Formal Properties of Agreeing Minimalist Grammars

Marina Ermolaeva*
Lomonosov Moscow State University
mail@mermolaeva.com

Gregory M. Kobele
University of Leipzig
gkobele@uni-leipzig.de

1 Introduction

Minimalist grammars (Stabler, 1997, MGs) are a formalization of core ideas of the minimalist program (Chomsky, 1995). One glaring discrepancy between minimalist practice and formalization has been agreement. Agree is taken in the linguistic community as a fundamental operation of grammar , whereas in the formal community it has been ignored.

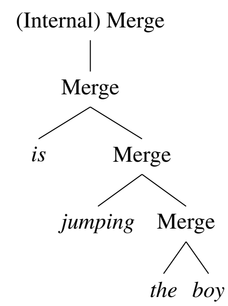
Ermolaeva and Kobele (2022) propose an extension of MGs which allow for agreement to take place between lexical items. Their implementation, in the spirit of post-syntactic approaches to Agree (Bobaljik, 2008), divorces agreement from the syntax proper, and mediates it via annotations on lexical items. A metaphorical perspective views the morphological information as a liquid, derivationally established dependencies between lexical items as channels, and the lexical annotations as sluices controlling the flow of liquid through channels. The formal properties of MGs with agreement have not been investigated. This paper aims to fill that gap.

2 Minimalist grammars

Minimalist grammars define lexical items (LIs) as strings annotated with syntactic feature bundles. We use a version of this formalism that has only one structure-building operation, Merge. It checks matching features of opposite polarities – *positive*, of the form +x, and *negative*, of the form −x. Merge can only target the first unchecked feature in any LI's bundle.

If an expression starts with a positive feature and contains a sub-expression with a matching negative one, the operation applies as internal Merge, or Move. Otherwise it combines two expressions, and the one with the positive feature becomes the head of the new expression. A complete expression has no features other than −t on its head. An example MG is shown in (1), along with the derivation tree of the complete expression *the boy is jumping* generated by it.

(1) *the* :: +n −d −k
boy :: −n
is :: +g +k −t
jumping :: +d −g



3 Agreeing minimalist grammars

To define an agreeing MG, one specifies for each LI a (finite) set of morphological features, which can either have values or be undefined (⊥), and for each of its syntactic features whether it permits incoming and/or outgoing morphological information.

Emitting channels are denoted by $^{\rightarrow}$ and can be annotated with lexically specified morphological features. *Receiving* channels are denoted by $_{\leftarrow}$. In the example lexicon (2), ϕ-features (person and number) and c(ase) values are transmitted across the {+n, −n} and {+k, −k} dependencies, enforcing subject-verb agreement.

(2) $\left[\text{THE}\right] :: +\mathtt{n}^{\rightarrow}_{\leftarrow}\ -\mathtt{d}\ -\mathtt{k}^{\rightarrow}_{\leftarrow}$
$\left[\text{BOY}, \begin{smallmatrix}\phi\text{:3s}\\ \text{c:}\bot\end{smallmatrix}\right] :: -\mathtt{n}^{\rightarrow\phi\text{:3s}}_{\leftarrow}$
$\left[\text{BOY}, \begin{smallmatrix}\phi\text{:3p}\\ \text{c:}\bot\end{smallmatrix}\right] :: -\mathtt{n}^{\rightarrow\phi\text{:3p}}_{\leftarrow}$
$\left[\text{I}, \begin{smallmatrix}\phi\text{:1s}\\ \text{c:}\bot\end{smallmatrix}\right] :: -\mathtt{d}\ -\mathtt{k}^{\rightarrow\phi\text{:1s}}_{\leftarrow}$
$\left[\text{THEY}, \begin{smallmatrix}\phi\text{:3p}\\ \text{c:}\bot\end{smallmatrix}\right] :: -\mathtt{d}\ -\mathtt{k}^{\rightarrow\phi\text{:3p}}_{\leftarrow}$
$\left[\text{BE}, \phi\text{:}\bot\right] :: +\mathtt{g}\ +\mathtt{k}^{\rightarrow\text{c:nom}}_{\leftarrow}\ -\mathtt{t}$
$\left[\text{JUMPING}\right] :: +\mathtt{d}\ -\mathtt{g}$

*This author was supported by the Fellowship from Noncommercial Foundation for the Advancement of Science and Education INTELLECT.

Proceedings of the Society for Computation in Linguistics (SCiL) 2023, pages 385-388.
Amherst, Massachusetts, June 15-17, 2023

Morphological equations (3) map sets of semantic and morphological features to strings.

(3)

$[\text{I}, \text{c:nom}] = I$	$[\text{BE}, \phi\text{:1s}] = am$
$[\text{THEY}, \text{c:nom}] = they$	$[\text{BE}, \phi\text{:3s}] = is$
$[\text{BOY}, {}^{\phi:3s}_{\text{c:nom}}] = boy$	$[\text{BE}, \phi\text{:3p}] = are$
$[\text{BOY}, {}^{\phi:3p}_{\text{c:nom}}] = boys$	$[\text{THE}] = the$
$[\text{JUMPING}] = jumping$	

Morphological features pass over channels in the following way:

(4) **Agreement assumptions**

i. Lexically specified features supersede those received from other LIs;

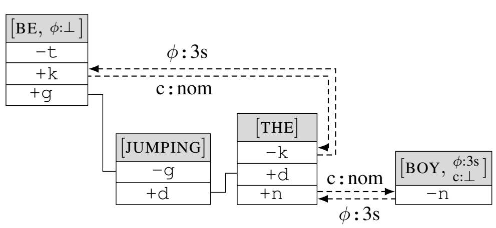
ii. Emitting channels send out the last version of all features received through any receiving channel of the LI, modified by (i) and (iii);

iii. No information is sent back along the syntactic dependency it was received through.

The condition (ii) is the overwriting principle used in (Ermolaeva and Kobele, 2022) to ensure that features received along later channels take priority. Additionally, (i) prevents values specified in the lexicon from being overwritten, and (iii) avoids breaking agreement in cases when a syntactic feature can both emit and receive information.

The syntactic dependencies between the LIs of *the boy is jumping*, and the paths taken by transmitted morphological features, are schematically represented in (5).

(5)



4 Relation to attribute grammars

The MG implementation of agreement can be viewed from the perspective of attribute grammars (Knuth, 1968, AGs), which were originally developed as a means of semantically interpreting context-free grammars. An AG consists of a context-free grammar, where each node (terminal and non-terminal) has a set of *attributes*, the values of which are determined by equations attached to the rules. The derivation trees of any MG are given by a context-free grammar whose non-terminals are tuples of feature bundles (Michaelis, 2001). An alternative but equivalent representation for MG derivations (Kobele, 2012), reminiscent of TAG derivation trees, is more useful for our present purposes. These can be obtained from the graphs above by removing all edges between movement features (at which point they become trees). In general, lexical items will be associated with a family of context-free rules in Greibach normal form, where the lexical item itself is the terminal element. For example, the lexical item BE is associated with a context-free rule in GNF which has the form $A \to be\ C$:

(6) $\langle \texttt{-t} \rangle \to \text{BE}\ \langle \texttt{-g};\ \texttt{-k} \rangle$

This rule mirrors the node BE in the dependency structure of (5), minus the movement edges. In that structure, the node BE has a single dependent, namely the node JUMPING. This node, upon being connected to BE heads a tree whose open features are $\texttt{-g}$ and $\texttt{-k}$. This corresponds to the non-terminal $\langle \texttt{-g}, \texttt{-k} \rangle$ in the rule. After satisfying its positive features, the tree rooted in the node BE in (5) has open features $\texttt{-t}$. This corresponds to the left-hand side non-terminal $\langle \texttt{-t} \rangle$. Rules for the other nodes of this tree are given below.

(7) $\langle \texttt{-g};\ \texttt{-k} \rangle \to \text{JUMPING}\ \langle \texttt{-d -k} \rangle$

(8) $\langle \texttt{-d -k} \rangle \to \text{THE}\ \langle \texttt{-n} \rangle$

(9) $\langle \texttt{-n} \rangle \to \text{BOY}$

To systematize the translation, each node has as attributes n copies of the morphological features of the grammar, one copy for each component of its tuple. The attribute equations for this rule are determined by the channel annotations, with the proviso that the left-hand side and the lexical item on the right-hand side represent the same node of the dependency tree. In (6)—as the lexical feature bundle for BE, namely $\texttt{+g+k}\overset{\to \text{c:nom}}{\leftarrow}\texttt{-t}$, has its annotations on the $\texttt{+k}$ feature—the only source and target for features is the C non-terminal. It must be the source of BE's lexically unspecified ϕ-features, and the target of its lexically specified case feature:[1]

(10)
$$be.\phi_1 := C.\phi_2$$
$$C.\text{c} := \text{nom}$$

[1] The attributes are here written using record notation instead of function notation ($A.f$ instead of $f(A)$).

These equations can be notationally incorporated into the body of the rule itself using co-indexed variables:

(11) $\langle\texttt{-t}\rangle \rightarrow \textsc{be}^{\alpha}\ \langle\texttt{-g};\ \texttt{-k}^{\text{nom}}_{\alpha}\rangle$

As there are only finitely many possible morphological feature bundles, rule (11) can be treated as a schema, and instantiated with all possible values for its variables. The resulting instantiated set of rules is the translation of a different, non-agreeing, MG Lexicon into this format, where features have been refined into triples of feature name, incoming morphological feature values, and outgoing feature values. This is reminiscent of the early minimalist "checking theory" (Chomsky, 1995), according to which verbs were inserted into syntax fully inflected. The unpacking algorithm in the next section can be thought of as a way to deforest the two step process described here.

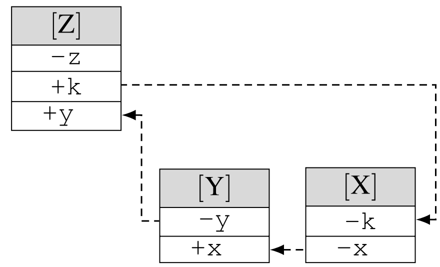
Just as there are multiple MCFG rules needed per lexical entry (because the MG operations hide information that the MCFG rule notation must make explicit), there can be multiple ways to annotate a given rule with agreement information, corresponding to different ways in which agreement information might be passed through the channels. While not visible in this particular example, in general we might need different versions of a rule depending on whether a particular attribute is inherited or synthesized.

4.1 A note on cyclicity

The overwriting principle is realized as a semantic function combining attribute values. A major topic in the AG literature is avoiding cyclic dependencies between attributes. Some static restrictions on such grammars have been proposed to guarantee non-cyclicity (such as synthesized attributes only, or left-to-right evaluation, etc). However, agreeing MGs do not implement these by virtue of their architecture. Indeed, it is simple to construct a MG (12) which gives rise to cyclic dependencies (see the dependency graph in (13)), although it does not seem linguistically natural.

(12) [X] :: $\texttt{-x}^{\rightarrow}\ \texttt{-k}_{\leftarrow}$
[Y] :: $\texttt{+x}_{\leftarrow}\ \texttt{-y}^{\rightarrow}$
[Z] :: $\texttt{+y}_{\leftarrow}\ \texttt{+k}^{\rightarrow}\ \texttt{-z}$

(13)



5 Unpacking a lexicon

LIs in an annotated lexicon are underspecified, in the sense that each of them can be thought of as a set of LIs where the information transmitted across agreement channels is fully instantiated. An agreeing MG can be converted into a standard MG with fully specified lexical items via the following *unpacking algorithm*.

Step 1. We unpack each underspecified LI by generating its *augments* – the set of LIs with all possible combinations of morphological feature annotations for each available channel.

Step 2. Each unpacked LI is checked against the agreement assumptions (4), which serve as a *lexical filter*. The filter is defined locally, strictly in terms of properties of individual LIs. This allows each augment to be inspected for internal consistency in isolation, without considering its context in a derivation. LIs that fail the check are removed from the unpacked lexicon. For example, (14) is a valid augment of THE :: $\texttt{+n}^{\rightarrow}_{\leftarrow}\ \texttt{-d}\ \texttt{-k}^{\rightarrow}_{\leftarrow}$, whereas (15)–(17) are not.

(14) [THE] :: $\texttt{+n}^{\rightarrow\text{c:nom}}_{\leftarrow\phi\text{:3s}}\ \texttt{-d}\ \texttt{-k}^{\rightarrow\phi\text{:3s}}_{\leftarrow\text{c:nom}}$

(15) [THE] :: $\texttt{+n}^{\rightarrow\text{c:acc}}_{\leftarrow\phi\text{:3s}}\ \texttt{-d}\ \texttt{-k}^{\rightarrow\phi\text{:1s}}_{\leftarrow\text{c:nom}}$
(emits features it did not receive)

(16) [THE] :: $\texttt{+n}^{\rightarrow\text{c:nom}}_{\leftarrow\phi\text{:3s}}\ \texttt{-d}\ \texttt{-k}^{\rightarrow}_{\leftarrow\text{c:nom}}$
(does not emit all received features)

(17) [THE] :: $\texttt{+n}^{\rightarrow\text{c:nom},\phi\text{:3s}}_{\leftarrow\text{c:nom},\phi\text{:3s}}\ \texttt{-d}\ \texttt{-k}^{\rightarrow\text{c:nom},\phi\text{:3s}}_{\leftarrow\text{c:nom},\phi\text{:3s}}$
(sends features back along the same dependency)

Step 3. For each remaining augment, we obtain its string realization from the morphological equations. Syntactic features are subscripted with their channel annotations. For clarity, we also remove any useless LIs (that cannot be part of any complete expression) by converting the grammar into an equivalent (multiple) context-free grammar (following the construction of (Michaelis, 2001)) and removing non-generating and unreachable rules. The standard MG obtained from (2) in this way is given in (18).

(18) *the* :: $-\mathtt{n}_{[\phi:3s][c:nom]}$ $-\mathtt{d}$ $-\mathtt{k}_{[\phi:3s][c:nom]}$
the :: $-\mathtt{n}_{[\phi:3p][c:nom]}$ $-\mathtt{d}$ $-\mathtt{k}_{[\phi:3p][c:nom]}$
boy :: $-\mathtt{n}_{[\phi:3s][c:nom]}$
boys :: $-\mathtt{n}_{[\phi:3p][c:nom]}$
I :: $-\mathtt{d}$ $-\mathtt{k}_{[\phi:1s][c:nom]}$
they :: $-\mathtt{d}$ $-\mathtt{k}_{[\phi:3p][c:nom]}$
jumping :: $+\mathtt{d}$ $-\mathtt{g}$
am :: $+\mathtt{g}$ $-\mathtt{k}_{[\phi:1s][c:nom]}$ $-\mathtt{t}$
is :: $+\mathtt{g}$ $-\mathtt{k}_{[\phi:3s][c:nom]}$ $-\mathtt{t}$
are :: $+\mathtt{g}$ $-\mathtt{k}_{[\phi:3p][c:nom]}$ $-\mathtt{t}$

6 Discussion

Agreeing MGs are naturally formalized in terms of AGs. As morphological features are finite-valued, the attributes can be unpacked into the lexicon. Normally AGs are used to provide a value to the entire tree, but here we just send information to leaves. We can easily add attributes to nodes which construct the derived sentence, where the words used depend on the morphological features inherited. These are two logically different things, as the attributes for word order are only synthesized (bottom-up).

References

Jonathan David Bobaljik. 2008. Where's phi? Agreement as a post-syntactic operation. *Phi-Theory: Phi features across interfaces and modules*, 4410.

Noam Chomsky. 1995. *The Minimalist Program*.

Marina Ermolaeva and Gregory M Kobele. 2022. Agree as information transmission over dependencies. *Syntax*, 25(4).

Donald E Knuth. 1968. Semantics of context-free languages. *Mathematical systems theory*, 2(2).

Gregory M. Kobele. 2012. Idioms and extended transducers. In *Proceedings of the Eleventh International Workshop on Tree Adjoining Grammars and Related Frameworks (TAG+11)*, Paris, pages 153–161.

Jens Michaelis. 2001. *On formal properties of minimalist grammars*. Ph.D. thesis, U of Potsdam.

Edward P. Stabler. 1997. Derivational minimalism. In Christian Retoré, editor, *Selected Papers from LACL '96*, pages 68–95.