Lexicon Embedded Syntax

Alain Polguère
ATILF UMR 7118, CNRS–Université de Lorraine
alain.polguere@univ-lorraine.fr

Depling 2015, Uppsala, August 24 2015
1

Lexicon vs. Grammar
A) Blind-eye attitude

- Widespread trend to equate the rule-system of a language with its grammar
  - Names of contemporary linguistic theories are generally denoting types of “Grammars”, with a capital G
  - But attitude that goes far back – cf. INALCO’s creation decree of 1795

- Lexicon seen as only a side-kick to grammar, and it is basically ignored
B) Continuum fallacy

“Classical” view of the lexicon ~ grammar dichotomy
– cf. Jepersen (1924)

- Lexicon: repository of words and word-related facts
- Grammar: repository of general facts, that are not relevant to specific words

But presence of intermediate elements

Hence, presupposition that there is no true distinction between lexical and grammatical information, no true distinction between modeling lexicons and grammars
Hot-cold metaphor
“It is one thing to establish a graded rather than absolute boundary, and quite another to conclude that it means the absence of that boundary. To take this tack is to commit the continuum fallacy, which involves arguing that if two extremes are connected by small intermediate differences and if at no step can one indicate a decisive difference, then the extremes are the same. To use an analogy, inability to specify at what temperature cold turns to hot should not lead to the conclusion that cold is really the same as hot.”
(Szcześniak, 2015: 78–79)
C) Temperate modular view

- Both lexicon and grammar exist
- They are truly distinct
- They share many elements
- They are functionally interwound
- Lexicon comes first
Lexical systems

- Formal lexical models presented in Polguère (2014), together with the corresponding lexicography

- Four properties
  1. Oriented graphs
     - Nodes: mainly, lexical units – i.e. lexemes and idioms
     - Edges: mainly, lexical function relations (Mel’čuk, 1996)
  2. Nodes are non-atomic: equivalent to lexicographic articles
  3. Possess topological properties of small-world networks
  4. Include measures of confidence
fr- and en-Lexical Networks

- Lexicography of virtual dictionaries – **Dicet** editor for lexical systems
- Two significant models (Gader et al., 2014)
  - fr-LN – manually produced
  - en-LN – compiled from the Princeton WordNet
- Other languages presently explored: es, kr, ru, ar
- What follows mainly concerns the fr-LN
2

Syntax of Idioms
Idioms in lexical systems

- Syntactic “prefabs”, with potential superficial variations – lexical content, morphology, linearization
- In the fr-LN, description of the **lexico-syntactic structure** of each idiom
SUCRER LES FRAISES
lit. `to sugar strawberries'
'to tremble because of old age'

locution verbale

SUCRER, FRAISE¹, LE Art

\[ SUCRER(v) \]
\[ \text{direct object} \]
\[ \text{FRAISE } \]¹⁺¹\( \text{(N)} \)
\[ \text{determinative} \]
\[ \text{LE(Art)} \]

\[ \alpha_1(v) \]
\[ \text{direct object} \]
\[ \alpha_2(\text{N}) \]
\[ \text{determinative} \]
\[ \alpha_3(\text{Art}) \]

V Art NC
On-going work

- Set of Surface-Syntactic dependencies
- Incorporate constraints on grammemes
- Connection between idioms’ lexico-syntactic structures (idiom template + individual properties of embedded lexical units) and behavior in texts
- Extraction of generalizations
3
Syntax of Collocations
Functional notion of collocation

- Semi-phraseological phrase such as *to run a fever*
- Compositional ⇒ it is not an idiom
- Made up of two elements
  - **Base** of the collocation – cf. *fever*
  - **Collocate** of the base – cf. *to run*
- A collocation is a combinatorial property of its base – it is a lexico-syntactic structure that belongs to the lexicon
Standard lexical functions

- Generalizations of lexical links
  - Paradigmatic (semantic) or syntagmatic (combinatorial)
  - Recurrent and universally present in all natural languages
  - Often expressed by morphological means

- Modeled as functions that apply to lexical units and return set of values, for given semantico-syntactic contents
Intensifying collocate: Magn

- Magn(fever) = high < raging
- Magn(headache) = bad, severe < terrible, violent < pounding, splitting

Application of a syntagmatic lexical function can return a value corresponding to a paradigmatic link: fusion

- Magn(rain_v) = hard, heavily, // pour down

System of lexical functions is implemented in lexical systems, such as the fr-LN, as part of their metalinguistic components
Standard syntagmatic lexical functions as grammar rules

- Illustration with $\text{Real}_1$
  - Meaning of $\text{Real}_1(L)$: ‘to realize L’, ‘to do what is supposed to be done as regards to L’
  - Syntax of $\text{Real}_1(L)$: takes L as first complement – and the first deep-syntactic actant of L as grammatical subject

- Cf. $\text{BALLOON}_N 2$ in the en-LN
Corresponding Deep-Syntactic structures

“Normal” application

```
Real_1(L)
```

```
X
```

```
L
```

```
I
```

```
II
```

“Fused” application

```
// Real_1(L)
```

```
X
```

```
I
```

```

X
```
Possible Surface derivations from lexicographic data

FLY\textsubscript{V} 3

subjectal  \quad direct objectal

X  \quad BALLOON\textsubscript{N} 2

\downarrow \quad \text{determinative}

\quad \alpha

-----

PILOT\textsubscript{V} 1

subjectal  \quad direct objectal

X  \quad BALLOON\textsubscript{N} 2

\downarrow \quad \text{determinative}

\quad \alpha

-----

BALLOON\textsubscript{V} 1

subjectal

X
To conclude: exploitation of lexicon embedded syntax

- Idioms and collocations are omnipresent in texts
  ⇒ their syntax represents a non-trivial part of the syntax that has to be dealt with in corpora
- “Lexicalized grammars” the other way round
References


