17a), and disallow (17b):

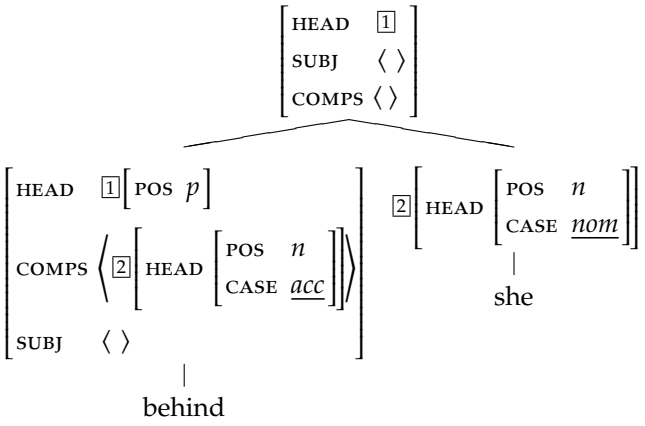
(17) a. (They sat) behind her.

b. (They sat) *behind she.

(18)



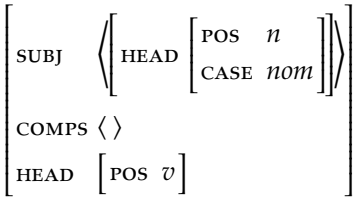
(19)



The structure 2 cannot be both HEAD | CASE *nom* and HEAD | CASE *acc*.

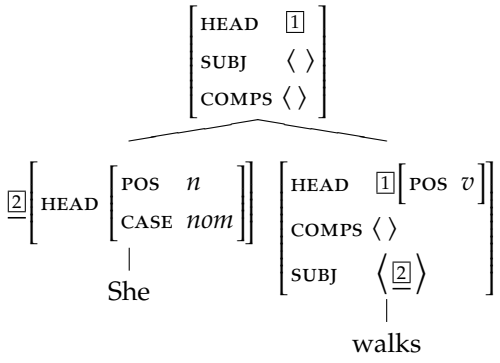
Similarly, if the lexical entry for *walks* is specified as in (20), we will allow (21a) and disallow (21b):

(20)

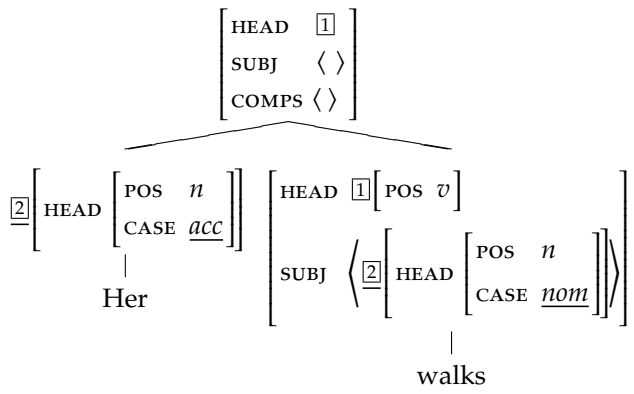


(21) a. She walks.
b. *Her walks.

(22)



(23)



We can account for the contrast in (24), if we assume there is a head feature PFORM, with values like *on*, *in*, *behind*, etc.

(24) Sam relies on Sandy (*at/*under/*near Sandy),

The lexical entry for *on* will be along the lines of (25):

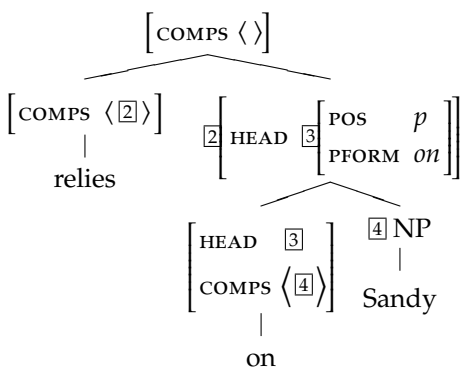
(25) $$\begin{bmatrix} \textsc{comps} & \left\langle \left[\textsc{head} \begin{bmatrix} \textsc{pos} & n \\ \textsc{case} & acc \end{bmatrix} \right] \right\rangle \\ \textsc{subj} & \langle\,\rangle \\ \textsc{head} & \begin{bmatrix} \textsc{pos} & p \\ \textsc{pform} & on \end{bmatrix} \end{bmatrix}$$

The lexical entry for *rely (on)* will be as in (26) (ignoring some details):

(26) $$\begin{bmatrix} \textsc{comps} & \left\langle \left[\textsc{head} \begin{bmatrix} \textsc{pos} & p \\ \textsc{pform} & on \end{bmatrix} \right] \right\rangle \\ \textsc{subj} & \langle \ldots \rangle \\ \textsc{head} & \begin{bmatrix} \textsc{pos} & v \end{bmatrix} \end{bmatrix}$$

These will allow structures like (27), but not structures where there is a preposition with a different PFORM value

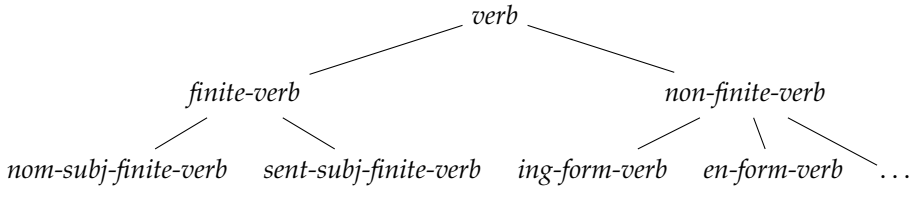
(27)



The obvious problem with this sort of approach is that it seems to lead to massive redundancy – it looks as though *every* finite verb will have to be individually specified as taking a nominative subject, and this is missing a generalisation.

But this is not a real problem.

Probably the simplest approach is assume an inheritance hierarchy of types.



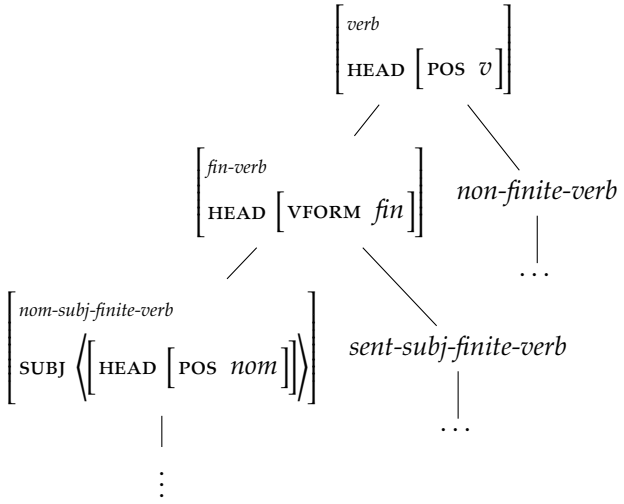
For example:

All verbs are [HEAD | POS *v*], all finite verbs are [HEAD | VFORM *fin*], a subtype of finite verbs are those that have nominative subjects (e.g. others have sentential subjects):

(28) a. $\textit{verb} \Rightarrow \left[\textsc{head}\ \left[\textsc{pos}\ \textit{v}\right]\right]$

b. $\textit{fin-verb} \Rightarrow \left[\textsc{head}\ \left[\textsc{vform}\ \textit{fin}\right]\right]$

c. $\textit{nom-subj-finite-verb} \Rightarrow \left[\textsc{subj}\ \left\langle\left[\textsc{head}\ \left[\textsc{pos}\ \textit{nom}\right]\right]\right\rangle\right]$



Now for a particular verb, we simply have to say at what point(s) in this hierarch it belongs. Other than this all we need to specify for a particular verb is its idiosyncratic properties (e.g. its phonology and its meaning).

This is a simplification, of course, because in general we do not want to classify (e.g.) verbs in only one way. We want to classify them in several dimensions – e.g. by their complementation patterns and their case assignment properties. So in general, we will have *multiple* inheritance.

3 Agreement

3.1 The Phenomenon

> "The term agreement commonly refers to some systematic covariance between a semantic or formal property of one element and a formal property of another." Steele (1978, 610),

(29) a. She is/*are/*am foolish. (Subj-Verb)
b. She is a fool/*fools. (predicate nominals)
c. These cats/*cat chased a *mice/mouse. (Det-N)
d. He$_i$ came in, and he$_i$/*she$_i$/*they$_i$ sat down. (Pronoun-Antecedent)
e. She admires herself/*themsleves. (Reflexive Pronoun-Antecedent)
f. This is [the man who/*which I saw] (Relative Pronoun)
g. What are those? (pointing to trousers) (?pragmatic?)
h. What is that? (pointing to a shirt)

Nowadays it is common to distinguish three different kinds of agreement (Pollard and Sag (1994),

Kathol (1999a), Wechsler and Zlatić (2003)):

- 'concord' (e.g. case, number, and gender agreement of Det and N)
- 'index' agreement (e.g. person, number, gender agreement of Subj and Verb, or pronoun and antecedent)
- 'pragmatic' (not really grammatical, e.g. honorific agreement)

3.2 Approaches

- Purely semantic approaches Dowty and Jacobson (1989, e.g.);
- Derivational 'full-feature' Approaches
- Constraint Based Approaches ('Unification': LFG, HPSG)

3.2.1 Purely Semantic Approach

Dowty and Jacobson (1989) suggest a purely semantic account.

No mechanisms of linguistic theory at all are used to link controller and agreeing form. They are just interpreted separately, and agreement failure will lead to semantic strangeness. The property denoted by *is foolish/a fool* can only be apply to singular entities (not collections), *washed herself* can only be predicated of an NP that denotes a female.

This will not cover all forms of agreement (see below).

3.2.2 Derivational Approaches

- agreement features are inherent in the agreement controller, . . .
- . . . where they are fully specified,
- they are copied or otherwise moved to the agreement target.

A *fully specified* bundle of (morpho-syntactic) agreement features is copied from agreement controller to target.

This leads to redundancy in the lexicon, and missed generalizations.

In French, adjectives agree with subjects in number and gender:

(30) Il est heureux / *heureuse
He is happy (masc) / happy (fem)

This means that *Je* must have two lexical entries:

(31) a. Je suis heureux (masc)
b. Je suis heureuse (fem)

Cf. also some languages will require multiple empty pronouns to trigger different kinds of agreement.

This will lead to a huge, unmotivated, explosion in languages where agreement controllers do not vary in form. Pollard and Sag (1994, 62ff)

3.3 Constraint Based Approaches

3.3.1 In General

Constraint based approaches rely on the possibilities of underspecification and unification: it does not matter where the information originates, if it stated that the agreement features on a noun and its determiner are the same, then they will become so.

In some cases it will seem as if information is moving from controller to target, in others it will seem to be the reverse, but this appearance is an artifact.

(32) a. The salmon swims.

b. The salmon swim.
c. These salmon swim.
d. This salmon swims.

(33) a. The salmon [$_S$ that had been . . .] was . . .
b. The salmon [$_S$ that was . . .] had been . . .

(e.g. Sag and Wasow (1999), Bender et al. (2003))

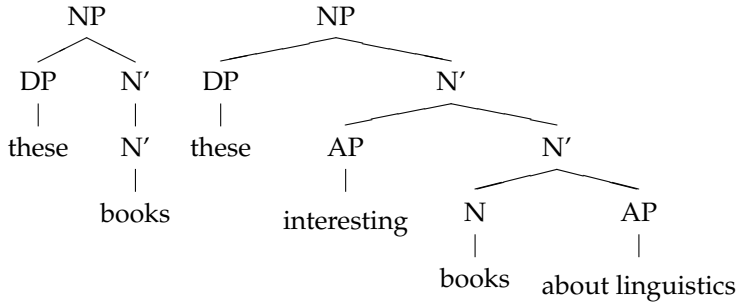
3.3.2 Determiner-Noun Agreement

In English, determiners and nouns agree for number:

(34) a. this book/*books
b. these books/*book

In many languages determiners and nouns (and/or adjectives) show agreement for gender and case (case concord). In (35), both *these* and *books* are plural.

(35)



To handle this, all we have to do is ensure is:

- that NUMBER is shared between a head and its mother (i.e. it is head feature); and
- that singular Ns select singular specifiers (similarly for plural).

That is, that nouns are required to have the following specification:

(36)
$$\begin{bmatrix} \textsc{head} \begin{bmatrix} \textsc{pos} & \textit{noun} \\ \textsc{num} & \boxed{1} \end{bmatrix} \\ \textsc{spr} \quad \left\langle \left[\textsc{head} \begin{bmatrix} \textsc{pos} & \textit{det} \\ \textsc{num} & \boxed{1} \end{bmatrix} \right] \right\rangle \end{bmatrix}$$

(Alternatively, we could assume that that Ns and their specifiers share NUMBER, but using SPR selection is more consistent with what we have assumed so far).

In other languages GENDER and CASE behave in the same way, it makes sense to group these features together as AGR values. Determiners do not show this behaviour in English, but we can see that some nouns (specifically pronouns) can have different values for number, gender, person, and case: cf *you* vs *he* vs *she*, *him*, *they*.

(37)
$$\begin{bmatrix} \textsc{head} \mid \textsc{agr} \begin{bmatrix} \textsc{num} & \textit{num} \\ \textsc{per} & \textit{per} \\ \textsc{gend} & \textit{gend} \\ \textsc{case} & \textit{case} \end{bmatrix} \end{bmatrix}$$

(38) $\text{XP} \rightarrow \text{SPR}\left[\textsc{head} \mid \textsc{agr} \ \boxed{1}\right], \quad \text{H}\left[\textsc{spr} \left\langle \left[\textsc{head} \mid \textsc{agr} \ \boxed{1}\right] \right\rangle\right]$

In English, the possible values of NUMBER are *sg* and *pl*. Notice that determiners that show no variation in number can be underspecified (we make sure *sg* and *pl* have a common supertype):

(39) *number*

sg *pl*

So, e.g. *no* (*no baby/babies*), *the* (*the baby/babies*), *whose* (*whose baby/babies*) can be (under) specified as just [HEAD|AGR|NUM *num*]

We will also want to specify the possible subtypes of *per*, *gender* and *case*:

(40) *person*: 1st 2nd 3rd; *gender*: masc fem neut; *case*: nom acc

Notice how this gives us just one lexical entry for e.g. *you*, rather than 8 (2*2*3 for each of the values of number, case, and gender):

(41)
$$\begin{bmatrix} \text{PHON} & \textit{/you/} \\ \text{HEAD} & \begin{bmatrix} \text{POS} & \textit{noun} \\ \text{AGR} & \begin{bmatrix} \text{PERSON} & 2 \\ \text{NUMBER} & \textit{number} \\ \text{GENDER} & \textit{gender} \\ \text{CASE} & \textit{case} \end{bmatrix} \end{bmatrix} \\ \text{SPR} & \langle\,\rangle \\ \text{COMPS} & \langle\,\rangle \end{bmatrix} \begin{bmatrix} \text{PHON} & \textit{/you/} \\ \text{HEAD} & \begin{bmatrix} \text{POS} & \textit{noun} \\ \text{AGR} & \begin{bmatrix} \text{PERSON} & 2 \end{bmatrix} \end{bmatrix} \\ \text{SPR} & \langle\,\rangle \\ \text{COMPS} & \langle\,\rangle \end{bmatrix}$$

(42)
$$\begin{bmatrix} \text{PHON} & \textit{/baby/} \\ \text{HEAD} & \begin{bmatrix} \text{POS} & \textit{noun} \\ \text{AGR} & \boxed{1}\begin{bmatrix} \text{NUM} & \textit{pl} \end{bmatrix} \end{bmatrix} \\ \text{SPR} & \left\langle \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{POS} & \textit{det} \\ \text{AGR} & \boxed{1} \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{COMPS} & \langle\,\rangle \end{bmatrix} \begin{bmatrix} \text{PHON} & \textit{/baby/} \\ \text{HEAD} & \begin{bmatrix} \text{POS} & \textit{noun} \\ \text{AGR} & \boxed{1}\begin{bmatrix} \text{NUM} & \textit{sg} \end{bmatrix} \end{bmatrix} \\ \text{SPR} & \left\langle \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{POS} & \textit{det} \\ \text{AGR} & \boxed{1} \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{COMPS} & \langle\,\rangle \end{bmatrix}$$

3.3.3 Subject Verb Agreement

We could extend this directly to subject-verb agreement (in fact, the rule we gave above actually applies to any spr-verb combination, hence to subj-verb combinations), but it is not clear that this is correct.

One reason is that, if we assume that CASE is an AGR feature, then we will be lead to assume that CASE is also a feature on verbs. A second reason is that what seems to matter for subject verb agreement is not the formal properties of the subject, but the interpretation.

Problems for Syntactic Accounts: Agreement Mismatches

- Reference Transfer
  - (43) a. The hash-browns at table five *is/are getting cold.
    b. The hash-browns at table five is/*are getting angry.
- Wh-*ever* Constructions: agreement is with the referent of the NP ('singular'), not the NP itself (which seems to be plural – cf the head N; in fact (44a) and (44b) seem similar, perhaps because the subject NPs have similar interpretations):
  - (44) a. ((Whoever's dogs) are running around) is/*are in trouble.
    b. Anyone whose dogs are running around is/*are in trouble.
- Relative Pronouns (±HUMAN)
  - (45) a. The soldiers *who* were trained at Sandhurst . . .
    b. The soldiers *which* were made of lead . . .
- Singular Plurals

(46) Steak and chips appeals/?appeal to me.

- Collectives
  (47) a. The committee is/are settling the issue to its/their satisfaction.
  b. *A new committee have been set up.

(There is dialect variation between British and American English here.)

This sort of consideration lead Pollard and Sag (1994) to treat subj-verb agreement as involving semantic indices. Semantic indices correspond to logical variables (x in 'For all x if x is a man, then x is mortal') or discourse markers – they are means by which we can keep track of what noun phrases 'refer' to (speaking loosely, NPs with the same index 'refer to' the same entity).

Pollard and Sag (1994) assume that indices are not atomic. Instead, they assume they are feature structures, with PERSON, NUMBER, and GENDER values.

The lexical entry for *baby* will then be as in (48); *babies* will be the same except with *pl* in place of *sg*:

(48)
$$\begin{bmatrix} \textsc{phon} & \textit{/baby/} \\ \textsc{head} & \begin{bmatrix} \textsc{pos} & \textit{noun} \\ \textsc{agr} & \begin{bmatrix} \textsc{num} & \textit{sg} \end{bmatrix} \end{bmatrix} \\ \textsc{spr} & \left\langle \begin{bmatrix} \textsc{head} & \begin{bmatrix} \textsc{pos} & \textit{det} \\ \textsc{agr} & \begin{bmatrix} \textsc{num} & \textit{sg} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \\ \textsc{comps} & \langle\,\rangle \\ \textsc{sem} & \begin{bmatrix} \textsc{index} & \begin{bmatrix} \textsc{num} & \textit{sg} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The semantics of *she*, will involve a value like the following:

(49)
$$\begin{bmatrix} \textsc{sem}\,|\,\textsc{index} & \boxed{1} \begin{bmatrix} \textit{index} & \\ \textsc{per} & \textit{3rd} \\ \textsc{num} & \textit{sing} \\ \textsc{gen} & \textit{fem} \end{bmatrix} \end{bmatrix}$$

To say that a verb is third person singular (like *swims*) is then just to say it selects a subject (SPR) whose index is third person singular:

(50)
$$\begin{bmatrix} \textsc{phon} & \textit{/swims/} \\ \textsc{head} & \begin{bmatrix} \textsc{pos} & \textit{verb} \\ \textsc{agr} & \begin{bmatrix} \textsc{num} & \textit{sg} \end{bmatrix} \end{bmatrix} \\ \textsc{spr} & \left\langle \begin{bmatrix} \textsc{head}\,|\,\textsc{pos} & \textit{noun} \\ \textsc{sem}\,|\,\textsc{index}\,|\,\textsc{num} & \textit{pl} \end{bmatrix} \right\rangle \\ \textsc{comps} & \langle\,\rangle \end{bmatrix}$$

This will allow (51a) and exclude (51b):

(51) a. No baby swims.
b. *All babies swims.

3.3.4 Pronoun-Antecedent agreement

Pronoun Antecedent agreement occurs because binding involves identity of indices (co-indexation), hence person, number and gender features.

(52) a. She$_i$ loves her$_i$/*his$_i$ mother
b. Every girl$_i$ loves her$_i$ mother.
c. John$_i$ loves her$_i$ mother.

(Example (52b) indicates why it is misleading to talk about co-*reference* in relation to pronouns and their antecedents).

3.4 Implications

This gives us an account of what happens when a collective noun that allows either singular or plural verb agreement appears with a determiner.

(53) a. This team are not going to win anything this year.
b. *These team are not going to win anything this year.

(AGR and SEM|IND values can diverge, subj-verb agreement depends on semantics ('mode of individuation'), NP internal agreement involves AGR).

It also gives an account 'hybrid nouns' (Corbett, 1991), which can trigger different kinds of agreement on different targets within the same clause.

Example: in languages like Spanish and French, if a title like 'Majesty' refers to a male, it triggers

- masculine agreement on a predicative adjective, but
- feminine agreement on an attributive adjective.

(54) Su Majestad$_i$ Suprema esta contento. (Éli$_i$ ...)
His Majesty Supreme$_{fem}$ is happy$_{masc}$. (He$_{masc}$ ...)

3.5 Questions

There are quite a lot of interesting issues.

- Coordination and agreement, e.g. agreement of verb and coordinate subjects, closest conjunct agreement vs resolved agreement.
- The section of Huddleston and Pullum (2002) on agreement has many problematic examples.

References

Emily M. Bender, Ivan A. Sag, and Thomas Wasow. *Instructor's Manual for* Syntactic Theory: A Formal Introduction. CSLI Publications, Stanford, CA, 2003. URL `http://hpsg.stanford.edu/book/`.

Greville G. Corbett. *Gender*. Cambridge University Press, Cambridge, UK, 1991.

David Dowty and Pauline Jacobson. Agreement as a semantic phenomenon. In Joyce Powers and K de Jong, editors, *Proceedings of Fifth Eastern States Conference on Linguistics (ECSOL '88)*, pages 95–101, Ohio State University, 1989.

G. Gazdar, E. Klein, and G.K. Pullum, editors. *Order Concord and Constituency*. Foris Pub., Dordrecht, 1983.

Rodney Huddleston and Geoffrey K Pullum, editors. *The Cambridge grammar of the English language*. Cambridge University Press, Cambridge, 2002.

Andreas Kathol. Agreement and the syntax-morphology interface in HPSG. In Robert D. Levine and Georgia Green, editors, *Studies in Contemporary Phrase Structure Grammar*, pages 209–260. Cambridge University Press, Cambridge and New York, 1999a.

Andreas Kathol. Agreement and the syntax-morphology interface in HPSG. In Robert Levine and Georgia Green, editors, *Studies in Current Phrase Structure Grammar*, pages 223–274. Cambridge University Press, 1999b.

Jong-Bok Kim and Peter Sells. *English Syntax: An Introduction*. CSLI Pub., Stanford, Ca., 2008.

Carl J. Pollard and Ivan A. Sag. An information-based theory of agreement. In D. Brentari and al, editors, *Papers on the Parasession on Agreement from the 24th Regional Meeting of the Chicago Linguistic Society*, volume 24, Chicago, Illinois, 1988. CLS, CLS. Also published in slightly revised form as CSLI Report 88-132. Stanford: Center for the Study of Language and Information.

Carl J. Pollard and Ivan A. Sag. *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago, 1994.

Ivan A Sag and Thomas Wasow. *Syntactic Theory: a Formal Introduction*. Number 92 in CSLI Lecture Notes. CSLI Publications, 1999.

Susan Steele. Word order variation: a typological study. In Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik, editors, *Universals of Human Language: IV: Syntax*, pages 585–623. Stanford University Press, Stanford, 1978.

Stephen Wechsler and Larisa Zlatić. Case realization and identity. *Lingua*, pages 539–560, 2001.

Stephen Wechsler and Larisa Zlatić. *The Many Faces of Agreement*. CSLI Publications, Stanford, CA, 2003.