

Semantic expressive capacity with bounded memory

Antoine Venant and Alexander Koller

Department of Language Science and Technology
Saarland University

{[venant](mailto:venant@coli.uni-saarland.de)|[koller](mailto:koller@coli.uni-saarland.de)}@coli.uni-saarland.de

Abstract

We investigate the capacity of mechanisms for compositional semantic parsing to describe relations between sentences and semantic representations. We prove that in order to represent certain relations, mechanisms which are syntactically projective must be able to remember an unbounded number of locations in the semantic representations, where nonprojective mechanisms need not. This is the first result of this kind, and has consequences both for grammar-based and for neural systems.

1 Introduction

Semantic parsers which translate a sentence into a semantic representation compositionally must recursively compute a partial semantic representation for each node of a syntax tree. These partial semantic representations usually contain placeholders at which arguments and modifiers are attached in later composition steps. Approaches to semantic parsing differ in whether they assume that the number of placeholders is bounded or not. Lambda calculus (Montague, 1974; Blackburn and Bos, 2005) assumes that the number of placeholders (lambda-bound variables) can grow unboundedly with the length and complexity of the sentence. By contrast, many methods which are based on unification (Copestake et al., 2001) or graph merging (Courcelle and Engelfriet, 2012; Chiang et al., 2013) assume a fixed set of placeholders, i.e. the number of placeholders is bounded.

Methods based on bounded placeholders are popular both in the design of hand-written grammars (Bender et al., 2002) and in semantic parsing for graphs (Peng et al., 2015; Groschwitz et al., 2018). However, it is not clear that all relations between language and semantic representations can be expressed with a bounded number of placeholders. The situation is particularly challenging

when one insists that the compositional analysis is *projective* in the sense that each composition step must combine adjacent substrings of the input sentence. In this case, it may be impossible to combine a semantic predicate with a distant argument immediately, forcing the composition mechanism to use up a placeholder to remember the argument position. If many predicates have distant arguments, this may exceed the bounded “memory capacity” of the compositional mechanism.

In this paper, we show that there are relations between sentences and semantic representations which can be described by compositional mechanisms which are bounded and non-projective, but not by ones which are bounded and projective. To our knowledge, this is the first result on expressive capacity with respect to semantics – in contrast to the extensive literature on the expressive capacity of mechanisms which describe just the string languages.

More precisely, we prove that tree-adjoining grammars can describe string-graph relations using the HR graph algebra (Courcelle and Engelfriet, 2012) with two sources (bounded, non-projective) which cannot be described using linear monadic context-free tree grammars and the HR algebra with k sources, for any fixed k (bounded, projective). This result is especially surprising because TAG and linear monadic CFTGs describe the same string languages; thus the difference lies only in the projectivity of the syntactic analysis.

We further prove that given certain assumptions on the alignment between tokens in the sentence and edges in the graph, *no* generative device for projective syntax trees can simulate TAG with two sources. This has practical consequences for the design of transition-based semantic parsers (whether grammar-based or neural).

Plan of the paper. We will first explain the linguistic background in Section 2 and lay the formal

foundations in Section 3. We will then prove the reduced semantic expressive capacity for aligned generative devices in Section 4 and for CFTGs in Section 5. We conclude with a discussion of the practical impact of our findings (Section 6).

2 Compositional semantic construction

The Principle of Compositionality, which is widely accepted in theoretical semantics, states that the meaning of a natural-language expression can be determined from the meanings of its immediate subexpressions and the way in which the subexpressions were combined. Implementations of this principle usually assume that there is some sort of syntax tree which describes the grammatical structure of a sentence. A semantic representation is then calculated by bottom-up evaluation of this syntax tree, starting with semantic representations of the individual words and then recursively computing a semantic representation for each node from those of its children.

2.1 Compositional mechanisms

Mechanisms for semantic composition will usually keep track of places at which semantic arguments are still missing or modifiers can still be attached. For instance, when combining the semantic representations for “John” and “sleeps” in a derivation of “John sleeps”, the “subject” argument of “sleeps” is filled with the meaning of “John”. The compositional mechanism therefore assigns a semantic representation to “sleeps” which has an unfilled *placeholder* for the subject.

The exact nature of the placeholder depends on the compositional mechanism. There are two major classes in the literature. *Lambda-style* compositional mechanisms use a *list* of placeholders. For instance, lambda calculus, as used e.g. in Montague Grammar (Montague, 1974), CCG (Steedman, 2001), or linear-logic-based approaches in LFG (Dalrymple et al., 1995) might represent “sleeps” as $\lambda x.\text{sleep}(x)$. Placeholders are lambda-bound variables (here: x).

By contrast, *unification-style* compositional mechanisms use *names* for placeholders. For example, a simplified form of the Semantic Algebra used in HPSG (Copestake et al., 2001) might represent “sleeps” as the feature structure [subj: $\boxed{1}$, sem:[pred:sleep, agent: $\boxed{1}$]]. This is unified with [subj:John]. The placeholders are *holes* with *labels* from a fixed set of argument names

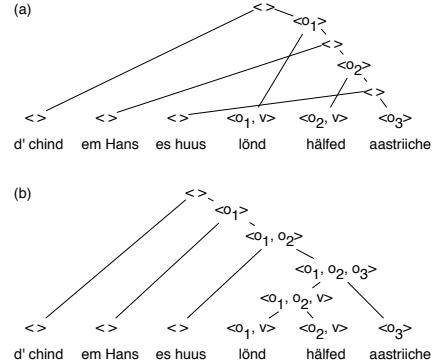


Figure 1: (a) Nonprojective and (b) projective analysis.

(e.g. subj). Named placeholders are also used in the HR algebra (Courcelle and Engelfriet, 2012) and its derivatives, like Hyperedge Replacement Grammars (Drewes et al., 1997; Chiang et al., 2013) and the AM algebra (Groschwitz et al., 2018).

2.2 Boundedness and projectivity

A fundamental difference between lambda-style and unification-style compositional mechanisms is in their “memory capacity”: the number of placeholders in a lambda-style mechanism can grow unboundedly with the length and complexity of the sentence (e.g. by functional composition of lambda terms), whereas in a unification-style mechanism, the placeholders are fixed in advance.

There is an informal intuition that unbounded memory is needed especially when an unbounded number of semantic predicates can be far away from their arguments in the sentence, and the syntax formalism does not allow these predicates to combine immediately with the arguments. For illustration, consider the two derivations of the following Swiss German sentence from Schieber (1985) in Fig. 1:

- (1) (dass) (mer) d' chind em Hans
 (that) (we) the-children-ACC Hans-DAT
 es huus lönd hälfed aastriche
 the-house-ACC let help paint
 ‘(that we) let the children help Hans paint the house’

The lexical semantic representation of each verb comes with a placeholder for its object (o_1, o_2, o_3) and, in the case of “lönd” and “hälfed”, also one for its verb complement (v). The derivation in Fig. 1a immediately combines each verb with its complements; the placeholders that are used at each node never grow beyond the ones the verbs

originally had. However, this derivation combines verbs with nouns which are not adjacent in the string, which is not allowed in many grammar formalisms. If we limit ourselves to combining only adjacent substrings (*projectively*, see Fig. 1b), we must remember the placeholders for all the verbs at the same time if we want to obtain the correct predicate-argument structure. Thus, the number of placeholders grows with the length of the sentence; this is only possible with a lambda-style compositional mechanism.

There is scattered evidence in the literature for this tension between bounded memory and projectivity. Chiang et al. (2013) report (of a compositional mechanism based on the HR algebra, unification-style) that a bounded number of placeholders suffices to derive the graphs in the AMR version of the Geoquery corpus, but Groschwitz et al. (2018) find that this requires non-projective derivations in 37% of the AMRBank training data (Banarescu et al., 2013). Approaches to semantic construction with tree-adjoining grammar either perform semantic composition along the TAG derivation tree using unification (non-projective, unification-style) (Gardent and Kallmeyer, 2003) or along the TAG derived tree using linear logic (projective, lambda-style) (Frank and van Genabith, 2001). Bender (2008) discusses the challenges involved in modeling the predicate-argument structure of a language with very free word order (Wambaya) with projective syntax. While the Wambaya noun phrase does not seem to require the projective grammar to collect unbounded numbers of unfilled arguments as in Fig. 1b, Bender notes that her projective analysis still requires a more flexible handling of semantic arguments than the HPSG Semantic Algebra (unification-style) supports.

In this paper, we define a notion of semantic expressive capacity and prove the first formal results about the relationship between projectivity and bounded memory.

3 Formal background

Let $\mathbb{N}_0 = \{0, 1, \dots\}$ be the nonnegative integers. A *signature* is a finite set Σ of function symbols f , each of which has been assigned a nonnegative integer called its *rank*. We write Σ_n for the symbols of rank n . Given a signature Σ , we say that all constants $a \in \Sigma_0$ are *trees* over Σ ; further, if $f \in \Sigma_n$ and t_1, \dots, t_n are trees over Σ , then

$f(t_1, \dots, t_n)$ is also a tree. We write \mathcal{T}_Σ for the set of all trees over Σ . We define the *height* $\text{ht}(t)$ of a tree $t = f(t_1, \dots, t_n)$ to be $1 + \max \text{ht}(t_i)$, and $\text{ht}(c) = 1$ for $c \in \Sigma_0$.

Let $X \notin \Sigma$, and let $\Sigma_X = \Sigma \cup \{X\}$ (with X as a constant of rank 0). Then we call a tree $C \in \mathcal{T}_{\Sigma_X}$ a *context* if it contains exactly one occurrence of X , and write \mathcal{C}_Σ for the set of all contexts. A context can be seen as a tree with exactly one hole. If $t \in \mathcal{T}_\Sigma$, we write $C[t]$ for the tree in \mathcal{T}_Σ that is obtained by replacing X with t .

Given a string $w \in W^*$, we write $|w|_a$ for the number of times that $a \in W$ occurs in w .

3.1 Grammars for strings and trees

We take a very general view on how semantic representations for strings are constructed compositionally. To this end, we define a notion of “grammar” which encompasses more devices for describing languages than just traditional grammars, such as transition-based parsers.

We say that a *tree grammar* G over the signature Σ is any finite device that defines a language $L(G) \subseteq \mathcal{T}_\Sigma$. For instance, regular tree grammars (Comon et al., 2007) are tree grammars, and context-free grammars can also be seen as tree grammars defining the language of parse trees.

We say that a *string grammar* $\mathcal{G} = (G, \text{yd})$ over the signature Σ and the alphabet W is a pair consisting of a tree grammar G over Σ and a *yield function* $\text{yd} : \mathcal{T}_\Sigma \rightarrow W^*$ which maps trees to strings over W (Weir, 1988). A string grammar defines a language $L(\mathcal{G}) = \{\text{yd}(t) \mid t \in L(G)\} \subseteq W^*$. We call the trees $t \in L(G)$ *derivations*.

A particularly common yield function is the function yd_{pr} , defined as $\text{yd}_{\text{pr}}(f(t_1, \dots, t_n)) = \text{yd}_{\text{pr}}(t_1) \dots \cdot \text{yd}_{\text{pr}}(t_n)$ if $n > 0$ and $\text{yd}_{\text{pr}}(c) = c$ if c has rank 0. This yield function simply concatenates the words at the leaves of t . Applied to the phrase-structure tree t in Fig. 2c, $\text{yd}_{\text{pr}}(t)$ is the Swiss German sentence in (1). Context-free grammars can be characterized as string grammars that combine a regular tree grammar with yd_{pr} . By contrast, we can model tree-adjoining grammars (TAG, Joshi and Schabes, 1997) by choosing a tree grammar G that describes derivation trees as in Fig. 2b. The yd function could then substitute and adjoin the elementary trees as specified by the derivation tree (see Fig. 2a) and then read off the words from the resulting derived tree in Fig. 2c.

We say that a string grammar is *projective* if

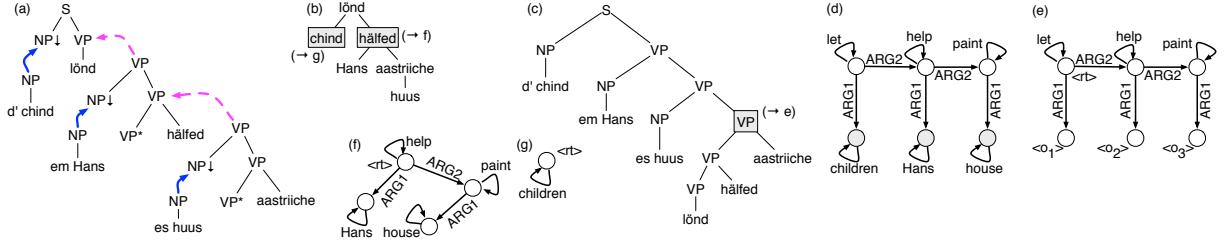


Figure 2: Semantic construction with TAG: (a) TAG derivation, (b) derivation tree, (c) derived tree, (d) semantic graph. (e) s-graph interpretations of the boxed node in (c); (f,g) s-graph interpretations at the boxed nodes in (b).

its yield function is yd_{pr} . Context-free grammars as construed above are clearly projective. Tree-adjoining grammars are *not* projective: For instance, the yield of the subtree below “aastriche” in Fig. 2b consists of the two separate strings “es Huus” and “aastriche”, which are then wrapped around “lönd hälfed” further up in the derivation.

If the grammar is projective, then for any context C there exist two strings $\text{left}(C)$ and $\text{right}(C)$ such that for any tree t , $yd(C[t]) = \text{left}(C) \cdot yd(t) \cdot \text{right}(C)$.

3.2 Context-free tree languages

Below, we will talk about *linear monadic context-free tree grammars* (LM-CFTGs; Rounds (1969), Comon et al. (2007)). An LM-CFTG is a quadruple $G = (N, \Sigma, R, S)$, where N is a ranked signature of nonterminals of rank at most one, Σ is a ranked signature of terminals, $S \in N_0$ is the start symbol, and R is a finite set of production rules of one of the forms

- $A \rightarrow t$ with $A \in N_0$ and $t \in \mathcal{T}_V$
- $A(t) \rightarrow C[t]$ with $A \in N_1$ and $C \in \mathcal{C}_V$,

where $V = N \cup \Sigma$. The trees in $L(G) \subseteq \mathcal{T}_\Sigma$ are obtained by expanding S with production rules. Nonterminals of rank zero are expanded by replacing them with trees. Nonterminals of rank one must have exactly one child in the tree; they are replaced by a context, and the variable in the context is replaced by the subtree below the child.

We can extend an LM-CFTG G to a string grammar $\mathcal{G} = (G, yd_{pr})$. Then LM-CFTG is weakly equivalent to TAG (Kepser and Rogers, 2011); that is, LM-CFTG and TAG generate the same class of string languages. Intuitively, the weakly equivalent LM-CFTG directly describes the language of derived trees of the TAG grammar (cf. Fig. 2c). Notice that LM-CFTG is projective.

Below, we will make crucial use of the following pumping lemma for LM-CFTLs:

Lemma 1 (Maibaum (1978)). *Let G be an LM-CFTG. There exists a constant $p \in \mathbb{N}_0$ such that for any $t \in L(G)$ with $\text{ht}(t) > p$, there exists a decomposition $t = C_1[C_2[C_3[C_4[t_5]]]]$ with $\text{ht}(C_2[C_3[C_4[X]]]) \leq p$ and $\text{ht}(C_2) + \text{ht}(C_4) > 0$ such that for any $i \in \mathbb{N}_0$, $C_1[v^i[t_5]] \in L(G)$, where we let $v^0 = C_3$ and $v^{i+1} = C_2[v^i[C_4[X]]]$.*

We call p the pumping height of $L(G)$.

3.3 The HR algebra

The specific unification-style semantic algebra we use in this paper is the *HR algebra* (Courcelle and Engelfriet, 2012). This choice encompasses much of the recent literature on compositional semantic parsing with graphs, based e.g. on Hyperedge Replacement Grammars (Chiang et al., 2013; Peng et al., 2015; Koller, 2015) and the AM algebra (Groschwitz et al., 2018).

The values of the HR algebra are *s-graphs*: directed, edge-labeled graphs, some of whose nodes may be designated as *sources*, written in angle brackets. S-graphs can be combined using the forget, rename, and merge operations. Rename $\text{ren}_{a \rightarrow b}$ changes an a -source node into a b -source node. Forget f_a makes it so the a -source node in the s-graph is no longer a source node. Merge \parallel combines two s-graphs while unifying nodes with the same source annotation. For instance, the s-graphs $\langle rt \rangle \xrightarrow{\text{ARG1}} \langle o \rangle$ and $\langle o \rangle \xrightarrow{\text{Hans}} \langle o \rangle$ are merged into $\langle rt \rangle \xrightarrow{\text{ARG0}} \langle o \rangle \xrightarrow{\text{Hans}} \langle o \rangle$.

The HR algebra uses operation symbols from a ranked signature Δ to describe s-graphs syntactically. Δ contains symbols for merge (rank 2) and the forget and rename operations (rank 1). It also contains constants (symbols of rank 0) which denote s-graphs of the form $\langle a \rangle \xrightarrow{f} \langle b \rangle$ and $\langle a \rangle \xleftarrow{f}$, where a, b are sources and f is an edge label. Terms $t \in \mathcal{T}_\Delta$ over this signature evaluate recursively to s-graphs $\llbracket t \rrbracket$, as usual in an algebra. Each instance of the HR algebra uses a fixed, finite set of k source names which can be used in the con-

stant s-graphs and the rename and forget operations. The class of graphs which can be expressed as values of terms over the algebra increases with k . We write \mathcal{H}_k for the HR algebra with k source names (and some set of edge labels).

Let G be an s-graph, and let G' be a subgraph of G , i.e. a subset of its edges. We call a node a *boundary node* of G' if it is incident both to an edge in G' and to an edge that is not in G' . For instance, the s-graph in Fig. 2e is a subgraph of the one in Fig. 2d; the boundary nodes are drawn shaded in (d). The following lemma holds:

Lemma 2. *Let $G = \llbracket C[t] \rrbracket$ be an s-graph, and let G' be a subgraph of G such that the s-graph $\llbracket t \rrbracket$ contains the same edges as G' . Then every boundary node in G' is a source in $\llbracket t \rrbracket$.*

3.4 Grammars with semantic interpretations

Finally, we extend string grammars to compositionally relate strings with semantic representations. Let $\mathcal{G} = (G, \text{yd})$ be a string grammar. The tree grammar G generates a language $L(G) \subseteq \mathcal{T}_\Sigma$ of trees. We will map each tree $t \in L(G)$ into a term $h(t)$ over some algebra \mathcal{A} over a signature Δ using a linear tree homomorphism (LTH) $h : \mathcal{T}_\Sigma \rightarrow \mathcal{T}_\Delta$ (Comon et al., 2007), i.e. by compositional bottom-up evaluation. This defines a relation between strings and values of \mathcal{A} :

$$\mathcal{REL}(\mathcal{G}, h, \mathcal{A}) = \{(\text{yd}(t), \llbracket h(t) \rrbracket_{\mathcal{A}}) \mid t \in L(G)\}$$

For instance, \mathcal{A} could be some HR algebra \mathcal{H}_k ; then $\mathcal{REL}(\mathcal{G}, h, \mathcal{H}_k)$ will be a binary relation between strings and s-graphs. In this case, we abbreviate $\llbracket h(t) \rrbracket$ as $\text{graph}(t)$.

If we look at an entire class \mathbb{G} of string grammars and a fixed algebra, this defines a class of such relations:

$$\mathcal{L}(\mathbb{G}, \mathcal{A}) = \{ \mathcal{REL}(\mathcal{G}, h, \mathcal{A}) \mid \mathcal{G} \in \mathbb{G}, h \text{ LTH} \}.$$

In the example in Fig. 2, we can define a linear homomorphism h to map the derivation tree t in (b) to a term $h(t)$ which evaluates to the s-graph shown in (d). At the top of this term, the s-graphs at the ‘‘chind’’ and ‘‘hälfed’’ (f,g) nodes are combined into (d) by $h(\text{lönd})$:

$$\begin{aligned} h(\text{lönd}) &= \langle \text{rt} \rangle \xrightarrow{\text{let}} \\ &\quad \parallel \text{f}_o(\langle \text{rt} \rangle \xrightarrow{\text{ARG1}} \langle o \rangle \parallel \text{ren}_{\text{rt} \rightarrow o}(G_{(f)})) \\ &\quad \parallel \text{f}_o(\langle \text{rt} \rangle \xrightarrow{\text{ARG2}} \langle o \rangle \parallel \text{ren}_{\text{rt} \rightarrow o}(G_{(g)})) \end{aligned}$$

This non-projective derivation produces the s-graph in (d) using only two sources, rt and o . By

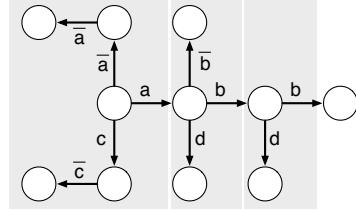


Figure 3: The CSD graph for $((2), (1, 0), (1), (0, 0))$; blocks indicated by gray boxes.

contrast, a homomorphic interpretation of the projective tree (c) has to use at least four sources, as the intermediate result in (e) illustrates.

4 Projective cross-serial dependencies

We will now investigate the ability of projective grammar formalisms $(\mathbb{G}, \mathcal{H}_k)$ to express $\mathcal{L}(\text{TAG}, \mathcal{H}_2)$. We will define a relation $\text{CSD} \in \mathcal{L}(\text{TAG}, \mathcal{H}_2)$ and prove that CSD cannot be generated by projective grammar formalisms with bounded k . We show this first for arbitrary projective \mathbb{G} , under certain assumptions on the alignment of words and graph edges. In Section 5, we drop these assumptions, but focus on $\mathbb{G} = \text{LM-CFTG}$.

4.1 The relation CSD

To construct CSD , consider the string language $\text{CSD}_s = \{A^n B^m C^n D^m \mid m, n \geq 1\}$, where

$$A = \{a \langle^k \bar{a}^k \rangle^k \mid k \geq 0\},$$

and analogously for B, C, D . An example string in CSD_s is $a \langle \langle \bar{a}a \rangle \rangle b \langle \bar{b} \rangle b \ c \langle \bar{c} \rangle d \ d$. Note that k can be chosen independently for each segment.

Every string $w \in \text{CSD}_s$ can be uniquely described by m, n , and a sequence $K(w) = (K^{(a)}, K^{(b)}, K^{(c)}, K^{(d)})$ of numbers specifying the k 's used in each segment, where $K^{(a)}, K^{(c)}$ each contain n numbers and $K^{(b)}, K^{(d)}$ contain m numbers. In the example, we have $n = 1, m = 2$, and $K(w) = ((2), (1, 0), (1), (0, 0))$.

We associate a graph G_w with each string $w \in \text{CSD}_s$ by the construction illustrated in Fig. 3. For each $1 \leq i \leq n$, we define the i -th a -block to be the graph consisting of nodes $u \xrightarrow{a} v$ with a further outgoing a -edge from u . In addition, u connects to a linear chain of $K_i^{(a)}$ edges with label \bar{a} , and v to a linear chain of $K_i^{(c)}$ \bar{c} -edges. G_w consists of a linear chain of the n a -blocks, followed by the m b -blocks (defined analogously). We let $\text{CSD} = \{(w, G_w) \mid w \in \text{CSD}_s\}$.

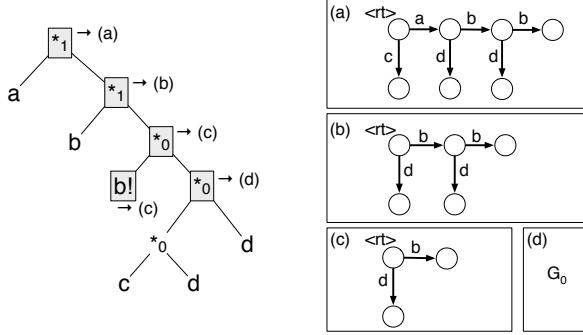


Figure 4: An derivation of $((0), (0,0), (0), (0,0))$.

Note that CSD is a more intricate version of the cross-serial dependency language. CSD can be generated by a TAG grammar along the lines of the one from Section 3.4, using a HR algebra with two sources; thus $\text{CSD} \in \mathcal{L}(\text{TAG}, \mathcal{H}_2)$.

4.2 CSD with bounded blocks

The characteristic feature of CSD is that edges which are close together in the graph (e.g. the a and c edge in an a -block) correspond to symbols that can be distant in the string (e.g. a and c tokens). Projective grammars cannot combine predicates (a) and arguments (c) directly because of their distance in the string; intuitively, they must keep track of either the c 's or the a 's for a long time, which cannot be done with a bounded k .

Before we go into exploiting this intuition, we first note that its correctness depends on the details of the construction of CSD, in particular the ability to select arbitrary and independent $K^{(x)}$ for the different $x \in \{a, b, c, d\}$. Consider the derivation t on the left of Fig. 4 with its projective yield $abbcdd$; this is the case $((0), (0,0), (0), (0,0))$ of CSD, corresponding to the CSD graph G_1 shown in Fig. 4 (a). We can map t to this graph by applying the following linear tree homomorphism h into \mathcal{H}_2 :

$$\begin{aligned} h(*_1) &= f_s(x_1 \parallel \text{ren}_{rt \rightarrow s}(x_2)) & h(*_0) &= x_1 \\ h(b) &= d \leftarrow b\langle rt \rangle \rightarrow \langle s \rangle & h(b!) &= d \leftarrow b\langle rt \rangle \\ h(a) &= c \leftarrow a\langle rt \rangle \rightarrow \langle s \rangle \end{aligned}$$

A derivation of the form $*_0(t_1, t_2)$ evaluates to the same graph as t_1 ; the graph value of t_2 is ignored. Thus if we assume that the subtree of t for cdd evaluates to some arbitrary graph G_0 , the complete derivation t evaluates to G_1 . Some intermediate results are shown on the right of Fig. 4.

If we let CSD_0 be the subset of CSD where all $K^{(x)}$ are zero, we can generalize this construction into an LM-CFTG which generates CSD_0 . Thus,

CSD_0 can be generated by a projective grammar that is interpreted into \mathcal{H}_2 . But note that the derivation in Fig. 4 is unnatural in that the symbols in the string are not generated by the same derivation steps that generate the graph nodes that intuitively correspond to them; for instance, the graphs generated for the d tokens are completely irrelevant. Below, we prevent unnatural constructions like this in two ways. We will first assume that string symbols and graph nodes must be *aligned* (Thm. 1). Then we will assume that the $K^{(x)}$ can be arbitrary, which allows us to drop the alignment assumption (Thm. 2).

4.3 k -distant trees

Let $\mathbf{R} \supseteq \text{CSD}_0$ be some relation containing at least the string-graph pairs of CSD_0 , e.g. CSD itself. Assume that \mathbf{R} is generated by a projective grammar (\mathcal{G}, h) with $\mathcal{G} = (G, \text{yd}_{\text{pr}})$ and a fixed number k of sources, i.e. we have $\mathbf{R} = \mathcal{REL}(\mathcal{G}, h, \mathcal{H}_k)$. We will prove a contradiction.

Given a pair $(w, G_w) \in \mathbf{R}$, we say that two edges e, f in G_w are *equivalent*, $e \equiv f$, if they belong to the same block. We call a derivation tree $t \in \mathbf{T} = L(G)$ k -distant if t has a subtree t' such that we can find k edges $e_1, \dots, e_k \in \text{graph}(t')$ with $e_i \not\equiv e_j$ for all $i \neq j$ and k further edges $e'_1, \dots, e'_k \in G_w \setminus \text{graph}(t')$ such that $e_i \equiv e'_i$ for all i . For such trees, we have the following lemma.

Lemma 3. *A k -distant tree has a subtree t' such that $\text{graph}(t')$ has at least k sources.*

Proof. Let BK_i be the i -th block in G_w ; we let $1 \leq i \leq m+n$ and do not distinguish between a - and b -blocks. Let t' be the subtree of t claimed by the definition of distant trees. For each i , let $E'_i = \text{BK}_i \cap \text{graph}(t')$ be the edges in the i -th block generated by t' , and let $E_i = \text{BK}_i \setminus E'_i$.

By definition, E_i and E'_i are both non-empty for at least k blocks. Each of these blocks is weakly connected, and thus contains at least one node u_i which is incident both to an edge in E_i and in E'_i . This node is a boundary node of $\text{graph}(t')$. Because u_1, \dots, u_k are all distinct, it follows from Lemma 2 that $\text{graph}(t')$ has at least k sources. \square

We also note the following lemma about derivations of projective string grammars, which follows from the inability of projective grammars to combine distant tokens. We write $\text{Sep} = \{a/c, c/a, b/d, d/b\}$.

Lemma 4. *Let $\mathcal{G} = (G, \text{yd})$ be a projective string grammar. For any $r \in \mathbb{N}_0$ there exists $s \in \mathbb{N}_0$ such*

that any $t \in L(G)$ with $\text{yd}(t) \in a^*b^sc^sd^*$ has a subtree t' such that $\text{yd}(t')$ contains r occurrences of x and no occurrences of y , for some $x/y \in \text{Sep}$.

4.4 Projectivity and alignments

A consequence of Lemma 3 is that if certain string-graph pairs in CSD_0 can only be expressed with $k+1$ -distant trees, then \mathbf{R} (which contains these pairs as well) is not in $\mathcal{L}(\mathcal{G}, \mathcal{H}_k)$, because \mathcal{H}_k only admits k sources.

However, as we saw in Section 4.2, pairs in CSD_0 can have unexpected projective derivations which make do with a low number of sources. So let's assume for now that the string grammar \mathcal{G} and the tree homomorphism h produce tokens and edge labels that fit together. Let us call \mathcal{G}, h aligned if for all constants $c \in \Sigma_0$, $\text{graph}(c)$ is a graph containing a single edge with label $\text{yd}(c)$. The derivation in Fig. 4 cannot be generated by an aligned grammar because the graph for the token b contains a d -edge. We write $\mathcal{L}_{\leftrightarrow}(\mathbb{G}, \mathcal{A}) = \{\mathcal{REL}(\mathcal{G}, h, \mathcal{A}) \mid \mathcal{G} \in \mathbb{G} \text{ and } \mathcal{G}, h \text{ aligned}\}$ for the class of string-semantics relations which can be generated with aligned grammars.

Under this assumption, it is easy to see that any relation including CSD_0 (hence, CSD) cannot be expressed with a projective grammar.

Theorem 1. *Let \mathbb{G} be any class of projective string grammars and $\mathbf{R} \supseteq \text{CSD}_0$. For any k , $\mathbf{R} \notin \mathcal{L}_{\leftrightarrow}(\mathbb{G}, \mathcal{H}_k)$.*

Proof. Assume that there is a $\mathcal{G} = (G, \text{yd}_{\text{pr}}) \in \mathbb{G}$ and an LTH h such that $\mathbf{R} = \mathcal{REL}(\mathcal{G}, h, \mathcal{H}_k)$. Given k , choose $s \in \mathbb{N}_0$ such that every tree $t \in \mathbf{T} = L(G)$ with $\text{yd}(t) = a^s b^s c^s d^s$ has a subtree t' such that $\text{yd}(t')$ contains $k+1$ occurrences of x and no occurrences of y , for some $x/y \in \text{Sep}$. Such an s exists according to Lemma 4. We can choose t such that $(\text{yd}(t), \text{graph}(t)) \in \text{CSD}_0$.

Because \mathcal{G}, h are aligned, $\text{graph}(t')$ contains no y -edge and at least $k+1$ x -edges. Each of these x -edges is non-equivalent to all the others, and equivalent to a y -edge in $\text{graph}(t) \setminus \text{graph}(t')$, so t is $k+1$ -distant. It follows from Lemma 3 that $\text{graph}(t')$ has $k+1$ sources, in contradiction to the assumption that \mathcal{G}, h uses only k sources. \square

5 Expressive capacity of LM-CFTG

Thm. 1 is a powerful result which shows that CSD cannot be generated by any device for generating projective derivations using bounded placeholder

memory – if we can assume that tokens and edges are aligned. We will now drop this assumption and prove that CSD cannot be generated using a fixed set of placeholders using LM-CFTG, regardless of alignment. The basic proof idea is to enforce a weak form of alignment through the interaction of the pumping lemma with very long \bar{x} -chains. The result is remarkable in that LM-CFTG and TAG are weakly equivalent; they only differ in whether they must derive the strings projectively or not.

Theorem 2. $\text{CSD} \notin \mathcal{L}(\text{LM-CFTG}, \mathcal{H}_k)$, for any k .

5.1 Asynchronous derivations

Assume that $\text{CSD} = \mathcal{REL}(\mathcal{G}, h, \mathcal{H}_k)$, for some k , with $\mathcal{G} = (G, \text{yd})$ an LM-CFTG. Proving that this is a contradiction hinges on a somewhat technical concept of *asynchronous* derivations, which have to do with how the nodes generating edge labels such as \bar{a} are distributed over a derivation tree. We prove that all asynchronous derivations of certain elements of CSD are distant (Lemma 5), and that all LM-CFTG derivations of CSD are asynchronous (Lemma 6), which proves Thm. 2.

In what follows, Let $\mathbf{T} = L(G)$. let us write for any tree or context t and symbol x , n_x^t as a shorthand for $|\text{yd}(t)|_x$, e_x^t for the number of x -edges in $\text{graph}(t)$ and m_x^t for the maximum length of a string in \bar{x}^* which is also substring of $\text{yd}(t)$.

Definition 1 (x, y, l -asynchronous derivation). *Let $x/y \in \text{Sep}$, $l > 0$, $t \in \mathbf{T}$, We call t an x, y, l -asynchronous derivation iff there is a decomposition $t = C[t']$ such that*

$$\begin{aligned} e_y^{t'} &\geq n_y^t - n_x^t l - m_y^t \\ e_x^{t'} &\leq n_x^t l + m_x^t(l+1). \end{aligned}$$

We call the pair (C, t') an x, y, l -asynchronous split of t .

Lemma 5. *For any $k, l > 0$, there is a pair $o_{k,l} = (w_{k,l}, G_{k,l}) \in \text{CSD}$ such that every x, y, l -asynchronous t with $o_{k,l} = (\text{yd}(t), \text{graph}(t))$ is k -distant.*

Proof. For $x \in \{a, b, c, d\}$ and $m \in \mathbb{N}_0$, let $\bar{x}^{(m)}$ denote the word $\langle^m \bar{x}^m \rangle^m$. Let $r = s = 3l+k+1$ and $o_{k,l} = (w_{k,l}, G_{k,l})$ be the unique element of CSD such that

$$w_{k,l} = (a\bar{a}^{(s)})^r (b\bar{b}^{(s)})^r (c\bar{c}^{(s)})^r (d\bar{d}^{(s)})^r.$$

Let t be an a, c, l -asynchronous derivation of $o_{k,l}$; other choices of $x/y \in \text{Sep}$ are analogous.

By definition, we can split $t = C[t']$ such that $\text{graph}(t')$ has at most $q_a = lr + (l+1)s = (2l+1)s$ \bar{a} -edges and at least $q_c = rs - rl - s = (2l+k+1)s$ \bar{c} -edges. Notice first that $\text{graph}(t')$ contains at most $2l + 1$ different complete a -blocks of $G_{k,l}$, because each a -block contains s \bar{a} -edges. Having $2l + 2$ of them would require $(2l + 2)s$ \bar{a} -edges, which is more than the q_a \bar{a} -edges that $\text{graph}(t')$ can contain.

Next, consider $2l + k$ distinct a -blocks of $G_{k,l}$. These blocks contain a total of $(2l + k)s < (2l + k + 1)s = q_c$ \bar{c} -edges. Hence, the \bar{c} -edges of $\text{graph}(t')$ cannot be contained within only $2l + k$ distinct blocks.

So we can find at least $2l + k + 1$ \bar{c} -edges in $\text{graph}(t')$ which are pairwise non-equivalent. There are at least k edges among these which are equivalent to an edge in $G_{k,l} \setminus \text{graph}(t')$, because $\text{graph}(t')$ contains at most l complete a -blocks of $G_{k,l}$. Thus, t is k -distant. \square

5.2 LM-CFTG derivations are asynchronous

So far, we have not used the assumption that \mathbf{T} is an LM-CFTL. We will now exploit the pumping lemma to show that all derivation trees of an LM-CFTG for CSD must be asynchronous.

Lemma 6. *If \mathbf{T} is an LM-CFTL, then there exists $l_0 \in \mathbb{N}_0$ such that for every $t \in \mathbf{T}$, there exists $x/y \in \text{Sep}$ such that t is x, y, l_0 -asynchronous.*

We prove this lemma by appealing to a class of derivation trees in which predicate and argument tokens are generated in separate parts.

Definition 2 (x, y, l -separated derivation). *Let $x/y \in \text{Sep}$. A tree $t \in \mathcal{T}_\Delta$ is x, y, l -separated if we can write $t = C_x[C_0[t_y]]$ such that $|\text{yd}(t_y)|_x = 0$ and $|\text{yd}(C_x)|_y = 0$ and $|\text{yd}(C_0)|_x \leq l$. The triple (C_x, C_0, t_y) is called an l -separation of t . We call an l -separation minimal if there is no other l -separation of t with a smaller C_0 .*

Intuitively, we can use the pumping lemma to systematically remove some contexts from a $t \in \mathbf{T}$. From the shape of CSD, we can conclude certain alignments between the strings and graphs generated by these contexts and establish bounds on the number of \bar{x} - and \bar{y} -edges generated by the lower part of a separated derivation. The full proof is in the appendix; we sketch the main ideas here.

Let p denote the pumping height of \mathbf{T} . There is a maximal number of string tokens and edges that a context of height at most p can generate under

a given yield and homomorphism. We call this number l_0 in the rest of the proof.

Lemma 7. *For $t \in \mathbf{T}$, let $r_{\bar{y}}^t$ be the length of the maximal substring of $\text{yd}(t)$ consisting in only \bar{y} -tokens and containing the rightmost occurrence of \bar{y} in $\text{yd}(t)$. If t is x, y, l_0 -separated, there exists a minimal x, y, l_0 -separation $D_x[D_0[t_y]]$ of t such that, letting $t_0 = D_0[t_y]$, $e_{\bar{y}}^{t_0} \geq n_{\bar{y}}^t - n_x^t l_0 - r_{\bar{y}}^t$.*

Moreover, for any x, y, l_0 -separation $t = E_x[E_0[t_y^1]]$, letting $t_1 = E_0[t_y^1]$, $e_{\bar{x}}^{t_1} \leq n_{\bar{x}}^{t_1} + n_x^t l_0$.

Proof (sketch). Both statements must be achieved in separated inductions on the height of t , although they mostly follow similar steps. We therefore focus here only on the crucial parts of the (slightly trickier) bound on $e_{\bar{y}}^{t_0}$. Let $D_x[D_0[t_y]]$ be a minimal x, y, l_0 -separation of t and $t_0 = D_0[t_y]$.

Base Case If $\text{ht}(t) \leq p$, we have $n_{\bar{y}}^t \leq l_0$. We also have $n_x^t > 0$, so $n_{\bar{y}}^t - n_x^t l_0 - r_{\bar{y}}^t \leq 0 \leq e_{\bar{y}}^{t_0}$.

Induction step If $\text{ht}(t) > p$, we apply Lemma 1 to t to yield a decomposition $t = C_1[C_2[C_3[C_4[t_5]]]]$, where $t' = C_1[C_3[t_5]] \in \mathbf{T}$, $\text{ht}(t') < \text{ht}(t)$ and $\text{ht}(C_2[[C_3]C_4]) \leq p$. We first observe that t' is x, y, l_0 -separated. By induction, there exists a minimal separation $t' = C_x[C_0[t'_y]]$ with $t'_0 = C_0[t'_y]$ validating the bound on $e_{\bar{y}}^{t'_0}$. Because of pumping considerations, we need to distinguish only three configurations of C_2 and C_4 . We present only the most difficult case here.

In this case C_2 and C_4 generate only one kind of bar symbol, \bar{y} , and brackets. One needs to examine all possible ways C_2 , C_4 and t_0 may overlap. We detail the reasoning in the case where t_0 does not overlap with C_2 or C_4 . Then, since all y -tokens are generated by t_0 , projectivity of the yield and the definition of CSD impose that the generated \bar{y} -tokens contribute to the rightmost y -chain i.e. $r_{\bar{y}}^t = r_{\bar{y}}^{t'} + n_{\bar{y}}^{C_2[C_4]}$. Hence $e_{\bar{y}}^{t_0} \geq e_{\bar{y}}^{t'_0} \geq n_{\bar{y}}^t - n_{\bar{y}}^{C_2[C_4]} + n_x^t l_0 - r_{\bar{y}}^{t_0} + n_{\bar{y}}^{C_2[C_4]}$. \square

Lemma 8. *For any $t \in \mathbf{T}$, if t is x, y, l_0 -separated then t is x, y, l_0 -asynchronous.*

Proof. By Lemma 7 there is a minimal x, y, l_0 -separation $t = D_x[D_0[t_y]]$ such that, for $t_0 = D_0[t_y]$, the bound on $e_{\bar{y}}^{t_0}$ and the bound on $e_{\bar{x}}^{t_0}$ both obtain. Observe that $r_{\bar{y}}^t \leq m_{\bar{y}}^t$ by definition, and since t_0 generates at most l_0 x -tokens, by projectivity it generates at most $(l_0 + 1)m_{\bar{x}}^t$ \bar{x} -tokens (one sequence of $m_{\bar{x}}^t$ between each occurrence of x and the next, plus possibly one before the first and one after the last). Thus t is x, y, l_0 -asynchronous. \square

Lemma 9. For any $t \in \mathbf{T}$, t is x, y, l_0 -separated for some $x/y \in \text{Sep}$.

Proof (sketch). The proof proceeds by induction on the height of t .

If $\text{ht}(t) \leq p$, then $|\text{yd}(t)|_z \leq l_0$ for any $z \in \{a, b, c, d\}$, hence t is trivially x, y, l_0 -separated for some $x/y \in \text{Sep}$.

If $h(t) > p$, Lemma 1 yields a decomposition $t = C_1[C_2[C_3[C_4[t_5]]]]$, where $t' = C_1[C_3[t_5]] \in \mathbf{T}$, $\text{ht}(t') < \text{ht}(t)$ and $\text{ht}(C_2[C_3[C_4]]) \leq p$. By induction t' is x, y, l_0 -separated for some $x/y \in \text{Sep}$. Let us assume $x/y = a/c$, other cases are analogous. The challenge is to conclude to the x, y, l separation of t , after reinsertion of C_2 and C_4 in t' .

If C_2 and C_4 generate no a - or c -token, the distribution of a - and c -tokens in the tree is not affected, hence t is a, c, l_0 -separated. Otherwise, due to pumping considerations, we need to distinguish three possible configurations regarding the shape of the yields of C_2 and C_4 . We present one here, see the appendix for the others; they are in the same spirit.

We consider the case where $\text{left}(C_2)$ contains some a -token and no b, c, d -tokens, and $\text{yd}(C_4)$ contains some c -token. Assume $\text{left}(C_4)$ contains some c . It follows that all b -tokens are generated by C_3 . So t has less than l_0 b -tokens, by definition of CSD it has then also less than l_0 d -tokens, so $(C_1, C_2[C_3[C_4]], t_5)$ is a d, b, l_0 -separation. Assume now that $\text{right}(C_4)$ contains some c . It follows that t_5 generates no d -token and C_1 generates no b -token. Hence $(C_1, C_2[C_3[C_4]], t_5)$ is a b, d, l_0 -separation. \square

This concludes the proof of Lemma 6 and Thm. 2.

6 Conclusion

We have established a notion of expressive capacity in compositional semantic parsing. We have proved that non-projective grammars can express sentence-meaning relations with bounded memory that projective ones cannot. This answers an old question in the design of compositional systems: assuming projective syntax, lambda-style compositional mechanisms can be more expressive than unification-style ones, which have bounded “memory” for unfilled arguments.

From a theoretical perspective, the stronger result of this paper is perhaps Thm. 2, which shows without further assumptions that weakly equivalent grammar formalisms can differ in their se-

mantic expressive capacity. However, Thm. 1 may have a clearer practical impact on the development of compositional semantic parsers. Consider, for instance, the case of CCG, a lexicalized grammar formalism that has been widely used in semantic parsing (Bos, 2008; Artzi et al., 2015; Lewis et al., 2016). While a potentially infinite set of syntactic categories can be used in the parses of a single CCG grammar, CCG derivations are still projective in our sense. Thus, if one assumes that derivations should be aligned (which is natural for a lexicalized grammar), Thm. 1 implies that CCG with lambda-style semantic composition is more semantically expressive than with unification-style composition. Indeed, lambda-style compositional mechanisms are the dominant approach in CCG (Steedman, 2001; Baldridge and Kruijff, 2002; Artzi et al., 2015).

Furthermore, under the alignment assumptions of Section 4, no unification-style compositional mechanism can describe string-meaning relations like CSD. This includes neural models. For instance, most transition-based parsers (Nivre, 2008; Andor et al., 2016; Dyer et al., 2016) are projective, in that the parsing operations can only concatenate two substrings on the top of the stack if they are adjacent in the string. Such transition systems can therefore not be extended to transition-based semantic parsers (Damonte et al., 2017) without (a) losing expressive capacity, (b) giving up compositionality, (c) adding mechanisms for non-projectivity (Gómez-Rodríguez et al., 2018), or (d) using a lambda-style semantic algebra. Thus our result clarifies how to build an effective and accurate semantic parser. \square

We have focused on whether a grammar formalism is projective or not, while holding the semantic algebra fixed. In the future, it would be interesting to explore how a unification-style compositional mechanism can be converted to a lambda-style mechanism with unbounded placeholders. This would allow us to specify and train semantic parsers using such abstractions, while benefiting from the efficiency of projective parsers.

Acknowledgments

We are grateful to Emily Bender, Guy Emerson, Meaghan Fowlie, Jonas Groschwitz, and the participants of the DELPH-IN workshop 2018 for fruitful discussions, and to the anonymous reviewers for their insightful feedback.

References

- Daniel Andor, Chris Alberti, David Weiss, Aliaksei Severyn, Alessandro Presta, Kuzman Ganchev, Slav Petrov, and Michael Collins. 2016. Globally normalized transition-based neural networks. In *Proceedings of ACL*.
- Yoav Artzi, Kenton Lee, and Luke Zettlemoyer. 2015. Broad-coverage CCG Semantic Parsing with AMR. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*.
- Jason Baldridge and Geert-Jan M. Kruijff. 2002. Coupling CCG and Hybrid Logic Dependency Semantics. In *Proceedings of the 40th ACL*.
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for Sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*.
- Emily M. Bender. 2008. Radical non-configurationality without shuffle operators: An analysis of Wambaya. In *Proceedings of the 15th International Conference on HPSG*.
- Emily M. Bender, Dan Flickinger, and Stephan Oepen. 2002. The Grammar Matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In *Proceedings of the COLING Workshop on Grammar Engineering and Evaluation*.
- Patrick Blackburn and Johan Bos. 2005. *Representation and Inference for Natural Language*. CSLI Publications.
- Johan Bos. 2008. Wide-coverage semantic analysis with Boxer. In *Semantics in Text Processing. STEP 2008 Conference Proceedings*. College Publications.
- David Chiang, Jacob Andreas, Daniel Bauer, Karl Moritz Hermann, Bevan Jones, and Kevin Knight. 2013. Parsing graphs with hyperedge replacement grammars. In *Proceedings of the 51st ACL*.
- Hubert Comon, Max Dauchet, Rémi Gilleron, Florent Jacquemard, Denis Lugiez, Sophie Tison, Marc Tommasi, and Christof Löding. 2007. *Tree Automata techniques and applications*. Published online at <http://tata.gforge.inria.fr/>.
- Ann Copestake, Alex Lascarides, and Dan Flickinger. 2001. An algebra for semantic construction in constraint-based grammars. In *Proceedings of the 39th ACL*.
- Bruno Courcelle and Joost Engelfriet. 2012. *Graph Structure and Monadic Second-Order Logic, a Language Theoretic Approach*. Cambridge University Press.
- Mary Dalrymple, John Lamping, Fernando C. N. Pereira, and Vijay Saraswat. 1995. Linear logic for meaning assembly. In *Proceedings of the Workshop on Computational Logic for Natural Language Processing*.
- Marco Damonte, Shay B. Cohen, and Giorgio Satta. 2017. An incremental parser for Abstract Meaning Representation. In *Proceedings of the 15th EACL*.
- Frank Drewes, Hans-Jörg Kreowski, and Annegret Habel. 1997. Hyperedge replacement graph grammars. In G. Rozenberg, editor, *Handbook of Graph Grammars and Computing by Graph Transformation*, pages 95–162. World Scientific.
- Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A. Smith. 2016. Recurrent neural network grammars. In *Proceedings of NAACL*.
- Anette Frank and Josef van Genabith. 2001. GlueTag: Linear logic based semantics for LTAG. In *Proceedings of the LFG Conference*.
- Claire Gardent and Laura Kallmeyer. 2003. Semantic construction in feature-based TAG. In *Proceedings of EACL*.
- Jonas Groschwitz, Matthias Lindemann, Meaghan Fowlie, Mark Johnson, and Alexander Koller. 2018. AMR dependency parsing with a typed semantic algebra. In *Proceedings of ACL*.
- Carlos Gómez-Rodríguez, Tianze Shi, and Lillian Lee. 2018. Global transition-based non-projective dependency parsing. In *Proceedings of ACL*.
- Aravind K. Joshi and Yves Schabes. 1997. Tree-Adjoining Grammars. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 3. Springer-Verlag.
- Stephan Kepser and James Rogers. 2011. The equivalence of tree adjoining grammars and monadic linear context-free tree grammars. *Journal of Logic, Language and Information*, 20(3):361–384.
- Alexander Koller. 2015. Semantic construction with graph grammars. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 228–238.
- Mike Lewis, Kenton Lee, and Luke Zettlemoyer. 2016. LSTM CCG Parsing. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- T. Maibaum. 1978. Pumping lemmas for term languages. *Journal of Computer and System Sciences*, pages 319–330.
- Richard Montague. 1974. The proper treatment of quantification in ordinary English. In R. Thomason, editor, *Formal philosophy: Selected papers of Richard Montague*. Yale University Press, New Haven.

Joakim Nivre. 2008. Algorithms for deterministic incremental dependency parsing. *Computational Linguistics*, 34(4):513–553.

Xiaochang Peng, Linfeng Song, and Daniel Gildea. 2015. A synchronous hyperedge replacement grammar based approach for AMR parsing. In *Proceedings of the 19th Conference on Computational Language Learning*.

William Rounds. 1969. Context-free grammars on trees. In *Proceedings of the First Annual ACM Symposium on Theory of Computing (STOC)*.

Stuart M. Shieber. 1985. Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, 8(3):333–343.

Mark Steedman. 2001. *The Syntactic Process*. MIT Press, Cambridge, MA.

David Weir. 1988. *Characterizing mildly context-sensitive grammar formalisms*. Ph.D. thesis, University of Pennsylvania.

A Details of the proof of Theorem 1

Lemma 4. Let $\mathcal{G} = (G, \text{yd})$ be a projective string grammar. For any $r \in \mathbb{N}_0$ there exists $s \in \mathbb{N}_0$ such that any $t \in L(\mathcal{G})$ with $\text{yd}(t) \in a^*b^sc^sd^*$ has a subtree t' such that $\text{yd}(t')$ contains r occurrences of x and no occurrences of y , for some $x/y \in \text{Sep}$.
Proof. Depending on yd , one can always choose $s > r$ such that any t with $|\text{yd}(t)| > 2s$ has at least one strict subtree t' with $|\text{yd}(t')| \geq 2r$.

The lemma follows by induction over the height of t . It is trivially true for height 1. For the induction step, consider that $w' = \text{yd}(t')$ must have at least r occurrences of some letter because of projectivity and the shape of $\text{yd}(t)$; assume it is a , the other cases are analogous. If w' has no occurrences of c , we are done. Otherwise, by projectivity, w' contains all the b 's, i.e. $w' \in a^*b^sc^+d^*$. In this case, either w' contains $s > r$ occurrences of b and no occurrences of d , in which case we are again done. Or it contains an occurrence of d ; then $w' \in a^*b^sc^sd^*$ is in the shape required by the lemma, and we can apply the induction hypothesis to identify a subtree t'' of t' with r occurrences of some x and none of the corresponding y ; and t'' is also a subtree of t . \square

B Details of the proof of Theorem 2

In all of the following, we assume that for some $k \in \mathbb{N}_0$ we have $\text{CSD} = \mathcal{REL}(\mathcal{G}, h, \mathcal{H}_k)$, where $\mathcal{G} = (G, \text{yd})$ is an LM-CFTG (hence projective, i.e. $\text{yd} = \text{yd}_{\text{pr}}$). We let $\mathbf{T} = L(\mathcal{G})$ and p be the pumping height of \mathbf{T} .

B.1 Terminology

Let us extend the domain of yd to contexts: for a context C , we let $\text{yd}(C) = \text{left}(C) \cdot \text{right}(C)$.

We say that a string s is *balanced* if, for any $z \in \{\bar{a}, \bar{b}, \bar{c}, \bar{d}\}$ and any position i in s such that $s_i = z$ there are two encompassing positions $k \leq i \leq l$ such that $s_{[k,l]} \in \{\langle n\bar{z} \rangle^n \mid n \in \mathbb{N}_0\}$. We say that a tree or a context is balanced if its yield is balanced. By construction, all trees of \mathbf{T} are balanced.

For $t \in \mathbf{T}$, a *pumping decomposition* of t is a 5-tuple $(C_1, C_2, C_3, C_4, t_5)$, consisting in 4 contexts $C_1\text{-}C_4$ and one tree t_5 such that $t = C_1[C_2[C_3[C_4[t_5]]]]$, $\text{ht}(C_2[C_3[C_4]]) \leq p$, $\text{ht}(C_2) + \text{ht}(C_4) > 0$ and for any $i \in \mathbb{N}_0$, $C_1[v^i[t_5]] \in \mathbf{T}$, where we let $v^0 = C_3$ and $v^{i+1} = C_2[v^i[C_4[X]]]$.

B.2 Pumping considerations

Lemma 10. Let $t \in \mathbf{T}$ with $\text{ht}(t) > p$, and consider a pumping decomposition $t = C_1[C_2[C_3[C_4[t_5]]]]$. Let $s = \text{left}(C_2) \cdot \text{left}(C_4) \cdot \text{right}(C_4) \cdot \text{right}(C_2) = \text{yd}(C_2[C_4])$. The two following propositions obtain:

- For any $(x, y) \in \{(a, c), (b, d)\}$, $|s|_x = |s|_y$.
- Let $t' = C_1[C_3[t_5]]$. For $u \in \mathcal{T}_\Delta$ and $z \in \{a, b, c, d, \bar{a}, \bar{b}, \bar{c}, \bar{d}\}$ let e_z^u denote the number of z -edges in $\text{graph}(u)$. It holds for any $z \in \{a, b, c, d, \bar{a}, \bar{b}, \bar{c}, \bar{d}\}$ that $e_z^t = e_z^{t'} + |s|_z$.

Proof. Let $(x, y) \in \{(a, c), (b, d)\}$. $t \in \mathbf{T}$ so $\text{yd}(t) \in \text{CSD}_s$ which entails

$$|\text{yd}(t)|_x = |\text{yd}(t)|_y. \quad (1)$$

since $t' \in \mathbf{T}$ by construction, we have $\langle \text{yd}(t'), \text{graph}(t') \rangle \in \text{CSD}$. From there

$$|\text{yd}(t')|_x = |\text{yd}(t')|_y. \quad (2)$$

But $|\text{yd}(t)|_{x,y} = |\text{yd}(t')|_{x,y} + |s|_{x,y}$. Plugging this into (1) yields

$$|\text{yd}(t')|_x + |s|_x = |\text{yd}(t')|_y + |s|_y.$$

Simplifying using (2) we find $|s|_x = |s|_y$ which establishes the first point. For the second point, we have from $\langle \text{yd}(t), \text{graph}(t) \rangle \in \text{CSD}$ and by definition of CSD $e_z^t = |\text{yd}(t)|_z = |\text{yd}(t')|_z + |s|_z$. Similarly since $\langle \text{yd}(t'), \text{graph}(t') \rangle \in \text{CSD}$ we have $e_z^{t'} = |\text{yd}(t')|_z$. Hence $e_z^t = e_z^{t'} + |s|_z$. \square

We will now present a pair of lemmas stating, in formal terms, that decompositions $t = C_1[C_2[C_3[C_4[t_5]]]]$ provided by the pumping lemma all fall within a small number of configurations:

- First, in the case where the ‘pumpable’ contexts C_2 and C_4 generate only ‘bar’ tokens and brackets in $\{\bar{a}, \bar{b}, \bar{c}, \bar{d}, \langle, \rangle\}^*$, we show that $\text{yd}(C_2) \in \{\langle, \rangle\}^*$, so that only C_4 is actually pumping ‘bar’ tokens of some kind. Moreover, t_5 generates only ‘bar’ tokens and brackets as well.
- Second, we explore the alternative, where the ‘pumpable’ contexts generate some of the ‘core’ tokens in $\{a, b, c, d\}$, say – for the sake of this informal presentation – some a -tokens. By lemma 10, they must generate as many c -tokens, for which we can again distinguish three possible configurations: 1. a 's and c 's are respectively generated on different sides of a single context (C_2 and/or C_4), but then neither C_2 nor C_4 generate any b or d -tokens. 2. C_2 generate both a and d -tokens (on the left and right sides respectively) and no b and c -tokens, while C_4 ensures generation of corresponding b and c -tokens (on the left and right sides respectively). 3. Or else, one of C_2, C_4 generates the a -tokens and no c, b or d while the other generates the corresponding c -tokens and no a, b or d .

Below follows the formal presentation of these lemmas:

Lemma 11. *Let $t \in \mathbf{T}$ with $\text{ht}(t) > p$, and consider a pumping decomposition $t = C_1[C_2[C_3[C_4[t_5]]]]$ such that for all $z \in \{a, b, c, d\}$, $|\text{yd}(C_2[C_4])|_z = 0$. There is $\bar{x} \in \{\bar{a}, \bar{b}, \bar{c}, \bar{d}\}$ such that all of the following holds:*

1. $\text{yd}(C_2) \in \{\langle, \rangle\}^*$ and $\text{yd}(C_4) \in \{\langle, \bar{x}, \rangle\}^*$.
2. Either $\text{left}(C_4) \in \{\bar{x}\}^*$ and $|\text{left}(C_3)|_z = 0$ for any $z \in \{a, b, c, d\}$, or symmetrically, $\text{right}(C_4) \in \{\bar{x}\}^*$ and $|\text{right}(C_3)|_z = 0$ for any $z \in \{a, b, c, d\}$.
3. $|\text{yd}(t_5)|_z = 0$ for any $z \in \{a, b, c, d\}$

Proof. **First point:** Let $s = \text{left}(C_1)$ and $n_0 = |s|_\langle$. Let $y \in \{\bar{a}, \bar{b}, \bar{c}, \bar{d}\}$ and assume $y \in \text{left}(C_2)$. Pumping C_2 - C_4 $n_0 + 1$ -times yields a tree $t_{n_0+1} \in \mathbf{T}$ such that $s \cdot \text{left}(C_2)^{n_0+1}$ is a prefix of $\text{yd}(t_{n_0+1})$. We thus see that t_{n_0+1} is not balanced, which is in contradiction with $t_{n_0+1} \in \mathbf{T}$. A symmetric argument establishes that $y \notin \text{right}(C_2)$.

Assume now that there are two distinct $\bar{x}, \bar{y} \in \{\bar{a}, \bar{b}, \bar{c}, \bar{d}\}$ such that $\bar{x} \in \text{yd}(C_4)$ and $\bar{y} \in \text{yd}(C_4)$. Notice that, since C_4 does not contain non-bar tokens, if \bar{x} and \bar{y} occur on the same side of C_4 (for instance $\text{left}(C_4) = \langle \bar{x} \rangle \langle \bar{y} \rangle$) then $t \notin \mathbf{T}$ be-

cause no string in CSD_S admits $\text{left}(C_4)$ as a substring, whereas $\text{yd}(t)$ does. So \bar{x} and \bar{y} must occur on distinct sides. It follows that C_4 does not generate tokens in $\{\langle, \rangle\}$ either: if for instance $\text{left}(C_4) = u \cdot \langle \cdot \bar{x} \cdot v$ for some strings u and v in $\{\bar{x}\}^*$, $u \cdot \langle \cdot \bar{x} \cdot v \cdot u \cdot \langle \cdot \bar{x} \cdot v$ would be a substring of $C_1[C_2[C_3[C_4[C_4[t_5]]]]] \in \mathbf{T}$ which again is a contradiction. Let now $n_1 = |\text{yd}(t_5)|_\langle$. Pumping C_2 - C_4 $n_1 + 1$ times yields a tree $t_{n_1+1} \in \mathbf{T}$ with a substring of the form $\bar{x}^{n_1+1} \text{yd}(t_5) \bar{y}^{n_1+1}$ (up to x/y symmetry) which cannot be balanced, yielding a final contradiction.

Second point: $\text{yd}(C_4) \notin \{\langle, \rangle\}^*$, because otherwise pumping C_2 and C_4 more times than the maximum number of occurrences of a bar token in $\text{yd}(t)$ would yield an unbalanced tree. So there is a \bar{x} such that $\bar{x} \in \text{left}(C_4)$ or $\bar{x} \in \text{right}(C_4)$. Assume for contradiction that any different token occurs on the same side of C_4 then $C_1[C_2[C_3[C_4[C_4[t_5]]]]] \in \mathbf{T}$ contains a substring that cannot be found in any string of CSD yielding a contradiction. So $\text{left}(C_4) \in \{\bar{x}\}^*$ or $\text{right}(C_4) \in \{\bar{x}\}^*$. Assume $\text{left}(C_4) \in \{\bar{x}\}^*$, the other case is symmetric. Assume for contradiction that $|\text{left}(C_3)|_z > 0$ for some $z \in \{a, b, c, d\}$. Let $n_2 = |\text{yd}(C_3)|_\langle$. Pumping C_2 - C_4 $n_2 + 1$ times yields a tree $t^{n_2+1} \in \mathbf{T}$ such that (by projectivity) $\text{yd}(t^{n_2+1})$ has a substring of the form $z \cdot u \cdot \bar{x}^{n_2+1}$ where $|u|_\langle \leq n_2$. Hence $t^{n_2+1} \in \mathbf{T}$ is not balanced, yielding a contradiction.

Third point: Assume for contradiction that $|\text{yd}(t_5)|_z > 0$. Assume that $\text{left}(C_4) \in \{\bar{x}\}^*$, the case $\text{right}(C_4) \in \{\bar{x}\}^*$ is symmetric, and point 2 ensures that these two cases are exhaustive. Let $n_3 = |t_5|_\langle$ and consider the tree $t^{n_3+1} \in \mathbf{T}$ obtained by pumping C_2 - C_4 $n_3 + 1$ times. By projectivity, $\text{yd}(t^{n_3+1})$ has a substring of the form $\bar{x}^k \cdot u \cdot z \cdot v$ with $k \geq n_3 + 1$ and $|u|_\langle \leq n_3$. Hence t^{n_3+1} is not balanced and $t^{n_3+1} \notin \mathbf{T}$, yielding a contradiction. \square

Lemma 12. *let $t \in \mathbf{T}$ with $\text{ht}(t) > p$, and consider a pumping decomposition $t = C_1[C_2[C_3[C_4[t_5]]]]$. Let $(x, y, X, Y) \in \{(a, c, A, C), (b, d, B, D)\}$ such that $|\text{yd}(C_2[C_4])|_x \neq 0$. One of the following obtains:*

1. For some $(i, j) \in \{(2, 4), (4, 2)\}$, $\text{left}(C_i) \in X^+$, $\text{right}(C_i) \in Y^+$, $\text{left}(C_j) \in X^*$ and $\text{right}(C_j) \in Y^*$.
2. $\text{left}(C_2) \in A^+$, $\text{right}(C_2) \in D^+$, $\text{left}(C_j) \in$

B^+ and $\text{right}(C_j) \in C^+$.

3. Either $\text{left}(C_2) \in X^+$, $\text{right}(C_2) = \epsilon$ and $\text{left}(C_4) \cdot \text{right}(C_4) \in Y^+$, or symmetrically $\text{left}(C_2) = \epsilon$, $\text{right}(C_2) \in Y^+$ and $\text{left}(C_4) \cdot \text{right}(C_4) \in X^+$.

Proof. All these observations follow easily from the first point of Lemma 10 (governing the relative number of occurrences of a, c -tokens on one hand and b, d -tokens on the other hand), projectivity, and the following observation: only one side of C_2 or C_4 cannot generate two different kinds of tokens in $\{a, b, c, d\}$ or be unbalanced. Otherwise pumping would (from projectivity) ensure that the resulting tree has a substring of a shape impossible for CSD (for example, if both a and b -tokens occur on the same side of C_2 , pumping once produces a substring $a \cdot u \cdot b \cdot v \cdot a \cdot u \cdot b \cdot v$). \square

B.3 Separation

Lemma 13. *Let $t = C_1[C_2[C_3[C_4[t_5]]]] \in \mathbf{T}$ and $t' = C_1[C_3[t_5]] \in \mathbf{T}$. If t is x, y, l -separated then so is t' .*

Proof. Consider an x, y, l -separation of t : $t = D_x[D_0[t_y]]$. Let C_x, C_0 and t'_y be respectively obtained by removing all nodes from C_2 or C_4 from D_x, D_0 and t_y . One easily checks that $t' = C_x[C_0[t'_y]]$.

Moreover, $|\text{yd}(C_x)|_y \leq |\text{yd}(D_x)|_y = 0$, $|\text{yd}(C_0)|_x \leq |\text{yd}(D_0)|_x \leq l$ and $|\text{yd}(t'_y)|_x \leq |\text{yd}(t_y)|_x = 0$. Hence t' is x, y, l -separated. \square

B.4 Minimality argument

Lemma 14. *Let $t = C_1[C_2[C_3[C_4[t_5]]]] \in \mathbf{T}$ and $t' = C_1[C_3[t_5]] \in \mathbf{T}$ such that t is x, y, l -separated. By Lemma 13, t' is separated. Let $D_x[D_0[t_y]]$ be a minimal separation of t and $C_x[C_0[t'_y]]$ be a minimal separation of t' . $D_0[t_y]$ contains all nodes of $C_0[t'_y]$.*

Proof. Assume for contradiction that a node π of C_0 is not in D_0 . It must then be in D_x or t_y . Assume that it is in D_x , the case where it is in t_y is analogous. Since π is not in D_0 , there is a non-trivial subcontext D'_x of D_x rooted at π , i.e. $D_x = D''_x[D'_x]$ with $\text{ht}(D'_x) > 0$. Let C''_x, C'_x be obtained by removing all nodes from C_2 or C_4 from D''_x and D'_x respectively. By definition of D_x , $|\text{yd}(D''_x[D'_x])|_y = 0$, hence $|\text{yd}(C''_x[C'_x])|_y = 0$. Further observe that we have $C_x[C_0] = C''_x[C'_x[C'_0]]$ for some subcontext C'_0 of C_0 . Since π is not in C_2 or C_4 , $\text{ht}(C'_x) > 0$

thus $\text{ht}(C'_0) < \text{ht}(C_0)$. But letting $E_x = C''_x[C'_x]$, $E_x[C'_0[t'_y]]$ is then an x, y, l -separation of t which contradicts the assumed minimality of $C_x[C_0[t'_y]]$. \square

B.5 Inductive bounds

For any tree or context t and symbol x , let us write n_x^t as a shorthand for $|\text{yd}(t)|_x$, e_x^t for the number of x -edges generated by t and r_x^t the length of the rightmost maximal substring of $\text{yd}(t)$ consisting in only x -tokens (more formally, $r_x^t = |s|_x$, where s is the unique substring such that $\text{yd}(t) = u \cdot s \cdot v$ where $s \in x^*$, if u is non empty its last token is not x , and $|v|_x = 0$).

There is a maximal number of string tokens and edges that a context of height at most p can generate under the considered yield and homomorphism. We call l_0 this number and focus from now on l_0 -separated and l_0 -asynchronous derivations.

Below are the proofs of the two statements of Lemma 7 of the main paper (respectively, 7-1 and 7-2).

Lemma 7-1. *If $t \in \mathbf{T}$ is x, y, l_0 -separated and $t = D_x[D_0[t_y]]$ is an x, y, l_0 -separation of t , then for $t_0 = D_0[t_y]$ we have*

$$e_x^{t_0} \leq n_x^{t_0} + n_x^t l_0. \quad (\text{x bound})$$

Proof. We prove the result by induction over the pair $(\text{ht}(t_0), \text{ht}(t))$ (with lexicographic ordering).

Base Case Assume $\text{ht}(t_0) \leq p$. Then $e^{t_0} \leq l_0$. Since $\text{yd}(t) \in \text{CSD}_s$, $n_x^t > 0$, thus $n_x^{t_0} + n_x^t l_0 \geq l_0$ which ensures the bound.

Induction step If $\text{ht}(t_0) > p$ then $h(t) \geq h(t_0) > p$. We apply Lemma 1 to t to yield a decomposition $t = C_1[C_2[C_3[C_4[t_5]]]]$, where $t' = C_1[C_3[t_5]] \in \mathbf{T}$, $\text{ht}(t') < \text{ht}(t)$ and $\text{ht}(C_2[C_3[C_4]]) \leq p$. Notice that t_0 cannot overlap with $C_2[C_3[C_4]]$ without overlapping with C_1 or t_5 as well, for otherwise $h(t_0) \leq p$.

As in the proof of Lemma 13, letting C_x, C_0, t'_y be obtained by removing all nodes from C_2 and C_4 from D_x, D_0 and t_y respectively, we obtain an x, y, l -separation $t' = C_x[C_0[t'_y]]$. We let $t'_0 = C_0[t'_y]$ and distinguish between possible configurations for C_2 and C_4 :

Case 0 If neither C_2 or C_4 generate any \bar{x} -token, we find by induction

$$e_{\bar{x}}^{t'_0} \leq n_{\bar{x}}^{t'_0} + n_x^{t'} l_0.$$

Moreover, we have $e_{\bar{x}}^{t'_0} = e_{\bar{x}}^{t_0}$, $n_{\bar{x}}^{t'_0} = n_{\bar{x}}^{t_0}$ and $n_x^{t'} \leq n_x^{t_0}$ which concludes.

Case 1 In this case Lemma 11 applies i.e. C_2 and C_4 generate only some \bar{z} -tokens and brackets. The only subcase not already covered by Case 0 is the one where $\bar{z} = \bar{x}$. Notice that $n_x^{t'} = n_x^t$. By induction,

$$e_{\bar{x}}^{t_0} \leq n_{\bar{x}}^{t_0} + n_x^{t'} l_0.$$

If t_0 does not overlap with C_2 or C_4 , we have $e_{\bar{x}}^{t_0} = e_{\bar{x}}^{t_0}$ and $n_{\bar{x}}^{t_0} = n_{\bar{x}}^{t_0}$ which ensures the bound.

Otherwise t_0 overlaps with C_4 . If all nodes of t_0 are contained in $C_4[t_5]$, then by Lemma 11, t_0 generate no y -token. By separation, neither does t which contradicts $t \in \text{CSD}$. Hence t_0 contains all nodes of C_4 . Then by lemma 11 again, $n_{\bar{x}}^{C_2[C_4]} = n_{\bar{x}}^{C_4}$, hence $n_{\bar{x}}^{t_0} = n_{\bar{x}}^{t_0} + n_{\bar{x}}^{C_2[C_4]}$ and $e_{\bar{x}}^{t_0} \leq e_{\bar{x}}^{t_0} + n_{\bar{x}}^{C_2[C_4]}$ which yields

$$e_{\bar{x}}^{t_0} \leq e_{\bar{x}}^{t_0} + n_{\bar{x}}^{C_2[C_4]} \leq n_{\bar{x}}^{t_0} + n_x^t l_0.$$

Case 2 In this case Lemma 12 applies and at least one of C_2-C_4 generate some token $z \in \{a, b, c, d\}$. The only subcase not already dealt with in Case 0 is the one where we can set $z = x$. We thus get inductively:

$$e_{\bar{x}}^{t_0} \leq n_{\bar{x}}^{t_0} + n_x^{t'} l_0.$$

Since C_2 or C_4 generate at least some x -token, we have $n_x^t \geq n_x^{t'} + 1$. Moreover $e_{\bar{x}}^{t_0} \leq e_{\bar{x}}^{t_0} + l_0$ since $C_2[C_4]$ generate at most l_0 \bar{x} -edges, and $n_{\bar{x}}^{t_0} \geq n_{\bar{x}}^{t_0}$. So we have $e_{\bar{x}}^{t_0} \leq n_{\bar{x}}^{t_0} + n_x^{t'} l_0 + l_0 \leq n_{\bar{x}}^{t_0} + n_x^t l_0$ concluding the proof. \square

Lemma 7-2. *If $t \in \mathbf{T}$ is x, y, l_0 -separated then t there exists a minimal x, y, l_0 -separation $D_x[D_0[t_y]]$ of t is such that, letting $t_0 = D_0[t_y]$, we have*

$$e_{\bar{y}}^{t_0} \geq n_{\bar{y}}^t - n_x^t l_0 - r_{\bar{y}}^t \quad (\text{y bound})$$

Proof. We prove the result by induction over the height of t .

t is x, y, l_0 -separated so let us consider $D_x[D_0[t_y]]$ a minimal x, y, l_0 -separation of t . Let $t_0 = D_0[t_y]$.

Base Case Assume $\text{ht}(t) \leq p$. Then. $n_{\bar{y}}^t \leq l_0$. Since $\text{yd}(t) \in \text{CSD}_s$, $n_x^t > 0$. Moreover, $0 \leq e_{\bar{y}}^{t_0}$ and $n_{\bar{y}}^t \leq l_0$. So $n_{\bar{y}}^t - n_x^t l_0 - r_{\bar{y}}^t \leq 0 \leq e_{\bar{y}}^{t_0}$ which ensures the bound.

Induction step If $\text{ht}(t) > p$, we apply Lemma 1 to t to yield a decomposition $t = C_1[C_2[C_3[C_4[t_5]]]]$, where $t' = C_1[C_3[t_5]] \in \mathbf{T}$, $\text{ht}(t') < \text{ht}(t)$ and $\text{ht}(C_2[C_3[C_4]]) \leq p$. By

Lemma 13, t' is x, y, l_0 -separated. We let $t' = C_x[C_0[t'_y]]$ be a minimal separation of t' verifying the bound and $t'_0 = C_0[t'_y]$. In other words, we have:

$$e_{\bar{y}}^{t_0} \geq n_{\bar{y}}^{t'} - n_x^{t'} l_0 - r_{\bar{y}}^{t'}. \quad (3)$$

By Lemma 14, $t_0 = D_0[t_y]$ contains all nodes of t'_0 . We distinguish cases according to Lemmas 11 and 12.

Case 1 Consider first the case where Lemma 11 applies i.e. C_2 and C_4 generate only one kind of bar token, \bar{z} , and brackets. We now distinguish cases depending on the value of \bar{z} . Before this, we emphasize that in all subcases it holds that $n_x^t = n_x^{t'}$.

subcase i) $\bar{z} \neq \bar{y}$. Since all nodes of t'_0 are contained in t_0 , we have $e_{\bar{x}}^{t_0} \geq e_{\bar{x}}^{t_0}$. Since C_2 and C_4 generate no \bar{y} -token, we have $n_{\bar{y}}^t = n_{\bar{y}}^{t'}$ and $r_{\bar{y}}^t = r_{\bar{y}}^{t'}$. Injecting into inequation (3) concludes.

subcase ii) $\bar{z} = \bar{y}$. We distinguish the different possible overlap of C_2 and C_4 with t_0 . Notice first that, by minimality, if any C_i , $i \in \{2, 4\}$ overlaps with t_0 then t_0 contains all nodes of C_i , for otherwise we would have $D_0 = D'_0[D''_0]$ with D'_0 a subcontext of C_i such that $\text{ht}(D'_0) > 0$, and in that case $(D_x[D'_0], D''_0, t_y)$ would be a smaller x, y, l -separation of t since C_i (hence D'_0) does not generate y -tokens.

Hence, in the case where t_0 overlaps with C_2 , t_0 contains all nodes of C_2 and C_4 . Since t_0 also contains all nodes of t'_0 , $e^{t_0} \geq e^{t'_0} + e^{C_2[C_4]} = e^{t'_0} + n_{\bar{y}}^t - n_{\bar{y}}^{t'}$. Moreover, $r_{\bar{y}}^t \geq r_{\bar{y}}^{t'}$. We can then conclude using inequation 3.

Consider now the case where t_0 does not overlap with C_2 or C_4 . Since all y -tokens are generated by t_0 , projectivity of the yield and the definition of CSD impose that $r_{\bar{y}}^t = r_{\bar{y}}^{t'} + n_{\bar{y}}^{C_2[C_4]}$. We further have $e_{\bar{y}}^{t_0} \geq e_{\bar{y}}^{t'_0}$, and injecting into inequation 3 yields $e_{\bar{y}}^{t_0} \geq n_{\bar{y}}^t - n_{\bar{y}}^{C_2[C_4]} + n_x^t l_0 - r_{\bar{y}}^{t_0} + n_{\bar{y}}^{C_2[C_4]}$ which simplifies into the desired y bound.

Finally, in the case where C_2 does not overlap with t_0 but C_4 does, all nodes of C_2 are contained in D_x and all nodes of C_4 are contained in t_0 . We must then have $|\text{yd}(C_3)|_y > 0$. Otherwise, there would exist an x, y, l -separation $E_x[E_0[t_y]]$ with $E_x = C_1[C_2[C_3[C_4]]]$, and $\text{ht}(E_0) < \text{ht}(D_0)$. Assume $|\text{yd}(C_3)|_x > 0$. Lemma 11, point 2, ensures that $|\text{left}(C_3)|_{x,y} = 0$ or $|\text{right}(C_3)|_{x,y} = 0$. Assume $|\text{right}(C_3)|_{x,y} = 0$ (the other case is symmetric). We then have both a x and a y generated on the left of C_3 . Since neither $C_1[C_2$

nor t_5 generate any y -token, projectivity imposes $r_{\bar{y}}^t = r_{\bar{y}}^{t'} + n_{\bar{y}}^{C_2[C_4]}$ and we can conclude as in the previous case. The only remaining subcase is when $|\text{yd}(C_3)|_x = 0$, in which case t is 0-separated, and considering the (minimal) 0-separation $(C_1, X, C_2[C_3[C_4]])$ we can use the same argument as in the case where t_0 encompasses all nodes of C_2 and C_4 .

Case 2 Consider now the remaining case where Lemma 12 applies. If neither C_2 or C_4 generate some x or y -token, they don't generate \bar{x} or \bar{y} -tokens either, and the same reasoning as Case 1 subcase i) applies. Otherwise $C_2[C_4]$ generate at least some x -token. We then have $n_x^t \geq n_x^{t'} + 1$. Since t_0 contains all nodes from t'_0 we further have $e_{\bar{y}}^{t_0} \geq e_{\bar{y}}^{t'_0}$. Finally $n_{\bar{y}}^t \leq n_{\bar{y}}^{t'} + l_0$. We conclude using inequation 3. \square

B.6 Conclusion

Lemma 8. *For any $t \in \mathbf{T}$, if t is x, y, l_0 -separated then t is x, y, l_0 -asynchronous.*

Proof. By Lemma 7-2, there is a minimal x, y, l_0 -separation $t = D_x[D_0[t_y]]$ such that the y bound obtains for $t_0 = D_0[t_y]$. By lemma 7-1 the x bound obtains for t_0 as well. Observe finally, that $r_{\bar{y}}^t \leq m_{\bar{y}}^t$ and since t_0 generates at most l_0 x -tokens, by projectivity and definition of CSD, it generates at most $(l_0 + 1)m_{\bar{x}}^t$ \bar{x} -tokens (one sequence of $m_{\bar{x}}^t$ between each occurrence of x and the next, plus possibly one in front of the first and one after the last). Hence,

$$\begin{aligned} e_{\bar{y}}^{t_0} &\geq n_{\bar{y}}^t - n_x^t l_0 - m_{\bar{y}}^t \\ e_{\bar{x}}^{t_0} &\leq n_x^t l + m_{\bar{x}}^t (l_0 + 1). \end{aligned}$$

and t is x, y, l_0 -asynchronous. \square

Lemma 9. *For any $t \in \mathbf{T}$, t is x, y, l_0 -separated for some $x/y \in \text{Sep}$.*

Proof. The proof proceeds by induction on the height of t .

If $\text{ht}(t) \leq p$. Then $|\text{yd}(t)|_z \leq l_0$ for any $z \in \{a, b, c, d\}$, hence t is trivially x, y, l_0 -separated for some $x/y \in \text{Sep}$.

If $h(t) > p$, Lemma 1 yields a decomposition $t = C_1[C_2[C_3[C_4[t_5]]]]$, where $t' = C_1[C_3[t_5]] \in \mathbf{T}$, $\text{ht}(t') < \text{ht}(t)$ and $\text{ht}(C_2[C_4]) \leq p$. By induction t' is x, y, l_0 -separated for some $x/y \in \text{Sep}$. For sake of succinctness, let us present the inductive step for $x/y = a/c$, the reasoning for other cases is analogous. Let us examine the different possible configurations of C_2 and C_4 .

Case 1 If Lemma 11 applies *i.e.* C_2 and C_4 generate only one kind of bar token, \bar{z} , and brackets, one checks easily that inserting C_2 and C_4 does not change the distribution of a and c -tokens in the tree, hence t is a, c, l_0 -separated.

Case 2 If Lemma 12 applies, note first that if C_2 and C_4 generate no a or c -token, we can conclude as in Case 1 as the distribution of a and c -tokens in the tree is not changed either. Otherwise, we assume that C_2 or C_4 generate some a or c -token and distinguish between subcases 1-3 of Lemma 12:

Subcase 1 in this case for some $i \in \{2, 4\}$ $\text{left}(C_i)$ contains an a -token and no b, c or d -token while $\text{right}(C_i)$ contains some c -token and no a, b or d -token. Assume $i = 2$, the case where $i = 4$ is similar. By projectivity and definition of CSD_s follows that all b -tokens are generated in $C_3[C_4[t_5]]$ and all c -tokens in C_1 . t is therefore $b, d, 0$ -separated, hence b, d, l_0 -separated.

Subcase 2 in this case, $\text{left}(C_2)$ contains some a -token and no b, c, d -token, $\text{right}(C_3)$ contains some d -token and no a, c, d -token, $\text{left}(C_4)$ contains some b -token and no a, c, d -token, $\text{right}(C_4)$ contains some c -token and no a, b, d -token. It follows that t_5 generate no occurrence of a and C_1 no occurrence of c . Since $|\text{yd}(C_2[C_3[C_4]])|_a \leq l_0$, $(C_1, C_2[C_3[C_4]], t_5)$ is an a, c, l_0 -separation.

Subcase 3 Assume $\text{left}(C_2)$ contains some a -token and no b, c, d -token and that $\text{left}(C_4)$ contains some c -token. It follows that all b -tokens are generated by C_3 . So $\text{yd}(t)$ contains less than l_0 b -tokens, by definition of CSD it also contains less than l_0 d -tokens, so $(C_1, C_2[C_3[C_4]], t_5)$ is a b, d, l_0 -separation.

Assume now $\text{left}(C_2)$ contains some a -token and no b, c, d -token and that $\text{right}(C_4)$ contains some c -token. It follows that t_5 generate no d -token and C_1 generate no b -token. Hence $(C_1, C_2[C_3[C_4]], t_5)$ is b, d, l_0 -separation.

The remaining cases are symmetric exchanging c with a, d with b , and left with right everywhere. \square