

SYNTACTIC SIMPLIFICATION

SYNTACTIC COMPLEXITY in natural language

This project processes syntactic complexity introduced by:

•Coordination

John [went to the shop **and** bought a book].

•Subordination

[John, **from London**, visited his friends.]

Prevalent signs of syntactic complexity

Introducing reported speech

Introducing more information about the people mentioned

Lists of alternate symptoms and treatment options

Prevalent functions of these signs

Left boundary of a subordinate clause

Right boundary of a subordinate noun phrase

News Health

ESMN
ESMAdvP
CLN

SSEV
CMV1
CMN1

SSCM

Coordinating clauses

News : who
or
and that
, and
;
Literature

AUTOMATIC PROCESSING OF SYNTACTIC COMPLEXITY

Simplification pipeline

•Tokenisation

Vance Yeboah is believed to have fled
the country and has yet to be tried

•Part of speech tagging

Vance Yeboah is believed to have fled
the country and has yet to be tried

•Sign detection and classification

Vance Yeboah is believed to have fled
the country and has yet to be tried

•Sentence rewriting algorithm

Vance Yeboah is believed to have fled the country
Vance Yeboah has yet to be tried

Statistical/machine learning approaches will be investigated

- Identifies subordinate constituents and conjoins of coordinated constituents
- Current approach is rule-based

Sentence rewriting algorithm

```

Input: Sentence containing at least one sign of syntactic complexity,  $s_0$ 
Output: Simplified sentence,  $A$ 
1  $A \leftarrow \emptyset;$ 
2  $S \leftarrow \{s_0\};$ 
3 while  $S \neq \emptyset$  do
4    $s_i \leftarrow \text{pop}(S);$ 
5   if  $s_i$  contains a sign of syntactic complexity then
6      $\{s_{i1}, s_{i2}\} \leftarrow \text{simplify}(s_i);$ 
7      $S \leftarrow S \cup \{s_{i1}, s_{i2}\};$ 
8   else
9      $A \leftarrow A \cup \{s_i\}$ 
10 end
11 end

```

Example rules:

$\bullet B [W_{(VBD|VBP)} C] s_{CMV1} [W_{RB} * W_V * D]$
The judge accepted that the chemists had exemplary characters and were each at the threshold of professional careers.

$\bullet B [W_{(VBD|VBP)} C]$
The judge accepted that the chemists had exemplary characters

$\bullet B [W_{RB} * W_V * D]$
The judge accepted that the chemists were each at the threshold of professional careers.

Extrinsic evaluation

A preliminary version of the syntactic simplification process was integrated into an information extraction system (Evans, 2011)

Template slot	IGNORE	MAJORITY	PATTERNS	STANFORD	MBL	HYBRID	KEY
finding	0.5845	0.7584	0.5556	0.7971	0.7971	0.8019	0.8696
technique	0.7729	0.7681	0.7778	0.7778	0.7778	0.7874	0.8019
system	0.7536	0.8357	0.7391	0.8309	0.8406	0.8454	0.8744
qualifier	0.6812	0.8164	0.5797	0.7971	0.8261	0.8213	0.8261
location	0.8696	0.8889	0.8985	0.8985	0.9034	0.9082	0.9324
ALL	0.7324	0.8135	0.7101	0.8203	0.8200	0.8328	0.8609

Template slot	IGNORE	MAJORITY	PATTERNS	STANFORD	MBL	HYBRID	KEY
finding	0.9420	0.8068	0.8454	0.8744	0.8551	0.8696	0.9275
technique	0.7729	0.7681	0.8116	0.7778	0.7778	0.7874	0.8019
system	0.7536	0.8357	0.8406	0.8309	0.8406	0.8454	0.8744
qualifier	0.6811	0.8164	0.6135	0.7971	0.8261	0.8213	0.8261
location	0.8696	0.8889	0.9130	0.8985	0.9034	0.9082	0.9324
ALL	0.8039	0.8232	0.8048	0.8357	0.8406	0.8464	0.8725

EVALUATION ISSUES

Intrinsic evaluation

- MBL method to predict class labels of signs:
- Accuracy of **68.7%**
- Automatic simplification of noun post-modifiers:
- F1 scores of **24.32% - 59.02%**

Challenges

- Automatic intrinsic evaluation requires the development of manually simplified documents. This involves:
 - recognition and preservation of important information
 - recognition of relevant accessibility issues
- User-focused evaluation may exploit:
 - Reading comprehension testing**
 - Tasks such as manual summarisation, MCQ and short answer tests, and document titling
 - Instrumental methods**
 - Eye tracking and F-MRI
- Automatic syntactic simplification may remove one type of complexity but introduce another (Arya et al., 2011)
 - This motivates exploration of methods to re-establish temporal relations, discourse links, and inference chains broken by the simplification process