Computational properties of Unification Grammars

Daniel Feinstein

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE MASTER DEGREE

University of Haifa

Faculty of Social Science

Department of Computer Science

October, 2004

Computational properties of Unification Grammars

By: Daniel Feinstein

Supervised By: Dr. Shuly Wintner

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE MASTER DEGREE

University of Haifa

Faculty of Social Science

Department of Computer Science

October, 2004

Approved by: ___________________________________ Date:_________

(supervisor)

Approved by: ___________________________________ Date:_________

(Chairman of M.A Committee)

Acknowledgment

The research was done under the supervision of Dr. Shuly Wintner in the department of Computer Science.

I would like to thank Shuly for his excellent guidance, assistance and support throughout this work. Shuly is a great adviser and a wonderful person to be with and thanks to him this whole experience was of great benefit to me.

Of course my special thanks to my parents Berta and Josef Feinstein for their unconditional support. Last but not least my thanks to my sweet girlfriend Nataly for her patience.

Contents

Computational properties of Unification Grammars

Daniel Feinstein

Abstract

There is currently considerable interest among computational linguists in grammatical formalisms with highly restricted generative power. This is based on the argument that a grammar formalism should not merely be viewed as a notation, but as part of the linguistic theory. It is now generally accepted that CFGs lack the generative power needed for this purpose. Unification grammars have the ability to describe phonological, morphological, syntactic and semantic properties of languages and thus they are linguistically plausible for modeling natural languages. However, unification grammars are Turing equivalent in their generative capacity: the recognition problem for unification grammars is undecidable in the general case. It is therefore important to constrain the expressivity of unification grammars in a way that would still permit an account of natural languages.

Mildly context-sensitive languages are a natural class of languages for characterizing natural languages. These formalisms were proved to have recognition algorithms with polynomial time complexity and there is no evidence that any natural language is outside of the mildly context-sensitive class of languages. In this work we define a constraint on unification grammars which ensures that grammars satisfying the constraint generate all and only the mildly context-sensitive languages. We thus provide a linguistically plausible formalism which is computationally tractable.

List of Figures

Chapter 1

Introduction

There is currently considerable interest among computational linguists in grammatical formalisms with highly restricted generative power. This is based on the argument that a grammar formalism should not merely be viewed as a notation, but as part of the linguistic theory. It should make predictions about the structure of natural language and its value is lessened to the extent that it supports both good and bad analyses. In order for a grammar formalism to have such predictive power its generative capacity must be constrained. This has led to interest in the use of context-free grammars (CFG) as a notation with which to express linguistic theories. However, it is now generally accepted that CFGs lack the generative power needed for this purpose (Huybregts, 1984; Shieber, 1985; Culy, 1985). Typical natural language constructions that require trans-context-free power are:

- reduplication, leading to constructions of the form $\{ww \mid w \in \Sigma^*\}$
- multiple agreement, corresponding to constructions of the form $\{a^n b^n a^j \mid 0 < j \leq n\}$,
- crossed agreement, as modeled by $\{a^n b^m c^n d^m \mid n, m > 0\}$,

As a result there is substantial interest in the development and study of constrained grammar formalisms whose generative power exceeds CFG.

1.1 Mildly context-sensitive grammars

Several linguistic formalisms have been proposed as capable of modeling the above mentioned phenomena. The class of *mildly context-sensitive (MCS) languages* is defined by Joshi (1985) as a class including all formalisms which properly extend CFG, can express "limited cross-serial dependencies", exhibit the constant growth property and can be parsed in polynomial time. We consider four mildly context-sensitive formalisms here: Linear Indexed Grammars (LIG), Head Grammars (HG), Tree Adjoining Grammars (TAG) and Combinatory Categorial Grammars (CCG). The four formalisms under consideration were developed independently and superficially differ considerably from one another.

- LIG (Gazdar, 1988) can be viewed as a generalization of CFG in which each nonterminal is associated with an unbounded stack of items drawn from some finite set. Rules are permitted to push items onto, pop items from, and copy the stack. For example,[1] a rule $A[..] \rightarrow aB[i..]$ is similar to the CFG rule $A \rightarrow aB$, except that it copies the stack of A to B, pushing the element i onto B's stack.

- HG (Pollard, 1984) can be viewed as a generalization of CFG in which a wrapping operation is used in addition to concatenation. The nonterminals of a CFG derive strings of terminals $(w_1 \dots w_k)$; the nonterminals of HG derive *headed strings*, which are pairs of terminal strings $(w_1 \dots w_i, w_{i+1} \dots w_k)$, denoted $w_1 \dots w_{i\uparrow} w_{i+1} \dots w_k$. The rules of HG are similar to those of CFG, but where CFG only defines concatenation of the daughters in each rule, HG allows an additional operation, *wrapping*, to be defined over the (two) daughters: $W(s_{\uparrow}t, u_{\uparrow}v) = (su_{\uparrow}vt)$. Derivations in HG are simple rewritings which apply either concatenation or wrapping, as specified in the rules.

- TAG (Joshi, 1985; Joshi, 2003) is a tree manipulation system. A grammar consists of two sets of trees, *initial* and *auxiliary*. TAG defines two operations on trees: *substitution*, which replaces a node labeled A in a tree by a tree whose root is labeled A; and *adjunction*, which takes a tree τ in which some internal node is labeled A, and an auxiliary tree in which A labels both the root

[1] $A[..]$ denotes a nonterminal symbol A with any stack content.

and some node on the frontier, and splices the auxiliary tree in τ, replacing the node labeled A by the entire auxiliary tree. The closure of the set of initial trees with respect to these two operations defines the tree language of a grammar, and the string language is defined as the set of all terminal yields of the tree language.

- Categorial grammars (CG) define a finite set of primitive categories. Each terminal symbol is assigned a finite number of primitive or complex categories, the latter obtained from the former by means of the operators $\backslash$ and $/$. For example, if the set of primitive categories is $\{N, NP, S\}$, then complex categories include $S\backslash NP$, NP/N, $NP\backslash(S/NP)$ etc. The intuition behind having two directional slashes is that it allows one to code the syntactic order of (for example) arguments of a verb in a lexicalized grammar: a transitive verb, which takes an NP to the left and an NP to the right and yields a sentence, could be written as $NP\backslash(S/NP)$. There are only two category-combination rules in CG: $\alpha_1/\alpha_2 \cdot \alpha_2 \rightarrow \alpha_1$ and $\alpha_2 \cdot \alpha_1\backslash\alpha_2 \rightarrow \alpha_1$ where α_i is a (complex or primitive) category. Combinatory CG (CCG, Steedman (2000)) adds a few more combination rules, the motivation being coordination and other complex linguistic phenomena. These include functional composition: $\alpha_1/\alpha_2 \cdot \alpha_2/\alpha_3 \rightarrow \alpha_1/\alpha_3$ and $\alpha_1\backslash\alpha_2 \cdot \alpha_3\backslash\alpha_1 \rightarrow \alpha_3\backslash\alpha_2$.

Informally, differences between the formalisms can be explained in terms of the way in which they can be seen to extend CFG. For example, in addition to string concatenation, HG introduces a wrapping operation with which one pair of strings can be wrapped around another. In other respects HG are identical to CFG since the derivation process involves context-free rewriting of members of a finite set of non-terminal symbols. Both CCG and LIG, on the other hand, use only string concatenation. However, they differ from CFG in that their derivation process involves rewriting of unbounded stack-like structures. The status of TAG, a tree manipulating system, is ambiguous since it is possible to interpret TAG as extending CFG in either of these ways.

Despite these differences, all four formalisms are weakly equivalent (Vijay-Shanker and Weir, 1994). We use the term *mildly context-sensitive* in this paper to refer to the class of the languages that the four formalism defined. These formalisms were proved to have recognition algorithms with time complexity $O(n^6)$, considering the size of the grammar a constant factor (Vijay-Shanker and Weir, 1990; Satta, 1994). As a result of the weak equivalence of four independently developed (and linguistically motivated) extensions of CFG, the class of mildly context-sensitive languages is considered to be linguistically meaningful. There is no evidence that any natural language is outside of the mildly context-sensitive class of languages. Mildly context-sensitive languages, therefore, are a natural class of languages for characterizing natural languages.

1.2 Unification grammars

Unification grammars (Shieber, 1986; Shieber, 1992; Carpenter, 1992) have originated as an extension of context-free grammars, the basic idea being to augment the context-free rules with non context-free annotations (feature structures) in order to express some additional information. Unification grammars have the ability to describe phonological, morphological, syntactic and semantic properties of languages and thus they are linguistically plausible for modeling natural languages. Today, several formulations of unification grammars exist, some of which do not assume an explicit context-free backbone. They are used extensively by computational linguists to describe the structure of a variety of natural languages.

We assume familiarity with theories of feature structures as formulated, e.g., by Shieber (1992) or Carpenter (1992). We summarize below the few concepts that are needed for the rest of this paper in order to set up notation, adapting the description of Jaeger, Francez, and Wintner (2004). We begin with a formal definition of attribute-value matrices (AVM).

Definition 1 (AVMs). *Given a signature consisting of a finite set* ATOMS *of atoms and a finite set* FEATS *of features, the set* AVMS *of AVMs is the least set such that*

1. $\boxed{i}\ a \in \text{AVMS}$ *for every variable* $\boxed{i}$ *and* $a \in \text{ATOMS}$*;*
2. $\boxed{i}\ [\,] \in \text{AVMS}$ *for every variable* $\boxed{i}$ *;*
3. *for every variable* $\boxed{i}$, $\text{F}_1, \ldots, \text{F}_n \in \text{FEATS}$ *and* $\mathsf{A}_1, \ldots, \mathsf{A}_n \in \text{AVMS}$*,* $n \geq 1$*,*

$$\mathsf{A} = \boxed{i} \begin{bmatrix} \text{F}_1 : \mathsf{A}_1 \\ \vdots \\ \text{F}_n : \mathsf{A}_n \end{bmatrix} \in \text{AVMS}$$

The ***value*** *of the feature* F_j *in* A*, denoted* $val(\mathsf{A}, \text{F}_j)$*, is* A_j*.*

Definition 2 (multi-AVM). *Given a signature consisting of a finite set* ATOMS *of atoms and a finite set* FEATS *of features, a* ***multi-AVM*** *of* ***length*** n *is a sequence* $\langle \mathsf{A}_1, \ldots, \mathsf{A}_n \rangle$ *such that for each* i*,* $1 \leq i \leq n$*,* A_i *is an AVM over the signature.*

Meta-variables A, B range over feature structures and σ, ρ over multi-AVMs. An multi-AVMs σ can be viewed as an ordered sequence $\langle \mathsf{A}_1, \ldots, \mathsf{A}_n \rangle$ of (not necessarily disjoint) feature structures. We identify multi-AVMs of length 1 with feature structures.

We now define another representation of feature structures called *abstract feature structures,* which is easier to work with mathematically. We start with pre- abstract feature structures, which consist of three components: a set Π of paths, corresponding to the paths defined in the intended feature graphs (taken as sequences of features); a function Θ that labels the end points of some of the paths (corresponding to the labeling of some of the sinks in graphs); and an equivalence relation specifying what sets of paths lead to the same node in the intended graph, without an explicit specification of the node's identity. Abstract feature structures are pre- abstract feature structures with some additional constraints imposed on them, which guarantee that the specification indeed corresponds to some concrete feature graph. We denote the set of all paths as PATHS (PATHS $=$ FEATS*).

Definition 3 (Abstract feature structures). *A* ***pre- abstract feature structure*** *(pre-AFS) is a triple* $\langle \Pi, \Theta, \approx \rangle$*, where*

- $\Pi \subseteq$ PATHS *is a non-empty set of paths*
- $\Theta : \Pi \to$ ATOMS *is a partial function, assigning an atom to some of the paths*
- $\approx \; \subseteq \Pi \times \Pi$ *is a relation specifying reentrancy*

An ***abstract feature structure*** *(AFS) is a pre-AFS* A *for which the following requirements hold:*

- Π *is prefix-closed: if* $\pi \cdot \alpha \in \Pi$ *then* $\pi \in \Pi$ *(where* $\pi, \alpha \in$ PATHS*)*
- A *is fusion-closed: if* $\pi \cdot \alpha \in \Pi$ *and* $\pi \approx \pi'$ *then* $\pi' \cdot \alpha \in \Pi$ *and* $\pi' \cdot \alpha \approx \pi \cdot \alpha$
- $\approx$ *is an equivalence relation with a finite index (with* $[\approx]$ *the set of its equivalence classes) including at least the pair* $\langle \epsilon, \epsilon \rangle$

- Θ *is defined only for maximal paths: if* $\Theta(\pi)\downarrow$ *then there exists no path* $\pi \cdot \alpha \in \Pi$ *such that* $\alpha \neq \varepsilon$

- Θ *respects the equivalence: if* $\pi_1 \approx \pi_2$ *then either both undefined or both are defined and* $\Theta(\pi_1) = \Theta(\pi_2)$

A *non-reentrant feature structure* is a feature structure whose reentrancy relation contains only pairs of equal paths. Let NRFSS be the set of all non-reentrant feature structures over this signature. An *abstruct multi-rooted structure* (AMRS) of length n is a sequence of n abstract feature structures, with possible reentrancies among elements of the sequence.

Definition 4 (Abstract multi-rooted structures). *A* **pre-abstract multi rooted structure** *(pre-AMRS) is a quadruple* $\sigma = \langle Ind, \Pi, \Theta, \approx\rangle$*, where:*

- $Ind \in \mathbb{N}$ *is the number of* ***indices*** *of* σ

- $\Pi \subseteq \{1, 2, \ldots, Ind\} \times$ PATHS *is a set of* ***indexed paths****, such that for each* i*,* $1 \leq i \leq Ind$*, there exists some* $\pi \in$ PATHS *with* $(i, \pi) \in \Pi$

- $\Theta : \Pi \rightarrow$ ATOMS *is a partial function, assigning an atom to some of the paths*

- $\approx\ \subseteq \Pi \times \Pi$ *is a relation specifying reentrancy*

An **abstract multi-rooted structure** *(AMRS) is a pre-AMRS* σ *for which the following requirements, naturally extending those of AFSs, hold:*

- Π *is prefix-closed: if* $\langle i, \pi\alpha\rangle \in \Pi$ *then* $\langle i, \pi\rangle \in \Pi$

- σ *is fusion-closed: if* $\langle i, \pi\alpha\rangle \in \Pi$ *and* $\langle i, \pi\rangle \approx \langle i', \pi'\rangle$ *then* $\langle i, \pi\alpha\rangle \in \Pi$ *and* $\langle i, \pi\alpha\rangle \approx \langle i', \pi'\alpha\rangle$

- $\approx$ *is an equivalence relation with a finite index (with* $[\approx]$ *the set of its equivalence classes) including at least the pairs* $\{\langle i, \epsilon\rangle \approx \langle i, \epsilon\rangle \mid 1 \leq i \leq Ind\}$*, and if* $\langle i, \epsilon\rangle \approx \langle j, \epsilon\rangle$ *then* $i = j$

- Θ *is defined only for maximal paths: if* $\Theta(\langle i, \pi\rangle)\downarrow$ *then there exists no pair* $\langle i, \pi\alpha\rangle \in \Pi$ *such that* $\alpha \neq \epsilon$

- Θ *respects the equivalence: if* $\langle i_1, \pi_1\rangle \approx \langle i_2, \pi_2\rangle$ *then* $\Theta(\langle i_1, \pi_1\rangle) = \Theta(\langle i_2, \pi_2\rangle)$

In the sequel, given a feature structure A, we write $\langle \Pi_A, \Theta_A, \approx_A \rangle$ for its abstract representation. Similarly, an AMRS σ is written as $\langle Ind_\sigma, \Pi_\sigma, \Theta_\sigma, \approx_\sigma \rangle$. For any AMRS σ, we denote a reentrancy relation between paths $(i, \pi_1), (j, \pi_2) \in \Pi_\sigma$, where $i, j \leq Ind$ and $((i, \pi_1), (j, \pi_2)) \in \approx_\sigma$, by $(i, \pi_1) \overset{\sigma}{\leftrightsquigarrow} (j, \pi_2)$.

Feature structures and AMRSs are partially ordered by *subsumption*, denoted '$\sqsubseteq$'. The least upper bound with respect to subsumption is the *unification* operator, denoted '$\sqcup$' (we use the term 'unification' both for the operator and for the result of its application). Unification is a partial operator; when A $\sqcup$ B is undefined we say that the unification *fails* and denote it as A $\sqcup$ B $= \top$. Unification is lifted to AMRSs: given two AMRSs σ and ρ, it is possible to unify the i-th element of σ with the j-th element of ρ. This operation, called *unification in context* and denoted $(\sigma, i) \sqcup (\rho, j)$, yields two modified variants of σ and ρ: as the unification is done *in the context* of the entire AMRSs, other elements might be affected. Hence, the result of unification in context (when it is defined) is a pair (σ', ρ').

One of the advantages resulting from the representation of linguistic information by means of abstract feature structures is the relative simplicity of *subsumption* and *unification*. The subsumption relation becomes not much more than *set* inclusion; and unification is basically set union.

Definition 5 (AFS subsumption). *Let $\sqsubseteq$ be a relation over AFSs such that* A $\sqsubseteq$ B *iff the following three conditions hold:*

- $\Pi_A \subseteq \Pi_B$
- $\approx_A \ \subseteq \ \approx_B$
- *if* $\Theta_A(\pi) \downarrow$ *then* $\Theta_B(\pi) \downarrow$ *and* $\Theta_A(\pi) = \Theta_B(\pi)$.

Namely, A is more general than B if and only if all the paths of A are also paths in B, if a (maximal) path is labeled in A then it is labeled identically in B and every reentrancy in A is a reentrancy in B. The unification of two AFSs can be defined in terms of set union and the closure operations:

Definition 6 (AFS unification). *The unification of two AFSs* A *and* B *(denoted* A $\sqcup$ B*) is defined only if for every path π which is defined in both* A *and* B, *either* $\Theta_A(\pi)$ *and* $\Theta_B(\pi)$ *are both defined and* *equal, or neither $\Theta_A(\pi)$ nor $\Theta_B(\pi)$ is defined, or only one is defined and π is a maximal path in the other. The closure operations are:*

- *$Cl(\mathsf{A})$ is the least fusion-closed pre-AFS that extends A;*
- *$Eq(\mathsf{A})$ is the least extension of A in which $\approx$ is an equivalence relation;*
- *$Ty(\mathsf{A})$ is the least extension of A in which Θ respects the $\approx$ relation.*

If the unification is defined, $\mathsf{A} \sqcup \mathsf{B} = Ty(Eq(Cl(\mathsf{C})))$, where

- $\Pi_C = \Pi_A \cup \Pi_B$
- $\approx_C = \approx_A \cup \approx_B$
- $\Theta_C(\pi) = \begin{cases} \Theta_A(\pi) & \textit{if } \Theta_B(\pi)\uparrow \\ \Theta_B(\pi) & \textit{if } \Theta_A(\pi)\uparrow \\ \Theta_B(\pi) & \textit{if } \Theta_A(\pi) = \Theta_B(\pi) \\ \textit{undefined} & \textit{otherwise} \end{cases}$

The unification fails if there exists a path π in both A and B, such that $\Theta_A(\pi) \neq \Theta_B(\pi)$, or if $\Theta_A(\pi)\downarrow$ and π is not a maximal path in B, or if $\Theta_B(\pi)\downarrow$ and π is not a maximal path in A. Otherwise, its result is obtained by first computing C, by union of the paths and the reentrancies of A and B, taking care of the types of the atoms; and then applying the closure operations: Cl adds necessary paths and reentrancies, Eq completes the resulting pre-AFS to one in which $\approx$ is an equivalence relation and finally, Ty sets the types of the added paths. Trivially, the result is an AFS.

While formally we manipulate *abstract* feature structures and MRSs, we depict them using the common AVM notation to facilitate readability. We use the terms feature structures and MRSs for both representations in the sequel.

Definition 7. ***Unification grammars*** *are defined over a signature consisting of a finite set* ATOMS *of atoms; a finite set* FEATS *of features and a finite set* WORDS *of words. A* ***unification grammar*** *is a tuple $G^u = \langle \mathcal{R}^u, \mathcal{L}, \mathsf{A}^s \rangle$ where:*

- *$\mathcal{R}^u$ is a finite set of rules, each of which is an MRS of length $n \geq 1$, with a designated first element, the **head** of the rule, followed by its **body**. The head and body are separated by an arrow ($\rightarrow$).*

- *$\mathcal{L}$ is a **lexicon**, which associates with every word $w \in$ WORDS a finite set of feature structures, $\mathcal{L}(w)$.*

- *A^s is a feature structure, the **start symbol**.*

We use meta-variables G^u (with or without subscripts) to denote unification grammars.

Example 1 (Unification grammar). *Let G^u_{ww} be a unification grammar over the signature $\langle$ATOMS, FEATS, WORDS$\rangle$, where FEATS $= \{$LIST, HD, TL$\}$, ATOMS $= \{s, elist, ta, tb\}$ and WORDS $= \{a, b\}$. The grammar has two rules, each an MRS of length 3, and two lexical entries, one for each element of* WORDS.

$$\mathsf{A}^s = \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: s \\ \text{TL}: elist \end{bmatrix} \end{bmatrix}$$

$$\mathcal{R}^u = \left\{ \begin{array}{l} \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: s \\ \text{TL}: elist \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \text{LIST}: \boxed{3} \end{bmatrix} \quad \begin{bmatrix} \text{LIST}: \boxed{3} \end{bmatrix} \\ \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: \boxed{1} \\ \text{TL}: \boxed{2} \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \text{LIST}: \boxed{2} \end{bmatrix} \quad \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: \boxed{1} \\ \text{TL}: elist \end{bmatrix} \end{bmatrix} \end{array} \right\}$$

$$\mathcal{L}(a) = \left\{ \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: ta \\ \text{TL}: elist \end{bmatrix} \end{bmatrix} \right\} \qquad \mathcal{L}(b) = \left\{ \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: tb \\ \text{TL}: elist \end{bmatrix} \end{bmatrix} \right\}$$

We extend the definition of unification to AMRSs. The input to the operation is a pair of AMRSs, with two indices pointing to the elements that are to be unified, and the output is a pair of AMRSs.

Definition 8 (Unification in context). *Let σ, ρ be two AMRSs of lengths n_σ, n_ρ, respectively. The unification of the i-th element in σ with the j-th element in ρ, denoted $(\sigma, i) \sqcup (\rho, j)$, is defined only if $i \leq n_\sigma$ and $j \leq n_\rho$, in which case it is a pair of AMRSs, $\langle \sigma'', \rho'' \rangle = \langle Ty(Eq(Cl(\sigma'))), Ty(Eq(Cl(\rho'))) \rangle$,* *where σ' and ρ' are defined as follows:*

$$Ind_{\sigma'} = Ind_\sigma$$
$$\Pi_{\sigma'} = \Pi_\sigma \cup \{\langle i,\pi\rangle \mid \langle j,\pi\rangle \in \Pi_\rho\}$$
$$\approx_{\sigma'} = \approx_\sigma \cup \{(\langle i_1,\pi_1\rangle, \langle i_2,\pi_2\rangle) \mid \langle j_1,\pi_1\rangle \approx_\rho \langle j_2,\pi_2\rangle\}$$
$$\Theta_{\sigma'}(\langle k,\pi\rangle) = \begin{cases} \Theta_\sigma(\langle k,\pi\rangle) & \textit{if } k \neq i \\ \Theta_\sigma(\langle k,\pi\rangle) & \textit{if } k = i \textit{ and } \Theta_\sigma(\langle i,\pi\rangle)\downarrow \\ \Theta_\rho(\langle j,\pi\rangle) & \textit{if } k = i \textit{ and } \Theta_\rho(\langle j,\pi\rangle)\downarrow \textit{ and } \Theta_\sigma(\langle i,\pi\rangle)\uparrow \\ \textit{undefined} & \textit{otherwise} \end{cases}$$
$$Ind_{\rho'} = Ind_\rho$$
$$\Pi_{\rho'} = \Pi_\rho \cup \{\langle j,\pi\rangle \mid \langle i,\pi\rangle \in \Pi_\sigma\}$$
$$\approx_{\rho'} = \approx_\rho \cup \{(\langle j_1,\pi_1\rangle, \langle j_2,\pi_2\rangle) \mid \langle i_1,\pi_1\rangle \approx_\sigma \langle i_2,\pi_2\rangle\}$$
$$\Theta_{\rho'}(\langle k,\pi\rangle) = \begin{cases} \Theta_\rho(\langle k,\pi\rangle) & \textit{if } k \neq j \\ \Theta_\rho(\langle k,\pi\rangle) & \textit{if } k = j \textit{ and } \Theta_\rho(\langle j,\pi\rangle)\downarrow \\ \Theta_\sigma(\langle i,\pi\rangle) & \textit{if } k = j \textit{ and } \Theta_\sigma(\langle i,\pi\rangle)\downarrow \textit{ and } \Theta_\rho(\langle j,\pi\rangle)\uparrow \\ \textit{undefined} & \textit{otherwise} \end{cases}$$

*The unification **fails** if there exists a path π such that $\Theta_\sigma(\langle i,\pi\rangle)\downarrow$ and $\Theta_\rho(\langle j,\pi\rangle)\downarrow$ but $\Theta_\sigma(\langle i,\pi\rangle) \neq \Theta_\rho(\langle j,\pi\rangle)$; or if there exist paths π,α, where $\alpha \neq \epsilon$, such that either $\Theta_\sigma(\langle i,\pi\rangle)\downarrow$ but $\langle j,\pi\alpha\rangle \in \Pi_\rho$, or $\Theta_\rho(\langle j,\pi\rangle)\downarrow$ but $\langle i,\pi\alpha\rangle \in \Pi_\sigma$.*

Compare the above definition to definition 6 and observe that the differences are minor. The unification returns two AMRSs, σ'' and ρ'', which are extensions (with respect to the closure operations Ty, Eq and Cl) of σ' and ρ', respectively.

To define the *language* generated by a unification grammar G^u, we extend the notion of *forms*: a form is simply an MRS. A form $\sigma_A = \langle \mathsf{A}_1,\ldots,\mathsf{A}_k\rangle$ *immediately derives* another form $\sigma_B = \langle \mathsf{B}_1,\ldots,\mathsf{B}_m\rangle$ (denoted by $\sigma_A \overset{1}{\Longrightarrow}_u \sigma_B$) iff there exists a rule $r^u \in \mathcal{R}^u$ of length n that licenses the derivation. The head of the rule is matched against some element A_i in σ_A using unification in context: $(\sigma_A, i) \sqcup (r^u, 0) = (\sigma'_A, r')$. If the unification does not fail, σ_B is obtained by replacing the i-th element of σ'_A with the body of r'. The reflexive transitive closure of '$\overset{1}{\Longrightarrow}_u$' is denoted by '$\overset{*}{\Longrightarrow}_u$'. An empty derivation sequence means that an empty sequence of rules is applied to the source

MRS and is denoted by '$\overset{0}{\Longrightarrow}_u$', for example $\sigma_A \overset{0}{\Longrightarrow}_u \sigma_A$.

Definition 9. *The **language** of a unification grammar G^u is $L(G^u) = \{s \in \textsc{Words}^* \mid s = w_1 \cdots w_n$ and $\mathsf{A}^s \overset{*}{\Longrightarrow}_u \sigma_l$ such that σ_l is unifiable with $\langle \mathsf{A}_1, \ldots, \mathsf{A}_n \rangle\}$, where $\mathsf{A}_i \in \mathcal{L}(w_i)$ for $1 \leq i \leq n$.*

Example 2 (Derivation sequence). *As an example, consider again the grammar G^u_{ww} of example 1. The following is a derivation sequence for the string $baba$ with this grammar. Note that the scope of variables is limited to a single MRS (so that multiple occurrences of the same tag in a single form denote reentrancy, whereas across forms they are unrelated).*

$$\mathsf{A}^s = \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: s \\ \text{TL}: elist \end{bmatrix} \end{bmatrix} \qquad \textit{apply rule 1 to the single element of the form}$$

$$\sigma_1 = \begin{bmatrix} \text{LIST}: \boxed{3} \end{bmatrix} \begin{bmatrix} \text{LIST}: \boxed{3} \end{bmatrix} \qquad \textit{apply rule 2 to the second element}$$

$$\sigma_2 = \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: \boxed{1} \\ \text{TL}: \boxed{2} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \text{LIST}: \boxed{2} \end{bmatrix} \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: \boxed{1} \\ \text{TL}: elist \end{bmatrix} \end{bmatrix} \qquad \textit{apply rule 2 to the first element}$$

$$\sigma_3 = \begin{bmatrix} \text{LIST}: \boxed{2} \end{bmatrix} \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: \boxed{1} \\ \text{TL}: elist \end{bmatrix} \end{bmatrix} \begin{bmatrix} \text{LIST}: \boxed{2} \end{bmatrix} \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: \boxed{1} \\ \text{TL}: elist \end{bmatrix} \end{bmatrix}$$

Now consider the MRS obtain by concatenating (the single elements of) $\langle \mathcal{L}(b), \mathcal{L}(a), \mathcal{L}(b), \mathcal{L}(a) \rangle$:

$$\sigma_l = \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: tb \\ \text{TL}: elist \end{bmatrix} \end{bmatrix} \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: ta \\ \text{TL}: elist \end{bmatrix} \end{bmatrix} \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: tb \\ \text{TL}: elist \end{bmatrix} \end{bmatrix} \begin{bmatrix} \text{LIST}: \begin{bmatrix} \text{HD}: ta \\ \text{TL}: elist \end{bmatrix} \end{bmatrix}$$

Since σ_l and σ_3 are unifiable, the string $baba$ is in $L(G^u_{ww})$. In fact, $L(G^u_{ww}) = \{ww \mid w \in \{a,b\}^+\}$.

Unification grammars are Turing equivalent in their generative capacity: determining whether a given string is generated by a given grammar is as hard as deciding whether a Turing machine halts on the empty input (Johnson, 1988). Therefore, the recognition problem for unification grammars is undecidable in the general case. In order to ensure decidability of the recognition problem, several constraints on unification grammars, commonly known as the *off-line parsability (OLP) constraints*, were suggested, such that the recognition problem is decidable for off-line parsable unification grammars (see Jaeger, Francez, and Wintner (2004) for a survey).

The idea behind all the OLP definitions is to rule out grammars which license trees in which unbounded amount of material is generated without expanding the frontier word. This can happen due to two kinds of rules: ϵ-rules (whose bodies are empty) and unit rules (whose bodies consist of a single element). When grammars are context-free, it is always possible to remove grammar rules which can cause such unbounded growth of the trees: in particular, one can always remove cyclic sequences of unit rules (which can be applied unboundedly, without expanding the yield of the tree). However, with unification grammars such a procedure turns out to be more problematic. It is not trivial to determine when a sequence of unit-rules is, indeed, cyclic; and when a rule is redundant.

Recently, Jaeger, Francez, and Wintner (2004) defined a novel OLP constraint which is shown to be effectively testable. However, even grammars which are OLP according to their definition are not guaranteed to have a polynomial parsing time.

1.3 Research objectives

The main objective of this work is to define constraints on unification grammars which will guarantee efficient (polynomial) processing. There are naïve constraints which restrict the expressiveness of unification grammars in a way which ensures polynomial parsing time, but they are too strong. One example is Generalized Phrase Structure Grammar (GPSG) (Gazdar et al., 1985). Among current syntactic theories, GPSG provides an appealing solution for describing natural languages with its modular system of composite categories, rules, constraints and feature propagation principles. GPSG is known to be equivalent to CFG, thus inducing a polynomial, $O(n^3)$, recognition parsing time. Another example of such a constraint (which we show in this paper as the first step towards a more interesting constraint) is to disallow reentrancies in feature structures. In both cases above the resulting formalisms are equivalent to CFG which, as we mentioned above, is not enough for describing natural languages.

Our main goal in this work is to define an effectively testable syntactic constraint on unification grammars which will ensure that grammars satisfying the constraint generate all and only the mildly context-sensitive languages. This is beneficial for both theoretical and practical reasons:

- From a theoretical point of view, constraining unification grammars to generate exactly the class of mildly context-sensitive languages will result in a grammatical formalism which is, on one hand, powerful enough for linguists to express linguistic generalizations in, and on the other hand cognitively adequate;

- Practically, such a constraint can provide an efficient recognition time algorithm for the limited class of unification grammars.

In this work we show the solution for the theoretical aspect of the problem by defining a mapping from unification grammars to one of mildly context-sensitive formalisms, Linear Indexed Grammar.

Chapter 2

Context-free Unification Grammars

In this section we define a constraint on unification grammars which ensures that grammars satisfying it generate all and only the context-free languages. This constraint disallows *any* reentrancies in the rules of the grammar. When rules are non-reentrant, applying a rule implies that an exact copy of the body of the rule is inserted into the generated (sentential) form, not affecting neighboring elements of the form the rule is applied to. The only difference between rule application in non-reentrant unification grammars and the analog operation in context-free grammars is that the former requires unification whereas the latter only calls for identity check. In this section we show that this small difference does not affect the generative power of the formalisms.

Let $G^{cf} = \langle V_N, V_t, \mathcal{R}^{cf}, S^{cf} \rangle$ be a context-free grammar. For the sake of simplicity, in this section we assume that the start symbol of G^{cf} occurs only on the left side of rules. If this is not the case, rename the original start symbol to S^{cf}_{old} and introduce a new start symbol, S^{cf}, and an additional unit rule $S^{cf} \rightarrow S^{cf}_{old}$. We also assume, for simplicity, that the grammar is given in a normal form, where each rule has either a sequence of (zero or more) non-terminals in its body or a single terminal. The set of all such context-free grammars is denoted CFGS.

Definition 10 (Non-reentrant unification grammar). *A unification grammar $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle$ over the signature $\langle \text{ATOMS}, \text{FEATS}, \text{WORDS} \rangle$ is* ***non-reentrant*** *iff for any rule $r^u \in \mathcal{R}^u$, r^u is non-reentrant. Let* UG_{nr} *be the set of all non-reentrant unification grammars.*

We show that the class of languages generated by non-reentrant unification grammars is exactly the class of context-free languages. The trivial direction is to map a CFG to a non-reentrant unification grammar, since every CFG is, trivially, such a unification grammar.

Definition 11 (Mapping from $\textsc{Cfgs}$ **to** UG_{nr}**).** *Let cfg2ug* $: \textsc{Cfgs} \mapsto \mathrm{UG}_{nr}$ *be a mapping of* $\textsc{Cfgs}$ *to* UG_{nr}*, such that if* $G^{cf} = \langle V_N, V_t, \mathcal{R}^{cf}, S^{cf} \rangle$ *and* $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle = \textit{cfg2ug}(G^{cf})$ *then* G^u *is over the signature* $\langle \textsc{Atoms}, \textsc{Feats}, \textsc{Words} \rangle$ *and:*

- $\textsc{Atoms} = V_N \cup V_t$
- $\textsc{Feats} = \emptyset$
- $\textsc{Words} = V_t$
- $\mathsf{A}^s = S^{cf}$
- *For all* $t_j \in V_t$*, for all* $A \to t_j \in \mathcal{R}^{cf}$*,* $\mathsf{A} \in \mathcal{L}(t_j)$
- *If* $B_0 \to B_1 \ldots B_n \in \mathcal{R}^{cf}$ *then* $B_0 \to B_1 \ldots B_n \in \mathcal{R}^u$*.*

Theorem 1. *Let* G^{cf} *be a CFG grammar. Then* $L(G^{cf}) = L(\textit{cfg2ug}(G^{cf}))$*.*

Proof. Since all feature structures are atomic, unification in $\textit{cfg2ug}(G^{cf})$ is reduced to identity check. As there is a one-to-one correspondence between feature structures in $\textit{cfg2ug}(G^{cf})$ and terminal and non-terminal symbols in G^{cf}, rule application in both grammars is identical. Hence both grammars induce the same derivation relation on forms, and therefore generate the same language. □

We now define a mapping from UG_{nr} to $\textsc{Cfgs}$. The non-terminal symbols of a context-free grammar in the image of the mapping are the set of all feature structures defined in the source unification grammar.

Definition 12 (Mapping from UG_{nr} **to** $\textsc{Cfgs}$**).** *Let ug2cfg* $: \mathrm{UG}_{nr} \mapsto \textsc{Cfgs}$ *be a mapping of* UG_{nr} *to* $\textsc{Cfgs}$*, such that if* $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle$ *is over the signature* $\langle \textsc{Atoms}, \textsc{Feats}, \textsc{Words} \rangle$ *and* $G^{cf} = \langle V_N, V_t, \mathcal{R}^{cf}, S^{cf} \rangle = \textit{ug2cfg}(G^u)$*, then:*

- $V_N = \{\mathsf{A}^s\} \cup \{\mathsf{A}_i \mid \mathsf{A}_0 \to \mathsf{A}_1 \ldots \mathsf{A}_n \in \mathcal{R}^u, 1 \leq i \leq n\} \cup \{\mathsf{A} \mid \mathsf{A} \in \mathcal{L}(a), a \in \textsc{Atoms}\}$*.* V_N *is the set of all the feature structures occurring in any of the rules or the lexicon of* G^u*.*

- $V_t = \text{WORDS}$
- $S^{cf} = \mathsf{A}^s$
- $\mathcal{R}^{cf}$ *consists of the following rules:*

  1. *Let* $\mathsf{A}_0 \rightarrow \mathsf{A}_1 \ldots \mathsf{A}_n \in \mathcal{R}^u$ *and* $\mathsf{B} \in \mathcal{L}(b)$. *If for some* i, $1 \leq i \leq n$, $\mathsf{A}_i \sqcup \mathsf{B} \neq \top$, *then* $\mathsf{A}_i \rightarrow b \in \mathcal{R}^{cf}$
  2. *If* $\mathsf{A}_0 \rightarrow \mathsf{A}_1 \ldots \mathsf{A}_n \in \mathcal{R}^u$ *and* $\mathsf{A}^s \sqcup \mathsf{A}_0 \neq \top$ *then* $S^{cf} \rightarrow \mathsf{A}_1 \ldots \mathsf{A}_n \in \mathcal{R}^{cf}$.
  3. *Let* $r_1^u = \mathsf{A}_0 \rightarrow \mathsf{A}_1 \ldots \mathsf{A}_n$ *and* $r_2^u = \mathsf{B}_0 \rightarrow \mathsf{B}_1 \ldots \mathsf{B}_m$, *where* $r_1^u, r_2^u \in \mathcal{R}^u$. *If for some* i, $1 \leq i \leq n$, $\mathsf{A}_i \sqcup \mathsf{B}_0 \neq \top$, *then the rule* $\mathsf{A}_i \rightarrow \mathsf{B}_1 \ldots \mathsf{B}_m \in \mathcal{R}^{cf}$

Example 3 (Mapping from UG_{nr} **to** CFGS**).** *Let* $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle$ *be a non-reentrant unification grammar for the language* $\{a^n b^n \mid 0 \leq n\}$ *over the signature* $\langle \text{ATOMS}, \text{FEATS}, \text{WORDS} \rangle$, *such that:*

- $\text{ATOMS} = \{v, u, w\}$
- $\text{FEATS} = \{\text{F}_1, \text{F}_2\}$
- $\text{WORDS} = \{a, b\}$
- $\mathsf{A}^s = \begin{bmatrix} \text{F}_1 : w \\ \text{F}_2 : w \end{bmatrix}$
- *The lexicon is defined as* $\mathcal{L}(a) = \{ \begin{bmatrix} \text{F}_2 : v \end{bmatrix} \}$ *and* $\mathcal{L}(b) = \{ \begin{bmatrix} \text{F}_2 : u \end{bmatrix} \}$
- *The set of rules* $\mathcal{R}^u$ *is defined as:*

  1. $\begin{bmatrix} \text{F}_1 : w \\ \text{F}_2 : w \end{bmatrix} \rightarrow \varepsilon$
  2. $\begin{bmatrix} \text{F}_2 : w \end{bmatrix} \rightarrow \begin{bmatrix} \text{F}_1 : u \\ \text{F}_2 : v \end{bmatrix} \begin{bmatrix} \text{F}_2 : w \end{bmatrix} \begin{bmatrix} \text{F}_1 : v \\ \text{F}_2 : u \end{bmatrix}$

Then the context-free grammar $G^{cf} = \langle V_N, V_t, \mathcal{R}^{cf}, S^{cf} \rangle = ug2cfg(G^u)$ *is defined as:*

- $V_N =$

$$\left\{ \begin{bmatrix} \text{F}_2 : v \end{bmatrix}, \begin{bmatrix} \text{F}_2 : u \end{bmatrix}, \begin{bmatrix} \text{F}_2 : w \end{bmatrix}, \begin{bmatrix} \text{F}_1 : w \\ \text{F}_2 : w \end{bmatrix}, \begin{bmatrix} \text{F}_1 : u \\ \text{F}_2 : v \end{bmatrix}, \begin{bmatrix} \text{F}_1 : v \\ \text{F}_2 : u \end{bmatrix} \right\}$$

- $V_t = \textsc{Words} = \{a, b\}$
- $S^{cf} = \mathsf{A}^s = \begin{bmatrix} \text{F}_1 : w \\ \text{F}_2 : w \end{bmatrix}$
- *The set of rules $\mathcal{R}^{cf}$ is defined as:*

  1. $\begin{bmatrix} \text{F}_1 : u \\ \text{F}_2 : v \end{bmatrix} \rightarrow a$
  2. $\begin{bmatrix} \text{F}_1 : v \\ \text{F}_2 : u \end{bmatrix} \rightarrow b$
  3. $\begin{bmatrix} \text{F}_1 : w \\ \text{F}_2 : w \end{bmatrix} \rightarrow \epsilon$
  4. $\begin{bmatrix} \text{F}_2 : w \end{bmatrix} \rightarrow \epsilon$
  5. $\begin{bmatrix} \text{F}_1 : w \\ \text{F}_2 : w \end{bmatrix} \rightarrow \begin{bmatrix} \text{F}_1 : u \\ \text{F}_2 : v \end{bmatrix} \begin{bmatrix} \text{F}_2 : w \end{bmatrix} \begin{bmatrix} \text{F}_1 : v \\ \text{F}_2 : u \end{bmatrix}$
  6. $\begin{bmatrix} \text{F}_2 : w \end{bmatrix} \rightarrow \begin{bmatrix} \text{F}_1 : u \\ \text{F}_2 : v \end{bmatrix} \begin{bmatrix} \text{F}_2 : w \end{bmatrix} \begin{bmatrix} \text{F}_1 : v \\ \text{F}_2 : u \end{bmatrix}$

Note that the size of *ug2cfg*(G^u) is polynomial in the size of G^u: $|\mathcal{R}^{cf}| \leq |\mathcal{R}^u| \times |\mathcal{R}^u|$. The following lemma shows that non-reentrant unification grammars are very limited, and in particular cannot "add information" beyond that which exists in the rules: if A_i is an element of a sentential form induced by such a grammar, then A_i is an element in the body of some grammar rule.

Lemma 2. *Let $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle$ be a non-reentrant unification grammar over the signature $\langle \textsc{Atoms}, \textsc{Feats}, \textsc{Words} \rangle$ and $\mathsf{A}^s \overset{*}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$ be a derivation sequence. Then for all A_i there exist a rule $r^u \in \mathcal{R}^u$ such that $r^u = \mathsf{B}_0 \rightarrow \mathsf{B}_1 \ldots \mathsf{B}_m$ and an index j, $0 < j \leq m$, for which $\mathsf{B}_j = \mathsf{A}_i$.*

Proof. We prove by induction on the length of the derivation sequence. The induction hypothesis is that if $\mathsf{A}^s \overset{k}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$ then for all A_i, where $1 \leq i \leq n$, there are a rule $r^u = \mathsf{B}_0 \rightarrow \mathsf{B}_1 \ldots \mathsf{B}_m$, $r^u \in \mathcal{R}^u$ and an index j such that $\mathsf{B}_j = \mathsf{A}_i$. If $k = 1$, then there is a rule $\mathsf{C} \rightarrow \mathsf{A}_1 \ldots \mathsf{A}_n$, $\mathsf{A}^s \sqcup \mathsf{C} \neq \top$, and all A_i are part of the rule's body because a non-reentrant rule does not propagate information from the rule head to the body. Assume that the hypothesis holds for every l, $0 < l < k$; let the length of the derivation sequence be k. If $\mathsf{A}^s \stackrel{k-1}{\Longrightarrow}_u \mathsf{D}_1 \ldots \mathsf{D}_m \stackrel{1}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$ then there exist an index j and a rule $r^u = \mathsf{C} \rightarrow \mathsf{A}_j \ldots \mathsf{A}_{n-m+j} \in \mathcal{R}^u$ such that:

1. $\mathsf{C} \sqcup \mathsf{D}_j \neq \top$
2. $\mathsf{D}_i = \begin{cases} \mathsf{A}_i & i < j \\ \mathsf{A}_{i+n-m} & i > j \end{cases}$

By the induction hypothesis for all A_i, where $i < j$ or $i > n - m + j$, there is a rule that contains A_i in its body. For A_i, where $j \leq i \leq n - m + j$, the rule r^u completes the proof. □

With this lemma we can now prove the main result of this chapter.

Theorem 3. *Let $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle$ be a non-reentrant unification grammar over the signature $\langle$*ATOMS, FEATS, WORDS*$\rangle$ and $G^{cf} = \langle V_N, V_t, \mathcal{R}^{cf}, S^{cf} \rangle =$ ug2cfg(G^u). Then $L(G^{cf}) = L(G^u)$.*

Proof. We prove by induction on the length of a derivation sequence that $\mathsf{A}^s \stackrel{*}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$ iff $S^{cf} \stackrel{*}{\Longrightarrow}_{cf} \mathsf{A}_1 \ldots \mathsf{A}_n$.

Assume that $\mathsf{A}^s \stackrel{*}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$. The induction hypothesis is that if $\mathsf{A}^s \stackrel{k}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$ then $S^{cf} \stackrel{k}{\Longrightarrow}_{cf} \mathsf{A}_1 \ldots \mathsf{A}_n$. If $k = 1$, then there is a rule $\mathsf{C} \rightarrow \mathsf{A}_1 \ldots \mathsf{A}_n$, $\mathsf{A}^s \sqcup \mathsf{C} \neq \top$, and by the definition of *ug2cfg*, $S^{cf} \rightarrow \mathsf{A}_1 \ldots \mathsf{A}_n \in \mathcal{R}^{cf}$. Then $S^{cf} \stackrel{k=1}{\Longrightarrow}_{cf} \mathsf{A}_1 \ldots \mathsf{A}_n$. Assume that the hypothesis holds for every l, $0 < l < k$; let the length of the derivation sequence be k. If $\mathsf{A}^s \stackrel{k-1}{\Longrightarrow}_u \mathsf{D}_1 \ldots \mathsf{D}_m \stackrel{1}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$ then there exist an index j and a rule $r^u_1 = \mathsf{C} \rightarrow \mathsf{A}_j \ldots \mathsf{A}_{n-m+j} \in \mathcal{R}^u$ such that:

1. $\mathsf{C} \sqcup \mathsf{D}_j \neq \top$
2. $\mathsf{D}_i = \begin{cases} \mathsf{A}_i & i < j \\ \mathsf{A}_{i+n-m} & i > j \end{cases}$

By lemma 2 there is some rule $r^u_2 \in \mathcal{R}^u$ such that D_j is an element of its body. Hence, by definition 12 there is a rule $r_3 = \mathsf{D}_j \rightarrow \mathsf{A}_j \ldots \mathsf{A}_{n-m+j} \in \mathcal{R}^{cf}$ which is a result of combining r^u_1 and r^u_2. By the induction hypothesis $S^{cf} \stackrel{k-1}{\Longrightarrow}_{cf} \mathsf{D}_1 \ldots \mathsf{D}_n$, and by application of the rule r_3 we obtain:

$$S^{cf} \stackrel{k}{\Longrightarrow}_{cf} \mathsf{D}_1 \ldots \mathsf{D}_{j-1} \mathsf{A}_j \ldots \mathsf{A}_{j+n-m} \mathsf{D}_{j+1} \ldots \mathsf{D}_m = \mathsf{A}_1 \ldots \mathsf{A}_n$$

Assume $S^{cf} \overset{*}{\Longrightarrow}_{cf} \mathsf{A}_1 \ldots \mathsf{A}_n$.[1] The induction hypothesis is that if $S^{cf} \overset{k}{\Longrightarrow}_{cf} \mathsf{A}_1 \ldots \mathsf{A}_n$ then $\mathsf{A}^s \overset{k}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$. If $k = 1$, then there is a rule $S^{cf} \rightarrow \mathsf{A}_1 \ldots \mathsf{A}_n \in \mathcal{R}^{cf}$ and by definition of *ug2cfg* (note that S^{cf} is not a part of any rule body in $\mathcal{R}^{cf}$), $\mathsf{C} \rightarrow \mathsf{A}_1 \ldots \mathsf{A}_n \in \mathcal{R}^u$, where $\mathsf{A}^s \sqcup \mathsf{C} \neq \top$. Then $\mathsf{A}^s \overset{k=1}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$. Assume that the hypothesis holds for every i, $0 < i < k$; let the length of the derivation sequence be k. If $S^{cf} \overset{k-1}{\Longrightarrow}_{cf} \mathsf{D}_1 \ldots \mathsf{D}_m \overset{1}{\Longrightarrow}_{cf} \mathsf{A}_1 \ldots \mathsf{A}_n$ then there exist an index j and a rule $r_1 = \mathsf{D}_j \rightarrow \mathsf{A}_j \ldots \mathsf{A}_{n-m+j} \in \mathcal{R}^{cf}$ such that:

$$\mathsf{D}_i = \begin{cases} \mathsf{A}_i & i < j \\ \mathsf{A}_{i+n-m} & i > j \end{cases}$$

By definition 11 there are rules $r_2^u = \mathsf{B}_0 \rightarrow \mathsf{B}_1 \ldots \mathsf{B}_p$, $r_3^u = \mathsf{C} \rightarrow \mathsf{A}_j \ldots \mathsf{A}_{n-m+j}$ in $\mathcal{R}^u$ and an index t, $1 \leq t \leq p$, such that $\mathsf{B}_t = \mathsf{D}_j$ and $\mathsf{C} \sqcup \mathsf{D}_j \neq \top$.

By the induction hypothesis, $\mathsf{A}^s \overset{k-1}{\Longrightarrow}_u \mathsf{D}_1 \ldots \mathsf{D}_n$, and by application of the rule r_3^u we obtain:

$$\mathsf{A}^s \overset{k}{\Longrightarrow}_u \mathsf{D}_1 \ldots \mathsf{D}_{j-1}\mathsf{A}_j \ldots \mathsf{A}_{j+n-m}\mathsf{D}_{j+1} \ldots \mathsf{D}_m = \mathsf{A}_1 \ldots \mathsf{A}_n$$

In sum, $\mathsf{A}^s \overset{*}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$ iff $S^{cf} \overset{*}{\Longrightarrow}_{cf} \mathsf{A}_1 \ldots \mathsf{A}_n$. Hence, $L(G^{cf}) = L(\mathit{ug2cfg}(G^u))$. □

Corollary 4. *The class of languages generated by non-reentrant unification grammars (UG_{nr}) is equivalent to the class of context-free languages.*

Proof. Immediate from theorem 1 and theorem 3. □

Definition 13 (Atomic Unification Grammars (AUG)). *A unification grammar $G^u = \langle \mathcal{R}^u, \mathcal{L}, \mathsf{A}^s \rangle$ is* ***atomic*** *if all rules in $\mathcal{R}^u$ contains only atomic feature structures (feature structures define by case 1 of definition 1).*

Since AUG is just a notational variant of CFG it does emphasize the idea that non-reentrant feature structures add nothing of substance to UG, at least in terms of weak generative capacity.

[1] Recall that all elements of V_N are feature structures, and therefore all the elements of a (CFG) sentential form can be represented as A_i, where A_i is a feature structure.

Chapter 3

Mildly Context Sensitive Unification Grammars

In this section we define a constraint on unification grammars which ensures that grammars satisfying it generate all and only the mildly context-sensitive languages. In section 3.1 we recall one of the mildly context-sensitive formalisms, Linear Indexed Grammar. Section 3.2 defines the constraint, namely one-reentrant unification grammars. Then, in section 3.3 we show the mapping of LIGs to one-reentrant unification grammars. The mapping of one-reentrant unification grammars to LIGs is shown in section 3.4.

3.1 Linear Indexed Grammars

In this section we use the definition of Vijay-Shanker and Weir (1994). In a Linear indexed grammar (LIG), strings are derived from nonterminals with an associated stack denoted $A[l_1 \dots l_n]$, where A is a nonterminal, each l_i is a stack symbol for $1 \leq i \leq n$, and l_1 is the top of the stack. $A[\,]$ denotes the nonterminal A associated with the empty stack. Since stacks can grow to be of unbounded size during a derivation, some way of partially specifying unbounded stacks in LIG productions is needed. We use $A[l_1 \dots l_n ..]$ to denote the nonterminal A associated with any stack η whose top n symbols are $l_1, l_2 \dots, l_n$ where $0 \leq n$. The set of all nonterminals in V_N, associated with stacks whose symbols come from V_s, is denoted $V_N[V_s^*]$.

Definition 14. *A **Linear Indexed Grammar** is a five tuple* $G^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, S\rangle$ *where*

- V_N *is a finite set of nonterminals,*
- V_t *is a finite set of terminals,*
- V_s *is a finite set of indices (stack symbols),*
- $S \in V_N$ *is the start symbol and*
- $\mathcal{R}^{li}$ *is a finite set of productions, having one of the following two forms:*

  1. *Production with a **fixed** stack at the head:* $N_i[p_1 \dots p_n] \rightarrow \alpha$
  2. *Production with an **unbounded** stack at the head:* $N_i[p_1 \dots p_n..] \rightarrow \alpha N_j[q_1 \dots q_m..]\beta$

  where $N_i, N_j \in V_N$, $p_1 \dots p_n, q_1 \dots q_m \in V_s$, $n, m \geq 0$ *and* $\alpha, \beta \in (V_t \cup V_N[V_s^*])^*$.

Definition 15. *Given a LIG* $G^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, S\rangle$*, the **derivation relation** '*$\Longrightarrow_{li}$*' is defined as follows:*

- *If* $N_i[p_1 \dots p_n] \rightarrow \alpha \in \mathcal{R}^{li}$ *then for all* $\Psi_1, \Psi_2 \in (V_N[V_s^*] \cup V_t)^*$,

$$\Psi_1 N_i[p_1 \dots p_n]\Psi_2 \Longrightarrow_{li} \Psi_1 \alpha \Psi_2$$

- *If* $N_i[p_1 \dots p_n..] \rightarrow \alpha N_j[q_1 \dots q_m..]\beta \in \mathcal{R}^{li}$ *then for all* $\Psi_1, \Psi_2 \in (V_N[V_s^*] \cup V_t)^*$ *and* $\eta \in V_s^*$,

$$\Psi_1 N_i[p_1 \dots p_n \eta]\Psi_2 \Longrightarrow_{li} \Psi_1 \alpha N_j[q_1 \dots q_m \eta]\beta \Psi_2$$

where $N_i, N_j \in V_N$, $p_1 \dots p_n, q_1 \dots q_m \in V_s$, $n, m \geq 0$ *and* $\alpha, \beta \in (V_t \cup V_N[V_s^*])^*$. *The **language**,* $L(G^{li})$*, generated by* G^{li}*, is* $\{w \in V_t^* \mid S[\,] \overset{*}{\Longrightarrow}_{li} w\}$*, where '*$\overset{*}{\Longrightarrow}_{li}$*' is the reflexive, transitive closure of '*$\Longrightarrow_{li}$*'.*

We change the definition above by adding the following production form with an *unbounded* stack at the head:

$$N_i[p_1 \dots p_n..] \rightarrow \alpha$$

where $N_i \in V_N$, $p_1 \ldots p_n \in V_s$, $0 \leq n$ and $\alpha \in (V_t \cup V_N[V_s^*])^*$. The derivation relation '$\Longrightarrow_{li}$' for the added production form is defined as follows:

$$\Psi_1 N_i[p_1 \ldots p_n \eta]\Psi_2 \Longrightarrow_{li} \Psi_1 \alpha \Psi_2$$

where $\Psi_1, \Psi_2 \in (V_N[V_s^*] \cup V_t)^*$. It is easy to see that such productions can be simulated by the two production forms given in definition 14, so the extended formalism is (weakly) equivalent to the original one.

LIG is one of the four formalisms that are known to be mildly context-sensitive. The following languages are known to be MCS:

- $L_1 = \{ww^R ww^R \mid w \in \{a, b\}\}$
- $L_2 = \{ww \mid w \in \{a, b\}\}$
- $L_3 = \{a^n b^n c^n d^n \mid 0 \leq n\}$

To demonstrate the expressiveness of this class of languages we provide below a grammar for L_2.

Example 4 (LIG for L_2). *Let $G_2^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, S\rangle$, where:*

- $V_N = \{S, N_2, N_3\}$
- $V_t = \{a, b\}$
- $V_s = V_t$
- $\mathcal{R}^{li} = \{r_1, r_2, r_3, r_4, r_5, r_6, r_7\}$, *where*
  1. $r_1 = S[\,] \rightarrow N_2[\,]$
  2. $r_2 = N_2[..] \rightarrow N_2[a..]a$
  3. $r_3 = N_2[..] \rightarrow N_2[b..]b$
  4. $r_4 = N_2[..] \rightarrow N_3[..]$
  5. $r_5 = N_3[a..] \rightarrow aN_3[..]$
  6. $r_6 = N_3[b..] \rightarrow bN_3[..]$

7. $r_7 = N_3[\,] \rightarrow \epsilon$

It is easy to see that $L(G_2^{li}) = L_2$. *For example, a derivation of the word* $abbabb$ *is*

$$\begin{aligned} S \Longrightarrow_{li} & \quad N_2[\,] & r_1 \\ \Longrightarrow_{li} & \quad N_2[b]b & r_3 \\ \Longrightarrow_{li} & \quad N_2[bb]bb & r_3 \\ \Longrightarrow_{li} & \quad N_2[abb]abb & r_2 \\ \Longrightarrow_{li} & \quad N_3[abb]abb & r_4 \\ \Longrightarrow_{li} & \quad aN_3[bb]abb & r_5 \\ \Longrightarrow_{li} & \quad abN_3[b]abb & r_6 \\ \Longrightarrow_{li} & \quad abbN_3[\,]abb & r_6 \\ \Longrightarrow_{li} & \quad abbabb & r_7 \end{aligned}$$

In contrast, seemingly similar languages are beyond mildly context-sensitive and hence cannot be generated by LIG (Vijay-Shanker and Weir, 1994):

- $L_4 = \{www \mid w \in \{a, b\}\}$
- $L_5 = \{a^{n^2} \mid 0 \leq n\}$
- $L_6 = \{a^n b^n c^n d^n e^n \mid 0 \leq n\}$

A crucial characteristic of LIG is that only *one* copy of the stack can be copied to a *single* element in the body of a rule. Once more than one copy is allowed, the expressive power grows beyond MCS. This is demonstrated by the following definition and examples.

Definition 16. *Linear indexed grammar 2 (LIG2) is an extension of LIG. The difference is in the definition of the productions set, where one more rule form is allowed:*

$$N_i[p_1 \ldots p_n..] \rightarrow N_j[q_1 \ldots q_m..]N_k[r_1 \ldots r_l..]$$

Where $N_i, N_j, N_k \in V_N$ *and* $p_1 \ldots p_n, q_1 \ldots q_m, r_1 \ldots r_l \in V_s$. *The derivation relation '*$\Longrightarrow_{li}$*' for the production form is defined as follows:*

$$\Psi_1 N_i[p_1 \ldots p_n \eta]\Psi_2 \Longrightarrow_{li} \Psi_1 N_j[q_1 \ldots q_m \eta]N_k[r_1 \ldots r_l \eta]\Psi_2$$

where $\Psi_1, \Psi_2 \in (V_N[V_s^*] \cup V_t)^*$.

We demonstrate the additional expressiveness by providing LIG2 grammars for L_4 and L_5, which are trans-MCS. Note that the grammar of example 5 is obtained from the grammar of example 4 by adding a single rule, r_4.

Example 5 (LIG2 for L_4). *Let* $G_4^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, S \rangle$, *where:*

- $V_N = \{S, N_2, N_3\}$
- $V_t = \{a, b\}$
- $V_s = V_t$
- $\mathcal{R}^{li} = \{r_1, r_2, r_3, r_4, r_5, r_6, r_7\}$, *where*

  1. $r_1 = S \to N_2[\,]$
  2. $r_2 = N_2[..] \to N_2[a..]a$
  3. $r_3 = N_2[..] \to N_2[b..]b$
  4. $r_4 = N_2[..] \to N_3[..]N_3[..]$
  5. $r_5 = N_3[a..] \to aN_3[..]$
  6. $r_6 = N_3[b..] \to bN_3[..]$
  7. $r_7 = N_3[\,] \to \epsilon$

It is easy to see that $L(G_4^{li}) = L_4$. *For example, a derivation of the word* $abbabbabb$ *is*

$$\begin{array}{rcl} S \Longrightarrow_{li2} & N_2[\,] & r_1 \\ \Longrightarrow_{li2} & N_2[b]b & r_3 \\ \Longrightarrow_{li2} & N_2[bb]bb & r_3 \\ \Longrightarrow_{li2} & N_2[abb]abb & r_2 \\ \Longrightarrow_{li2} & N_3[abb]N_3[abb]abb & r_4 \\ \Longrightarrow_{li2} & aN_3[bb]N_3[abb]abb & r_5 \end{array}$$

$$
\begin{aligned}
&\Longrightarrow_{li2} && abN_3[b]N_3[abb]abb && r_6 \\
&\Longrightarrow_{li2} && abbN_3[\,]N_3[abb]abb && r_6 \\
&\Longrightarrow_{li2} && abbN_3[abb]abb && r_7 \\
&\Longrightarrow_{li2} && abbaN_3[bb]abb && r_5 \\
&\Longrightarrow_{li2} && abbabN_3[b]abb && r_6 \\
&\Longrightarrow_{li2} && abbabbN_3[\,]abb && r_6 \\
&\Longrightarrow_{li2} && abbabbabb && r_7
\end{aligned}
$$

Example 6 (LIG2 for L_5). *Let* $G_5^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, S\rangle$*, where:*

- $V_N = \{S, N_2, N_3\}$
- $V_t = \{a\}$
- $V_s = V_t$
- $\mathcal{R}^{li} = r_1, r_2, r_3, r_4, r_5, r_6, r_7$*, where*
  1. $r_1 = S \rightarrow N_2[a]$
  2. $r_2 = N_2[..] \rightarrow N_3[..]N_2[aa..]$
  3. $r_3 = N_2[..] \rightarrow N_3[..]$
  4. $r_4 = N_3[a..] \rightarrow aN_3[..]$
  5. $r_5 = N_3[\,] \rightarrow \epsilon$

The grammar is based on the observation that $n^2 = 1 + 3 + 5 + \ldots + (2n - 2)$*. It is easy to see that* $L(G_5^{li}) = L_5$*. For example, a derivation of the word* $aaaa$ *is*

$$
\begin{aligned}
S &\Longrightarrow_{li2} && N_2[a] && r_1 \\
&\Longrightarrow_{li2} && N_3[a]N_2[aaa] && r_2 \\
&\Longrightarrow_{li2} && N_3[a]N_3[aaa] && r_3 \\
&\overset{*}{\Longrightarrow}_{li2} && aN_3[\,]aaaN_3[\,] && r_4 \times 4 \\
&\overset{*}{\Longrightarrow}_{li2} && aaaa && r_5 \times 2
\end{aligned}
$$

3.2 One-reentrant Unification Grammars

In this section we define a constrained variant of unification grammars, namely *one-reentrant unification grammars*, that generates exactly the mildly context-sensitive class of languages. The major constraint on unification grammars is that each rule can include at most one reentrancy, reflecting the LIG situation where stacks can be copied to exactly one daughter in each rule.

Definition 17 (One-reentrant unification grammar). *A unification grammar $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle$ over the signature $\sigma = \langle \textsc{Atoms}, \textsc{Feats}, \textsc{Words} \rangle$ is* ***one-reentrant*** *iff for every rule $r^u \in \mathcal{R}^u$, r^u includes at most one reentrancy, between the head of the rule and some element of the body.*

Let UG_{1r} be the set of all one-reentrant unification grammars.

Example 7 (One-reentrant unification grammar). *Let $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle$ be a one-reentrant unification grammar over the signature $\langle \textsc{Atoms}, \textsc{Feats}, \textsc{Words} \rangle$, such that*

- $\textsc{Atoms} = \{s, t, u, v\}$
- $\textsc{Feats} = \{\textsc{f}, \textsc{g}\}$
- $\textsc{Words} = \{a, b, c, d\}$;
- $\mathsf{A}^s = \begin{bmatrix} \textsc{f} : s \\ \textsc{g} : s \end{bmatrix}$
- *The lexicon is defined as $\mathcal{L}(a) = \{s\}$, $\mathcal{L}(b) = \{t\}$, $\mathcal{L}(c) = \{u\}$ and $\mathcal{L}(d) = \{v\}$.*
- *The set of productions $\mathcal{R}^u$ is defined as*

  1. $\begin{bmatrix} \textsc{f} : s \\ \textsc{g} : \boxed{1} \end{bmatrix} \rightarrow s \begin{bmatrix} \textsc{f} : s \\ \textsc{g} : \begin{bmatrix} \textsc{f} : s \\ \textsc{g} : \boxed{1} \end{bmatrix} \end{bmatrix} v$
  2. $\begin{bmatrix} \textsc{f} : s \\ \textsc{g} : \boxed{1} \end{bmatrix} \rightarrow \begin{bmatrix} \textsc{f} : t \\ \textsc{g} : \boxed{1} \end{bmatrix}$
  3. $\begin{bmatrix} \textsc{f} : t \\ \textsc{g} : \begin{bmatrix} \textsc{f} : s \\ \textsc{g} : \boxed{1} \end{bmatrix} \end{bmatrix} \rightarrow t \begin{bmatrix} \textsc{f} : t \\ \textsc{g} : \boxed{1} \end{bmatrix} u$

4. $\begin{bmatrix} \text{F} : t \\ \text{G} : s \end{bmatrix} \rightarrow \epsilon$

Then $L(G^u) = \{a^n b^n c^n d^n \mid n \geq 0\}$.

One-reentrant unification grammars induce highly constrained (sentential) forms: in such forms, there are no reentrancies whatsoever, neither between distinct elements nor within a single element.

Lemma 5. *If τ is a sentential form induced by a one-reentrant grammar then there are no reentrancies between elements of τ or within an element of τ.*

Proof. By simple induction on the length of a derivation sequence. The proposition follows directly from the fact that rules in a one-reentrant unification grammar have no reentrancies between elements of their body. □

Since all the feature structures in forms induced by a one-reentrant unification grammar are non-reentrant, unification is simplified. The following property is phrased in terms of abstract feature structures (see definition 3):

Lemma 6. *Let* A *and* B *be unifiable non-reentrant feature structures. Then* $\mathsf{C} = \mathsf{A} \sqcup \mathsf{B}$ *is defined as follows:*

- $\Pi_C = \Pi_A \cup \Pi_B$
- $\Theta_C(\pi) = \begin{cases} \Theta_A(\pi) & \textit{if } \Theta_A(\pi)\downarrow \\ \Theta_B(\pi) & \textit{if } \Theta_A(\pi)\uparrow \textit{ and } \Theta_B(\pi)\downarrow \\ \textit{undefined} & \textit{otherwise} \end{cases}$
- $\approx_C = \{(\pi, \pi) \mid \pi \in \Pi_C\}$

Crucially, C is also a *non-reentrant* feature structure whose set of paths is the union of Π_A and Π_B.

Proof. Immediate from the definition of unification. □

To simplify the construction of a mapping from LIG to UG, we first define a simplified variant of one-reentrant unification grammars, which we presently prove to be equivalent to the original definition.

Definition 18 (Simplified one-reentrant unification grammars). *A one-reentrant unification grammar* $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle$ *over the signature* $\sigma = \langle \textsc{Atoms}, \textsc{Feats}, \textsc{Words} \rangle$ *is* ***simplified*** *iff the lexical categories of words are inconsistent with any feature structure (except themselves). Formally, if* τ *is a sentential form induced by* G^u *and* τ^i *is an element of* τ *then for each word* $a \in \textsc{Words}$*,* $\mathcal{L}(a) = \{\mathsf{A}\}$*, where* $\mathsf{A} \sqcup \tau^i \neq \top$ *iff* $\mathsf{A} = \tau^i$*.*

Definition 19 (Lexicon simplification procedure). *Let lexSmp be a mapping of one-reentrant UGs to simplified one-reentrant UGs such that if* $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle$ *over the signature* $\langle \textsc{Atoms}, \textsc{Feats}, \textsc{Words} \rangle$ *is a one-reentrant UG and lexSmp*$(G^u) = \widehat{G^u} = \langle \widehat{\mathcal{R}^u}, \mathsf{A}^s, \widehat{\mathcal{L}} \rangle$*, then* $\widehat{G^u}$ *is over the signature* $\langle \widehat{\textsc{Atoms}}, \textsc{Feats}, \textsc{Words} \rangle$ *where:*

- $\widehat{\textsc{Atoms}} = \textsc{Atoms} \cup \textsc{Words}$
- *If* $a \in \textsc{Words}$ *then* $\widehat{\mathcal{L}}(a) = \{a\}$*. Note that* a *is a word, whereas* $\{a\}$ *is a set of feature structures that includes a single feature structure consisting of the single atom* a*.*
- $\widehat{\mathcal{R}^u} = \mathcal{R}^u \cup \{\mathsf{A} \to a \mid \mathsf{A} \in \mathcal{L}(a)\}$

Trivially, *lexSmp*(G^u) is a simplified one-reentrant unification grammar. It is also easy to verify that $L(G^u) = L(\textit{lexSmp}(G^u))$. In the rest of this section we restrict the discussion to simplified one-reentrant unification grammars.

3.3 Mapping of Linear Indexed Grammars to one-reentrant Unification Grammars

In order to simulate a given LIG with a unification grammar, a dedicated signature is defined based on the parameters of the LIG.

Definition 20. *Given a LIG* $G^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, S \rangle$*, let* τ *be* $\langle \textsc{Atoms}, \textsc{Feats}, \textsc{Words} \rangle$*, where*

- $\textsc{Atoms} = V_N \cup V_s \cup \{elist\}$*;*
- $\textsc{Feats} = \{\textsc{head}, \textsc{tail}\}$*;*
- $\textsc{Words} = V_t$*;*

We use τ throughout this section as the signature over which unification grammars are defined. We will use feature structures over the signature τ to represent and simulate LIG symbols. In particular, feature structures will encode lists in the natural way, hence the features HEAD and TAIL. For the sake of brevity, we use standard list notation when feature structures encode lists. Thus,

1. If $A = \begin{bmatrix} \text{HEAD} : p_1 \\ \text{TAIL} : \begin{bmatrix} \ddots & \\ & \begin{bmatrix} \text{HEAD} : p_n \\ \text{TAIL} : \; elist \end{bmatrix} \end{bmatrix} \end{bmatrix}$, then A is depicted as $\langle p_1, \ldots, p_n \rangle$.

   Note that an empty list, $\langle\rangle$ depicts the feature structure $\begin{bmatrix} \text{TAIL} : elist \end{bmatrix}$.

2. If $A = \begin{bmatrix} \text{HEAD} : p_1 \\ \text{TAIL} : \begin{bmatrix} \ddots & \\ & \begin{bmatrix} \text{HEAD} : p_n \\ \text{TAIL} : \; \boxed{\text{i}} \; [\,] \end{bmatrix} \end{bmatrix} \end{bmatrix}$, then A is depicted as $\langle p_1, \ldots, p_n, \boxed{\text{i}} \; \rangle$.

   Note that the list, $\langle \; \boxed{\text{i}} \; \rangle$ depicts the feature structure $\begin{bmatrix} \text{TAIL} : \boxed{\text{i}} \; [\,] \end{bmatrix}$.

With this list representation, LIG symbols are mapped to feature structures as follows.

Definition 21 (Mapping of LIG symbols to feature structures). *Let toFs be a mapping of a linear indexed grammar symbols to feature structures, such that:*

1. *If $t \in V_t$ then toFs$(t) = \langle t \rangle$*

2. *If $N \in V_N$ and $\eta \in V_s^*$, then toFs$(N[\eta]) = \langle N \rangle \cdot \eta$*

Example 8. *Let $G^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, S \rangle$ be a LIG such that $V_N = \{S\}$, $V_t = \{t_1, t_2\}$ and $V_s = \{s_1, s_2\}$. Then*

$$\begin{aligned} \textit{toFs}(S[s_1]) &= \langle S, s_1 \rangle \\ \textit{toFs}(t_1) &= \langle t_1 \rangle \\ \textit{toFs}(S[s_2, s_1, s_1]) &= \langle S, s_2, s_1, s_1 \rangle \end{aligned}$$

When feature structures that are images of LIG symbols are concerned, unification is reduced to identity, as the following lemma shows.

Lemma 7. *Let* $X_1, X_2 \in V_N[V_s^*] \cup V_t$*. If* $\textit{toFs}(X_1) \sqcup \textit{toFs}(X_2) \neq \top$ *then* $\textit{toFs}(X_1) = \textit{toFs}(X_2)$*.*

Proof. Simple induction on the length of X_1. $\square$

When a feature structure which is represented as an unbounded list (a list that is not terminated by $elist$) is unifiable with an image of a LIG symbol, the former is a prefix of the latter.

Lemma 8. *Let* $X \in V_N[V_s^*] \cup V_t$ *and* $\mathsf{C} = \langle p_1, \ldots, p_n, \boxed{i} \rangle$ *be a non-reentrant feature structure, where* $p_1, \ldots, p_n \in V_s$*. Then* $\mathsf{C} \sqcup \textit{toFs}(X) \neq \top$ *iff* $\textit{toFs}(X) = \langle p_1, \ldots, p_n \rangle \cdot \alpha$*, where* $\alpha \in V_s^*$*.*

Proof. Assume that $\mathsf{C} \sqcup \textit{toFs}(X) \neq \top$. By definition 21, $\textit{toFs}(X)$ is a feature structure that is represented as a list, terminated by $elist$, whose elements are atoms. Hence by definition of unification, the prefix of length n of $\textit{toFs}(X)$ equals to $\langle p_1, \ldots, p_n \rangle$. Therefore, $\textit{toFs}(X) = \langle p_1, \ldots, p_n \rangle \cdot \alpha$, where $\alpha \in V_s^*$.

Assume that $\mathsf{C} \sqcup \textit{toFs}(X) = \top$. By definition of unification for all $\alpha \in V_s^*$, $\langle p_1, \ldots, p_n, \boxed{i} \rangle \sqcup \langle p_1, \ldots, p_n \rangle \cdot \alpha \neq \top$. Therefore, we obtain that for all $\alpha \in V_s^*$, $\textit{toFs}(X) \neq \langle p_1, \ldots, p_n \rangle \cdot \alpha$. $\square$

The mapping *toFs* is extended to sequences of symbols in the natural way, by setting $\textit{toFs}(\alpha\beta) = \textit{toFs}(\alpha)\textit{toFs}(\beta)$, where $\alpha, \beta \in (V_N[V_s^*] \cup V_t)^*$. Note that the mapping is one to one because the LIG symbol can be deterministicly restored from its image (the feature structure). If the list contains only a single element then the LIG symbol is either a terminal symbol or a non-terminal symbol with an empty stack. When the list representation of a feature structure consists of more than one element, the first element of the list is a non-terminal symbol and the remainder of the list is the non-terminal stack content.

To simulate LIG with a unification grammar we represent each LIG symbol in the grammar as a feature structure, encoding the stack of LIG non-terminals as lists. Rules that propagate stacks (from mother to daughter) are simulated by means of value sharing (reentrancy) in the unification grammar.

Definition 22 (Mapping from LIGS **to** UG_{1r}**).** *Let lig2ug be a mapping of* LIGS *to* UG_{1r}*, such that if* $G^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, S \rangle$ *and* $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle = \textit{lig2ug}(G^{li})$ *then* G^u *is over the signature* τ *(definition 20) and:*

- $\mathsf{A}^s = \textit{toFs}(S[\,])$

- *For all* $t \in V_t$, $\mathcal{L}(t) = \{\textit{toFs}(t)\}$.

- $\mathcal{R}^u$ *is defined by:*

  1. *A LIG rule of the type* $X_0 \to \alpha$*, where* $X_0 \in V_N[V_s^*]$ *and* $\alpha \in (V_N[V_s^*] \cup V_t)^*$*, is mapped to the unification rule:*

$$\textit{toFs}(X_0) \to \textit{toFs}(\alpha)$$

  2. *A LIG rule of the type* $N_i[p_1, \ldots, p_n..] \to \alpha\ N_j[q_1, \ldots, q_m..]\ \beta$*, where* $\alpha, \beta \in (V_N[V_s^*] \cup V_t)^*$*,* $N_i, N_j \in V_N$ *and* $p_1, \ldots, p_n, q_1, \ldots, q_m \in V_s$*, is mapped to the unification rule:*

$$\langle N_i, p_1, \ldots, p_n, \boxed{1}\ \rangle \to \textit{toFs}(\alpha)\ \langle N_j, q_1, \ldots, q_m, \boxed{1}\ \rangle\ \textit{toFs}(\beta)$$

Evidently, $\textit{lig2ug}(G^{li}) \in \mathrm{UG}_{1r}$ for any LI grammar G^{li} because each of its rules has at most one reentrancy.

Example 9 (Mapping from LIGS **to** UG_{1r}**).** *We map the LIG* G_2^{li} *of example 4 above to* $G^u = \textit{lig2ug}(G^{li})$ *defined above the signature* τ *of definition 20, with the start symbol* $\textit{toFs}(S[\ ])$*. The lexicon is defined for the words* a *and* b *as* $\mathcal{L}(a) = \{\langle a \rangle\}$ *and* $\mathcal{L}(b) = \{\langle b \rangle\}$*. The set of productions* $\mathcal{R}^{li}$*, is defined as follows:*

1. $r_1^u = \langle S \rangle \to \langle N_2 \rangle$*, where the LIG rule is* $r_1 = S[\,] \to N_2[\,]$
2. $r_2^u = \langle N_2, \boxed{1}\ \rangle \to \langle N_2, a, \boxed{1}\ \rangle \langle a \rangle$*, where the LIG rule is* $r_2 = N_2[..] \to N_2[a..]a$
3. $r_3^u = \langle N_2, \boxed{1}\ \rangle \to \langle N_2, b, \boxed{1}\ \rangle \langle b \rangle$*, where the LIG rule is* $r_3 = N_2[..] \to N_2[b..]b$
4. $r_4^u = \langle N_2, \boxed{1}\ \rangle \to \langle N_3, \boxed{1}\ \rangle$*, where the LIG rule is* $r_4 = N_2[..] \to N_3[..]$
5. $r_5^u = \langle N_3, a, \boxed{1}\ \rangle \to \langle a \rangle \langle N_3, \boxed{1}\ \rangle$*, where the LIG rule is* $r_5 = N_3[a..] \to aN_3[..]$
6. $r_6^u = \langle N_3, b, \boxed{1}\ \rangle \to \langle b \rangle \langle N_3, \boxed{1}\ \rangle$*, where the LIG rule is* $r_6 = N_3[b..] \to bN_3[..]$
7. $r_7^u = \langle N_3 \rangle \to \epsilon$*, where the LIG rule is* $r_7 = N_3[\,] \to \epsilon$

Lemma 9. *The mapping lig2ug of definition 22 is one to one.*

Proof. Immediately follows from the fact that the mapping *toFs* is one-to-one. □

To show that the unification grammar *lig2ug*(G^{li}) correctly simulates the LIG grammar G^{li} we first prove that every derivation in the latter has a corresponding derivation in the former (theorem 10). Theorem 11 proves the reverse direction.

Theorem 10. *Let* $G^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, S\rangle$ *be a LIG and* $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L}\rangle$ *be lig2ug*(G^{li}). *If* $S[\,] \stackrel{*}{\Longrightarrow}_{li} \alpha$ *then* $\mathsf{A}^s \stackrel{*}{\Longrightarrow}_u$ *toFs*(α), *where* $\alpha \in (V_N[V_s^*] \cup V_t)^*$.

Proof. We prove by induction on the length of the derivation sequence. The induction hypothesis is that if $S[\,] \stackrel{k}{\Longrightarrow}_{li} \alpha$, then $\mathsf{A}^s \stackrel{k}{\Longrightarrow}_u$ *toFs*(α). If $k = 1$, then

1. $S[\,] \stackrel{k=1}{\Longrightarrow}_{li} \alpha$;
2. Hence, $S[\,] \to \alpha \in \mathcal{R}^{li}$;
3. By definition 22, *toFs*$(S) \to$ *toFs*$(\alpha) \in \mathcal{R}^u$;
4. Since $\mathsf{A}^s =$ *toFs*(S) we obtain that $\mathsf{A}^s \to$ *toFs*$(\alpha) \in \mathcal{R}^u$;
5. Therefore, $\mathsf{A}^s \stackrel{k=1}{\Longrightarrow}_u$ *toFs*(α)

Assume that the hypothesis holds for every i, $0 < i < k$; let the length of the derivation sequence be k.

1. Let $S[\,] \stackrel{k-1}{\Longrightarrow}_{li} \gamma_1\ N[p_1, \ldots, p_n]\ \gamma_2 \stackrel{1}{\Longrightarrow}_{li} \gamma_1\ \alpha\ \gamma_2$, where $\gamma_1, \gamma_2, \alpha \in (V_N[V_s^*] \cup V_t)^*$. Let $r \in \mathcal{R}^{li}$ be a LIG rule that is applied to $N[p_1, \ldots, p_n]$ at step k of the derivation.
2. By the induction hypothesis, $\mathsf{A}^s \stackrel{k-1}{\Longrightarrow}_u$ *toFs*$(\gamma_1\ N_i[p_1, \ldots, p_n]\ \gamma_2)$.
3. By definition 21,

$$\begin{aligned} \textit{toFs}(\gamma_1 N[p_1, \ldots, p_n]\gamma_2) = &\quad \textit{toFs}(\gamma_1)\ \textit{toFs}(N[p_1, \ldots, p_n])\ \textit{toFs}(\gamma_2) \\ = &\quad \textit{toFs}(\gamma_1)\ \langle N, p_1, \ldots, p_n, \rangle\ \textit{toFs}(\gamma_2) \end{aligned}$$

4. From (2) and (3), $\mathsf{A}^s \stackrel{k-1}{\Longrightarrow}_u$ *toFs*$(\gamma_1)\ \langle N, p_1, \ldots, p_n\rangle$ *toFs*(γ_2).
5. The rule r can be of either of two forms as follows:
   (a) Let r be $N[p_1, \ldots, p_n] \to \alpha$.

i. By definition 22, $\mathcal{R}^u$ includes the rule *toFs*$(N[p_1, \ldots, p_n]) \rightarrow$ *toFs*(α).

ii. This rule is applicable to the form in (4), providing $\mathsf{A}^s \overset{k}{\Longrightarrow}_u$ *toFs*(γ_1) *toFs*(α) *toFs*(γ_2).

iii. By definition 21, *toFs*(γ_1) *toFs*(α) *toFs*(γ_2) $=$ *toFs*$(\gamma_1\ \alpha\ \gamma_2)$. Hence, $\mathsf{A}^s \overset{k}{\Longrightarrow}_u$ *toFs*$(\gamma_1\ \alpha\ \gamma_2)$.

(b) Let r be $N[p_1, \ldots, p_x..] \rightarrow \alpha_1\ M[q_1, \ldots, q_m..]\ \alpha_2$, where $x \leq n$, $M \in V_N$, $q_1, \ldots, q_m \in V_s$ and $\alpha_1, \alpha_2 \in (V_N[V_s^*] \cup V_t)^*$.

i. By applying the rule r at the last derivation step in (1) we obtain:

$$S[\,] \overset{k-1}{\Longrightarrow}_{li} \gamma_1\ N[p_1, \ldots, p_n]\ \gamma_2 \overset{1}{\Longrightarrow}_{li} \gamma_1\ \alpha_1\ M[q_1, \ldots, q_m, p_{x+1}, \ldots, p_n]\ \alpha_2\ \gamma_2$$

ii. By definition 22, $\mathcal{R}^u$ includes the rule

$$\langle N, p_1, \ldots, p_x, \boxed{1}\ \rangle \rightarrow \textit{toFs}(\alpha_1)\ \langle M, q_1, \ldots, q_m, \boxed{1}\ \rangle\ \textit{toFs}(\alpha_2)$$

iii. By applying this rule to the form in (4) we obtain

$$\begin{aligned} \mathsf{A}^s \overset{k-1}{\Longrightarrow}_u &\quad \textit{toFs}(\gamma_1)\ \langle N, p_1, \ldots, p_n \rangle\ \textit{toFs}(\gamma_2) \\ \overset{1}{\Longrightarrow}_u &\quad \textit{toFs}(\gamma_1)\ \textit{toFs}(\alpha_1)\ \langle M, q_1, \ldots, q_m, p_{x+1}, \ldots, p_n \rangle\ \textit{toFs}(\alpha_2)\ \textit{toFs}(\gamma_2) \end{aligned}$$

iv. By definition 21,

$$\textit{toFs}(M[q_1, \ldots, q_m, p_{x+1}, \ldots, p_n]) = \langle M, q_1, \ldots, q_m, p_{x+1}, \ldots, p_n \rangle$$

Hence

$$\mathsf{A}^s \overset{k}{\Longrightarrow}_u \textit{toFs}(\gamma_1)\ \textit{toFs}(\alpha_1)\ \textit{toFs}(M[q_1, \ldots, q_m, p_{x+1}, \ldots, p_n])\ \textit{toFs}(\alpha_2)\ \textit{toFs}(\gamma_2)$$

v. Therefore, $\mathsf{A}^s \overset{k}{\Longrightarrow}_u$ *toFs*$(\gamma_1\ \alpha_1\ M[q_1, \ldots, q_m, p_{x+1}, \ldots, p_n]\ \alpha_2\ \gamma_2)$.

□

Theorem 11. *Let* $G^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, S \rangle$ *be a LIG and* $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle =$ *lig2ug*(G^{li}) *be a one-reentrant unification grammar. If* $\mathsf{A}^s \overset{*}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$ *then* $S[\,] \overset{*}{\Longrightarrow}_{li} X_1 \ldots X_n$ *such that for every* i*,* $1 \leq i \leq n$*,* $\mathsf{A}_i =$ *toFs*(X_i)*.*

Proof. We prove by induction on the length of the (unification) derivation sequence. The induction hypothesis is that if $\mathsf{A}^s \xRightarrow{k}_u \mathsf{A}_1 \ldots \mathsf{A}_n$, then $S[\,] \xRightarrow{k}_{li} X_1 \ldots X_n$ such that for every i, $1 \le i \le n$, $\mathsf{A}_i = \textit{toFs}(X_i)$. If $k = 1$, then $\mathsf{A}^s \xRightarrow{k=1}_u \mathsf{A}_1 \ldots \mathsf{A}_n$. Hence, $\mathsf{A}^s \rightarrow \mathsf{A}_1 \ldots \mathsf{A}_n \in \mathcal{R}^u$. By definition 22, $\mathsf{A}^s = \textit{toFs}(S[\,])$. Since *toFs* is a one-to-one mapping we obtain that the unification rule is created from the LIG rule $S^{li}[\,] \rightarrow X_1 \ldots X_n \in \mathcal{R}^{li}$, where for every i, $1 \le i \le n$, $\mathsf{A}_i = \textit{toFs}(X_i)$. Therefore, $S^{li}[\,] \xRightarrow{k}_{li} X_1 \ldots X_n$ and for every i, $1 \le i \le n$, $\mathsf{A}_i = \textit{toFs}(X_i)$.

Assume that the hypothesis holds for every l, $0 < l < k$; let the length of the derivation sequence be k.

1. Assume that $\mathsf{A}^s \xRightarrow{k}_u \mathsf{A}_1 \ldots \mathsf{A}_n$. Then $\mathsf{A}^s \xRightarrow{k-1}_u \mathsf{B}_1 \ldots \mathsf{B}_m \xRightarrow{1}_u \mathsf{A}_1 \ldots \mathsf{A}_n$.

2. The last step of the unification derivation is established through a rule $r^u = \mathsf{C}_0 \rightarrow \mathsf{C}_1 \ldots \mathsf{C}_{n-m+1}$, $r^u \in \mathcal{R}^u$, and an index j, such that:

$$(\langle \mathsf{B}_1, \ldots, \mathsf{B}_m \rangle, j) \sqcup (\langle \mathsf{C}_0, \ldots, \mathsf{C}_{n-m+1} \rangle, 0) =$$
$$(\langle \mathsf{B}_1, \ldots, \mathsf{B}_{j-1}, \mathsf{Q}, \mathsf{B}_{j+1}, \ldots, \mathsf{B}_m \rangle, \langle \mathsf{Q}, \mathsf{A}_j, \ldots, \mathsf{A}_{j+n-m} \rangle)$$

3. By lemma 5, the sentential form $\langle \mathsf{A}_1, \ldots, \mathsf{A}_n \rangle$ has no reentrancies between its elements, hence for every i, $1 \le i < j$, $\mathsf{A}_i = \mathsf{B}_i$ and for i, $j < i \le m$, $\mathsf{A}_{i+n-m} = \mathsf{B}_i$.

4. By the induction hypothesis, if $\mathsf{A}^s \xRightarrow{k-1}_u \mathsf{B}_1 \ldots \mathsf{B}_m$ then $S^{li}[\,] \xRightarrow{k-1}_{li} Y_1 \ldots Y_m$ and

$$\langle \mathsf{B}_1, \ldots, \mathsf{B}_m \rangle = \langle \textit{toFs}(Y_1), \ldots, \textit{toFs}(Y_m) \rangle$$

5. Hence, $\mathsf{A}^s \xRightarrow{k-1}_u \textit{toFs}(Y_1) \ldots \textit{toFs}(Y_m) \xRightarrow{1}_u \mathsf{A}_1 \ldots \mathsf{A}_n$ and from (3), for every i, $1 \le i < j$, $\mathsf{A}_i = \textit{toFs}(Y_i)$ and for i, $j < i \le m$, $\mathsf{A}_{i+n-m} = \textit{toFs}(Y_i)$.

6. By definition 22, the rule r^u is created from a LIG rule r. We now show that the rule r can be applied to the element Y_j of the LIG sentential form, $\langle Y_1, \ldots, Y_m \rangle$, and the resulting sentential form, $\langle X_1, \ldots, X_n \rangle$, for every i, $1 \le i \le n$, satisfies the equation $\mathsf{A}_i = \textit{toFs}(X_i)$. Since from (5), for every i, $1 \le i < j$, $\mathsf{A}_i = \textit{toFs}(Y_i)$ and for i, $j < i \le m$, $\mathsf{A}_{i+n-m} = \textit{toFs}(Y_i)$, we just need to show that $\mathsf{A}_i = \textit{toFs}(X_i)$ for every i, $j \le i \le n - m + j$.

7. By definition of LIG the rule r has one of the following forms:

(a) Let $r = N_i[p_1, \ldots, p_x] \rightarrow Z_1 \ldots Z_{n-m+1}$. Hence, by definition 22, the unification rule r^u is

$$\textit{toFs}(N_i[p_1, \ldots, p_x]) \rightarrow \textit{toFs}(Z_1) \ldots \textit{toFs}(Z_{n-m+1})$$

where $\mathsf{C}_0 = \textit{toFs}(N_i[p_1, \ldots, p_x])$ and for every i, $1 \leq i \leq n - m + 1$, $\mathsf{C}_i = \textit{toFs}(Z_i)$. Note that there are no reentrancies between the elements of the unification rule r^u and hence $\langle \mathsf{A}_j, \ldots, \mathsf{A}_{n-m+j} \rangle = \langle \mathsf{C}_1, \ldots, \mathsf{C}_{n-m+1} \rangle$.

We now show that the rule r can be applied to the element Y_j of the LIG sentential form. Since $\mathsf{C}_0 \sqcup \mathsf{B}_j = \mathsf{C}_0 \sqcup \textit{toFs}(Y_j) = \textit{toFs}(N_i[p_1, \ldots, p_x]) \sqcup \textit{toFs}(Y_j) \neq \top$ we obtain, by lemma 7, that

$$\textit{toFs}(Y_j) = \textit{toFs}(N_i[p_1, \ldots, p_x])$$

Since *toFs* is one-to-one mapping we obtain that $Y_j = N_i[p_1, \ldots, p_x]$. Hence the LIG rule r can be applied to Y_j.

We now show that $\mathsf{A}_i = \textit{toFs}(X_i)$ for every i, $j \leq i \leq n - m + j$. We apply the rule r to Y_j as follows:

$$Y_1 \ldots Y_j \ldots Y_m \overset{1}{\Longrightarrow}_{li} X_1 \ldots X_{j-1} Z_1 \ldots Z_{n-m+1} X_{n-m+j+1} \ldots X_n$$

Hence $\langle X_j, \ldots, X_{n-m+j} \rangle = \langle Z_1, \ldots, Z_{n-m+1} \rangle$. Therefore,

$$\begin{aligned} \langle \mathsf{A}_j, \ldots, \mathsf{A}_{n-m+j} \rangle &= \langle \mathsf{C}_1, \ldots, \mathsf{C}_{n-m+1} \rangle \\ &= \langle \textit{toFs}(Z_1), \ldots, \textit{toFs}(Z_{n-m+1}) \rangle \\ &= \langle \textit{toFs}(X_j), \ldots, \textit{toFs}(X_{n-m+j}) \rangle \end{aligned}$$

(b) Let $r = N_i[p_1, \ldots, p_x..] \rightarrow Z_1 \ldots Z_{e-1}\ N_f[q_1, \ldots, q_y..]\ Z_{e+1} \ldots Z_{n-m+1}$, where $1 \leq e \leq n - m + 1$. Hence, by definition 22, the unification rule r^u is defined as

$$\langle N_i, p_1, \ldots, p_x, \boxed{1} \rangle \rightarrow \textit{toFs}(Z_1 \ldots Z_{e-1})\, \langle N_f, q_1, \ldots, q_y, \boxed{1} \rangle\, \textit{toFs}(Z_{e+1} \ldots Z_{n-m+1})$$

where $\mathsf{C}_0 = \langle N_i, p_1, \ldots, p_x, \boxed{1} \rangle$, $\mathsf{C}_e = \langle N_f, q_1, \ldots, q_y, \boxed{1} \rangle$ and for every i, $i \neq e$, $\mathsf{C}_i = \textit{toFs}(Z_i)$. Note that there is a reentrancy between C_0 and C_e. We now calculate the information propagated from B_j to A_{j+e-1} during the last step of the unification derivation (see 2). Since $\mathsf{C}_0 \sqcup \mathsf{B}_j = \mathsf{C}_0 \sqcup \textit{toFs}(Y_j) \neq \top$ we obtain by lemma 8, that $\textit{toFs}(Y_j) = \langle N_i, p_1, \ldots, p_x, \gamma \rangle$, where $\gamma \in V_s^*$. Therefore, $\mathsf{A}_{j+e-1} = \langle N_f, q_1, \ldots, q_y, \gamma \rangle$.

We now show that the LIG rule r can be applied to the element Y_j of the LIG sentential form. Since *toFs* is one-to-one and *toFs*$(Y_j) = \langle N_i, p_1, \ldots, p_x, \gamma\rangle$ we obtain that $Y_j = N_i[N_i, p_1, \ldots, p_x, \gamma]$. Hence the LIG rule r can be applied to Y_j.

We now show that $\mathsf{A}_i =$ *toFs*(X_i) for every i, $j \leq i \leq n-m+j$. We apply the rule r to Y_j as follows:

$$Y_1 \ldots Y_j \ldots Y_m \overset{1}{\Longrightarrow}_{li}$$
$$X_1 \ldots X_{j-1}\; Z_1 \ldots Z_{e-1}\; N_f[q_1, \ldots, q_y, \gamma]\; Z_{e+1} \ldots Z_{n-m+1}\; X_{n-m+j+1} \ldots X_n$$

Hence $\langle X_j, \ldots, X_{n-m+j}\rangle = \langle Z_1, \ldots, Z_{e-1}, N_f[q_1, \ldots, q_y, \gamma], Z_{e+1}, \ldots, Z_{n-m+1}\rangle$. Therefore,

$$\begin{aligned}
&\langle \mathsf{A}_j, \ldots, \mathsf{A}_{j+e-1}, \ldots, \mathsf{A}_{n-m+j}\rangle \\
&= \mathsf{C}_1 \ldots \mathsf{C}_{e-1}\; \langle N_f[q_1, \ldots, q_y, \gamma]\rangle\; \mathsf{C}_{e+1} \ldots \mathsf{C}_{n-m+1} \\
&= \textit{toFs}(Z_1 \ldots Z_{e-1})\; \textit{toFs}(N_f[q_1, \ldots, q_y, \gamma])\; \textit{toFs}(Z_{e+1} \ldots Z_{n-m+1}) \\
&= \langle \textit{toFs}(X_j), \ldots, \textit{toFs}(X_{n-m+j})\rangle
\end{aligned}$$

□

Corollary 12. *If* $G^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, S^{li}\rangle$ *is a LIG then there exists a unification grammar* $G^u =$ *lig2ug*(G^{li}) *such that* $L(G^u) = L(G^{li})$.

Proof. Let $G^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, N\rangle$ be a LIG and $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L}\rangle =$ *lig2ug*(G^{li}). Then by theorem 10, if $S[\,] \overset{*}{\Longrightarrow}_{li} \alpha$ then $\mathsf{A}^s \overset{*}{\Longrightarrow}_u$ *toFs*(α), where $\alpha = w_1, \ldots, w_n \in V_t^*$. By definition 22, for every i, $\mathcal{L}(w_i) = \{$*toFs*$(w_i)\}$, hence *toFs*$(\alpha) =$ *toFs*$(w_1), \ldots,$ *toFs*(w_n). Hence $\mathsf{A}^s \overset{*}{\Longrightarrow}_u$ *toFs*$(w_1), \ldots,$ *toFs*$(w_n) \in L(G^u)$.

Assume that $\mathsf{A}^s \overset{*}{\Longrightarrow}_u \mathsf{A}_1, \ldots, \mathsf{A}_n$, where $\mathsf{A}_1, \ldots, \mathsf{A}_n$ is a pre-terminal sequence and $\mathsf{A}_1, \ldots, \mathsf{A}_n \overset{*}{\Longrightarrow}_u w_1, \ldots, w_n$. By theorem 11, there is the LIG derivation sequence such that $S[\,] \overset{*}{\Longrightarrow}_{li} X_1, \ldots, X_n$ and for all i, *toFs*$(X_i) = \mathsf{A}_i$. By definition 22, each entry $\mathcal{L}(w_i) = \{\mathsf{A}_i\}$ in the lexicon of G^u is created from a terminal rule $X_i \to w_i$ in $\mathcal{R}^{li}$. Therefore, $S[\,] \overset{*}{\Longrightarrow}_{li} X_1, \ldots, X_n \overset{*}{\Longrightarrow}_{li} w_1, \ldots, w_n$. □

3.4 Mapping of one-reentrant Unification Grammars to Linear Indexed Grammars

We are now interested in the reverse direction, namely mapping unification grammars to LIG. Of course, since unification grammars are more expressive than LIGs, only a subset of the former can be correctly simulated by the latter. The differences between the two formalisms can be summarized along three dimensions:

- *The basic elements*
  - UG manipulates feature structures; rules (and forms) are MRSs, whereas
  - LIG manipulates terminals and non-terminals with stacks of elements; rules (and forms) are sequences of such symbols.
- *Rule application*
  - In UG a rule is applied by *unification in context* of the rule and a sentential form, both of which are MRSs, whereas
  - In LIG, the head of a rule and the selected element of a sentential form must have the same non-terminal symbol and consistent stacks.
- *Propagation of information in rules*
  - In UG information is shared through reentrancies, whereas
  - In LIG, information is propagated by copying the stack from the head of the rule to one element of its body.

We will show that one-reentrant unification grammars, as defined in definition 17, can all be mapped correctly to LIG. For the rest of this section we fix a signature $\langle \textsc{Atoms}, \textsc{Feats}, \textsc{Words} \rangle$ over which unification grammars are defined.

One-reentrant unification grammars are highly constrained. They induce non-reentrant (sentential) forms, and unification of non-reentrant feature structures is highly simplified (see section 3.2, lemma 5 and lemma 6). Still, it is important to note that even with such grammars, feature structures can grow unboundedly deep, and representing them by means of LIG symbols is the greatest challenge of our solution.

Definition 23. *Let* A *be a feature structure with no reentrancies. The* ***height*** *of* A*, denoted* $|\mathsf{A}|$*, is the length of the longest path in* A*. This is well-defined since non-reentrant feature structures are acyclic.*

Definition 24. *Let* $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle \in \mathrm{UG}_{1r}$ *be a one-reentrant unification grammar. The* ***maximum height*** *of the grammar, maxHt*(G^u)*, is the height of the highest feature structure in the grammar, defined as:*

$$\mathit{maxHt}(G^u) = \max_{r^u \in \mathcal{R}^u} (\max_{0 \le i \le |r^u|} (|r_i^u|))$$

where r_i^u *is the* i*-th element of* r^u*. This is well defined since by definition of one-reentrant grammars all feature structures of the grammar are non-reentrant.*

The following lemma indicates an important property of one-reentrant unification grammars. Informally, in any feature structure that is an element of a sentential form induced by such grammars, if two paths are long (specifically, longer than the maximum height of the grammar), then they must have a long common prefix.

Lemma 13. *Let* $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle \in \mathrm{UG}_{1r}$ *be a one-reentrant unification grammar. Let* λ *be a sentential form derived by* G^u *and* A *be an element of* λ*. If* $\pi \cdot \langle \mathrm{F}_j \rangle \cdot \pi_1, \pi \cdot \langle \mathrm{F}_k \rangle \cdot \pi_2 \in \Pi_A$*, where* $\mathrm{F}_j, \mathrm{F}_k \in \mathrm{FEATS}$*,* $\mathrm{F}_j \neq \mathrm{F}_k$ *and* $|\pi_1| \le |\pi_2|$*, then* $|\pi_1| \le \mathit{maxHt}(G^u)$*.*

Proof. We prove by induction on the length of the derivation sequence that if $\mathsf{A}^s \overset{*}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$, then the lemma conditions hold. Let $h = \mathit{maxHt}(G^u)$.

The induction hypothesis is that if $\mathsf{A}^s \overset{k}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$, then the lemma conditions hold for any A_l, where $1 \le l \le n$. If $k = 0$, then by definition $\mathsf{A}^s \overset{0}{\Longrightarrow}_u \mathsf{A}^s$. Since $|\mathsf{A}^s| \le h$ then for any j,π and π_1, such that $\pi \cdot \langle \mathrm{F}_j \rangle \cdot \pi_1 \in \Pi_{\mathsf{A}^s}$, $|\pi \cdot \langle \mathrm{F}_j \rangle \cdot \pi_1| \le h$. Therefore, $|\pi_1| < h$.

Assume that the hypothesis holds for every i, $0 \le i < k$; let the length of the derivation sequence be k. Let $\mathsf{A}^s \overset{k-1}{\Longrightarrow}_u \mathsf{B}_1 \ldots \mathsf{B}_m \overset{1}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$. Then by definition of UG_{1r} derivation, there are an index j and a rule $r^u = \mathsf{C}_0 \to \mathsf{C}_1 \ldots \mathsf{C}_{n-m+1}$, $r^u \in \mathcal{R}^u$, such that

$$(\langle \mathsf{C}_0, \ldots, \mathsf{C}_{n-m+1} \rangle, 0) \sqcup (\langle \mathsf{B}_1, \ldots, \mathsf{B}_m \rangle, j) = (\langle \mathsf{Q}_0, \ldots, \mathsf{Q}_{n-m+1} \rangle, \langle \mathsf{B}_1, \ldots, \mathsf{B}_{j-1}, \mathsf{Q}_0, \mathsf{B}_{j+1}, \ldots, \mathsf{B}_m \rangle)$$

where

1. $\langle \mathsf{A}_1, \ldots, \mathsf{A}_{j-1} \rangle = \langle \mathsf{B}_1, \ldots, \mathsf{B}_{j-1} \rangle$

2. $\langle \mathsf{A}_j, \ldots, \mathsf{A}_{n-m+j} \rangle = \langle \mathsf{Q}_1, \ldots, \mathsf{Q}_{n-m+1} \rangle$

3. $\langle \mathsf{A}_{n-m+j+1}, \ldots, \mathsf{A}_n \rangle = \langle \mathsf{B}_{j+1}, \ldots, \mathsf{B}_m \rangle$

By the induction hypothesis, in cases (1) and (3) the lemma conditions hold for A_l, where $1 \leq l < j$ or $n - m + j + 1 \leq l \leq n$. We now analyze case (2). Since G^u is one-reentrant there are only two options for the rule r^u:

1. r^u has no reentrancies;

2. $(0, \pi_0) \overset{r^u}{\leftrightsquigarrow} (e, \pi_e)$, where $1 \leq e \leq n - m + 1$;

If r^u is non-reentrant, $\langle \mathsf{C}_1, \ldots, \mathsf{C}_{n-m+1} \rangle = \langle \mathsf{Q}_1, \ldots, \mathsf{Q}_{n-m+1} \rangle = \langle \mathsf{A}_j, \ldots, \mathsf{A}_{n-m+j} \rangle$. Hence for any l, $j \leq l \leq n - m + j$, $|\mathsf{A}_l| \leq h$. Hence, for any F, π and π_1, such that $\pi \cdot \langle \mathrm{F}_j \rangle \cdot \pi_1 \in \Pi_{\mathsf{A}_i}$, $|\pi \cdot \langle \mathrm{F}_j \rangle \cdot \pi_1| \leq h$. Therefore, $|\pi_1| < h$.

If $(0, \pi_0) \overset{r^u}{\leftrightsquigarrow} (e, \pi_e)$ then by the definition of unification, $\mathsf{Q}_l = \mathsf{C}_l$ if $1 \leq l < e$ or $e < l \leq n - m + 1$, hence $|\mathsf{Q}_l| \leq h$. Therefore, the lemma conditions hold for any Q_l, where $l \neq e$. We now check whether the lemma conditions hold for Q_e. The rule r^u, when applied to B_j, can result in modifying the body of the rule, $\mathsf{C}_1 \ldots \mathsf{C}_{n-m+1}$. However, due to the fact that r^u is one-reentrant, only a single element C_e can be modified. Furthermore, the only possible modifications to C_e are addition of paths and further specification of atoms (lemma 6). The latter has no effect on path length, so we focus on the former. The only way for a path $\pi_e \cdot \pi$ to be added is if some path $\pi_0 \cdot \pi$ already exists in B_j. Hence, let P be a set of paths such that:

$$P = \{\pi_e \cdot \pi \mid \pi_0 \cdot \pi \in \Pi_{\mathsf{B}_j}\}$$

By definition of unification $\Pi_{\mathsf{Q}_e} = P \cup \Pi_{\mathsf{C}_e}$. To check the lemma conditions we only need to check the pairs of paths where both members are longer than h, otherwise the conditions trivially hold. Since for any path π, $\pi \in \Pi_{\mathsf{C}_e}$, $|\pi| \leq h$, we check only the pairs of paths from P to evaluate the lemma conditions. Let $\pi_e \cdot \pi_1, \pi_e \cdot \pi_2 \in P \subseteq \Pi_{\mathsf{Q}_e}$, where $|\pi_1| \leq |\pi_2|$, π_1 and π_2 differ in the first feature. By definition of P, $\pi_0 \cdot \pi_1, \pi_0 \cdot \pi_2 \in \Pi_{\mathsf{B}_j}$. Hence, by the induction hypothesis $|\pi_1| \leq h$. Therefore, for any pair of paths in Π_{Q_e} the lemma conditions hold. $\square$

Lemma 13 provides an important property of one-reentrant unification grammars that facilitates a view of all the feature structures induced by a such grammar as (unboundedly long) lists of elements drawn from a finite, predefined set. The set consists of all features in FEATS and all the non-reentrant feature structures whose height is limited by the maximal height of the unification grammar. Note that even with one-reentrant unification grammars, feature structures can be unboundedly deep. What lemma 13 establishes is the fact that if a feature structure induced by a one-reentrant unification grammar is deep, then it can be represented as a *single* "core" path which is long, and all the sub-structures which "hang" from this core are depth-bounded. We use this property to encode such feature structures as *cords*.

Definition 25 (Cords). *Let* Ψ : NRFSS $\times$ PATHS $\mapsto$ (FEATS $\cup$ NRFSS)* *be a mapping of pairs of non-reentrant feature structures and paths to sequences of features and feature structures such that if* A *is a non-reentrant feature structure and* $\pi = \langle \text{F}_1, \ldots, \text{F}_n \rangle \in \Pi_A$, *then the* ***cord*** $\Psi(\mathsf{A}, \pi)$ *is* $\langle \mathsf{A}_1, \text{F}_1, \ldots, \mathsf{A}_n, \text{F}_n, \mathsf{A}_{n+1} \rangle$, *where for* $1 \leq i \leq n+1$, A_i *are non-reentrant feature structures such that:*

- $\Pi_{A_i} = \{\varepsilon\} \cup \{\langle \text{G} \rangle \cdot \pi \mid \text{G} \in \text{FEATS}, \pi \in \text{PATHS}, \langle \text{F}_1, \ldots, \text{F}_{i-1}, \text{G} \rangle \cdot \pi \in \Pi_A,$ *when* $i \leq n, \text{G} \neq \text{F}_i\}$
- *If* $\Theta_A(\langle \text{F}_1, \ldots, \text{F}_{i-1} \rangle \cdot \pi) \downarrow$ *then* $\Theta_{A_i}(\pi) = \Theta_A(\langle \text{F}_1, \ldots, \text{F}_{i-1} \rangle \cdot \pi)$, *otherwise* $\Theta_{A_i}(\pi)$ *is undefined.*

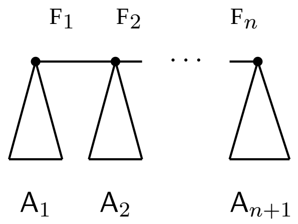
We also define two operators on cords $\Psi(\mathsf{A}, \pi)$ *as follows:*

- *last*$(\Psi(\mathsf{A}, \pi)) = \mathsf{A}_{n+1}$
- *butLast*$(\Psi(\mathsf{A}, \pi)) = \langle \mathsf{A}_1, \text{F}_1, \ldots, \mathsf{A}_n, \text{F}_n \rangle$

The ***height*** *of a cord is defined as* $|\Psi(\mathsf{A}, \pi)| = \max_{1 \leq i \leq n+1}(|\mathsf{A}_i|)$. *For each cord* $\Psi(\mathsf{A}, \pi)$ *we refer to* A *as the* ***base feature structure*** *and to* π *as the* ***base path***. *The* ***length*** *of a cord is the length of the base path.*

Example 10. *Let* A *be a non-reentrant feature structure and* $\pi = \langle \text{F}_1, \ldots, \text{F}_n \rangle \in \Pi_A$ *be a path. Then*

A may be represented as follows:



$\Psi(\mathsf{A}, \pi) = \langle \mathsf{A}_1, \text{F}_1, \ldots, \mathsf{A}_n, \text{F}_n, \mathsf{A}_{n+1} \rangle$, *where* $\mathsf{A}_1, \ldots, \mathsf{A}_{n+1}$ *are non-reentrant feature structures.*

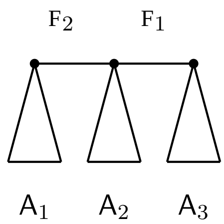
Example 11. *Let* A *be a non-reentrant feature structure over the signature* $\text{FEATS} = \{\text{F}_1, \text{F}_2, \text{F}_3\}$, $\text{ATOMS} = \{a, b\}$*:*

$$\mathsf{A} = \begin{bmatrix} \text{F}_1 : b \\ \text{F}_2 : \begin{bmatrix} \text{F}_1 : \begin{bmatrix} \text{F}_2 : \begin{bmatrix} \text{F}_3 : a \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{F}_3 : \begin{bmatrix} \text{F}_1 : [\,] \\ \text{F}_2 : a \\ \text{F}_3 : \begin{bmatrix} \text{F}_1 : [\,] \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

If $\pi = \langle \text{F}_2, \text{F}_1 \rangle$ *then the cord representation of* A *on the path* π *is* $\Psi(\mathsf{A}, \pi) = \langle \mathsf{A}_1, \text{F}_2, \mathsf{A}_2, \text{F}_1, \mathsf{A}_3 \rangle$, *where*

$$\mathsf{A}_1 = \begin{bmatrix} \text{F}_1 : b \\ \text{F}_3 : \begin{bmatrix} \text{F}_1 : [\,] \\ \text{F}_2 : a \\ \text{F}_3 : \begin{bmatrix} \text{F}_1 : [\,] \end{bmatrix} \end{bmatrix} \end{bmatrix} ;\ \mathsf{A}_2 = [\,] ;\ \mathsf{A}_3 = \begin{bmatrix} \text{F}_2 : \begin{bmatrix} \text{F}_3 : a \end{bmatrix} \end{bmatrix}$$

And the graph representation is



Note that the function Ψ is one to one. In other words, given $\Psi(\mathsf{A}, \pi)$, both A and π are uniquely determined. The path π is determined by the sequence of the features on the cord $\Psi(\mathsf{A}, \pi)$, in the order they occur in the cord. Since A is non-reentrant, all A_i in $\Psi(\mathsf{A}, \pi)$ are non-reentrant feature structures, i.e., trees. To see that A is uniquely determined, simply view π as a branch of a tree and "hang" the subtrees A_i on π, in the order determined by the features in the cord, to obtain a unique feature structure.

Lemma 14. *Let G^u be a one-reentrant unification grammar and let A be an element of a sentential form induced by G^u. Then there is a path $\pi \in \Pi_A$ such that the height of $\Psi(\mathsf{A}, \pi)$ is less then maxHt(G^u).*

Proof. An immediate corollary of lemma 13. □

Later in this section we manipulate cords: concatenate two cords into one cord and split a cord in two. Two cords can be concatenated by adding a feature between them. The only requirement is that the resulting cord be well defined: the added feature must not be present in the last element of the first cord. To split a cord into two cords we do the reverse process: remove one of the cord's features. This is similar to splitting a tree (a non-reentrant feature structure) by removing one of its arcs (a feature in the feature structure). The following lemma provides a formal base for these operations.

Lemma 15. *Let A and B be two non-reentrant feature structures. Let π_A, π_B be paths such that $\pi_A \in \Pi_A$, $\pi_B \in \Pi_B$ and last$(\Psi(\mathsf{A}, \pi_A)) \notin$ ATOMS. And let G be a feature such that $\langle \text{G} \rangle \notin \Pi_{last(\Psi(\mathsf{A}, \pi_A))}$. Then $\Psi(\mathsf{A}, \pi_A) \cdot \langle \text{G} \rangle \cdot \Psi(\mathsf{B}, \pi_B)$ is a cord.*

Let $\Psi(\mathsf{A}, \pi_A) = \langle \mathsf{A}_1, \text{F}_1, \ldots, \mathsf{A}_i, \text{F}_i, \mathsf{A}_{i+1}, \ldots, \text{F}_n, \mathsf{A}_{n+1} \rangle$. Then for any i, $1 \leq i \leq n$, the sequences $\langle \mathsf{A}_1, \text{F}_1, \ldots, \mathsf{A}_i \rangle$ and $\langle \mathsf{A}_{i+1}, \ldots, \text{F}_n, \mathsf{A}_{n+1} \rangle$ are cords.

Proof. Immediate from the definition of cords. □

Example 12. *Let A be a feature structure over the signature ATOMS $= \langle a, b \rangle$, FEATS $= \langle \text{F}_1, \text{F}_2 \rangle$, such that*

$$\mathsf{A} = \begin{bmatrix} \text{F}_1 : \begin{bmatrix} \text{F}_1 : b \\ \text{F}_2 : \begin{bmatrix} \text{F}_1 : \begin{bmatrix} \text{F}_2 : b \end{bmatrix} \\ \text{F}_2 : a \end{bmatrix} \end{bmatrix} \\ \text{F}_2 : a \end{bmatrix}$$

Then $\Psi(\mathsf{A}, \langle \text{F}_1, \text{F}_2, \text{F}_1 \rangle) = \langle \begin{bmatrix} \text{F}_2 : a \end{bmatrix}, \text{F}_1, \begin{bmatrix} \text{F}_1 : b \end{bmatrix}, \text{F}_2, \begin{bmatrix} \text{F}_2 : a \end{bmatrix}, \text{F}_1, \begin{bmatrix} \text{F}_2 : b \end{bmatrix} \rangle$. We can split this cord in two by removing one of the features. For example, removing the feature F_2 creates the following *two cords:* $\eta = \langle \begin{bmatrix} \text{F}_2 : a \end{bmatrix}, \text{F}_1, \begin{bmatrix} \text{F}_1 : b \end{bmatrix} \rangle$ *and* $\gamma = \langle \begin{bmatrix} \text{F}_2 : a \end{bmatrix}, \text{F}_1, \begin{bmatrix} \text{F}_2 : b \end{bmatrix} \rangle$*, where the base feature structure of the cord* η *is* $\begin{bmatrix} \text{F}_1 : \begin{bmatrix} \text{F}_1 : b \end{bmatrix} \\ \text{F}_2 : a \end{bmatrix}$.

We can concatenate the cords γ *and* η *with the feature* F_1 *as follows:*

$$\gamma \cdot \langle \text{F}_1 \rangle \cdot \eta = \langle \begin{bmatrix} \text{F}_2 : a \end{bmatrix}, \text{F}_1, \begin{bmatrix} \text{F}_2 : b \end{bmatrix}, \text{F}_1, \begin{bmatrix} \text{F}_2 : a \end{bmatrix}, \text{F}_1, \begin{bmatrix} \text{F}_1 : b \end{bmatrix} \rangle$$

The base feature structure of the cord $\gamma \cdot \langle \text{F}_1 \rangle \cdot \eta$ *is the feature structure* B*:*

$$\mathsf{B} = \begin{bmatrix} \text{F}_1 : \begin{bmatrix} \text{F}_1 : \begin{bmatrix} \text{F}_1 : \begin{bmatrix} \text{F}_1 : b \end{bmatrix} \\ \text{F}_2 : a \end{bmatrix} \\ \text{F}_2 : b \end{bmatrix} \\ \text{F}_2 : a \end{bmatrix}$$

Then $\gamma \cdot \langle \text{F}_1 \rangle \cdot \eta = \Psi(\mathsf{B}, \langle \text{F}_1, \text{F}_1, \text{F}_1 \rangle)$.

So far we have shown how to map non-reentrant feature structures to lists whose elements are drawn from a finite domain. This mapping resolves the first major difference between LIG and UG, by providing a representation of the *basic elements*. We use cords as the stack contents of LIG non-terminal symbols: cords can be unboundedly long, but so can LIG stacks; the crucial point is that cords are height limited, implying that they can be represented using a *finite* number of elements, which will be LIG stack symbols in our mapping.

We now investigate how to resolve the second major difference between LIG and UG: *rule application*. A unification rule can be applied to a sentential form only if the head of the rule, C_0, and some selected element in the form, D_j, are unifiable. In contrast, a linear indexed rule can be applied to a sentential form only if the head of the rule, X_0, and some selected element in the form, Y_j,

- have the same non-terminal symbol; and
- either the content of the stack of X_0 and Y_j are equal (*fixed* head of a LIG rule), or
- the stack of X_0 is unbounded and is a prefix of Y_j (*unbounded* head of a LIG rule).

We now show how to simulate, in LIG, the unification in context of a rule and a sentential form. The first step is to have exactly one non-terminal symbol; when all non-terminal symbols are identical, only the content of the stack has to be taken into account. Note that in order for a LIG rule to be applicable to a sentential form, the stack of the rule's head must be a *prefix* of the stack of the selected element in the form. The only question is whether the two stacks are equal (fixed rule head) or not (unbounded rule head). Since the contents of stacks are cords, we need a property relating two cords, on one hand, with unifiability of their base feature structures, on the other. Lemma 16 establishes such a property. Informally, if the base path of one cord is a prefix of the base path of the other cord and all feature structures along the common path of both cords are unifiable, then the base feature structures of both cords are unifiable. The reverse direction also holds.

Lemma 16. *Let* $\mathsf{A}, \mathsf{B} \in \textsc{NrFss}$ *be non-reentrant feature structures and* $\pi_1, \pi_2 \in \textsc{Paths}$ *be paths such that*

- $\pi_1 \in \Pi_B$,
- $\pi_1 \cdot \pi_2 \in \Pi_A$,
- $\Psi(\mathsf{A}, \pi_1 \cdot \pi_2) = \langle \mathsf{t}_1, \textsc{f}_1, \ldots, \textsc{f}_{|\pi_1|}, \mathsf{t}_{|\pi_1|+1}, \textsc{f}_{|\pi_1|+1}, \ldots, \mathsf{t}_{|\pi_1 \cdot \pi_2|+1} \rangle$,
- $\Psi(\mathsf{B}, \pi_1) = \langle \mathsf{s}_1, \textsc{f}_1, \ldots, \mathsf{s}_{|\pi_1|+1} \rangle$, *and*
- $\langle \textsc{f}_{|\pi_1|+1} \rangle \notin \Pi_{s_{|\pi_1|+1}}$

then for all i, $1 \leq i \leq |\pi_1| + 1$, $\mathsf{s}_i \sqcup \mathsf{t}_i \neq \top$ *iff* $\mathsf{A} \sqcup \mathsf{B} \neq \top$.

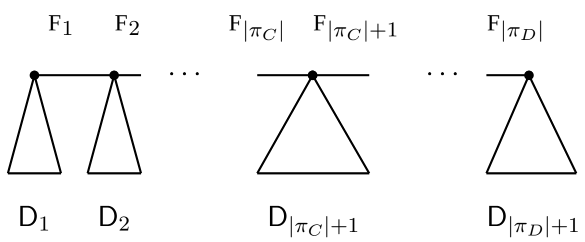
Proof. Assume that for every $1 \leq i \leq |\pi_1| + 1$, $\mathsf{s}_i \sqcup \mathsf{t}_i \neq \top$. Since the prefixes of $\Psi(\mathsf{B}, \pi_1)$ and $\Psi(\mathsf{A}, \pi_1 \cdot \pi_2)$ are consistent up to $\textsc{f}_{|\pi_1|+1}$ and the suffix of the cord $\Psi(\mathsf{A}, \pi_1 \cdot \pi_2)$ does not occur in $\Psi(\mathsf{B}, \pi_1)$, and hence does not contradict with B, the feature structures A and B are unifiable.

Assume that $\mathsf{A} \sqcup \mathsf{B} \neq \top$. Then all subtrees of the feature structures are consistent. Therefore, $\mathsf{s}_i \sqcup \mathsf{t}_i \neq \top$, for every $1 \leq i \leq |\pi_1| + 1$. □

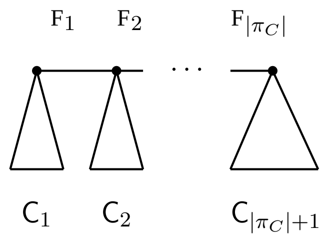
Given some one-reentrant unification grammar, the set of feature structures which are the heads of all rules in the grammar is a finite set. However, these feature structures are unified during derivation with elements of sentential forms induced by the grammar, and these can constitute an infinite set. We take advantage in our construction of the fact that although a potentially infinite set of feature structures is involved, these feature structures can all be represented as cords, with height-bounded feature structures hung on their base paths.

The length of a cord of an element of a sentential form induced by the grammar cannot be bounded, however the length of any cord representation of a rule head is limited by the grammar height. By lemma 16, unifiability of two feature structures can be reduced to a comparison of two cords representing them and only the prefix of the longer cord (as long as the shorter cord) affects the result. Since the cord representation of any grammar rule's head is limited by the height of the grammar we always choose it as the shorter cord in the comparison. Hence only a prefix of the cord in sentential forms, limited by the grammar height, affects unification and, therefore, rule application. Since the set of rule heads is finite, so is the set of their cord representations; each element of this set is a cord of a limited length. In a similar way, it is possible to construct a *finite* set of the prefixes of all the cord representations of all the feature structures which are elements of sentential forms induced by the grammar. We use the grammar height as the limit of cord length in this set.

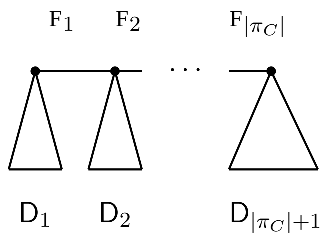
Example 13. *Let* D *be a selected element of a sentential form induced by a one-reentrant unification grammar* G^u. *Let* C *be the head of a unification rule applied to* D. *Let* $\Psi(\mathsf{D}, \pi_D)$ *be a cord whose height is limited by the grammar height, where* $\pi_D = \langle \mathrm{F}_1, \cdots, \mathrm{F}_{|\pi_D|} \rangle$. *Let* π_C *be the maximal prefix of* π_D *such that* $\pi_C \in \Pi_C$, $\pi_C = \langle \mathrm{F}_1, \cdots, \mathrm{F}_{|\pi_C|} \rangle$ *and* $\langle \mathrm{F}_1, \cdots, \mathrm{F}_{|\pi_C|+1} \rangle \notin \Pi_C$. *Such a prefix always exists because* ε *is a common prefix of all paths in* PATHS. *Note that the height of the cord* $\Psi(\mathsf{C}, \pi_C)$ *is limited by the grammar height because the height of* C *is limited by the grammar height. The cord* $\Psi(\mathsf{D}, \pi_D)$ *is graphically represented as:*



Whereas $\Psi(\mathsf{C}, \pi_C)$ *is similarly represented as:*



By lemma 16, $\mathsf{D} \sqcup \mathsf{C} \neq \top$ *iff* $\mathsf{D}_i \sqcup \mathsf{C}_i \neq \top$ *for all* i*,* $1 \leq i \leq |\pi_C|+1$*. Note that the feature structures* D_i*, where* $i > |\pi_C| + 1$*, do not affect the unifiability of* D *and* C*. In other words, to determine whether* C *is unifiable with some feature structure* D*, whose cord is* $\Psi(\mathsf{D}, \pi_D)$*, it is sufficient to check the unifiability of* C *with the feature structure* A*, where* $\Psi(\mathsf{A}, \pi_C)$ *is:*



Example 13 motivates the following corollary:

Corollary 17. *Let* $\Psi(\mathsf{A}, \pi_A)$*,* $\Psi(\mathsf{B}, \pi_B)$*,* $\Psi(\mathsf{C}, \pi_A)$ *be cords, where* $\Theta_A(\pi_a) \uparrow$*. Let* G *be a feature such that:*

- $\langle \text{G} \rangle \notin \Pi_{\textit{last}(\Psi(\mathsf{A}, \pi_A))}$ *and*
- $\langle \text{G} \rangle \notin \Pi_{\textit{last}(\Psi(\mathsf{C}, \pi_A))}$

Consider the cord $\Psi(\mathsf{A}, \pi_A) \cdot \langle \text{G} \rangle \cdot \Psi(\mathsf{B}, \pi_B)$ *(by lemma 15, this is well defined) and write it as* $\Psi(\mathsf{D}, \pi_A \cdot \langle \text{G} \rangle \cdot \pi_B)$*. Then* $\mathsf{C} \sqcup \mathsf{A} \neq \top$ *iff* $\mathsf{C} \sqcup \mathsf{D} \neq \top$*.*

We now define, for a feature structure C (which is a head of a rule) and some path π, the set that includes all feature structures that are both unifiable with C and can be represented as a cord whose height is limited by the grammar height and whose base path is π. We call this set the *compatibility set* of C and π. Latter in this section, we use the compatibility set to define the set of all possible prefixes of cords whose base feature structures are unifiable with C (see definition 27). Crucially, the compatibility set of C is finite for any feature structure C since the heights and the lengths of the cords are limited.

Definition 26 (Compatibility set). *Given a non-reentrant feature structure* C*, a path* $\pi = \langle F_1, \ldots, F_n \rangle \in \Pi_C$ *and a natural number* h*, the* ***compatibility set***, $\Gamma(\mathsf{C}, \pi, h)$*, is defined as the set of all feature structures* A *such that*

- $\mathsf{C} \sqcup \mathsf{A} \neq \top$,
- $\pi \in \Pi_A$*, and*
- $|\Psi(\mathsf{A}, \pi)| \leq h$

The compatibility set is defined for a feature structure and a given path (when h is taken to be the grammar height). We now define two similar sets, FIXEDHEAD and UNBOUNDEDHEAD, for a given feature structure, independently of a path. Later in this section, when we map rules of a one-reentrant unification grammar to LIG rules (definition 28), the set FIXEDHEAD will be used to define heads of fixed rules in LIG and the set UNBOUNDEDHEAD to define heads of unbounded rules. Each unification rule will be mapped to a *set* of LIG rules, each with a different head. The stack of the head will be some member of the sets FIXEDHEAD and UNBOUNDEDHEAD. Each such member is a prefix of the stack of potential elements of sentential forms that the LIG rule can be applied to.

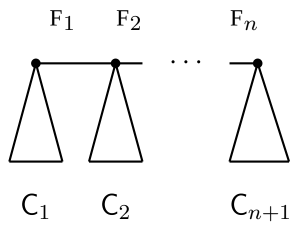
Definition 27. *Let* C *be a non-reentrant feature structure and* h *be a natural number. We define the* ***fixed rule head*** *set,* FIXEDHEAD(C, h)*, and the* ***unbounded rule head*** *set,* UNBOUNDEDHEAD(C, h) *as follows:*

$$\text{FIXEDHEAD}(\mathsf{C}, h) = \{\Psi(\mathsf{A}, \pi) \mid \pi \in \Pi_C, \mathsf{A} \in \Gamma(\mathsf{C}, \pi, h)\}$$

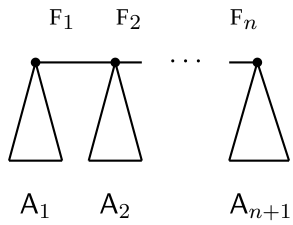
$$\text{UNBOUNDEDHEAD}(\mathsf{C}, h) =$$

$$\{\Psi(\mathsf{A}, \pi) \cdot \langle F \rangle \mid \Psi(\mathsf{A}, \pi) \in \text{FIXEDHEAD}(\mathsf{C}, h), \Theta_C(\pi)\uparrow, F \in \text{FEATS}, \langle F \rangle \notin \Pi_{last(\Psi(\mathsf{C} \sqcup \mathsf{A}, \pi))}\}$$

Informally, let C be a head of a unification rule and $\pi = \langle \textsc{f}_1, \ldots, \textsc{f}_n \rangle \in \Pi_C$ be a path such that $\Psi(\mathsf{C}, \pi) = \langle \mathsf{C}_1, \textsc{f}_1, \ldots, \mathsf{C}_{n+1} \rangle$. Graphically, this situation is depicted thus:

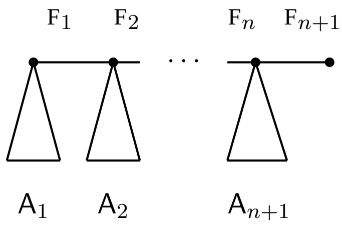


Then $\textsc{fixedHead}(\mathsf{C}, h)$ consists of cords like

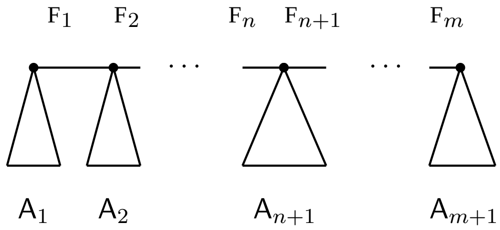


where $\mathsf{A}_i \sqcup \mathsf{C}_i$, $1 \leq i \leq n+1$. Let A be the base feature structure of the cord $\langle \mathsf{A}_1, \textsc{f}_1, \ldots, \mathsf{A}_{n+1} \rangle$. Then by definition of $\textsc{fixedHead}(\mathsf{C}, h)$, $\mathsf{A} \sqcup \mathsf{C} \neq \top$. When the LIG symbol $N[\mathsf{A}_1, \textsc{f}_1, \ldots, \mathsf{A}_{n+1}]$ occurs as the head of a rule, this rule is applicable only to a sentential form with an identical element. For each such A we create a LIG rule whose head is $N[\mathsf{A}_1, \textsc{f}_1, \ldots, \mathsf{A}_{n+1}]$. This is possible because the set $\textsc{fixedHead}(\mathsf{C}, h)$ is finite.

Similarly, $\textsc{unboundedHead}(\mathsf{C}, h)$ consists of cord prefixes like



where the value of the path $\langle \textsc{f}_{n+1} \rangle$ in $\mathsf{A}_{n+1} \sqcup \mathsf{C}_{n+1}$ is undefined. Let A be the base feature structure of the cord $\langle \mathsf{A}_1, \textsc{f}_1, \ldots, \mathsf{A}_{n+1} \rangle$ and let $\eta = \langle \mathsf{A}_{n+2}, \textsc{f}_{n+2}, \ldots, \mathsf{A}_{m+1} \rangle$ be a cord, where $m > n$. Then by corollary 17, the base feature structure D of the cord $\langle \mathsf{A}_1, \textsc{f}_1, \ldots, \mathsf{A}_{m+1} \rangle$ is unifiable with C for any cord η. In contrast to the previous case, a rule whose head is $N[\mathsf{A}_1, \textsc{f}_1, \ldots, \mathsf{A}_{n+1}, \textsc{f}_{n+1}..]$ is applicable to any element of the form $N[\mathsf{A}_1, \textsc{f}_1, \ldots, \mathsf{A}_{m+1}]$. Note that the content of the stack of such a LIG symbol is a cord of the form:



Example 14. *To illustrate the structure of the sets* FIXEDHEAD *and* UNBOUNDEDHEAD *we give here some examples of the elements in these sets for the feature structure* C*, the head of the rule* r_5^u *of example 9. Recall that*

$$\mathsf{C} = \begin{bmatrix} \text{HEAD} : N_3 \\ \text{TAIL} : \begin{bmatrix} \text{HEAD} : a \\ \text{TAIL} : [\,] \end{bmatrix} \end{bmatrix}$$

Let G_2^u *be the unification grammar of the example 9. The grammar height of* G^u *is 2 and the set of all paths in* C *is*

$$\Pi_C = \{\varepsilon, \langle\text{HEAD}\rangle, \langle\text{TAIL}\rangle, \langle\text{TAIL}, \text{HEAD}\rangle, \langle\text{TAIL}, \text{TAIL}\rangle\}$$

Hence there are five compatibility sets for C*, one for each path in* Π_C *as follows:* $\Gamma(\mathsf{C}, \varepsilon, 2)$*,* $\Gamma(\mathsf{C}, \langle\text{HEAD}\rangle, 2)$*,* $\Gamma(\mathsf{C}, \langle\text{TAIL}\rangle, 2)$*,* $\Gamma(\mathsf{C}, \langle\text{TAIL}, \text{HEAD}\rangle, 2)$ *and* $\Gamma(\mathsf{C}, \langle\text{TAIL}, \text{TAIL}\rangle, 2)$*. For example, here are some elements of the compatibility set* $\Gamma(\mathsf{C}, \langle\text{HEAD}\rangle, 2)$*:*

$$\begin{bmatrix} \text{HEAD} : N_3 \\ \text{TAIL} : \begin{bmatrix} \text{HEAD} : [\,] \\ \text{TAIL} : N_1 \end{bmatrix} \end{bmatrix} \textit{ and } \begin{bmatrix} \text{HEAD} : [\,] \\ \text{TAIL} : [\,] \end{bmatrix}$$

similarly, some examples of elements of the compatibility set $\Gamma(\mathsf{C}, \langle\text{TAIL}, \text{TAIL}\rangle, 2)$ *are*

$$\begin{bmatrix} \text{HEAD} : [\,] \\ \text{TAIL} : \begin{bmatrix} \text{HEAD} : a \\ \text{TAIL} : \begin{bmatrix} \text{HEAD} : \begin{bmatrix} \text{HEAD} : a \\ \text{TAIL} : N_1 \end{bmatrix} \\ \text{TAIL} : \begin{bmatrix} \text{HEAD} : N_1 \\ \text{TAIL} : [\,] \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \textit{ and } \begin{bmatrix} \text{HEAD} : [\,] \\ \text{TAIL} : \begin{bmatrix} \text{HEAD} : [\,] \\ \text{TAIL} : [\,] \end{bmatrix} \end{bmatrix}$$

An example of an element of the set $\text{FIXEDHEAD}(\mathsf{C}, 2)$ *is the sequence contributed by the feature structure* $\begin{bmatrix} \text{HEAD} : [\,] \\ \text{TAIL} : [\,] \end{bmatrix}$*, an element of* $\Gamma(\mathsf{C}, \langle\text{HEAD}\rangle, 2)$*:*

$$\langle \begin{bmatrix} \text{TAIL} : [\,] \end{bmatrix}, \text{HEAD}, [\,] \rangle$$

The same feature structure contributes also another sequence to $\text{FIXEDHEAD}(\mathsf{C}, 2)$ *when it is viewed as a member of the compatibility set* $\Gamma(\mathsf{C}, \langle\text{TAIL}\rangle, 2)$*:*

$$\langle \begin{bmatrix} \text{HEAD} : [\,] \end{bmatrix}, \text{TAIL}, [\,] \rangle$$

Finally, the set $\text{UNBOUNDEDHEAD}(\mathsf{C}, 2)$ *includes the following sequences that are contributed by the same feature structure above:*

$$\langle \begin{bmatrix} \text{TAIL} : [\,] \end{bmatrix}, \text{HEAD}, [\,], \text{TAIL} \rangle, \langle \begin{bmatrix} \text{TAIL} : [\,] \end{bmatrix}, \text{HEAD}, [\,], \text{HEAD} \rangle$$

$$\langle \begin{bmatrix} \text{HEAD} : [\,] \end{bmatrix}, \text{TAIL}, [\,], \text{TAIL} \rangle, \langle \begin{bmatrix} \text{HEAD} : [\,] \end{bmatrix}, \text{TAIL}, [\,], \text{HEAD} \rangle$$

We have shown that the two cases above, FIXEDHEAD and UNBOUNDEDHEAD, cover all the possible feature structures that are unifiable with a rule head C. This accounts for the second major difference between LIG and one-reentrant UG, namely *rule application*. We now investigate the last major difference: *propagation of information in rules*.

In one-reentrant unification grammars information is shared between the rule's head and a single element of the rule's body. We have shown above how the stack of a LIG rule, simulating some unification rule, is defined. We have also discussed the possible options for the stacks of candidate elements in potential sentential forms to which the LIG rule can be applied. We now discuss the body of the LIG rule, when the head – and in particular its stack – is known.

Without loss of generality, let $r^u = \langle \mathsf{C}_0, \ldots, \mathsf{C}_n \rangle$ be a unification rule such that $(0, \mu_0) \overset{r^u}{\leftrightsquigarrow} (e, \mu_e)$, where $1 \leq e \leq n$. This rule is mapped to a *set* of LIG rules. Let r be a member of this set, and let X_0 and X_e be the head and the e-th element of r, respectively. We now explain how structure sharing in the unification rule is modeled in the LIG rule. Consider X_0 first; it was created to reflect a potential unification between C_0, the head of r^u, and some feature structure D_j. The stack of X_0 is $\Psi(\mathsf{A}_0, \pi_0)$, where $\Psi(\mathsf{A}_0, \pi_0)$ is the maximal prefix of $\Psi(\mathsf{D}_j, \pi_j)$ such that $\mathsf{A}_0 \in \Gamma(\mathsf{C}_0, \pi_0, h)$ (notice that it follows from corollary 17 that unifiability is dependent on the prefix only, and the remainder of D_j can be safely ignored). In other words, X_0 was defined to reflect the unification of C_0 and A_0. Consider now the effect of this unification, namely

$$(\langle \mathsf{A}_0 \rangle, 0) \sqcup (r^u, 0) = (\langle \mathsf{P}_0 \rangle, \langle \mathsf{P}_0, \ldots, \mathsf{P}_e, \ldots, \mathsf{P}_n \rangle)$$

When the rule r^u is applied to A_0, information is shared between P_0 and P_e where the shared values are $val(\mathsf{P}_0, \mu_0)$ and $val(\mathsf{P}_e, \mu_e)$. X_0 can have two forms: either it has a fixed stack or an unbounded stack. If the stack of X_0 is fixed, the LIG rule can be applied only to an element (of a sentential form) with an identical stack, i.e., with the same cord. Therefore, X_e should be a LIG representation of P_e. Hence, the stack value of X_e can be defined as a cord whose base feature structure is P_e. The only caveat is the base path of this cord: we have to be careful to define the cord such that its height is limited by the grammar height. Observe that the height of the value of the path μ_e in P_e can exceed the grammar height and recall that due to the unification, $val(\mathsf{P}_e, \mu_e) = val(\mathsf{P}_0, \mu_0)$ (and, in particular, both are of the same height). What we know for certain, however, is that the stack of X_0 is obtained by $\Psi(\mathsf{A}_0, \pi_0)$. Furthermore, P_0 is obtained by unifying C_0 with A_0, so that $\Psi(\mathsf{P}_0, \pi_0)$ is well defined and, in particular, its height is limited by the grammar height. $\Psi(\mathsf{P}_0, \pi_0)$ is graphically depicted in figure 3.1.

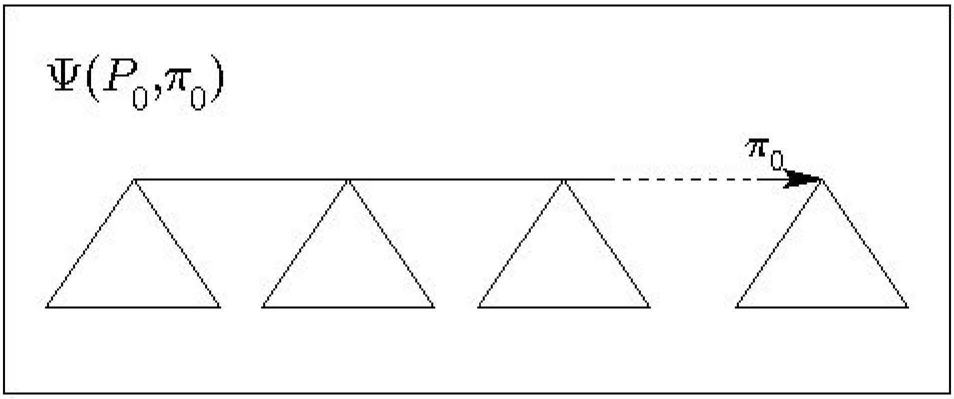


Figure 3.1: The cord $\Psi(\mathsf{P}_0, \pi_0)$.

Consider now the path μ_0 and, in particular, $val(\mathsf{P}_0, \mu_0)$. If μ_0 is *not* a prefix of π_0 then the height of $val(\mathsf{P}_0, \mu_0)$ is limited by the height of $\Psi(\mathsf{P}_0, \pi_0)$ and hence by the grammar height (see figure 3.2

and figure 3.3). Hence $val(\mathsf{P}_e, \mu_e)$ (which is identical to $val(\mathsf{P}_0, \mu_0)$) is also limited by the grammar height. In this case, we 'hang' P_e on the base path μ_e (see figure 3.4).

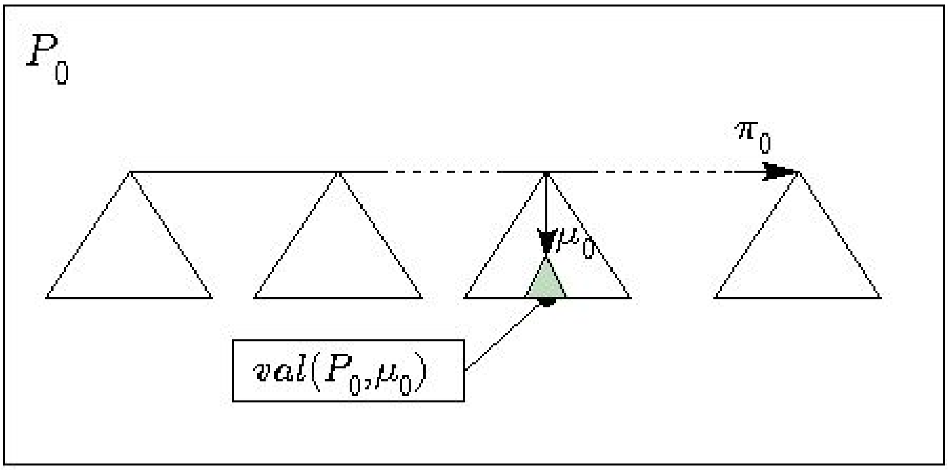


Figure 3.2: P_0, when the stack of X_0 is fixed and μ_0 is not a prefix of π_0.

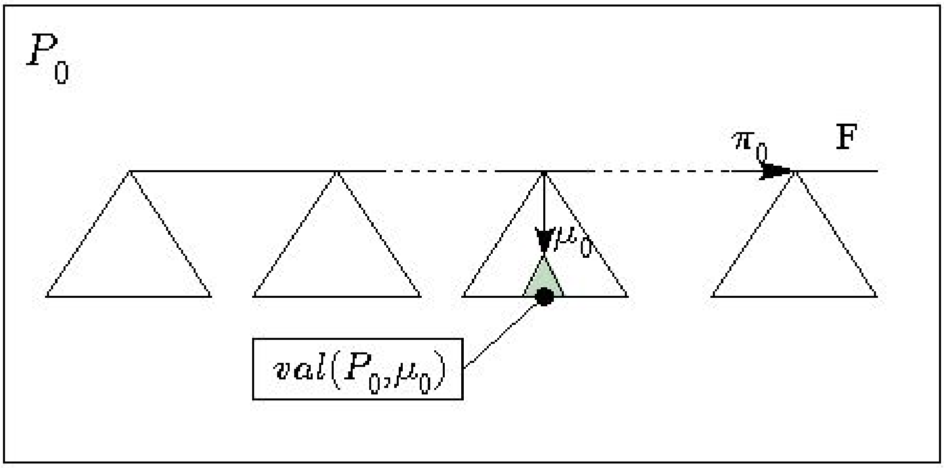


Figure 3.3: P_0, when the stack of X_0 is unbounded and μ_0 is not a prefix of π_0.

If μ_0 *is* a prefix of π_0, however, let $\pi_0 = \mu_0 \cdot \nu$. Again, consider two sub-cases. In the first sub-case, the stack of X_0 is fixed. This situation is graphically depicted in figure 3.5. In this case we can limit the height of $val(\mathsf{P}_0, \mu_0)$ by the height of the cord $\Psi(\mathsf{P}_0, \pi_0)$ and the length of the path ν, $|val(\mathsf{P}_0, \mu_0)| \leq |\Psi(\mathsf{P}_0, \pi_0)| + |\nu|$ and the same holds for $val(\mathsf{P}_e, \mu_e)$. Since the height of the cord $\Psi(\mathsf{P}_0, \pi_0)$ is limited by the grammar height, h, we obtain that $|val(\mathsf{P}_e, \mu_e)| \leq h + |\nu|$. In this case

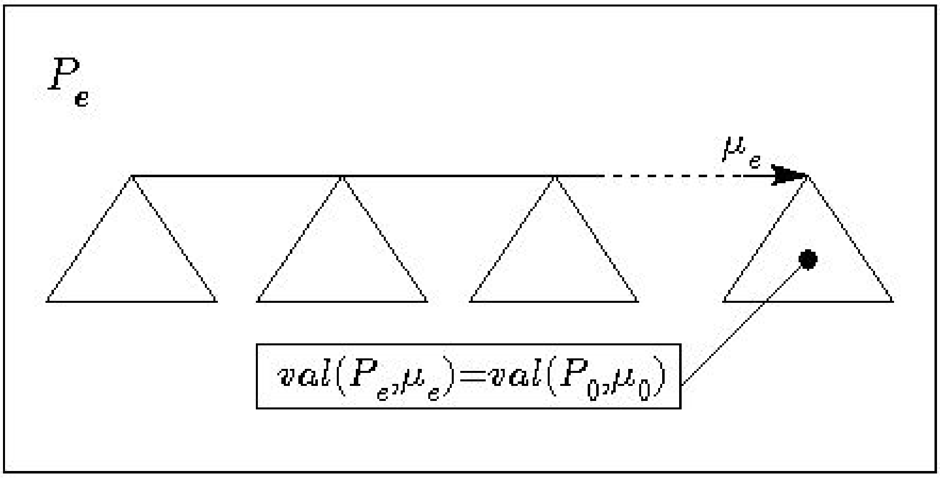


Figure 3.4: P_e, when the path μ_0 is not a prefix of π_0.

we use the path $\mu_e \cdot \nu$ as the base path on which P_e is 'hung' (see figure 3.6).

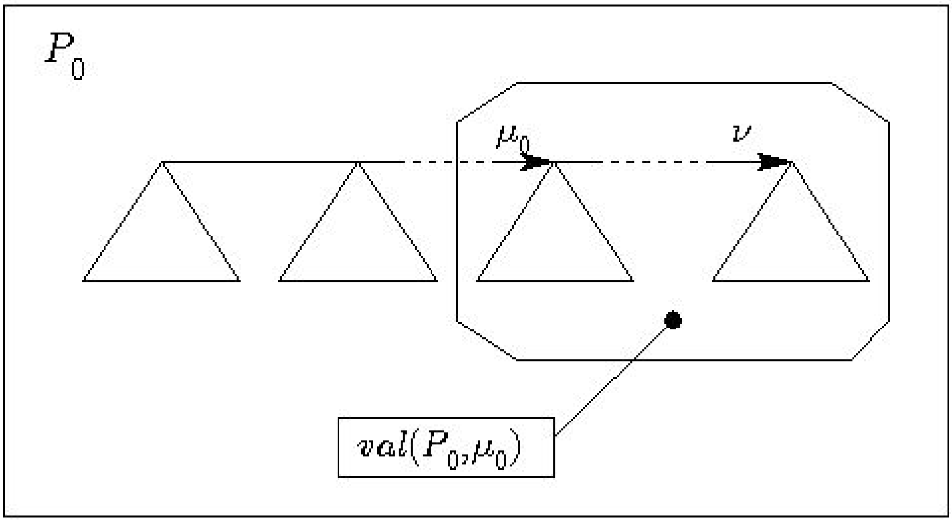


Figure 3.5: P_0, when the stack of X_0 is fixed and $\pi_0 = \mu_0 \cdot \nu$.

However, when the stack of X_0 is unbounded and $\pi_0 = \mu_0 \cdot \nu$, the fixed part of the stack contains not only a cord but also a feature (see definition 27); denote this feature by F. In this case the height of $val(\mathsf{P}_e, \mu_e)$ cannot be bounded because only a subset of the information that is propagated is known when the mapping is computed (see figure 3.7).

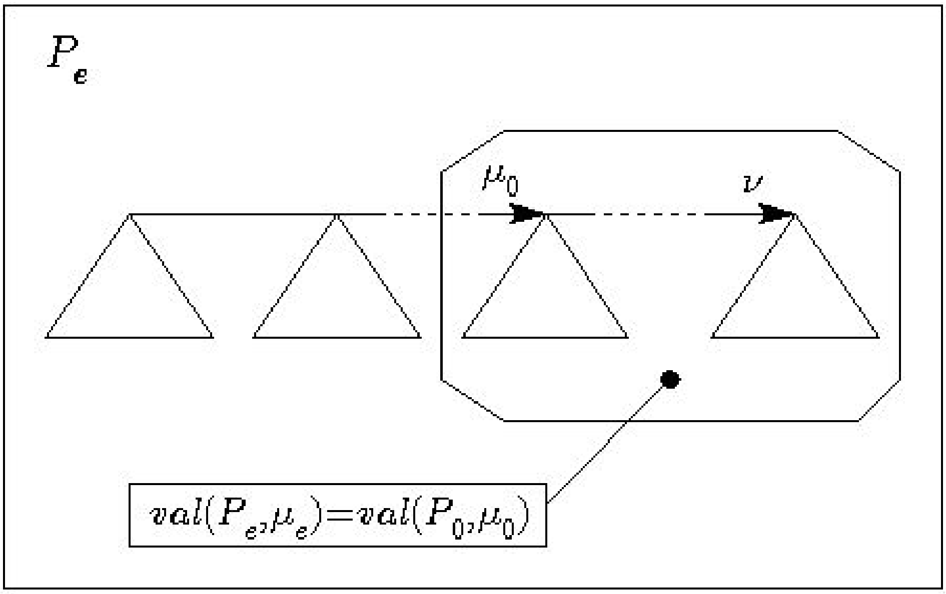


Figure 3.6: P_e, when the stack of X_0 is fixed and $\pi_0 = \mu_0 \cdot \nu$.

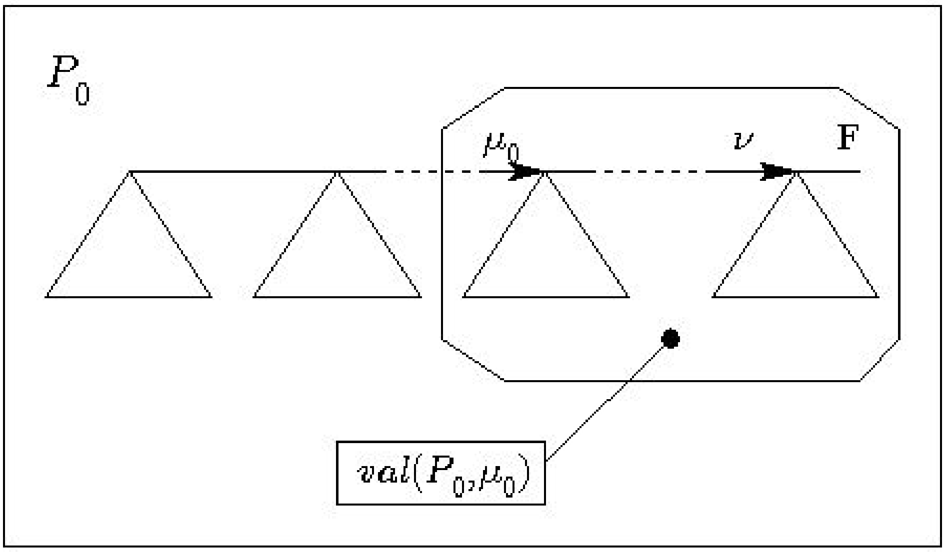


Figure 3.7: P_0, when the stack of X_0 is unbounded and $\pi_0 = \mu_0 \cdot \nu$.

Let D be an element of a sentential form and $r^u = \langle \mathsf{C}_0, \ldots, \mathsf{C}_e, \ldots, \mathsf{C}_n \rangle$ be a one-reentrant rule applicable to D such that $(0, \mu_0) \overset{r^u}{\leadsto} (e, \mu_e)$. Let $\langle \mathsf{Q}_0, \ldots, \mathsf{Q}_e, \ldots, \mathsf{Q}_n \rangle$ be defined as

$$(\langle \mathsf{D} \rangle, 0) \sqcup (r^u, 0) = (\langle \mathsf{Q}_0 \rangle, \langle \mathsf{Q}_0, \ldots, \mathsf{Q}_e, \ldots, \mathsf{Q}_n \rangle)$$

Let $\Psi(\mathsf{D}, \pi_D)$ be a cord of D whose height is limited by the grammar height and let π_0 be the maximal prefix of π_D such that $\pi_0 \in \Pi_{C_0}$ (recall that in our case μ_0 is a prefix of π_0, such that $\pi_0 = \mu_0 \cdot \nu$). We divide the cord $\Psi(\mathsf{Q}_0, \pi_D)$ into three parts as follows (see figure 3.8):

- The first part of the cord (I) is the prefix of the cord whose length is $|\mu_0|$. This part of the cord is not propagated to Q_e.
- The second part (II) is a prefix of the propagated cord that is affected by the unification of C_0 with D. By lemma 16, the length of this part of the cord is limited by the length of the cord $\Psi(\mathsf{C}_0, \pi_0)$ and hence by the grammar height.
- The third part of the cord (III), the suffix, is propagated to Q_e *unchanged*.

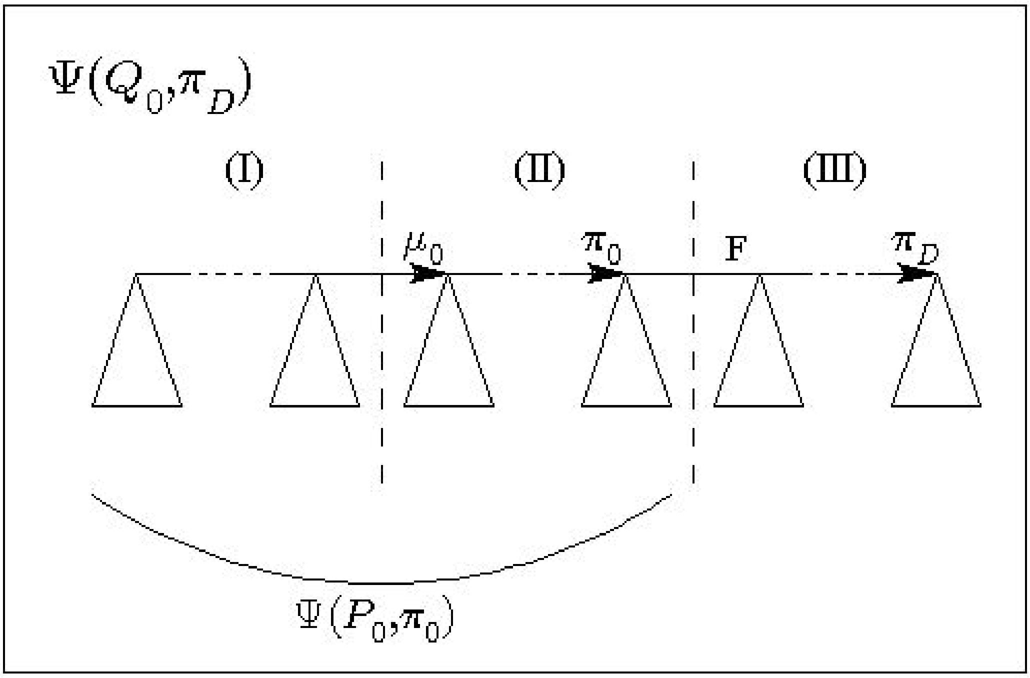


Figure 3.8: The three parts of the cord $\Psi(\mathsf{Q}_0, \pi_D)$.

The only problem is propagating the second part of the cord, since LIG has no provisions for propagating run-time changeable stacks. However, we know that the length of part II of the cord is limited by the grammar height. Therefore, we can calculate the unification of this part of the cord with all possible cords of all rules' heads of the grammar at "compile" time and use the result to define the contents of the stack of the e-th element of the LIG rule. Let P_0 be a feature structure whose cord $\Psi(\mathsf{P}_0, \pi_0)$ is a prefix of the cord $\Psi(\mathsf{Q}_0, \pi_D)$. Let A be a feature structure such that the cord $\Psi(\mathsf{A}, \pi_0)$ is a prefix of $\Psi(\mathsf{D}, \pi_D)$. Hence $\mathsf{P}_0 = \mathsf{C}_0 \sqcup \mathsf{A}$ (see figure 3.9).

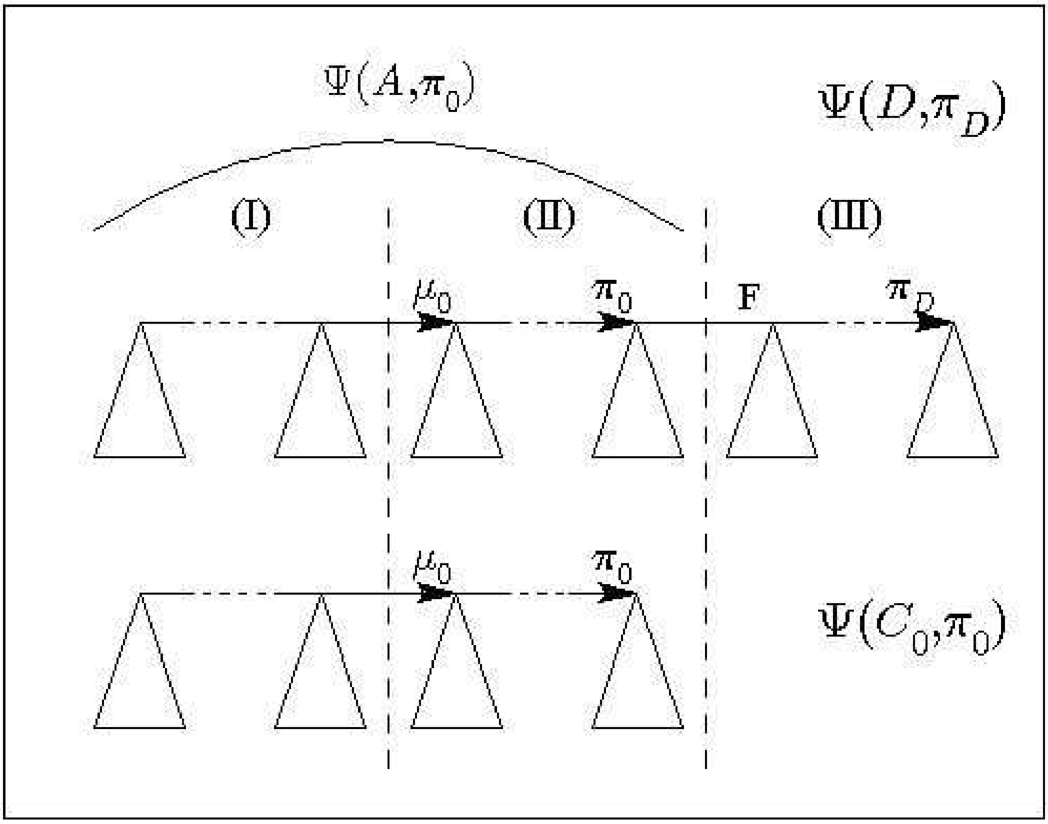


Figure 3.9: The three parts of the cord $\Psi(\mathsf{D}, \pi_D)$.

The information propagated from part II of the cord $\Psi(\mathsf{Q}_0, \pi_D)$ can be calculated by the unification in context of A with the rule r^u. Let the sequence $\langle \mathsf{P}_0, \ldots, \mathsf{P}_e, \ldots, \mathsf{P}_n \rangle$ be defined as

$$(\langle \mathsf{A} \rangle, 0) \sqcup (r^u, 0) = (\langle \mathsf{P}_0 \rangle, \langle \mathsf{P}_0, \ldots, \mathsf{P}_e, \ldots, \mathsf{P}_n \rangle)$$

The rule r^u has only one reentrancy, $(0, \mu_0) \overset{r^u}{\leftrightsquigarrow} (e, \mu_e)$, hence $\mathsf{P}_i = \mathsf{C}_i$ for all $i \neq 0$ and $i \neq e$. Let F be a feature such that $\langle \Psi(\mathsf{A}, \pi_0), \text{F} \rangle$ is a prefix of $\Psi(\mathsf{D}, \pi_D)$. Then in this case we map the rule r^u to

$$N[\Psi(\mathsf{A}, \pi_0), \text{F}..] \to N[\Psi(\mathsf{C}_1, \varepsilon)] \dots N[\Psi(\mathsf{P}_e, \mu_e \cdot \nu), \text{F}..] \dots N[\Psi(\mathsf{C}_n, \varepsilon)]$$

where $X_0 = N[\Psi(\mathsf{A}, \pi_0), \text{F}..]$ and $X_e = N[\Psi(\mathsf{P}_e, \mu_e \cdot \nu), \text{F}..]$. The resulting cord of P_e, $\Psi(\mathsf{P}_e, \mu_e \cdot \nu)$, is limited by the grammar height for the same reason as in the case of X_0 with a fixed stack (see figure 3.10).

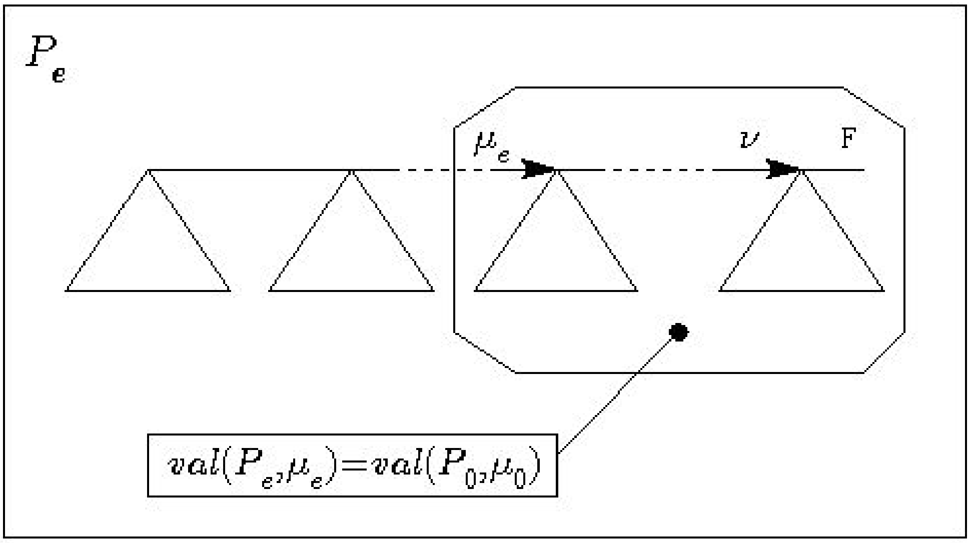


Figure 3.10: P_e, when the stack of X_0 is unbounded and $\pi_0 = \mu_0 \cdot \nu$.

The feature F at the end of the fixed part of the stack of the LIG rule head is added to avoid generation of ill-defined cords in the stacks of elements of LIG sentential forms (see example 15).

Example 15. *Let* $\mathsf{P}_e = \left[\text{F}_1 : \left[\text{F}_2 : \left[\text{F}_3 : a\right]\right]\right]$, $\mu_e = \langle \text{F}_1 \rangle$ *and* $\mu_e \cdot \nu = \langle \text{F}_1, \text{F}_2 \rangle$. *Then for some* $\text{G} \in \text{FEATS}$ *the* e*-th element of the LIG rule body is* $N[\Psi(\mathsf{P}_e, \mu_e \cdot \nu), \text{G}..]$. *If* $\text{G} = \text{F}_3$ *the sequence* $\langle \Psi(\mathsf{P}_e, \mu_e \cdot \nu), \text{G} \rangle$ *is not a valid cord prefix since* $val(\mathit{last}(\Psi(\mathsf{P}_e, \mu_e \cdot \nu)), \langle \text{F}_3 \rangle) \uparrow$.

We now combine all the solutions for the three major differences between one-reentrant unification grammars and LIG to define the mapping from the former to the latter. In a LIG simulating a one-reentrant UG, feature structures are represented as stacks of symbols. The set of stack symbols V_s, therefore, is defined as a set of height bounded non-reentrant feature structures. Also, all the features of the UG are stack symbols. The set V_s is finite due to the restriction on feature structures (no reentrancies and height-boundedness). The set of terminals, V_t, is the set of words of the UG. There are exactly two non-terminal symbols, N and S, the latter of which is the start symbol.

The set of rules can be divided to four: start rule, terminal rules, non-reentrant rules and one-reentrant rules. The *start rule* only applies once in a derivation. It simulates the situation in unification grammars of a rule whose head is unifiable with the start symbol. In LIG, the start rule applies to the start symbol S only; and once applied, it yields a sentential form of length 1, consisting of the non-terminal N with a stack representation of the unification grammar start symbol.

Since the source unification grammar is simplified (definition 18), *terminal rules* are just a straightforward implementation of the lexicon in terms of LIG. *Non-reentrant rules* are simulated in a similar way to how rules of a non-reentrant unification grammar are simulated by CFG (see section 2). The major difference is the head of the rule, X_0, which is defined as explained above. *One-reentrant rules* are simulated in a similar way to non-reentrant rules. The only difference is in the selected element of the rule body, X_e, which is defined as explained above.

Definition 28 (Mapping from UG_{1r} to LIGS). *Let ug2lig be a mapping of* UG_{1r} *to* LIGS*, such that if* $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle \in \text{UG}_{1r}$ *then ug2lig*$(G^u) = \langle V_N, V_t, V_s, \mathcal{R}^{li}, S \rangle$*, where:*

- $V_N = \{N, S\}$*, where* N *and* S *are fresh symbols.*
- $V_t = \text{WORDS}$
- $V_s = \text{FEATS} \cup \{\mathsf{A} \mid \mathsf{A} \in \text{NRFSS}, |\mathsf{A}| \leq \textit{maxHt}(G^u)\}$
- *The set of rules,* $\mathcal{R}^{li}$*, is defined as follows:*

  Let C_0 *be a non-reentrant feature structure, then the* ***rule head*** *set,* $\text{LIGHEAD}(\mathsf{C}_0)$*, is defined as:*

$$\begin{aligned}\text{LIGHEAD}(\mathsf{C}_0) = \quad & \{N[\eta] \mid \eta \in \text{FIXEDHEAD}(\mathsf{C}_0, \textit{maxHt}(G^u))\} \\ \cup \quad & \{N[\eta..] \mid \eta \in \text{UNBOUNDEDHEAD}(\mathsf{C}_0, \textit{maxHt}(G^u))\}\end{aligned}$$

  1. $S[\,] \rightarrow N[\Psi(\mathsf{A}^s, \varepsilon)]$
  2. *For every* $w \in \text{WORDS}$ *such that* $\mathcal{L}(w) = \{\mathsf{C}_0\}$ *and for every* $\pi_0 \in \Pi_{\mathsf{C}_0}$*, the rule* $N[\Psi(\mathsf{C}_0, \pi_0)] \rightarrow w$ *is in* $\mathcal{R}^{li}$*.*

3. *Let* $r^u = \langle \mathsf{C}_0, \ldots, \mathsf{C}_n \rangle \in \mathcal{R}^u$ *be a non-reentrant rule. Then for every* $X_0 \in \textsc{ligHead}(\mathsf{C}_0)$ *the rule* $X_0 \to N[\Psi(\mathsf{C}_1, \varepsilon)] \ldots N[\Psi(\mathsf{C}_n, \varepsilon)]$ *is in* $\mathcal{R}^{li}$.

4. *Let* $r^u = \langle \mathsf{C}_0, \ldots, \mathsf{C}_n \rangle \in \mathcal{R}^u$ *and* $(0, \mu_0) \overset{r^u}{\leftrightsquigarrow} (e, \mu_e)$*, where* $1 \leq e \leq n$*. Then for every* $X_0 \in \textsc{ligHead}(\mathsf{C}_0)$ *the rule*

$$X_0 \to N[\Psi(\mathsf{C}_1, \varepsilon)] \ldots N[\Psi(\mathsf{C}_{e-1}, \varepsilon)]\ X_e\ N[\Psi(\mathsf{C}_{e+1}, \varepsilon)] \ldots N[\Psi(\mathsf{C}_n, \varepsilon)]$$

is in $\mathcal{R}^{li}$*, where* X_e *is defined as follows. Let* π_0 *be the base path of* X_0 *and* A *be the base feature structure of* X_0*. Applying the rule* r^u *to* A*, we now examine the possible modifications to* C_e *by defining* $(\langle \mathsf{A} \rangle, 0) \sqcup (r^u, 0) = (\langle \mathsf{P}_0 \rangle, \langle \mathsf{P}_0, \ldots, \mathsf{P}_e, \ldots, \mathsf{P}_n \rangle)$.

   (a) *If* μ_0 *is not a prefix of* π_0 *then* $X_e = N[\Psi(\mathsf{P}_e, \mu_e)]$.

   (b) *If* $\pi_0 = \mu_0 \cdot \nu$, $\nu \in \textsc{Paths}$ *then*

      i. *If* $X_0 = N[\Psi(\mathsf{A}, \pi_0)]$ *then* $X_e = N[\Psi(\mathsf{P}_e, \mu_e \cdot \nu)]$.

      ii. *If* $X_0 = N[\Psi(\mathsf{A}, \pi_0), \textsc{f}..]$ *then* $X_e = N[\Psi(\mathsf{P}_e, \mu_e \cdot \nu), \textsc{f}..]$.

In order for the construction to be well defined, all cords must be shown to have heights limited by the grammar height. This was informally shown in the discussion above.

Example 16. *To illustrate the mapping of one-reentrant unification rules to LIG rules we give here some examples of LIG rules created from the rule* r_5^u *of the example 9. Recall that the rule* r_5^u *is defined as:*

$$\begin{bmatrix} \textsc{head}: N_3 \\ \textsc{tail}: \begin{bmatrix} \textsc{head}: a \\ \textsc{tail}: \boxed{1} \end{bmatrix} \end{bmatrix} \to \begin{bmatrix} \textsc{head}: a \\ \textsc{tail}: \ elist \end{bmatrix} \begin{bmatrix} \textsc{head}: N_3 \\ \textsc{tail}: \boxed{1} \end{bmatrix}$$

In this rule $\mu_0 = \langle \textsc{tail}, \textsc{tail} \rangle$ *and* $\mu_e = \mu_2 = \langle \textsc{tail} \rangle$.

- *Case 4a,* μ_0 *is not a prefix of* π_0*. Let* $\pi_0 = \langle \textsc{head} \rangle$ *and* $\mathsf{A} = \begin{bmatrix} \textsc{head}: [\,] \\ \textsc{tail}: \begin{bmatrix} \textsc{head}: a \\ \textsc{tail}: \ b \end{bmatrix} \end{bmatrix}$*. Then the LIG rule is*

$$N[\begin{bmatrix} \textsc{tail}: \begin{bmatrix} \textsc{head}: a \\ \textsc{tail}: \ b \end{bmatrix} \end{bmatrix}, \textsc{head}, [\,]\,] \to N[\begin{bmatrix} \textsc{head}: a \\ \textsc{tail}: \ elist \end{bmatrix}]\ N[\begin{bmatrix} \textsc{head}: N_3 \end{bmatrix}, \textsc{tail}, b\,]$$

- *Case 4(b)i, μ_0 is a prefix of π_0 and the stack of the head of the LIG rule is fixed. Let* $\pi_0 = \langle \text{TAIL}, \text{TAIL} \rangle$ *and* $\mathsf{A} = \begin{bmatrix} \text{HEAD} : [\,] \\ \text{TAIL} : \begin{bmatrix} \text{HEAD} : a \\ \text{TAIL} : \ b \end{bmatrix} \end{bmatrix}$. *Then the LIG rule is*

$$N[\begin{bmatrix}\text{HEAD} : [\,]\end{bmatrix}, \text{TAIL}, \begin{bmatrix}\text{HEAD} : a\end{bmatrix}, \text{TAIL}, b\,] \rightarrow$$
$$N[\begin{bmatrix}\text{HEAD} : a \\ \text{TAIL} : \ elist\end{bmatrix}]\, N[\begin{bmatrix}\text{HEAD} : N_3\end{bmatrix}, \text{TAIL}, b\,]$$

- *Case 4(b)ii, μ_0 is a prefix of π_0 and the stack of the head of the LIG rule is unbounded. Let* $\pi_0 = \langle \text{TAIL}, \text{TAIL} \rangle$ *and*

$$\mathsf{A} = \begin{bmatrix} \text{HEAD} : [\,] \\ \text{TAIL} : \begin{bmatrix} \text{HEAD} : a \\ \text{TAIL} : \begin{bmatrix}\text{HEAD} : b\end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Then the LIG rule is

$$N[\begin{bmatrix}\text{HEAD} : [\,]\end{bmatrix}, \text{TAIL}, \begin{bmatrix}\text{HEAD} : a\end{bmatrix}, \text{TAIL}, \begin{bmatrix}\text{HEAD} : b\end{bmatrix}, \text{TAIL}..] \rightarrow$$
$$N[\begin{bmatrix}\text{HEAD} : a \\ \text{TAIL} : \ elist\end{bmatrix}]\, N[\begin{bmatrix}\text{HEAD} : N_3\end{bmatrix}, \text{TAIL}, \begin{bmatrix}\text{HEAD} : b\end{bmatrix}, \text{TAIL}..]$$

Theorem 18. *Let $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle$ be a one-reentrant unification grammar and $\mathsf{A}^s \stackrel{*}{\Longrightarrow}_u \mathsf{A}_1 \dots \mathsf{A}_n$ be a derivation sequence. If $G^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, S \rangle = ug2lig(G^u)$ then there is a sequence of paths $\langle \pi_1, \dots, \pi_n \rangle$, such that $S[\,] \stackrel{*}{\Longrightarrow}_{li} N[\Psi(\mathsf{A}_1, \pi_1)] \dots N[\Psi(\mathsf{A}_n, \pi_n)]$.*

Proof. We prove by induction on the length of the derivation sequence. The induction hypothesis is that if $\mathsf{A}^s \stackrel{k}{\Longrightarrow}_u \mathsf{A}_1 \dots \mathsf{A}_n$, then there is a sequence of paths $\langle \pi_1, \dots, \pi_n \rangle$, such that $S[\,] \stackrel{k+1}{\Longrightarrow}_{li} N[\Psi(\mathsf{A}_1, \pi_1)] \dots N[\Psi(\mathsf{A}_n, \pi_n)]$. If $k = 0$, then

1. By the definition of derivation in UG, $\mathsf{A}^s \stackrel{0}{\Longrightarrow}_u \mathsf{A}^s$;

2. By definition 28 case (1), the rule $S[\,] \rightarrow N[\Psi(\mathsf{A}^s, \varepsilon)]$ is in $\mathcal{R}^{li}$.

3. Hence, $S[\,] \stackrel{1}{\Longrightarrow}_{li} N[\Psi(\mathsf{A}^s, \varepsilon)]$ and $N[\Psi(\mathsf{A}^s, \varepsilon)]$ is well defined since $\Psi(\mathsf{A}^s, \varepsilon) = \langle \mathsf{A}^s \rangle$, $|\mathsf{A}^s| \leq$ *maxHt*(G^u).

Assume that the hypothesis holds for every i, $0 \leq i < k$. Assume further that $\mathsf{A}^s \stackrel{k-1}{\Longrightarrow}_u \mathsf{D}_1 \ldots \mathsf{D}_m \stackrel{1}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$.

1. By definition of UG derivation, there are an index j and a rule $r^u = \mathsf{C}_0 \rightarrow \mathsf{C}_1 \ldots \mathsf{C}_{n-m+1}$, $r^u \in \mathcal{R}^u$, such that r^u is applicable to D_j:

$$(\langle \mathsf{C}_0, \ldots, \mathsf{C}_{n-m+1} \rangle, 0) \sqcup (\langle \mathsf{D}_1, \ldots, \mathsf{D}_m \rangle, j) = (\langle \mathsf{Q}_0, \ldots, \mathsf{Q}_{n-m+1} \rangle, \langle \mathsf{D}_1 \ldots \mathsf{D}_{j-1} \mathsf{Q}_0 \mathsf{D}_{j+1} \ldots \mathsf{D}_m \rangle)$$

where

- $\langle \mathsf{A}_1, \ldots, \mathsf{A}_{j-1} \rangle = \langle \mathsf{D}_1, \ldots, \mathsf{D}_{j-1} \rangle$
- $\langle \mathsf{A}_j, \ldots, \mathsf{A}_{n-m+j} \rangle = \langle \mathsf{Q}_1, \ldots, \mathsf{Q}_{n-m+1} \rangle$
- $\langle \mathsf{A}_{n-m+j+1}, \ldots, \mathsf{A}_n \rangle = \langle \mathsf{D}_{j+1}, \ldots, \mathsf{D}_m \rangle$

Note that it is only possible to write the MRS $\langle \mathsf{A}_1, \ldots, \mathsf{A}_n \rangle$ in such a way due to the fact that the grammar G^u is one-reentrant: by lemma 5, no reentrancies can occur among two elements in a sentential form.

2. Hence, $\mathsf{A}^s \stackrel{k}{\Longrightarrow}_u \mathsf{D}_1 \ldots \mathsf{D}_{j-1} \mathsf{Q}_1 \ldots \mathsf{Q}_{n-m+1} \mathsf{D}_{j+1} \ldots \mathsf{D}_m$

3. By the induction hypothesis there is a sequence of paths $\langle \nu_1, \ldots, \nu_m \rangle$ such that

$$S[\,] \stackrel{k}{\Longrightarrow}_{li} N[\Psi(\mathsf{D}_1, \nu_1)] \ldots N[\Psi(\mathsf{D}_m, \nu_m)]$$

4. We denote $\Psi(\mathsf{D}_j, \nu_j)$ as $\langle \mathsf{B}_1, \mathrm{F}_1, \ldots, \mathsf{B}_{|\nu_j|+1} \rangle$ (recall that j is the index of the selected element in the sentential form).

   We now want to show the existence of a rule $r \in \mathcal{R}^{li}$, created from r^u by the mapping *ug2lig*, which can be applied to j-th element of the LIG sentential form, $N[\Psi(\mathsf{D}_j, \nu_j)]$. We define the feature structure A to be a "bridge" between D_j and C_0 which together with a path π_0 (a prefix of the path ν_j) defines the head of the rule r.

5. Let π_0 be a maximal prefix of ν_j such that $\pi_0 \in \Pi_{C_0}$. Recall that $\langle \mathsf{B}_1, \mathrm{F}_1, \ldots, \mathsf{B}_{|\pi_0|+1} \rangle$ is a prefix of $\Psi(\mathsf{D}_j, \nu_j)$ because π_0 is a prefix of ν_j. Let A be such that $\Psi(\mathsf{A}, \pi_0) = \langle \mathsf{B}_1, \mathrm{F}_1, \ldots, \mathsf{B}_{|\pi_0|+1} \rangle$. By the induction hypothesis, $\mathsf{B}_i \leq \mathit{maxHt}(G^u)$, $1 \leq i \leq |\nu_j| + 1$. We will show that A is unifiable with both D_j and C_0.

6. We first show that $\mathsf{A} \in \Gamma(\mathsf{C}_0, \mathit{maxHt}(G^u))$. Since $\mathsf{D}_j \sqcup \mathsf{C}_0 \neq \top$ and A is a substructure of D_j we obtain that $\mathsf{A} \sqcup \mathsf{C}_0 \neq \top$. Since $\pi_0 \in \Pi_A$ and $|\mathsf{B}_i||leq\mathit{maxHt}(G^u)$, $1 \leq i \leq |\nu_j| + 1$, $\mathsf{A} \in \Gamma(\mathsf{C}_0, \mathit{maxHt}(G^u))$.

7. We now show that there is a LIG rule r, a mapping of r^u, which is applicable to $N[\Psi(\mathsf{D}_j, \nu_j)]$. There are two possibilities for the relation between π_0 and ν_j (recall that π_0 is a prefix of ν_j):

   - If $\nu_j = \pi_0$ then $\mathsf{A} = \mathsf{D}_j$ and $\Psi(\mathsf{A}, \pi_0) = \Psi(\mathsf{D}_j, \nu_j)$. Hence, every rule of the form $N[\Psi(\mathsf{A}, \pi_0)] \to \alpha$ is applicable to $\Psi(\mathsf{D}_j, \nu_j)$. Since $\mathsf{A} \in \Gamma(\mathsf{C}_0, \mathit{maxHt}(G^u))$ we obtain that $N[\Psi(\mathsf{A}, \pi_0)] \in \textsc{ligHead}(\mathsf{C}_0)$. Hence, the rule $N[\Psi(\mathsf{A}, \pi_0)] \to \alpha$ is in $\mathcal{R}^{li}$, where $\alpha \in (V_N[V_s^*] \cup V_t)^*$ is determined by r^u.

   - If $\nu_j \neq \pi_0$ then $\nu_j = \pi_0 \cdot \langle \mathrm{F}_{|\pi_0|+1}, \ldots, \mathrm{F}_{|\nu_j|} \rangle$. Recall that $val(\mathsf{B}_{|\pi_0|+1}, \langle \mathrm{F}_{|\pi_0|+1} \rangle) \uparrow$ because $\Psi(\mathsf{D}_j, \nu_j) = \langle \mathsf{B}_1, \mathrm{F}_1, \ldots, \mathsf{B}_{|\nu_j|+1} \rangle$ and $|\pi_0| + 1 < |\nu_j| + 1$. Since $\Psi(\mathsf{A}, \pi_0) = \langle \mathsf{B}_1, \mathrm{F}_1, \ldots, \mathsf{B}_{|\pi_0|+1} \rangle$, we obtain that every rule of the form $N[\Psi(\mathsf{A}, \pi_0), \mathrm{F}_{|\pi_0|+1}..] \to \alpha$ is applicable to $N[\Psi(\mathsf{D}_j, \nu_j)]$. Since $\mathsf{A} \in \Gamma(\mathsf{C}_0, \mathit{maxHt}(G^u))$ we obtain that

     $$N[\Psi(\mathsf{A}, \pi_0), \mathrm{F}_{|\pi_0|+1}..] \in \textsc{ligHead}(\mathsf{C}_0)$$

     Hence, the rule $N[\Psi(\mathsf{A}, \pi_0), \mathrm{F}_{|\pi_0|+1}..] \to \alpha$ is in $\mathcal{R}^{li}$, where $\alpha \in (V_N[V_s^*] \cup V_t)^*$ is determined by r^u.

8. The LIG rule r whose existence was established in (7) is applied to $N[\Psi(\mathsf{D}_j, \nu_j)]$ as follows:

$$\begin{aligned} S[\,] &\overset{k}{\Longrightarrow}_{li} N[\Psi(\mathsf{D}_1, \nu_1)] \ldots N[\Psi(\mathsf{D}_m, \nu_m)] \\ &\overset{1}{\Longrightarrow}_{li} N[\Psi(\mathsf{D}_1, \nu_1)] \ldots N[\Psi(\mathsf{D}_{j-1}, \nu_{j-1})]\; Y_1 \ldots Y_{n-m+1}\; N[\Psi(\mathsf{D}_{j+1}, \nu_{j+1})] \ldots N[\Psi(\mathsf{D}_m, \nu_m)] \end{aligned}$$

9. We now investigate the possible outcomes of applying the rule r to the selected element of the sentential form. Let $r = X_0 \to \alpha$, where $\alpha \in (V_N[V_s^*] \cup V_t)^*$. To complete the proof we have

to show that for some sequence of paths $\langle \pi_1, \ldots, \pi_{n-m+1} \rangle$,

$$\langle Y_1, \ldots, Y_{n-m+1} \rangle = \langle N[\Psi(\mathsf{Q}_1, \pi_1)], \ldots, N[\Psi(\mathsf{Q}_{n-m+1}, \pi_{n-m+1})] \rangle$$

where $\mathsf{Q}_1, \ldots \mathsf{Q}_{n-m+1}$ are determined by the unification grammar, see (1) above.

- Assume that r^u has no reentrancies. Hence, $\mathsf{Q}_i = \mathsf{C}_i$, $1 \leq i \leq n-m+1$. By definition 28 case (3), the LIG rule body is

  $$\alpha = \langle N[\Psi(\mathsf{C}_1), \varepsilon)], \ldots, N[\Psi(\mathsf{C}_{n-m+1}], \varepsilon)] \rangle = \langle N[\Psi(\mathsf{Q}_1), \varepsilon)], \ldots, N[\Psi(\mathsf{Q}_{n-m+1}], \varepsilon)] \rangle$$

  Since the rule r does not copy the stack, $\alpha = \langle Y_1, \ldots, Y_{n-m+1} \rangle$. Therefore,

  $$\langle Y_1, \ldots, Y_{n-m+1} \rangle = \langle N[\Psi(\mathsf{Q}_1, \varepsilon)], \ldots, N[\Psi(\mathsf{Q}_{n-m+1}, \varepsilon)] \rangle$$

- Assume that $(0, \mu_0) \overset{r^u}{\rightsquigarrow} (e, \mu_e)$, where $1 \leq e \leq n$. Hence, $\mathsf{Q}_i = \mathsf{C}_i$ and $Y_i = N[\Psi(\mathsf{Q}_i, \varepsilon)]$ is well defined for all i, $i \neq e$. By definition 28 case (4), the LIG rule body is

  $$\begin{aligned} \alpha &= \langle N[\Psi(\mathsf{C}_1, \varepsilon)], \ldots, N[\Psi(\mathsf{C}_{e-1}, \varepsilon)], X_e, N[\Psi(\mathsf{C}_{e+1}, \varepsilon)], \ldots, N[\Psi(\mathsf{C}_{n-m+1}, \varepsilon)] \rangle \\ &= \langle N[\Psi(\mathsf{Q}_1, \varepsilon)], \ldots, N[\Psi(\mathsf{Q}_{e-1}, \varepsilon)], X_e, N[\Psi(\mathsf{Q}_{e+1}, \varepsilon)], \ldots, N[\Psi(\mathsf{Q}_{n-m+1}, \varepsilon)] \rangle \end{aligned}$$

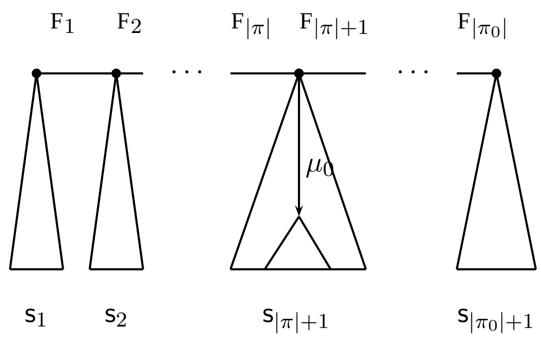
  This case is similar to the previous case, with the exception of X_e, which may be more complicated due to the propagation of the stack from X_0. We therefore focus on X_e (other elements of α are as above). Recall that by definition 28, $\langle \mathsf{P}_0, \ldots, \mathsf{P}_{n-m+1} \rangle$ is a sequence of feature structures such that

  $$(\langle \mathsf{A} \rangle, 0) \sqcup (r^u, 0) = (\langle \mathsf{P}_0 \rangle, \langle \mathsf{P}_0 \ldots \mathsf{P}_{n-m+1} \rangle)$$

  We now analyze all the possible values of X_e, according to definition 28 case (4):

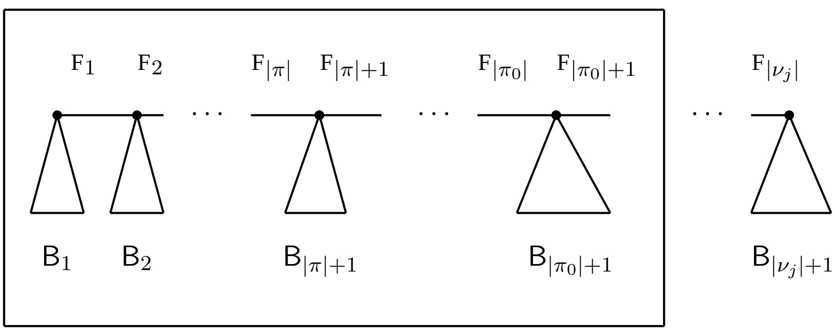
  (a) Case 4a: if μ_0 is not a prefix of π_0 then by definition 28, $X_e = N[\Psi(\mathsf{P}_e, \mu_e)]$. Let π be the maximal prefix of π_0 and μ_0 such that $\mu_0 = \pi \cdot \mu_0'$. We denote $\Psi(\mathsf{C}_0, \pi_0)$ as

$\langle \mathsf{s}_1, \mathrm{F}_1, \ldots, \mathsf{s}_{|\pi_0|+1} \rangle$, and graphically represent it as:



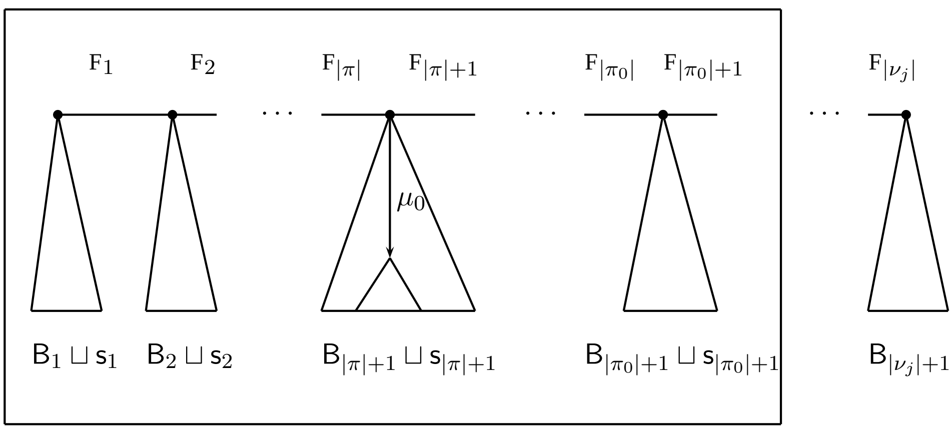
The cord $\Psi(\mathsf{D}_j, \nu_j)$ with its prefix $\Psi(\mathsf{A}, \pi_0)$ are represented as follows:

$\Psi(\mathsf{A}, \pi_0)$



Note that the case $\pi_0 = \nu_j$ is just a special case of the figure above. The cord $\Psi(\mathsf{D}_j \sqcup \mathsf{C}_0, \nu_j)$ with its prefix $\Psi(\mathsf{A} \sqcup \mathsf{C}_0, \pi_0)$ are represented as follows:

$\Psi(\mathsf{A} \sqcup \mathsf{C}_0, \pi_0)$



Hence, $val(\mathsf{A} \sqcup \mathsf{C}_0, \mu_0) = val(\mathsf{B}_{|\pi|+1} \sqcup \mathsf{s}_{|\pi|+1}, \mu_0') = val(\mathsf{D}_j \sqcup \mathsf{C}_0, \mu_0)$. By definition of unification in context $val(\mathsf{P}_e, \mu_e) = val(\mathsf{A} \sqcup \mathsf{C}_0, \mu_0)$ and $val(\mathsf{Q}_e, \mu_e) =$

$val(\mathsf{D}_j \sqcup \mathsf{C}_0, \mu_0)$. Hence, $val(\mathsf{P}_e, \mu_e) = val(\mathsf{Q}_e, \mu_e)$ and $\mathsf{Q}_e = \mathsf{P}_e$. Therefore,

$$\begin{aligned} \alpha = & \langle Y_1, \ldots, Y_{n-m+1} \rangle \\ = & \langle N[\Psi(\mathsf{Q}_1, \varepsilon)], \ldots, N[\Psi(\mathsf{Q}_e, \mu_e)], \ldots, N[\Psi(\mathsf{Q}_{n-m+1}, \varepsilon)] \rangle \end{aligned}$$

(b) Case 4b: if μ_0 is a prefix of π_0, let $\pi_0 = \mu_0 \cdot \nu$, $\nu \in$ PATHS. Then by definition 28, the following holds:

- Case 4(b)i:

  If $X_0 = N[\Psi(\mathsf{A}, \pi_0)]$ then $X_e = N[\Psi(\mathsf{P}_e, \mu_e \cdot \nu)]$. Since $N[\Psi(\mathsf{A}, \pi_0)]$ is applicable to $N[\Psi(\mathsf{D}_j, \nu_j)]$ we obtain that $\pi_0 = \nu_j$ and $\mathsf{A} = \mathsf{D}_j$. Hence, $\mathsf{P}_e = \mathsf{Q}_e$. Therefore,

$$\begin{aligned} \alpha = & \langle Y_1, \ldots, Y_{n-m+1} \rangle \\ = & \langle N[\Psi(\mathsf{Q}_1, \varepsilon)], \ldots, N[\Psi(\mathsf{Q}_e, \mu_e \cdot \nu)], \ldots, N[\Psi(\mathsf{Q}_{n-m+1}, \varepsilon)] \rangle \end{aligned}$$

- Case 4(b)ii:

  If $X_0 = N[\Psi(\mathsf{A}, \pi_0), \mathrm{F}_{|\pi_0|+1}..]$ then $X_e = N[\Psi(\mathsf{P}_e, \mu_e \cdot \nu), \mathrm{F}_{|\pi_0|+1}..]$. Let $\beta = \langle \mathsf{B}_{|\pi_0|+2}, \mathrm{F}_{|\pi_0|+2}, \ldots, \mathsf{B}_{|\nu_j|+1} \rangle$. By definition of A, $\Psi(\mathsf{D}_j, \nu_j) = \Psi(\mathsf{A}, \pi_0) \cdot \langle \mathrm{F}_{|\pi_0|+1} \rangle \cdot \beta$. We apply the LIG rule r to $N[\Psi(\mathsf{D}_j, \nu_j)]$ and obtain

$$\begin{aligned} & \langle Y_1, \ldots, Y_{n-m+1} \rangle \\ & = \langle N[\Psi(\mathsf{Q}_1, \varepsilon)], \ldots, N[\Psi(\mathsf{P}_e, \mu_e \cdot \nu), \mathrm{F}_{|\pi_0|+1}, \beta], \ldots, N[\Psi(\mathsf{Q}_{n-m+1}, \varepsilon)] \rangle \end{aligned}$$

  By definition of unification in context P_e differs from Q_e only in the value of the path $\mu_e \cdot \nu \cdot \langle \mathrm{F}_{|\pi_0|+1} \rangle$. The difference is in the value of the path $\mu_e \cdot \nu \cdot \langle \mathrm{F}_{|\pi_0|+1} \rangle$, it is not defined in P_e and equals β in Q_e. Hence, $\Psi(\mathsf{P}_e, \mu_e \cdot \nu) \cdot \langle \mathrm{F}_{|\pi_0|+1} \rangle \cdot \beta = \Psi(\mathsf{Q}_e, \mu_e \cdot \nu)$. Therefore,

$$\langle Y_1, \ldots, Y_{n-m+1} \rangle = \langle N[\Psi(\mathsf{Q}_1, \varepsilon)], \ldots, N[\Psi(\mathsf{Q}_e, \nu_j)], \ldots, N[\Psi(\mathsf{Q}_{n-m+1}, \varepsilon)] \rangle$$

Note that in this case Y_e is well defined because it was created by applying a LIG rule to a well defined non-terminal symbol.

□

Theorem 19. *Let* $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L}\rangle$ *be a one-reentrant unification grammar and* $G^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, N\rangle$ = *ug2lig*(G^u) *be LIG. If* $S[\,] \overset{*}{\Longrightarrow}_{li} Y_1 \dots Y_n$*, where* $Y_i \in V_N[V_s^*]$, $1 \le i \le n$*, then there are a sequence of paths* $\langle \pi_1, \dots, \pi_n\rangle$ *and a derivation sequence* $\mathsf{A}^s \overset{*}{\Longrightarrow}_u \mathsf{A}_1 \dots \mathsf{A}_n$*, such that* $Y_i = N[\Psi(\mathsf{A}_i, \pi_i)]$.

Proof. We prove by induction on the length of the LIG derivation sequence. The induction hypothesis is that if $S[\,] \overset{k}{\Longrightarrow}_{li} Y_1 \dots Y_n$, where $Y_i \in V_N[V_s^*]$, $1 \le i \le n$, then $\mathsf{A}^s \overset{k-1}{\Longrightarrow}_u \mathsf{A}_1 \dots \mathsf{A}_n$, such that for some sequence of paths $\langle \pi_1, \dots, \pi_n\rangle$, $Y_i = N[\Psi(\mathsf{A}_i, \pi_i)]$, $1 \le i \le n$. If $k = 1$, then

1. By definition 28, the only rule that may be applied to the start symbol S in G^{li} is the rule defined by case (1) of the definition: $S[\,] \to N[\Psi(\mathsf{A}^s, \varepsilon)]$.
2. Hence, for $k = 1$, the only derivation sequence is $S[\,] \overset{1}{\Longrightarrow}_{li} N[\Psi(\mathsf{A}^s, \varepsilon)]$
3. By the definition of derivation in UG, $\mathsf{A}^s \overset{0}{\Longrightarrow}_u \mathsf{A}^s$.

Assume that the hypothesis holds for every i, $1 \le i \le k$; let the length of the derivation sequence be $k + 1$.

1. Assume that $S[\,] \overset{k+1}{\Longrightarrow}_{li} Y_1 \dots Y_n$. Then $S[\,] \overset{k}{\Longrightarrow}_{li} Y'_1 \dots Y'_m \overset{1}{\Longrightarrow}_{li} Y_1 \dots Y_n$.
2. By the induction hypothesis, there exists a sequence of paths $\langle \nu_1, \dots, \nu_m\rangle$ and feature structures $\mathsf{D}_1, \dots, \mathsf{D}_m$, such that for $1 \le i \le m$, $Y'_i = N[\Psi(\mathsf{D}_i, \nu_i)]$, and $\mathsf{A}^s \overset{k-1}{\Longrightarrow}_u \mathsf{D}_1 \dots \mathsf{D}_m$. We therefore write:

$$S[\,] \overset{k}{\Longrightarrow}_{li} N[\Psi(\mathsf{D}_1, \nu_1)] \dots N[\Psi(\mathsf{D}_m, \nu_m)] \overset{1}{\Longrightarrow}_{li} Y_1 \dots Y_n$$

3. Furthermore, let $r = X_0 \to X_1 \dots X_{n-m+1}$ be the G^{li} rule used for the last derivation step, and j be the index of the element to which r is applied, such that

$$N[\Psi(\mathsf{D}_1, \nu_1)] \dots N[\Psi(\mathsf{D}_m, \nu_m)] \overset{1}{\Longrightarrow}_{li}$$
$$N[\Psi(\mathsf{D}_1, \nu_1)] \dots N[\Psi(\mathsf{D}_{j-1}, \nu_{j-1})] Y_j \dots Y_{n-m+j} N[\Psi(\mathsf{D}_{n-m+j+1}, \nu_{n-m+j+1})] \dots N[\Psi(\mathsf{D}_m, \nu_m)]$$

4. We denote $\Psi(\mathsf{D}_j, \nu_j)$ as $\langle \mathsf{t}_1, \mathtt{F}_1, \dots, \mathsf{t}_{|\nu_j|+1}\rangle$

5. By definition 28, the rules that may be applied to $N[\Psi(\mathsf{D}_j, \nu_j)]$ are created by cases (3) and (4) of the definition, because the rule created by case (1) is headed by the non-terminal symbol S and the rules created by case (2) do not derive non-terminal symbols. Let $r^u = \mathsf{C}_0 \rightarrow \mathsf{C}_1 \ldots \mathsf{C}_{n-m+1}$ be a rule in $\mathcal{R}^u$ such that the rule r is created from r^u. Note that there may be more than one such rule.

6. We now show that $\mathsf{C}_0 \sqcup \mathsf{D}_j \neq \top$. In both cases (3) and (4) of definition 28 the head of the rule r, X_0, is a member of $\text{LIGHEAD}(\mathsf{C}_0)$. Since r is applicable to $N[\Psi(\mathsf{D}_j, \nu_j)]$ we obtain that X_0 has one of the following forms:

   (a) $X_0 = N[\Psi(\mathsf{D}_j, \nu_j)]$. By definition 27, $\Psi(\mathsf{D}_j, \nu_j) \in \text{FIXEDHEAD}(\mathsf{C}_0, \textit{maxHt}(G^u))$. Since Ψ is a one-to-one mapping, we obtain that $\nu_j \in \Pi_{C_0}$ and $\mathsf{D}_j \in \Gamma(\mathsf{C}_0, \nu_j, \textit{maxHt}(G^u))$. By definition of Γ, $\mathsf{D}_j \sqcup \mathsf{C}_0 \neq \top$.

   (b) $X_0 = N[\eta..]$, where η is a prefix of $\Psi(\mathsf{D}_j, \nu_j)$. Hence, we obtain that

   $$\eta = \langle \mathsf{t}_1, \mathrm{F}_1, \ldots, \mathsf{t}_{|\pi_0|+1}, \mathrm{F}_{|\pi_0|+1} \rangle$$

   where π_0 is a prefix of ν_j. By definition 27, $\eta \in \text{UNBOUNDEDHEAD}(\mathsf{C}_0, \textit{maxHt}(G^u))$. Hence, there are a path $\pi_0 \in \Pi_{C_0}$ and a feature structure $\mathsf{A} \in \Gamma(\mathsf{C}_0, \pi_0, \textit{maxHt}(G^u))$ such that $\eta = \Psi(\mathsf{A}, \pi_0) \cdot \langle \mathrm{F}_{|\pi_0|+1} \rangle$. By definition of Γ, $\mathsf{A} \sqcup \mathsf{C}_0 \neq \top$. Therefore, by corollary 17, $\mathsf{C}_0 \sqcup \mathsf{D}_j \neq \top$.

7. Since $\mathsf{C}_0 \sqcup \mathsf{D}_j \neq \top$, the rule r^u is applicable to D_j as follows:

   $$\begin{aligned} \mathsf{A}^s \overset{k-1}{\Longrightarrow}_u &\quad \mathsf{D}_1 \ldots \mathsf{D}_m \\ \overset{1}{\Longrightarrow}_u &\quad \mathsf{D}_1 \ldots \mathsf{D}_{j-1} \mathsf{Q}_1 \ldots \mathsf{Q}_{n-m+1} \mathsf{D}_{n-m+j+1} \ldots \mathsf{D}_m \end{aligned}$$

   where $\mathsf{Q}_1, \ldots, \mathsf{Q}_{n-m+1}$ are feature structures.

8. From (6) above, X_0 uniquely defines π_0, A and $\mathrm{F}_{|\pi_0|+1}$. We denote $\Psi(\mathsf{C}_0, \pi_0)$ as $\langle \mathsf{s}_1, \mathrm{F}_1, \ldots, \mathsf{s}_{|\pi_0|+1} \rangle$. Recall that for every $1 \leq i \leq |\pi_0| + 1$, $\mathsf{s}_i \sqcup \mathsf{t}_i \neq \top$ because $\mathsf{A} \in \Gamma(\mathsf{C}_0, \pi_0, \textit{maxHt}(G^u))$. Let $\langle \mathsf{P}_0, \ldots, \mathsf{P}_{n-m+1} \rangle$ be the sequence of feature structures such that

   $$(\langle \mathsf{A} \rangle, 0) \sqcup (r^u, 0) = (\langle \mathsf{P}_0 \rangle, \langle \mathsf{P}_0, \ldots, \mathsf{P}_{n-m+1} \rangle)$$

9. Now we show that there is a sequence of paths $\langle \pi_1, \ldots, \pi_{n-m+1} \rangle$ such that

$$\langle Y_j, \ldots, Y_{n-m+j} \rangle = \langle N[\Psi(\mathsf{Q}_1, \pi_1)], \ldots, N[\Psi(\mathsf{Q}_{n-m+1}, \pi_{n-m+1})] \rangle$$

Without loss of generality, if r^u is reentrant we assume that its reentrant path is (e, μ_e), that is, $(0, \mu_0) \overset{r^u}{\leftrightsquigarrow} (e, \mu_e)$, where $1 \leq e \leq n$. By the definition of LIG there are two options for the rule r:

(a) The rule r does not copy the stack from the head to the body. Hence, $\langle X_1, \ldots, X_{n-m+1} \rangle = \langle Y_j, \ldots, Y_{n-m+j} \rangle$. Consider the possible sources of the rule r, according to definition 28:

- Case (3):

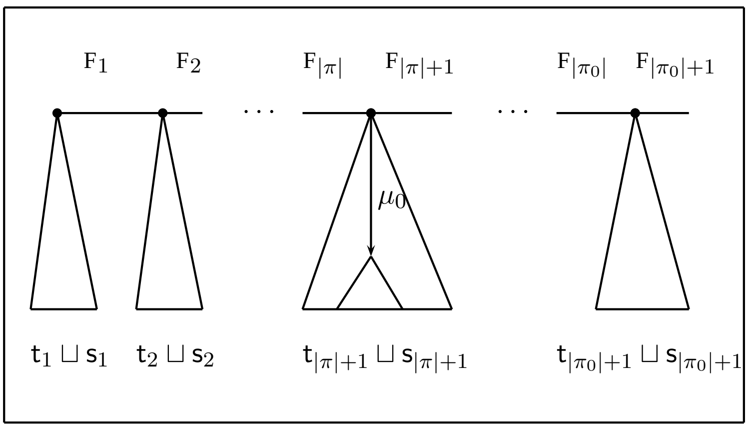
  The rule r^u is non-reentrant. Hence, for $1 \leq i \leq n-m+1$, $\mathsf{C}_i = \mathsf{Q}_i$ and $X_i = N[\Psi(\mathsf{C}_i, \varepsilon)] = N[\Psi(\mathsf{Q}_i, \varepsilon)]$. Therefore,

$$\langle Y_j, \ldots, Y_{n-m+j} \rangle = \langle N[\Psi(\mathsf{Q}_1, \varepsilon)], \ldots, N[\Psi(\mathsf{Q}_{n-m+1}, \varepsilon)] \rangle$$

- Case (4a):

  If μ_0 is not a prefix of π_0 then for all i, $i \neq e$, $X_i = Y_{i+j-1} = N[\Psi(\mathsf{C}_i, \varepsilon)] = N[\Psi(\mathsf{Q}_i, \varepsilon)]$, and $X_e = N[\Psi(\mathsf{P}_e, \mu_e)]$. Let π be the maximal prefix of π_0 and μ_0 such that $\mu_0 = \pi \cdot \mu'_0$. The cord $\Psi(\mathsf{D}_j \sqcup \mathsf{C}_0, \nu_j)$ is graphically represented as:

$\Psi(\mathsf{A} \sqcup \mathsf{C}_0, \pi_0)$



  Hence, $val(\mathsf{A} \sqcup \mathsf{C}_0, \mu_0) = val(\mathsf{t}_{|\pi|+1} \sqcup \mathsf{s}_{|\pi|+1}, \mu'_0) = val(\mathsf{D}_j \sqcup \mathsf{C}_0, \mu_0)$. By definition of unification in context $val(\mathsf{P}_e, \mu_e) = val(\mathsf{A} \sqcup \mathsf{C}_0, \mu_0)$ and $val(\mathsf{Q}_e, \mu_e) =$

$val(\mathsf{D}_j \sqcup \mathsf{C}_0, \mu_0)$. Hence, $val(\mathsf{P}_e, \mu_e) = val(\mathsf{Q}_e, \mu_e)$ and $\mathsf{Q}_e = \mathsf{P}_e$. Therefore, $X_e = N[\Psi(\mathsf{Q}_e, \mu_e)]$ and

$$\langle Y_j, \ldots, Y_{n-m+j}\rangle = \langle N[\Psi(\mathsf{Q}_1, \varepsilon)], \ldots, N[\Psi(\mathsf{Q}_e, \mu_e)], \ldots, N[\Psi(\mathsf{Q}_{n-m+1}, \varepsilon)]\rangle$$

- Case (4(b)i):

  If $\pi_0 = \mu_0 \cdot \nu$, $\nu \in$ PATHS then $X_e = N[\Psi(\mathsf{P}_e, \mu_e \cdot \nu)]$. Since $N[\Psi(\mathsf{A}, \pi_0)]$ is applicable to $N[\Psi(\mathsf{D}_j, \nu_j)]$ we obtain that $\pi_0 = \nu_j$ and $\mathsf{A} = \mathsf{D}_j$. Hence $\mathsf{P}_e = \mathsf{Q}_e$. Therefore, $X_e = N[\Psi(\mathsf{Q}_e, \mu_e \cdot \nu)]$ and

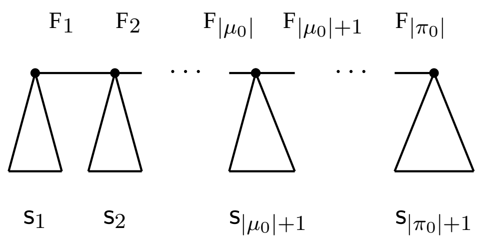
$$\langle Y_j, \ldots, Y_{n-m+j}\rangle = \langle N[\Psi(\mathsf{Q}_1, \varepsilon)], \ldots, N[\Psi(\mathsf{Q}_e, \mu_e \cdot \nu)], \ldots, N[\Psi(\mathsf{Q}_{n-m+1}, \varepsilon)]\rangle$$

(b) The rule r copies the stack from X_0 to X_e. By definition 28, r is created from a reentrant unification rule, r^u, by case (4(b)ii) of the definition 28. Let $\nu_j = \pi_0 \cdot \nu'_j$ and $\pi_0 = \mu_0 \cdot \nu$, $\nu'_j, \nu \in$ PATHS. By the definition for all i, $i \neq e$, $X_i = Y_{i+j-1} = N[\Psi(\mathsf{C}_i, \varepsilon)] = N[\Psi(\mathsf{Q}_i, \varepsilon)]$ and $X_e = N[\Psi(\mathsf{P}_e, \mu_e \cdot \nu), \mathrm{F}_{|\pi_0|+1}..]$. Hence we just have to show that for some path $\pi_e \in$ PATHS, $Y_{j+e-1} = N[\Psi(\mathsf{Q}_e, \pi_e)]$. We will show that this equation holds for $\pi_e = \mu_e \cdot \nu \cdot \nu'_j$. Since π_0, A and $\mathrm{F}_{|\pi_0|+1}$ are uniquely defined by X_0 we obtain the following:

- $\Psi(\mathsf{C}_0, \pi_0) = \langle \mathsf{s}_1, \mathrm{F}_1, \ldots, \mathsf{s}_{|\pi_0|+1}\rangle$
- $\Psi(\mathsf{D}_j, \nu_j) = \langle \mathsf{t}_1, \mathrm{F}_1, \ldots, \mathsf{t}_{|\nu_j|+1}\rangle$
- $\Psi(\mathsf{D}_j \sqcup \mathsf{C}_0, \nu_j) = \Psi(\mathsf{Q}_0, \nu_j) = \langle \mathsf{s}_1 \sqcup \mathsf{t}_1, \mathrm{F}_1, \ldots, \mathsf{s}_{|\pi_0|+1} \sqcup \mathsf{t}_{|\pi_0|+1}, \mathrm{F}_{|\pi_0|+1}, \ldots, \mathsf{t}_{|\nu_j|+1}\rangle$
- $\Psi(\mathsf{A} \sqcup \mathsf{C}_0, \pi_0) = \Psi(\mathsf{P}_0, \pi_0) = \langle \mathsf{s}_1 \sqcup \mathsf{t}_1, \mathrm{F}_1, \ldots, \mathsf{s}_{|\pi_0|+1} \sqcup \mathsf{t}_{|\pi_0|+1}\rangle$
- $\Psi(\mathsf{P}_e, \mu_e \cdot \nu) = \textit{butLast}(\Psi(\mathsf{C}_e, \mu_e)) \cdot \langle \mathsf{s}_{|\mu_0|+1} \sqcup \mathsf{t}_{|\mu_0|+1}, \mathrm{F}_{|\mu_0|+1}, \ldots, \mathsf{s}_{|\pi_0|+1} \sqcup \mathsf{t}_{|\pi_0|+1}\rangle$
- $\Psi(\mathsf{Q}_e, \pi_0 \cdot \nu'_j) =$

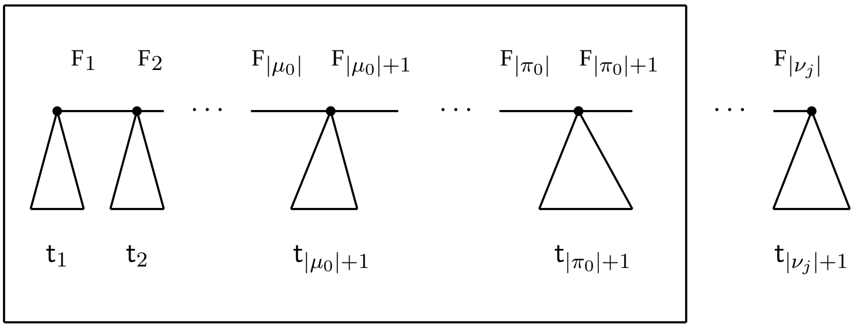
  $\textit{butLast}(\Psi(\mathsf{C}_e, \mu_e)) \cdot \langle \mathsf{s}_{|\mu_0|+1} \sqcup \mathsf{t}_{|\mu_0|+1}, \mathrm{F}_{|\mu_0|+1}, \ldots, \mathsf{s}_{|\pi_0|+1} \sqcup \mathsf{t}_{|\pi_0|+1}, \mathrm{F}_{|\pi_0|+1}, \ldots, \mathsf{t}_{|\nu_j|+1}\rangle$

The cord $\Psi(\mathsf{C}_0, \pi_0) = \langle \mathsf{s}_1, \mathrm{F}_1, \ldots, \mathsf{s}_{|\pi_0|+1}\rangle$ is graphically represented as:



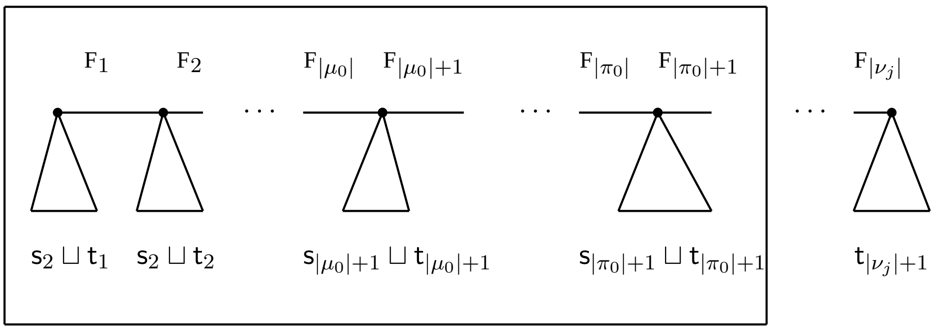
The cord $\Psi(\mathsf{D}_j, \nu_j) = \langle \mathsf{t}_1, \mathrm{F}_1, \ldots, \mathsf{t}_{|\nu_j|+1} \rangle$ whose prefix is the cord $\Psi(\mathsf{A}, \nu_j)$ is graphically represented as:

$\Psi(\mathsf{A}, \pi_0)$



The cord $\Psi(\mathsf{D}_j \sqcup \mathsf{C}_0, \nu_j) = \Psi(\mathsf{Q}_0, \nu_j) = \langle \mathsf{s}_1 \sqcup \mathsf{t}_1, \mathrm{F}_1, \ldots, \mathsf{s}_{|\pi_0|+1} \sqcup \mathsf{t}_{|\pi_0|+1}, \mathrm{F}_{|\pi_0|+1}, \ldots, \mathsf{t}_{|\nu_j|+1} \rangle$ whose prefix is the cord $\Psi(\mathsf{A} \sqcup \mathsf{C}_0, \pi_0) = \Psi(\mathsf{P}_0, \pi_0) = \langle \mathsf{s}_1 \sqcup \mathsf{t}_1, \mathrm{F}_1, \ldots, \mathsf{s}_{|\pi_0|+1} \sqcup \mathsf{t}_{|\pi_0|+1} \rangle$ is graphically represented as:

$\Psi(\mathsf{A} \sqcup \mathsf{C}_0, \pi_0)$



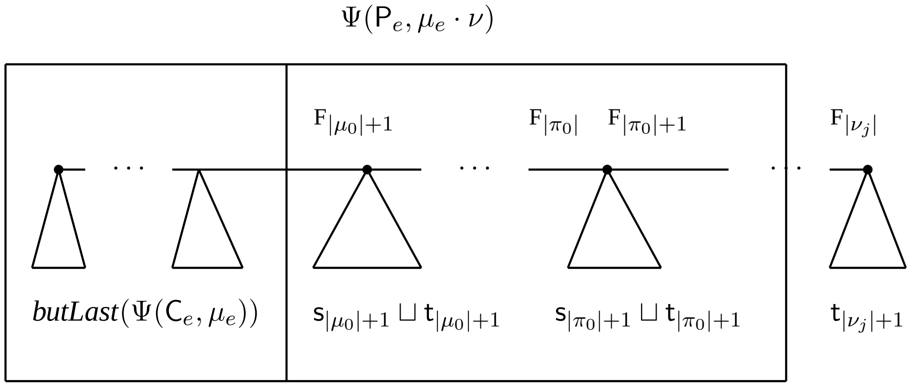
The relation between the cords

$$\Psi(\mathsf{P}_e, \mu_e \cdot \nu) = \mathit{butLast}(\Psi(\mathsf{C}_e, \mu_e)) \cdot \langle \mathsf{s}_{|\mu_0|+1} \sqcup \mathsf{t}_{|\mu_0|+1}, \mathrm{F}_{|\mu_0|+1}, \ldots, \mathsf{s}_{|\pi_0|+1} \sqcup \mathsf{t}_{|\pi_0|+1} \rangle$$

and

$$\Psi(\mathsf{Q}_e, \pi_0 \cdot \nu_j') =$$
$$\mathit{butLast}(\Psi(\mathsf{C}_e, \mu_e)) \cdot \langle \mathsf{s}_{|\mu_0|+1} \sqcup \mathsf{t}_{|\mu_0|+1}, \mathrm{F}_{|\mu_0|+1}, \ldots, \mathsf{s}_{|\pi_0|+1} \sqcup \mathsf{t}_{|\pi_0|+1}, \mathrm{F}_{|\pi_0|+1}, \ldots, \mathsf{t}_{|\nu_j|+1} \rangle$$

is graphically represented as:



Hence,

$$\begin{aligned} Y_e =& N[\Psi(\mathsf{P}_e, \mu_e \cdot \nu), \mathrm{F}_{|\pi_0|+1}, \mathsf{t}_{|\pi_0|+2}, \mathrm{F}_{|\pi_0|+2}, \ldots, \mathsf{t}_{|\nu_j|+1}]] \\ =& N[\textit{butLast}(\Psi(\mathsf{C}_e, \mu_e)), \mathsf{s}_{|\mu_0|+1} \sqcup \mathsf{t}_{|\mu_0|+1}, \mathrm{F}_{|\mu_0|+1}, \ldots, \mathsf{s}_{|\pi_0|+1} \sqcup \mathsf{t}_{|\pi_0|+1}, \mathrm{F}_{|\pi_0|+1}, \ldots, \mathsf{t}_{|\nu_j|+1}] \\ =& N[\Psi(\mathsf{Q}_e, \pi_0 \cdot \nu'_j)] \end{aligned}$$

Therefore,

$$\langle Y_j, \ldots, Y_{n-m+j} \rangle = \langle N[\Psi(\mathsf{Q}_1, \varepsilon)], \ldots, N[\Psi(\mathsf{Q}_e, \pi_0 \cdot \nu'_j)], \ldots, N[\Psi(\mathsf{Q}_{n-m+1}, \varepsilon)] \rangle$$

□

Corollary 20. *Let* $G^u \in \mathrm{UG}_{1r}$, *then* $L(G^u) = L(\textit{ug2lig}(G^u))$.

Proof. Let $G^u = \langle \mathcal{R}^u, \mathsf{A}^s, \mathcal{L} \rangle$ be a one-reentrant unification grammar and $G^{li} = \langle V_N, V_t, V_s, \mathcal{R}^{li}, N \rangle = \textit{ug2lig}(G^u)$. Then by theorem 18, there is a sequence of paths $\langle \pi_1, \ldots, \pi_n \rangle$ such that

$$\text{if } \mathsf{A}^s \overset{*}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n \text{ then } S[\,] \overset{*}{\Longrightarrow}_{li} N[\Psi(\mathsf{A}_1, \pi_1)] \ldots N[\Psi(\mathsf{A}_n, \pi_n])$$

Where $\mathsf{A}^s \overset{*}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$ is a pre-terminal sequence. Assume that $\mathsf{A}^s \overset{*}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n \overset{*}{\Longrightarrow}_u w_1, \ldots w_n$, where $w_i \in \textsc{Words}$, $1 \le i \le n$. Hence, $\mathcal{L}(w_i) = \{\mathsf{D}_i\}$ and $\mathsf{A}_i \sqcup \mathsf{D}_i \neq \top$. Since the grammar is a simplified unification grammar (definition 18), $\mathsf{A}_i = \mathsf{D}_i$. By definition 28 case (2), the rule $N[\Psi(\mathsf{A}_i, \pi_i)] \to w_i$ is in $\mathcal{R}^{li}$. Therefore, $S[\,] \overset{*}{\Longrightarrow}_{li} N[\Psi(\mathsf{A}_1, \pi_1)] \ldots N[\Psi(\mathsf{A}_n, \pi_n]) \overset{*}{\Longrightarrow}_{li} w_1, \ldots w_n$.

By theorem 19, if $S[\,] \overset{*}{\Longrightarrow}_{li} Y_1 \ldots Y_n$ then there are a sequence of paths $\langle \pi_1, \ldots, \pi_n \rangle$, and a derivation sequence $\mathsf{A}^s \overset{*}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n$ such that for $0 \le i \le n$, $Y_i = N[\Psi(\mathsf{A}_i, \pi_i)]$. Assume that $S[\,] \overset{*}{\Longrightarrow}_{li} N[\Psi(\mathsf{A}_1, \pi_1)] \ldots N[\Psi(\mathsf{A}_n, \pi_n]) \overset{*}{\Longrightarrow}_{li} w_1, \ldots w_n$, $w_i \in V_t$. Then the rules

$N[\Psi(\mathsf{A}_i, \pi_i)] \rightarrow w_i$ in $\mathcal{R}^{li}$, $1 \leq i \leq n$. By definition 28, each such rule is created from a lexicon entry $\mathcal{L}(w_i) = \{\mathsf{A}_i\}$. Hence, $\mathsf{A}^s \stackrel{*}{\Longrightarrow}_u \mathsf{A}_1 \ldots \mathsf{A}_n \stackrel{*}{\Longrightarrow}_u w_1, \ldots w_n$. $\square$

Chapter 4

Conclusions

In this work we explore the influence of reentrancies on the generative power of unification grammars. Our main contribution is the definition of two constraints on unification grammars which dramatically limit their expressivity. We prove that non-reentrant unification grammars generate exactly the class of context-free languages; and that one-reentrant unification grammars generate exactly the class of mildly context-sensitive languages. While these results do not characterize the classes of unification grammars that license context free languages and mildly context sensitive languages (because the restrictions are sufficient but not necessary), they provide two linguistically plausible constrained formalisms whose computational processing is tractable.

This work can be extended in a number of directions. We are well aware of the fact that our mapping of one-reentrant unification grammars to LIG is highly verbose. In particular, it results in LIGs with a huge number of rules, many of which will never participate in any derivation. We believe that it should be possible to optimize the mapping such that much smaller LIGs are generated. Furthermore, the equivalence proofs of section 3.4 are rather complex, perhaps owing to the choice of LIG as the target formalism. It would be interesting to experiment with a mapping of one-reentrant unification grammars to some other mildly context-sensitive formalism, notably TAG.

The two constraints on unification grammars (non-reentrant and one-reentrant) are parallel to the first two classes of the Weir hierarchy of languages (Weir, 1992). A possible extension of this work could be a definition of constraints on unification grammars that would generate all the classes of the hierarchy.

Another direction is an extension of one-reentrant unification grammars, where the reentrancy inside a grammar rule does not have to be between the head and one element in the body, but can also be, for example, between two elements of the body or within an element. We believe that a formalism of one-reentrant unification grammars, where the reentrancy inside a grammar rule can only be between two elements of the body, generates all and only context free languages. A formal characterization of the class of languages generated by such grammars is an interesting direction for future research. Then it is interesting to explore the power of two-reentrant unification grammars, possibly with limited kinds of reentrancies.

It may also be possible to extend one-reentrant UGs to multi-reentrant UGs without extending their generation power. One research direction is to allow some kind of disjoint reentrancies, where the reentrant paths have no common edges.

Chapter 5

Bibliography

References

Carpenter, Bob. 1992. *The Logic of Typed Feature Structures*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press.

Culy, Christopher. 1985. The complexity of the vocabuary of Bambara. *Linguistics and Philosophy*, 8:345–351.

Gazdar, Gerald. 1988. Applicability of indexed grammars to natural languages. In Uwe Reyle and Christian Rohrer, editors, *Natural Language Parsing and Linguistic Theories*. Reidel Publishing Company, Dordrecht, pages 69–94.

Gazdar, Gerald. E., Ewan Klein, Jeoffrey K. Pullum, and Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Harvard University Press, Cambridge, Mass.

Huybregts, Riny. 1984. The weak inadequacy of context-free phrase structure grammars. In Ger J. de Haan, Mieke Trommelen, and Wim Zonneveld, editors, *Van Periferie Naar Kern*. Foris Publications, Dordrecht, pages 81–99.

Jaeger, Efrat, Nissim Francez, and Shuly Wintner. 2004. Unification grammars and off-line parsability. *Journal of Logic, Language and Information*.

Johnson, Mark. 1988. *Attribute-Value Logic and the Theory of Grammar*, volume 16 of *CSLI Lecture Notes*. CSLI, Stanford, California.

Joshi, Aravind K. 1985. How much context-sensitivity is required to provide reasonable structural descriptions: Tree adjoining grammars. In D. Dowty, L. Karttunen, and A. Zwicky, editors, *Natural Language Processing: Psycholinguistic, Computational and Theoretical Perspectives*. Cambridge University Press.

Joshi, Aravind K. 2003. Tree-adjoining grammars. In Ruslan Mitkov, editor, *The Oxford handbook of computational linguistics*. Oxford university Press, chapter 26, pages 483–500.

Pollard, Carl. 1984. *Generalized phrase structure grammars, head grammars and natural language*. Ph.D. thesis, Stanford University.

Satta, Giorgio. 1994. Tree-adjoining grammar parsing and boolean matrix multiplication. In *Proceedings of the 20st Annual Meeting of the Association for Computational Linguistics*, volume 20.

Shieber, Stuart M. 1985. Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, 8:333–343.

Shieber, Stuart M. 1986. *An Introduction to Unification Based Approaches to Grammar*. Number 4 in CSLI Lecture Notes. CSLI.

Shieber, Stuart M. 1992. *Constraint-Based Grammar Formalisms*. MIT Press, Cambridge, Mass.

Steedman, Mark. 2000. *The Syntactic Process*. Language, Speech and Communication. The MIT Press, Cambridge, Mass.

Vijay-Shanker, K. and David J. Weir. 1990. Parsing constrained grammar formalisms. *Computational Linguistics*, 19(4):591–636.

Vijay-Shanker, K. and David J. Weir. 1994. The equivalence of four extensions of context-free grammars. *Mathematical systems theory*, 27:511–545.

Weir, David J. 1992. A geometric hierarchy beyond context-free languages. *Theoretical Computer Science*, 104:235–261.