Forced asymmetry and the role of features in language change

Elly van Gelderen

DGfS AG3, 6 March 2015

Chomsky’s latest linguistic work has problematized projection and labeling. With the earlier Phrase Structure Grammar and X’-bar theory, it is taken for granted that a phrase contains a specifier, head, and complement; the current work only assumes a labeling algorithm to meet requirements of the conceptual-intentional interface. Such an algorithm automatically rules out particular configurations, for instance, ones with an XP and YP as sisters, and this avoids the need for EPP features as a trigger for movement to the Spec TP. However, other features continue to play a role in the labelling because \{XP, YP\} can be labelled if they share a feature. In this talk, I point out how regular patterns of language change can be seen as resolutions to the labeling paradox and how they themselves can possibly shed light on the precise nature of this algorithm and the role of features therein. I focus on the change where specifiers are reanalyzed as heads.

Keywords: agreement, feature, head, label, negative, pronoun, specifier

1 Introduction

Early Phrase Structure Grammar (e.g. Chomsky 1965) and X’-bar theory (e.g. Jackendoff 1977) take for granted that a phrase is headed and expands to a maximal projection with a specifier, head, and complement. This X’-schema is seen by many as perhaps one of the greatest insights into syntactic structure. The spirit of the current Minimalist Program (Chomsky 1995 through the present), however, is to attribute as little as possible to the computation, restricting it to simple merge with a labeling algorithm needed for the conceptual-intentional interface. This labeling algorithm automatically rules out certain configurations, for instance, ones with a specifier and thereby minimizes “syntactically encoded featural triggers” (Ott & Šimík 2015) such as the EPP.

In this talk, I show how regular patterns of language change shed light on the precise nature of the labeling algorithm proposed in Chomsky (2013; 2014). Section 2 provides some background on phrase structure, in particular the recent problematization of projection and labeling. Section 3 considers the change from phrase to head in negatives, wh-elements, relative pronouns, and subject pronouns as an elimination of the labeling paradox and section 4 is a conclusion.
2 From phrase structure to the labeling algorithm

In early Generative Grammar (e.g. Chomsky 1965: 85), language-specific phrase structure rules, such as (1), are responsible for generating sentence structure. (1a) generates the basic sentence and (1b) the Verb Phrase. Chomsky (1970) and, especially, Jackendoff (1977: 17) reformulate these rules as a category-independent and language-independent schema, as given in (2).

\[
\begin{align*}
(1) & \quad a. \quad S & \rightarrow & \text{NP} & \text{VP} \\
& \quad b. \quad \text{VP} & \rightarrow & \text{V} & \text{NP} \\
(2) & \quad a. \quad \text{XP} & \rightarrow & \text{YP} & \text{X'} \\
& \quad b. \quad \text{X'} & \rightarrow & \text{X} & \text{ZP}
\end{align*}
\]

In the mid-1980s, the X’-schema of (2) is extended to grammatical categories, such as T, C, and D, and the result is the familiar structure in (3), again with the head determining the label of the higher phrase. Relevant to the current workshop is that the T in English has EPP features that trigger movement of a DP to Spec TP.

\[
\begin{array}{c}
\text{CP} \\
\quad \text{Spec} & \text{C'} \\
\quad \text{C} & \text{TP} \\
\quad \text{Spec} & \text{T'} \\
\text{John} & \text{vP} \\
\text{may} & \text{John eat apples}
\end{array}
\]

Taking the Minimalist Program seriously means attributing less and less to Universal Grammar, in particular to rules such as (2), and restricting the generative part
of a derivation to a computational operation called Merge. External Merge (EM) takes two objects and yields an unordered set \{X, Y\} without a label (Chomsky 2013: 42); Internal Merge (IM) takes an already formed syntactic object and takes part of that and merges it with the original syntactic object. Labeling the set is not part of Merge and should therefore be avoided and left to a requirement of the interface. The labeling algorithm (LA), stated in (4a), involves just a minimal search and “must take place at the phase level, as part of the Transfer operation” (Chomsky 2014: 4). It is like Agree, not Match, and part of Minimal Computation, i.e. a third factor effect. Rizzi (2014: 12) formulates it slightly differently, as in (4b).

(4) a. The **Labeling Algorithm** is “a special case of minimal search” seeking “heads H within its search domain” (Chomsky 2014: 4).


b. **Labeling Algorithm**: The category created by Merge receives the label of the closest head. Labelling must be complete at the interfaces. (Rizzi 2014: 12)

There are three potential sets in need of labels, namely \{X, YP\}, \{XP, YP\}, and \{X, Y\}. The first case is unproblematic - Chomsky says “trivial” - because the LA selects the head X. The other two are “interesting” because there is no unambiguous label. Subjects in English exemplify \{XP, YP\} and the resolution to their labelling, IM, forces movement without having to rely on EPP features, a desired consequence. Thus, in (5), a label cannot be found between XP and YP: both X and Y are as accessible to minimal search and therefore as appropriate as labels.

(5) \[ \text{vP} \]
\[ \text{DP (}=\text{XP}) \quad \text{v’ (}=\text{YP}) \]
\[ \text{D (}=\text{X}) \quad \text{v (}=\text{Y}) \quad \ldots \]

Chomsky (2013: 43) provides two solutions to labeling problems such as these: “There are, then, two ways in which [syntactic object] SO can be labeled: (A) modify SO
so that there is only one visible head, or (B) X and Y are identical in a relevant respect, providing the same label, which can be taken as the label of the SO. These are the two cases that are prominently found”.

Solution (A) applied in (5): the DP must move after which the vP can be labelled v. Other examples where the {XP, YP} set can be modified, i.e. case (A), is through movement of one of the maximal projections, as in successive cyclic movement of a phrase in a copula clause. Movement of one of the maximal projections in (6) would result in a structure that can be labeled. According to Chomsky (2013: 44), “[t]he intuitive idea is that the lower XP copy [in (6)] is invisible to LA, since it is part of a discontinuous element, so therefore \( \beta \) will receive the label of YP”.

(6) XP copula \{\beta \text{ XP, YP}\} \quad \text{(Chomsky (2013: 44)}

Although \( \beta \) receives a label in (6), as does v in (5), this results in another case of \{XP, YP\}. Assuming the next merge will be a T in (5) and that the copula is in T in (6), the result is the well-known issue that subjects in English face: they are drawn to Spec TP because of EPP features. Instead of EPP-features, labelling requirements in (5) and (6) force DP-movement. The result, given in (7), is an unlabeled \( \alpha \) because the subject internally merges to the T’ resulting in \{XP, YP\}.

(7) \( \alpha[\text{Tom T [ Tom v* read a book]]} \quad \text{(adapted from Chomsky 2014: 6-7)}

Here, solution (B) applies since the heads of these phrases share phi-features, and the set is successfully labeled \( \langle \text{phi, phi} \rangle \).

Certain results of merge cannot be labelled if we just take the categorical heads into consideration. The solution to the labeling problems comes from features that are shared and that then count as a head. This second solution to the labelling problem, i.e. case (B), can be exemplified by means of \textit{wh}-constructions as well, something that used to be called Spec-Head agreement, involving “\{XP, YP\} in situ without raising.” If “the most prominent feature of the \{XP, YP\} set “is shared”, labelling is not a problem. It will be labelled using “the interrogative feature Q, a feature of C and the head of \( \alpha \)” in (8a)
Sharing the Q-features between the PP and C in (8a) has the result that the PP does not move further, as the ungrammatical (8b) shows.

(8) a. They wondered [α in which Texas city [C [JFK was assassinated]]]
   b. *In which city did they wonder JFK was assassinated.

The fact that the wh-element cannot move further from (8a) to (8b) is called the 'halting problem' or 'criterial freezing' in Rizzi (2006; 2014), the basic intuition being that the wh-element included in the PP shares contradicting features: y/n for the embedded C and wh for the main clause.

Labelling resolutions also provide an account for the that-trace effect in (9a): α cannot be labelled by the phase head C if who has moved. When the phase-head C deletes, as in (9b), it transfers phasehood to T and who can remain in Spec TP until it is moved in the next phase.

(9) a. * [γ Who do you v* [ε think [δ C that [α t T read the book]]]]
   b. [γ Who do you v* [ε think [δ C [α t T read the book]]]]

Apart from the problems to label {XP, YP}, the set {X, Y} is problematic. Here Chomsky (2013: 47) says that this applies when one of the heads is a root and the other a functional element determining its category. If roots don’t count as labels, no problem arises. Chomsky (2014: 9) mentions another case of head-movement, namely to T and v* and here “T [is] affixed to V. More generally, the conventional theory of head-raising seems to have the story backwards: the host should be affixed to the raised element” so these are not cases of {X, Y} because “the affix is invisible to the labeling algorithm”. See Carstens, Hornstein, & Seely (2013) as well.

Labeling paradoxes can be resolved by having one of the XPs move, as in (5) and (6), or by ignoring one label (the root), or by feature-sharing in (7) and (8a). The first two solutions are worked out in Chomsky (2013), whereas the latter is the focus of Chomsky (2014). I now turn to some linguistic changes that may be accounted for by the
requirements of the labeling algorithm and I will account for why some change does not occur because it doesn’t present a problem for the labeling algorithm.

3 Specifier to Head in Language Change

The Merge of \{XP, YP\} results in problems for the labeling algorithm at the interface level and can be resolved in the ways we have seen above, namely by movement, ignoring the label, and by feature-sharing. There are other mechanisms that could resolve labeling problems, namely the change from phrase to head where the XP in \{XP, YP\} is reanalyzed as head and the change from (non-agreeing phase) head to affix. The first change would eliminate one of the offending phrases without needing movement or having to resort to feature-sharing. One would therefore expect that changes towards the form \{X, YP\} would be common. This is in fact the case, as shown in van Gelderen (2004: 2011), Jäger (2005; 2010), Weiß (2007), Willis (2007), Bayer & Brandner (2008), Bácskai-Atkári & Dekány (2014), and other work. The change to affix is frequent in grammaticalization but has not been explored in a Minimalist framework and I will just have a few things to say in the conclusion.

In section 3.1, I will mention some instances of phrase to head change, some interesting in that they don’t involve the typical phase heads D, C/T, and v* but also Neg. Then, in section 3.2, I turn to cases where the phrase does not change to a head and these will straightforwardly be accounted for because they do not involve a labeling paradox. Section 3.3 looks at relatives where change to a head is frequent. Finally, in section 3.4, I turn to pronominal subjects, which are claimed by Chomsky (2013: 46) not to be heads. Here, I will propose an approach based on features: certain subject pronouns are indistinguishable in their features from those of T and then resolve the labelling paradox by pointing to T, rather than D. This again results in a preference of head-like subjects over phrasal ones.

3.1 Phrasal negative to negative head

In this part, I examine changes involving the NegP and will argue that these can be accounted for by labeling needs. In Chomsky (2013; 2014), only the labelling of CP, TP,
v*P, and R(oott)P has been considered and therefore showing that phrase to head occurs
NegPs may make us aware of labeling needs of other phrases.

Jespersen (1917) was among the first to discuss changes in negatives with
examples from many languages and talks about weakening and strengthening tendencies.
Willis, Lucas, and Breitbarth (2013: 7, 21, 169) provide recent case studies on the well-
known cyclical change in negative marking. A typical chain of changes is given in (10),
from the history of English.

\[(10) \quad \text{earlyOE} > \quad \text{OE/ME} > \quad \text{earlyModE} > \quad \text{ColloqEnglish} \]

\[\quad \text{no/ne} \quad \quad \quad \quad \text{(ne)} \quad \quad \text{not} \quad \quad \quad \quad \quad \quad -n’t \quad \quad \quad -n’t \ldots \quad \text{nothing} \]

Examples of the stages are given in (11)\textsuperscript{1}: (11a) shows the use of a negative by itself,
(11b) of the determiner *nan* and the regular negative *ne*; (11c) shows a *ne* contracted with
the verb into *nes* and a negative adverbial *nawhit*, a contracted form of the negative
indefinite *na wiht* ‘no creature’, (11d) shows the adverbial *not* by itself, and (11e) shows
the cliticization of *not* onto the verb, and (11f) the reinforcement in colloquial English.

\[(11) \quad \text{a. } \quad \text{Men ne cunnun secgan to soðe ... hwa} \]
\[\text{Man not could tell to truth ... who} \]
\[\quad \text{‘No man can tell for certain ... who’. (Beowulf 50-2, Klaeber edition)} \]

\[\text{b. } \quad \text{ne fand þær nan þing buton ealde weallas & wilde wuda} \]
\[\text{not found there no thing except old walls and wild woods} \]
\[\quad \text{‘He found there nothing but old walls and wild woods’.} \]
\[\text{(Peterborough Chronicle, addition to year 963, Thorpe edition 220)} \]

\[\text{c. } \quad \text{Nes þis meiden nawhit heruore imenget in hire mod inwið.} \]
\[\text{not.was this maiden not herefo re troubled in her mind within} \]
\[\quad \text{‘This maiden was not troubled in her mind because of this.’} \]
\[\text{(Katerine, d’Ardenne edition 28, 21-22)} \]

\[\text{d. } \quad \text{Yit it semeth that He wolde not leve thee thus lightly} \]
\[\text{‘Yet it seems that he wanted not leave you thus lightly.’} \]

\[\text{\textsuperscript{1} The data are well-known and I have just used examples from well-known primary sources.} \]
(Cloud of Unknowing, 241-42)

e. And to þis I cannot answer þee bot þus: ‘I wote neuer.’
   ‘And to this I can’t answer you except thusly: I knew never.’
   (Cloud of Unknowing, 450-51,
   www.lib.rochester.edu/camelot/teams/cloud.htm)

f. I can’t do nothing for you either, Billy.
   (Ken Kesey, One flew over the Cuckoo’s Nest  p. 118)

The phenomenon shown in (11) is called the negative cycle. The motivation for it is most often seen as pragmatically driven. Thus, Kiparsky & Condoravdi (2006), in examining Jespersen’s Cycle in Greek, find no evidence for phonetic weakening and similarly suggest pragmatic and semantic reasons. A simple negative cannot be emphatic; in order for a negative to be emphatic, it needs to be reinforced, e.g. by a minimizer. Adapting ideas from Dahl (2001), they argue that, when emphatic negatives are overused, their semantic impact weakens and they become the regular negative and a new emphatic will appear. L’Arrivée (2010), examining the history of French negation, argues that a specific pragmatic function, namely accessibility of a proposition to the hearer, plays a role.

I will argue that the syntactic labelling mechanism favors the negative as head, e.g. just the ne or –n’t, but that the need to make the negative meaning obvious necessitates renewal in the form of an additional negative indefinite, which in turn is made into a head. This accounts for the cycle. The negative ne in (11ab) is a head because it precedes the verb and the additional negative DP in (11b) is an argument that is in negative concord. In (11c), the determiner and noun are written as one word and are no longer used as argument but as an adverb. This adverbial form is still phrasal and will ultimately result in the form not, as in (11d). In the stage of (11e), the negative is a head because it is joined by another head and therefore similar to (11a). This is followed next by a phrasal negative renewal, as in (11f), similar to (11b).

Van Gelderen (2004: 18) justifies principle (12), to account for these changes from phrase to head seen between (11bc) and (11e) (not in (11d) is ambiguous in status). This principle is at work in the internalized grammar either due to Universal Grammar or
due to general cognitive principles and holds for external merge (projection) as well as internal merge (movement). In accordance with the HPP, a speaker will build (13b) rather than (13a), if given evidence compatible with either.

(12) Head Preference Principle (HPP)
    Be a head, rather than a phrase.

(13) a. NegP  b. NegP
    nothing    Neg’        Neg    YP
                     nought/not
    Neg        YP

The change from (13a) to (13b) is immediately obvious from the labeling mechanisms: *nothing* in (13a) is an XP and Neg’ a YP so, unless they share features, the NegP cannot be labeled. If labeling is behind this change, the features in (13a) must be more opaque and less desirable to the language learner.

(13b) is the stage of (11a); the stage in (11bc) comes about through pragmatic strengthening, not something narrow syntax is responsible for. I’ll assume negative features connected to head and phrasal negation (but for our purposes here will remain agnostic about interpretability/valuation). These negative features allow for the stage in (11bc). However, this stage doesn’t appear stable as (11de) arise, evidence for another phrase to head reanalysis, as in (13). Stage (11f) involves further pragmatic strengthening, allowed under feature-sharing.

The Head Preference Principle of (12) is relevant to a number of historical changes: whenever possible, a word is seen as a head rather than a phrase, and this is expected if such a crucial mechanism as the labeling mechanism is the reason for it. Other examples are given in Table 1 but see the vast literature on grammaticalization (Hopper & Traugott 2003; Heine & Kuteva 2002).
Demonstrative pronoun *that* to complementizer
Negative adverb to negation marker
Adverb to complementizers (e.g. *till*)

| Demonstrative pronoun *that* to complementizer | Demonstrative pronoun to article |
| Negative adverb to negation marker | Adverb to aspect marker |
| Adverb to complementizers (e.g. *till*) | Full pronoun to agreement |

Table 1: Examples of the change from phrase to head (van Gelderen 2011: 14)

This change from phrase to head is slow since a child learning the language will continue to encounter a pronoun or a negative as both a phrase and a head. For instance, coordinated pronouns are phrases as are emphatic pronouns but clitic pronouns are not (see van Gelderen 2011 for more in this). If both remain in the input, phrases will continue to be triggered in the child's grammar.

This subsection has introduced how labelling can be involved in the phrase to head reanalysis. It is frequent in syntactic change and receives an analysis under the labeling paradox as presented in Chomsky (2013; 2014). I now turn an interesting construction where phrases are not reanalyzed as heads.

### 3.2 No need for phrase to head reanalysis

As seen, Chomsky argues that the \{XP, YP\} label in (8a) does not result in a labeling paradox because the interrogative feature is shared between the XP and the Y of the YP (Chomsky 2013: 45). I will now turn to this sharing as an explanation for why certain specifiers are not reanalyzed as heads – something that was hitherto a puzzle. Sentences where *whether* is an XP and the C’ a YP, as in (14), can be labeled because these elements share interrogative features. Hence, there must not be pressure for *whether* to reanalyze as head.

(14) I wonder [ whether [ C [ he’ll do it]]].

Q Q

The Indo European origin of the lexical item *whether* is as a phrasal pronoun, as in (15a). In Old English, it is also a *yes/no* marker, as in (15bc), and a complementizer, as in (15d).
(15)  

a.  \textit{Hwæðer}  \textit{þara}  \textit{twegra dyde}  \textit{þæs}  \textit{fæder}  \textit{willan}

`Who of the two did the father's will?'

(West Saxon Gospel Corpus, Matthew 21.31, Skeat’s edition)

b.  \textit{Hwæðer}  \textit{wille ge ðæt ic cume to eow, ðe mid gierde ðe mid monnðwere gæste?}

Whether will you that I come to you or with rod or with gentle spirit

`Do you want that I come to you, with a rod or with gentleness of spirit?'

(Alfred, Pastoral Care, Sweet’s edition 117.7-8).

c.  \textit{Hwæðer}  \textit{ic mote lybban oðdæt ic hine geseo}

Whether I might live until I see him

`Might I live until I see him?'


d.  \textit{þær se snotera bad} \textit{hwæþer} \textit{him alwalda}  \textit{æfre wille} ... \textit{wyrpe gefremman.}

there the wise waited whether he almighty ever would ... change accomplish

`There the wise one waited whether the almighty would ever grant him change'

(Beowulf 1313-5).

The original pronoun has a meaning of `who of the two’ and this meaning is reanalyzed as \textit{yes/no} marker and as interrogative complementizer. That reanalysis from a position in the VP to one higher in the CP happens before the Old English period. The \textit{yes/no} marking function appears in a specifier position in (15b) and in a head position in (15c) before it disappears between the 15th and 18th centuries (see OED). \textit{Whether} as \textit{yes/no} marker is probably lost because there is a good alternative in Subject-Auxiliary Inversion. This auxiliary is a head and therefore no labelling problems arise.

What is interesting is that the complementizer function of \textit{whether} remains in use up to the present but that it is in a specifier and not in a head position. It has never shown any inclination of becoming a head, as the impossibility of \textit{wh}-extraction in (16) shows and the co-occurrence of it and a head in (17).\textsuperscript{2}

\textsuperscript{2} For many speakers of (American) English, \textit{if} is a head since a \textit{wh}-element can be extracted, as in (i). I have not considered the history of \textit{if} and if it became a head.

(i)  \textit{What did I wonder if he’ll do?}
(16)  *What did you wonder whether he’ll do what?

(17)  
a. And it's a very good question about **whether if** you think that Barack Obama is an agent of change ... (COCA Spoken 2008)

b. I don't know **whether if** you think there's going to be a nuclear fallout and ... (BNC F7L)

Construction (17) occurs 56 times in the Corpus of Historical American English (COHA), 50 times in the Corpus of Contemporary American English (COCA), and 14 times in the British National Corpus (BNC). These and (16) show that *whether* is still a specifier of CP. So, to what can we ascribe *whether* not being reanalyzed? If *whether* in (14) and (15d) is like the PP in (8) and has transparent interrogative features that it shares with the C, there is no labeling conflict for the two XPs. This feature sharing therefore provides an account for why *whether*, unlike other phrases, is not reanalyzed as head.

3.3 Relative pronouns

Rizzi (2014: 4) mentions Q, R, TOP, and FOC as criterial features, i.e. those that halt further movement. The criterial head carries instructions for the interfaces about the correct label. Van Gelderen (2011) provides data about changes in relative clauses from specifier to head and, in this section, I’ll see what’s going on with the features. English is well-known for having either a wh-element or a *that* or nothing as relative marker, with the former in specifier and the latter in head position. Thus (18a) is ungrammatical, but (18bcd) fine. I add the optional preposition to show that the wh-element is indeed a phrase.

(18)  
a. *The book (about) which that I read.

b. The book (about) which I read.

c. The book that I read.

d. The book I read.
In Old English, both the specifier and head can be occupied, e.g., as in (19), which means they share features. In Middle English, this shifts to using just the head, as in (20), which is expected under the Labelling Algorithm.

(19)  
\[ \text{ðonne cymeð se man se þæt swiftoste hors hafað} \]
then comes the man that that fastest horse has

`Then comes the man who has the fastest horse.' (Orosius, 17.22, Bately edition)

(20)  
\[ \text{and suggeð feole þinges ... þat næuere nes i-wurðen} \]
and say many things that never not-was happened

`and say many things that never happened.'

(Layamon, Caligula 11472-3, Brook & Leslie edition)

What do the features that need to be shared in a criterial position, such as in (19), entail? I’ll speculate they are phi-features and that a major shift in the feature system around 1200 made the person features of the demonstrative less transparent. Demonstratives disappear not only as relative markers but also as referring to humans. Instead, he, she, and they start to be used that way and articles appear instead of demonstratives (see van Gelderen 2013). That loss of phi-features in the demonstrative may be responsible for the demise of (19).

Since the head in (20) is perhaps perceived as too bland for a construction that is often used in highly formal contexts, there is then an introduction of the wh element as specifier. Many have observed that the wh form is extended from interrogative to relative on the basis of Latin and French. Steinki (1932: 43) argues that the wh forms are introduced consciously.

(21)  
\[ \text{wyche schalle be on Wedynsday next, be þe grace of God, who preserue 3ow.} \]

`which shall be on next Wednesday by the grace of God, who keeps you.'

(Paston Letters 395, Davis p. 636)
3.4 Subject pronouns

As mentioned in section 2, Chomsky (2013: 44) accounts for the EPP, i.e. the fact that vP-internal subjects move to Spec TP, in a structural way. If the subject (the EA) remains in the Spec of vP, as in (22), this subject and the rest of the vP, β in (22), cannot be labeled. Therefore the subject must move out of the vP. This motivation for movement of the subject is a better explanation than the earlier stipulation that there are EPP features on T that make the subject move.

(22) $T [\beta \text{EA} [v^* [V \text{IA}]]]$ (Chomsky 2013: 44)

The next step after (22) is to (internally) merge the DP and T’ (α in (23)) and, again, these are both non-heads, as (23) shows, and can therefore not be labeled.

(23) $[C [\alpha \text{DP TP}]]$

Chomsky (2013: 45) therefore entertains the following feature-sharing for (23): “α receives its own interpretation and should be labeled -- in accord with the general principle that all SOs that reach the interfaces must be labeled. Perhaps that can be achieved by the device suggested for embedded interrogatives. NP and TP share prominent features, namely [phi]-features -- so-called ‘SPEC-Head agreement,’ known to have a variety of interesting properties”. However, “sharpening this condition requires a closer analysis of Agree, which would carry us too far afield”. He does sharpen the features in Chomsky (2014: 4) where he has a slightly different account of EPP (and ECP), namely one based on weak T. In English (but not Italian), the T is too weak “to serve as a label” but, with an overt subject in what used to be Spec TP, the construction can be labelled $<\text{phi, phi}>$.

A very straightforward escape from the labeling paradox in (23) would be to have a subject that has the status of a head. However, Chomsky (2013: 46) says that (pronoun) subjects cannot be heads because they would label the TP incorrectly, as D-headed, not T-headed. They do participate in feature-sharing. What I will therefore explore in the remainder of this section is the feature-sharing characteristic of the Agree-relationship
and, in the process, I will explain why pronouns change from phonologically fully independent phrases to agreement markers, as has happened in a number of languages, the most well-known case being French perhaps (see Lambrecht 1981; Roberts & Roussou 2003). First, some examples will be given of subject pronouns that have a more elaborate shape than the agreement markers showing that the latter derive from the former and, as they become agreement, they lose certain features. The latter is necessary, I argue, in order to be labelled as phi and therefore be acceptable to the labelling mechanism.

According to Tauli (1958: 99, the Basque verbal prefixes n-, g-, z- are identical to the pronouns ni ‘I’, gu ‘we’, and zu ‘you.’ As early as the 19th century, Proto Indo European verbal endings -mi, si, -ti are considered to arise from first, second, and third person pronouns (e.g. Bopp 1816). Hale (1973: 340) argues that in Pama-Nyungan inflectional markers are derived from independent pronouns: “the source of pronominal clitics in Walbiri is in fact independent pronouns”. Likewise, Mithun (1991) claims that Iroquoian agreement markers derive from Proto-Iroquoian pronouns and Haugen (2008) argues that Nahuatl agreement markers derive from earlier forms. Fuß (2005) and van Gelderen (2011) cite many additional examples. Many of these cases involve first and second person pronouns which I’ll account for now.

As mentioned earlier, Chomsky (2013) is not definite on what the matching features between the subject DP and T in (23) have to be so that the conditions of the labeling algorithm puts on the structure in (23) are met. Chomsky (2014) suggests these are phi, but not more is said. A full DP in English has person and number, marks definiteness, and may contain a deverbal noun with thematic structure, e.g. their painting of pictures. The features an English DP is sharing with T in (23) may be person and number but there are lots of features to make the DP a distinct unit. There are other languages where pronominal and clitic forms have fewer features, e.g. first and second person pronouns, such as those in French (24), are pure person and number markers.

(24)  

*J’ai deux livres sur ma table*  
I have two books on my table  
‘I have two books on my table.’
These person and number features are indistinguishable from those of the agreeing T and therefore label the $\alpha$ containing the DP and TP in (23) correctly as phi-headed. So, in (24), $je$ is a head with first person singular features that are the same as the phi-features of T and (24) is in accordance with the labelling algorithm.

Now, we have a reason for the reanalysis of pronouns to agreement markers. A fully phrasal pronoun (that can be coordinated and modified) cannot be seen as sharing the features of T but a head (that has to be adjacent to a verb) can be seen by the child acquiring French (or English) as similar to T. When the features of a pronoun overlap with those of the agreeing T, they may disappear and a structure as in (25) may be the result. This structure can of course receive a label.

(25)  
\[
\begin{array}{c}
TP \\
\hline
T & v^*P
\end{array}
\]

This account is very similar to accounts such as Roberts (2010) and van Gelderen (2011) who suggest the change from pronoun to agreement marker is due to a confusion as to whether the pronoun actually values the features of T or is itself in need of valuation. Taking Chomsky’s idea of feature-sharing, the preference for subjects that are heads with minimal features similarly makes sense.

If we look at which pronouns are the first ones to grammaticalize into agreement markers, they are typically the first person singular ones. For instance, since Lambrecht (1981), it has been accepted that French weak pronouns such as $je$ ‘I’ and $tu$ ‘you.SG’ are agreement markers on the verb and frequently doubled, as in (26a). What has also been known for a long time is that third person subject pronouns are slower to gain agreement status. To be an agreement marker, they would have to appear obligatorily and that is not the case, certainly with indefinite subject, as in (26b), which are rare with the weak pronoun.

(26)  
a. $\textit{moi, j'écoute tous le temps.}$  
Spoken French
‘Me, I listen all the time.’ (corpus d’entretiens spontanés)

b. si un: un Russe i va en France ...

if a a Russian he goes to France

‘If a Russian goes to France.’ (Fonseca-Greber 2000: 335)

The reason the third person is ‘slow’ is that there are more features to be shared, e.g. gender. Gender and number are in fact deleted when the pronoun becomes the agreement marker, as in (27), where i is marked for only third person singular although les tomates are feminine and plural.

(27) Les tomates, i sont encore vertes

‘The tomatoes, they are still green.’ (Lambrecht 1981: 40)

If we think that DP and TP in (23) must share a feature in order to be labelled and that pronominal features can be bleached so as to indistinguishable from those of T, we have an explanation for the change from pronoun to agreement marker and also for the light feature-load of those agreement markers. This is represented schematically in (28).

(28) \[ \text{DP} \quad \text{TP} \quad > \quad \text{D} \quad \text{TP} \]

\[ \text{ps, #, definite} \quad \text{ps} \]

4 Conclusion

The systematic change from phrase to head is accounted for if such reanalyses resolve labeling paradoxes of the [XP YP] kind by eliminating one of the two XPs. This change is well-attested and, in this paper, I discuss stages where the XP and YP share features to stages where only the head is left. As in Chomsky (2013; 2014), many questions remain unresolved. The main one is what the nature of the features shared between the XP and YP is such that both are legible at the interface. In one case considered in this paper, it is interrogative-features and, in another, phi-features. Future work should focus on these inventories.
The first change discussed involves negatives and here the challenge is the renewal through a specifier making us think about features in the NegP. The second case is that of *whether*, a specifier that shows no inclination of reanalyzing. Relatives show a change from Specifier to head that can be attributed to changes in the features of the demonstrative. The last change is that of subject pronouns reanalyzing as T and present another instance where a head is to be preferred. When the pronoun has features similar to those of T, there is again no labelling paradox, and change is towards that analysis.

References
Bayer, Josef 2012. Doubly-filled Comp, wh head-movement, and derivational economy. ms.
BNC, British National Corpus. Available at http://www.natcorp.ox.ac.uk.
Corpus d’entretiens spontANéS,
Relativization Structures. Linguistic Inquiry 42.4: 519-560.
Dahl, Osten 2001. Inflationary Effects in Language and Elsewhere. In Joan Bybee and
Paul Hopper, eds. Frequency and the Emergence of Linguistic Structure, 471-480.
Philadelphia: Benjamins.
Fonseca-Greber, Bonnibeth 2000. The Change from Pronoun to Clitic and the Rise of
Null Subjects in Spoken Swiss French. University of Arizona Diss.
Gelderen, Elly van 2013. The Diachrony of Pronouns and Demonstratives, In Search of
Universal Grammar: From Old Norse to Zoque, edited by Terje Lohndal, 195-
218. Amsterdam: Benjamins.
Hale, Ken 1973. Person marking in Warlbiri. Stephen Anderson & Paul Kiparsky (eds), A
Heine, Bernd & Tania Kuteva 2002. World Lexicon of Grammaticalization. Cambridge:
Cambridge University Press.
University Press.
24.2: 227-262.
Linguistische Berichte 224: 467-493.
Brian Joseph & A. Ralli (eds.). Proceedings of the 2nd International Conference


Ott, Dennis & Radek Šimík 2015. Alternatives to formal features. DGfS Workshop description.


Notes
In order to label \( \beta \) in the coordinate structure in (1a), either Z or W – both non-heads - must raise and this is shown in (1b).

(7) a. \([\alpha \text{ Conj } [\beta Z W]]\)
    b. \([\gamma Z [\alpha \text{ Conj } [\beta Z W]]]\) \hspace{1cm} \text{(Chomsky (2013: 46)}

In (1b), although \( \beta \) gets the label of W, the resulting \( \gamma \) is another \{XP, YP\} structure, hence unlabelable. Chomsky (2013: 46) accounts for this by arguing that “the label [of \( \gamma \) in (1b)] is not Conj but rather the label of Z, typically shared with W; if the coordinated expressions are APs, then \( \gamma \) is an AP, etc. It follows that Conj and the construction \( \alpha \) that it heads are not available as a label, so that \( \gamma \) receives the label of Z.” So, like the root, the Conj is stipulated not to count as label.

There are other approaches that account for labeling paradoxes, e.g. Ceccetto & Donati (2010) and Donati & Ceccetto (2011) discuss a resolution of the labelling paradox of a phrase in the specifier of the CP in that the structure either becomes a DP (a free relative) or CP (an interrogative).