Scattered Context Grammar

Petr Horáček, Eva Zámečníková and Ivana Burgetová

Department of Information Systems
Faculty of Information Technology
Brno University of Technology
Božetěchova 2, 612 00 Brno, CZ

FRVŠ MŠMT FR97/2011/G1

Outline

- **Introduction**
- **Transformational Scattered Context Grammars**
- **Scattered Context in English Syntax**

- **Introduction**
- **Transformational Scattered Context Grammars**
- **Scattered Context in English Syntax**

Definition

A scattered context grammar (SCG) G is a quadruple $G = (N, T, P, S)$, where

- N is a finite set of *nonterminals*,
- T is a finite set of *terminals*, $N \cap T = \emptyset$
- P is a finite set of *rules* of the form

$$(A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n),$$

  where $A_1, \dots, A_n \in N$, $x_1, \dots, x_n \in (N \cup T)^*$,
- $S \in N$ is the *start symbol*.

Derivation step

Let $G = (N, T, P, S)$ be an SCG. For $u, v \in (N \cup T)^*$, $p \in P$ we define $u \Rightarrow v\,[p]$, if there is a factorization of $u = u_1 A_1 \dots u_n A_n u_{n+1}$, $v = u_1 x_1 \dots u_n x_n u_{n+1}$ and $p = (A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n)$, where $u_i \in (N \cup T)^*$ for $1 \leq i \leq n$.

Motivation

- Many common English sentences contain expressions and words mutually depending on each other, although they are not adjacent to each other in the sentence.

Example

He usually goes to work early.

- The subject (*he*) and the predicator (*goes*) are related.

☹ *He usually go to work early.*
☹ *I usually goes to work early.*

- Ungrammatical sentences – the form of the predicator depends on the form of the subject.
    - *he. . . go*, *I. . . goes* – illegal combinations

Motivation

- Consider the scattered context rule:

$$(\text{He}, \text{goes}) \rightarrow (\text{We}, \text{go})$$

- This rule checks if the subject is the pronoun *he* and if the verb *go* is in 3rd person singular.
- If the sentence satisfies this property, it can be transformed.

Example

He usually goes to work early.
$\Rightarrow$ *We usually go to work early.*

- The related words may occur far away from each other.

Example

He almost regularly goes to work early.
$\Rightarrow$ *We almost regularly go to work early.*

He usually, but not always, goes to work early.
$\Rightarrow$ *We usually, but not always, go to work early.*

Classification of verbs

1. Auxiliary verbs
   - Modal verbs: *can*, *may*, *must*, *will*, *shall*, *ought*, *need*, *dare*
   - Non-modal verbs: *be*, *have*, *do*
2. Lexical verbs

- In reality, these classes may overlap.
  - For example, *do* appears as auxiliary verb in some sentences, as lexical verb in other sentences.
- Inflectional forms of verbs are called paradigms.

Form	Paradigm	Person	Example
Primary	Present	3rd sg	*She walks home.*
		Other	*They walk home.*
	Preterite		*She walked home.*
Secondary	Plain form		*They should walk home.*
	Gerund-participle		*She is walking home.*
	Past participle		*She has walked home.*

Paradigms of the Verb *be*

- The only exception in English: *be*
  - 9 paradigms in its neutral form.
  - All primary forms have their negative contracted counterparts.
  - *Irrealis* paradigm – in sentences of unrealistic nature.

I wish I were rich.

Form	Paradigm	Person	Neutral	Negative
Primary	Present	1st sg	*am*	*aren't*
		3rd sg	*is*	*isn't*
		Other	*are*	*aren't*
	Preterite	1st sg, 3rd sg	*was*	*wasn't*
		Other	*were*	*weren't*
	Irrealis	1st sg, 3rd sg	*were*	*weren't*
Secondary	Plain form		*be*	—
	Gerund-participle		*being*	—
	Past participle		*been*	—

Personal Pronouns

- Great amount of inflectional variation

Non-reflexive				Reflexive
Nominative	Accusative	Genitive		
Plain		Dependent	Independent	
I	me	my	mine	myself
you	you	your	yours	yourself
he	him	his	his	himself
she	her	her	hers	herself
it	it	its	its	itself
we	us	our	ours	ourselves
you	you	your	yours	yourselves
they	them	their	theirs	themselves

Topic

- Introduction
- **Transformational Scattered Context Grammars**
- Scattered Context in English Syntax

Definition

A transformational scattered context grammar G is a quadruple $G = (N, T, P, I)$, where

- N is a finite set of *nonterminals*,
- T is a finite set of *terminals*, called the *output vocabulary*, $N \cap T = \emptyset$
- P is a finite set of *rules* of the form

$$(A_1, \dots, A_n) \rightarrow (x_1, \dots, x_n),$$

  where $A_1, \dots, A_n \in N$, $x_1, \dots, x_n \in (N \cup T)^*$,
- $I \subseteq N \cup T$ is the *input vocabulary*.

Transformation

Let $G = (N, T, P, S)$ be a transformational SCG. The transformation T that G defines from $K \subseteq I^*$ is defined as:

$$T(G, K) = \{(x, y) \colon x \Rightarrow_G^* y, x \in K, y \in T^*\}$$

Transformational SCG – Example

Define the transformational SCG $G = (N, T, P, I)$, where
$N = \{A, B, C\}$, $T = \{a, b, c\}$, $I = \{A, B, C\}$ and
$P = \{(A, B, C) \to (a, bb, c)\}$

Example

$AABBCC \Rightarrow_G aABbbcC \Rightarrow_G aabbbbcc$

$(AABBCC, aabbbbcc) \in T(G, I^*)$

- If we restrict the input sentences to the language

$$L = \{A^n B^n C^n \colon n \geq 1\},$$

we get

$$T(G, L) = \{(A^n B^n C^n, a^n b^{2n} c^n) \colon n \geq 1\}$$

Topic

- Introduction
- Transformational Scattered Context Grammars
- **Scattered Context in English Syntax**

Notations

- T – the set of all English words including all their inflectional forms
- $T_V \subset T$ – the set of all verbs including all their inflectional forms
- $T_{VA} \subset T_V$ – the set of all auxiliary verbs including all their inflectional forms
- $T_{Vpl} \subset T_V$ – the set of all verbs in plain form
- $T_{PPn} \subset T$ – the set of personal pronouns in nominative

Verb paradigms:

- $\pi_{3rd}(v)$ – the verb v in 3rd person singular present
- $\pi_{pres}(v)$ – the verb v in present (other than 3rd person singular)
- $\pi_{pret}(v)$ – the verb v in preterite

- We assume here that the set of all English words T is finite and fixed.

- We want to negate the clause.

Example

Neither Thomas nor his wife went to the party.
$\Rightarrow$ *Both Thomas and his wife went to the party.*

Set $G = (N, T, P, I)$, where $N = I = \{\langle x \rangle : x \in T\}$ and P is defined as:

$$
\begin{aligned}
P &= \{(\langle \text{neither} \rangle, \langle \text{nor} \rangle) \rightarrow (\text{both}, \text{and})\} \\
&\cup \{(\langle x \rangle) \rightarrow (x) : x \in T - \{\text{neither}, \text{nor}\}\}
\end{aligned}
$$

Example

$\langle \text{neither} \rangle \langle \text{thomas} \rangle \langle \text{nor} \rangle \langle \text{his} \rangle \langle \text{wife} \rangle \langle \text{went} \rangle \langle \text{to} \rangle \langle \text{the} \rangle \langle \text{party} \rangle$
$\Rightarrow_G$ both $\langle \text{thomas} \rangle$ and $\langle \text{his} \rangle \langle \text{wife} \rangle \langle \text{went} \rangle \langle \text{to} \rangle \langle \text{the} \rangle \langle \text{party} \rangle$
$\Rightarrow_G$ both thomas and $\langle \text{his} \rangle \langle \text{wife} \rangle \langle \text{went} \rangle \langle \text{to} \rangle \langle \text{the} \rangle \langle \text{party} \rangle$
$\Rightarrow_G$ both thomas and his $\langle \text{wife} \rangle \langle \text{went} \rangle \langle \text{to} \rangle \langle \text{the} \rangle \langle \text{party} \rangle$
$\Rightarrow_G^5$ both thomas and his wife went to the party

- Existential clause = clause that indicates an existence.
- Usually formed using the dummy subject *there*.
- In some cases, however, the dummy subject is not mandatory.

Example

A nurse was present.
$\Rightarrow$ *There was a nurse present.*

Set $G = (N, T, P, I)$, where $N = I = \{\langle x \rangle : x \in T\} \cup \{X\}$ (X is a new symbol such that $X \notin T \cup I$) and P is defined as:

$$
\begin{aligned}
P \quad = \quad & \{(\langle x \rangle, \langle \text{is} \rangle) \rightarrow (\text{there is } x\, X, \varepsilon), \\
& (\langle x \rangle, \langle \text{are} \rangle) \rightarrow (\text{there are } x\, X, \varepsilon), \\
& (\langle x \rangle, \langle \text{was} \rangle) \rightarrow (\text{there was } x\, X, \varepsilon), \\
& (\langle x \rangle, \langle \text{were} \rangle) \rightarrow (\text{there were } x\, X, \varepsilon) : x \in T\} \\
\cup \quad & \{(X, \langle x \rangle) \rightarrow (X, x) : x \in T\} \\
\cup \quad & \{(X) \rightarrow (\varepsilon)\}
\end{aligned}
$$

$$
\begin{aligned}
P \;=\;& \{(\langle x\rangle, \langle \text{is}\rangle) \rightarrow (\text{there is } x\, X, \varepsilon),\\
& (\langle x\rangle, \langle \text{are}\rangle) \rightarrow (\text{there are } x\, X, \varepsilon),\\
& (\langle x\rangle, \langle \text{was}\rangle) \rightarrow (\text{there was } x\, X, \varepsilon),\\
& (\langle x\rangle, \langle \text{were}\rangle) \rightarrow (\text{there were } x\, X, \varepsilon) : x \in T\}\\
\cup\;& \{(X, \langle x\rangle) \rightarrow (X, x) : x \in T\}\\
\cup\;& \{(X) \rightarrow (\varepsilon)\}
\end{aligned}
$$

Example

$\langle$a$\rangle\langle$nurse$\rangle\langle$was$\rangle\langle$present$\rangle$
$\Rightarrow_G$ there was a X $\langle$nurse$\rangle\langle$present$\rangle$
$\Rightarrow_G$ there was a X nurse $\langle$present$\rangle$
$\Rightarrow_G$ there was a X nurse present
$\Rightarrow_G$ there was a nurse present

- Two ways of transforming declarative clauses into interrogative depending on the predicator.

1. Predicator is auxiliary verb – simply swap the subject and the predicator.

Example

He is mowing the lawn.
$\Rightarrow$ *Is he mowing the lawn?*

2. Predicator is lexical verb – add the dummy *do* (in the correct form) to the beginning of the clause.

Example

She usually gets up early.
$\Rightarrow$ *Does she usually get up early?*

$$
\begin{aligned}
P \quad = \quad & \{(\langle p \rangle, \langle v \rangle) \rightarrow (vp, X) \colon v \in T_{VA}, p \in T_{PPn}\} \\
\cup \quad & \{(\langle p \rangle, \langle \pi_{pret}(v) \rangle) \rightarrow (\text{did } p, vX), \\
& \ (\langle p \rangle, \langle \pi_{3rd}(v) \rangle) \rightarrow (\text{does } p, vX), \\
& \ (\langle p \rangle, \langle \pi_{pres}(v) \rangle) \rightarrow (\text{do } p, vX) \colon v \in T_{Vpl} - T_{VA}, p \in T_{PPn}\} \\
\cup \quad & \{(\langle x \rangle, X) \rightarrow (x, X), \\
& \ (X, \langle y \rangle) \rightarrow (X, y) \colon x \in T - T_V, y \in T\} \\
\cup \quad & \{(X) \rightarrow (\varepsilon)\}
\end{aligned}
$$

Example

⟨he⟩⟨is⟩⟨mowing⟩⟨the⟩⟨lawn⟩
$\Rightarrow_G$ is he X ⟨mowing⟩⟨the⟩⟨lawn⟩
$\Rightarrow_G$ is he X mowing ⟨the⟩⟨lawn⟩
$\Rightarrow_G$ is he X mowing the ⟨lawn⟩
$\Rightarrow_G$ is he X mowing the lawn
$\Rightarrow_G$ is he mowing the lawn

$$
\begin{aligned}
P \;=\; & \{(\langle p\rangle, \langle v\rangle) \rightarrow (vp, X) \colon v \in T_{VA}, p \in T_{PPn}\} \\
\cup\; & \{(\langle p\rangle, \langle \pi_{pret}(v)\rangle) \rightarrow (\text{did } p, vX), \\
& (\langle p\rangle, \langle \pi_{3rd}(v)\rangle) \rightarrow (\text{does } p, vX), \\
& (\langle p\rangle, \langle \pi_{pres}(v)\rangle) \rightarrow (\text{do } p, vX) \colon v \in T_{Vpl} - T_{VA}, p \in T_{PPn}\} \\
\cup\; & \{(\langle x\rangle, X) \rightarrow (x, X), \\
& (X, \langle y\rangle) \rightarrow (X, y) \colon x \in T - T_V, y \in T\} \\
\cup\; & \{(X) \rightarrow (\varepsilon)\}
\end{aligned}
$$

Example

⟨she⟩⟨usually⟩⟨gets⟩⟨up⟩⟨early⟩

$\Rightarrow_G$ does she ⟨usually⟩ get X ⟨up⟩⟨early⟩

$\Rightarrow_G$ does she usually get X ⟨up⟩⟨early⟩

$\Rightarrow_G$ does she usually get X up ⟨early⟩

$\Rightarrow_G$ does she usually get X up early

$\Rightarrow_G$ does she usually get up early

Generation of Grammatical Sentences

- So far, we have assumed that the set of English words is finite.
  - Reasonable assumption in practice – we all commonly use a finite and fixed vocabulary in everyday English.
- From theoretical point of view, the set of all well-formed English words is infinite.

Example

Your grandparents are all your grandfathers and all your grandmothers.

Your great-grandparents are all your great-grandfathers and all your great-grandmothers.

Your great-great-grandparents are all your great-great-grandfathers and all your great-great-grandmothers.

⋮

$L = \{\text{your } \{\text{great-}\}^i\text{grandparents are all your } \{\text{great}\}^i\text{grandfathers and all your } \{\text{great}\}^i\text{grandmothers} : i \geq 0\}$

Introduce the SCG $G = (N, T, P, S)$, where $T = \{\text{all}, \text{and}, \text{are}, \text{grandfathers}, \text{grandmothers}, \text{grandparents}, \text{great-}, \text{your}\}$, $N = \{S, \#\}$, and P consists of these three productions:

$$
\begin{array}{rcl}
(S) & \rightarrow & (\text{your } \#\text{grandparents are all your } \#\text{grandfathers} \\
 & & \quad \text{and all your } \#\text{grandmothers}), \\
(\#, \#, \#) & \rightarrow & (\#\text{great-}, \#\text{great-}, \#\text{great-}), \\
(\#, \#, \#) & \rightarrow & (\varepsilon, \varepsilon, \varepsilon)
\end{array}
$$

Example

$S \Rightarrow_G$ your #grandparents are all your #grandfathers and all your #grandmothers

$\Rightarrow_G$ your #great-grandparents are all your #great-grandfathers and all your #great-grandmothers

$\Rightarrow_G$ your great-grandparents are all your great-grandfathers and all your great-grandmothers

References

- Sheila A. Greibach, John E. Hopcroft:
  Scattered Context Grammars,
  Journal of Computer and System Sciences, 3:233-247, 1969
- Alexander Meduna, Jiří Techet:
  Scattered Context Grammars and their Applications,
  WIT Press, 2010