

# Annotating Signs of Syntactic Complexity to Support Sentence Simplification

Richard Evans (R.J.Evans@wlv.ac.uk) and Constantin Orasan (C.Orasan@wlv.ac.uk)

## Syntactic complexity

### Key aspects of syntactic complexity:

#### Coordination

She knew the risks and still insisted the operation should go ahead, Dr Addicott said.

#### Subordination

McKay, of Wark, Northumberland, denies five charges of contaminating food.

### Signs of syntactic complexity:

#### Coordinators (conjunctions and punctuation marks)

#### Subordination boundaries (complementisers, wh-words, and punctuation marks)

### Annotation scheme encodes:

#### linking functions of coordinators

#### bounding functions of subordination boundaries

#### information about false signs that do not have a coordinating or bounding function

## Research context: the FIRST project

Aims to improve accessibility of text for readers with autistic spectrum disorders (ASD). This process includes **syntactic simplification**.

**Syntactic simplification** can be framed as the conversion of long syntactically complex sentences into sequences of shorter sentences that convey the same meaning.

#### It can improve:

- the reliability of NLP applications (IE, MT, and syntactic parsing)
- the accessibility of written texts for people with reading difficulties

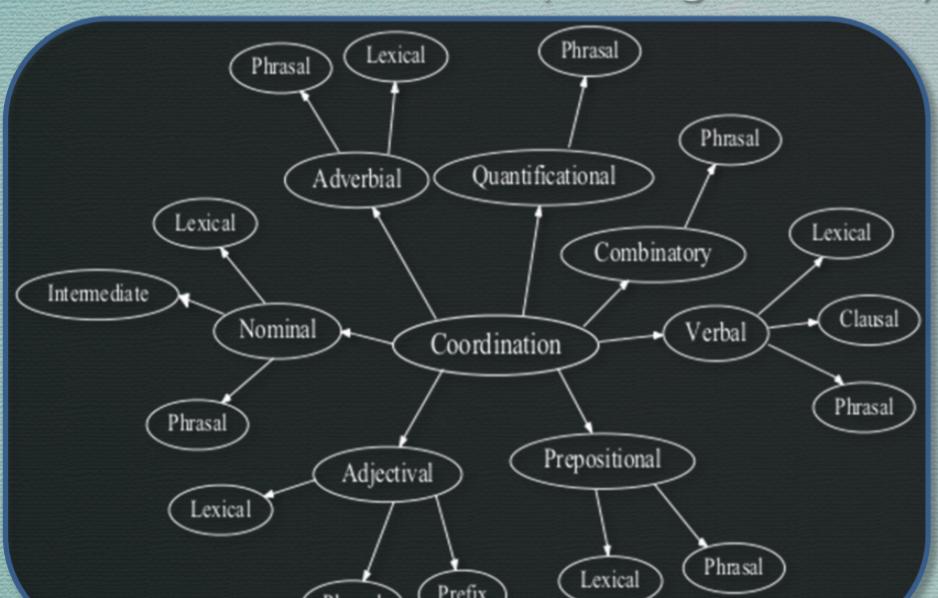
We are developing a system for syntactic simplification which operates in two steps:

- Identifying and classifying signs of syntactic complexity
- Rewriting sentences using manually crafted rules

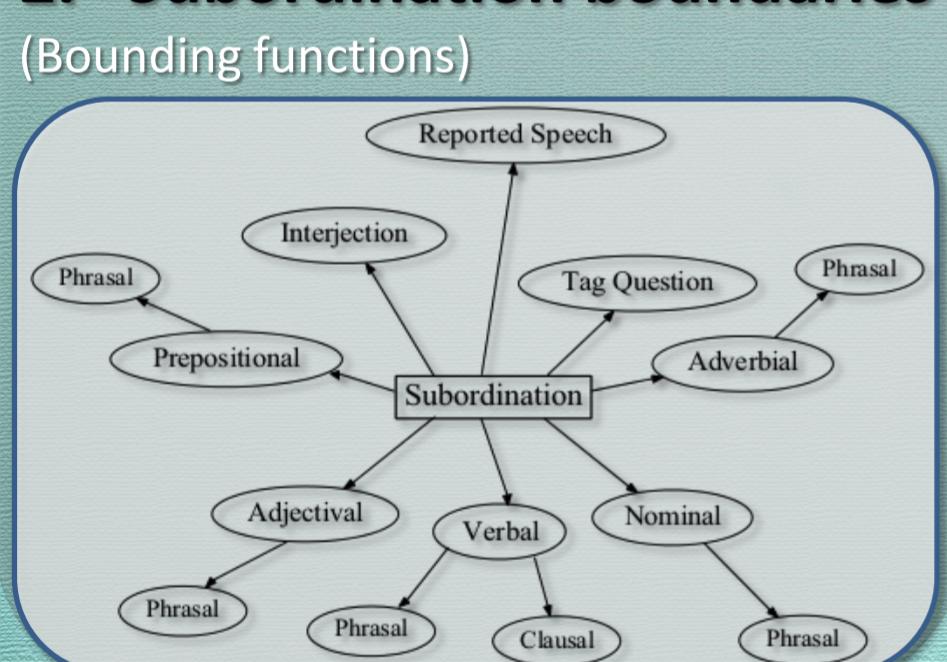
The scheme also includes class labels for **false signs** and **signs bounding speech-related elements**

## Signs of syntactic complexity

### 1. Coordinators (Linking functions)



### 2. Subordination boundaries (Bounding functions)



## Coordinators

### Class labels

{C|ES|SS} {P|L|I|M|E} {V|N|A|Adv|P|Q} {0-4}

Function  
(coordinating two conjoins or bounding subordinated constituents)

Projection level  
(lexical, intermediate maximal, extended)

### Examples

#### CLA

"He had a stable [and] loving family."

#### CMP

"But the melancholy experience of the courtrooms [and] of life is that people have a good character in some respects and not in others."

## Subordination boundaries

### Class labels

Grammatical category  
(nominal, verbal, etc.)

Optional key  
distinguishing between coordination of different types of nominal and verbal maximal projection

### Example

#### ESMV

"Being put into a psychiatric ward with people with long-term mental illnesses who are shaking with the drugs they are taking, there's no way you can feel normal and be OK with yourself," she told BBC TV's That's Esther programme with Esther Rantzen.

## Advantages of the scheme

Fine grained annotation enables flexibility in the design of syntactic simplification algorithms

Penn Treebank (Marcus et al., 1993):

- Does not directly encode specific linking/bounding functions of conjunctions and commas.
- Expensive to produce

Van Delden and Gomez (2002)

- Focus only on classification of commas
- Evans (2011)

Encodes restricted range of phenomena, focusing on coordination

Maier et al. (2012)

- Requires existing syntactic annotation

## Corpus analysis 1

### Sign and class distribution

Collection	Genre	#Docs	#Sents	#Words	#Signs
1. METER corpus	News	674	22,858	295,718	30,459 12,796
2. www.patient.co.uk	Healthcare	752	79,684	1,174,460	97,244 8,987
3. Gutenberg	Literature	24	4,468	87,712	11,031 11,031

