Multiword Expressions at the Grammar–Lexicon Interface

Timothy Baldwin
Talk Outline

1. Introduction

2. MWEs at the Grammar–Lexicon Interface

3. A Case Study: English Determinerless PPs
   - Problem Statement
   - Analyses

4. Summary
What are Multiword Expressions (MWEs)?

**Definition:** A **multiword expression** ("MWE") is:

1. decomposable into multiple simplex words
2. lexically, phonetically, morphosyntactically, semantically, and/or pragmatically idiosyncratic

Adapted from Baldwin and Kim [2010]
Some Examples

- Mt Rokko, ad hoc, by and large, Toy Story, kick the bucket, part of speech, in step, Hanshin Tigers, trip the light fantastic, telephone box, call (someone) up, take a walk, do a number on (someone), take advantage (of), pull strings, kindle excitement, fresh air, ....
Lexical Idiomaticity

- Lexical idiomaticity = one or more of the elements of the MWE does not have a usage outside of MWEs
- Examples of lexical idiomaticity:
  
  \(ad\) hominem, \(bok\) choy, \(a\) la \(mode\), \(to\) and \(fro\)

- Complications of lexical idiomaticity:
  
  ▶ cross-linguistic effects, e.g. \(ad\) is unmarked in Latin
  ▶ simple lexical occurrence outside of MWEs not sufficient, e.g. \(a\) la \(mode\)

Source(s): Bauer [1983], Trawiński et al. [2008]
Morphosyntactic Idiomaticity

- Morphosyntactic idiomaticity = the morphosyntax of the MWE differs from that of its components.
- Examples of morphosyntactic idiomaticity:
  - *cat’s cradle, yin hry “evil eye”*

- Examples of syntactic idiomaticity:

  ![Diagram of syntactic structure]

  - AdvP: `IN CC JJ` (by and large)
  - V[trans]: `VB[trans] CC VB[trans]`
    - `by` and `large`
    - `wine` and `dine`

Source(s): Katz and Postal [2004], Chafe [1968], Bauer [1983], Sag et al. [2002]
Talk Outline

1. Introduction

2. MWEs at the Grammar–Lexicon Interface

3. A Case Study: English Determinerless PPs
   - Problem Statement
   - Analyses

4. Summary
MWEs and Productivity

- MWEs populate a spectrum of productivity, from completely unproductive, lexicalised MWEs such as *by and large* to highly productive (but still semi-lexicalised) MWEs such as English verb–particle constructions.
- The first question in terms of the grammar–lexicon interface when it comes to MWEs, is the degree of productivity of a given MWE, to determine whether it is possible to enumerate all of the instances of a given MWE class, or if not, what the rules that govern the productivity of the MWE are.
The next question is what degree of morphosyntactic flexibility the (open or closed) class of MWEs has:

- if none, treat as a “word-with-spaces” and map onto a pre-existing type in the grammar, where possible
- if only morphological flexibility (e.g. *attorney general*), capture this appropriately and map to a type in the grammar
- if morphosyntactic flexibility, ascertain the level of flexibility; most MWEs are somewhat constrained in their morphosyntactic flexibility, but working out the precise limits of flexibility is notoriously difficult (suggest erring on the side of under-constraint)

Beware metalinguistic markers such as *proverbial*
Idioms

- Idioms are a fascinating class of MWE, in that they take many different forms lexically and syntactically, with syntactic flexibility determined through “decomposability” [Nunberg et al., 1994, Baldwin et al., 2003]:
  - if the semantics of the idiom can be (idiosyncratically) ascribed to parts of the MWE, those parts will tend to undergo (almost) unlimited syntactic flexibility
    - *spill the beans* = *reveal’ (secret’)* ≈ *spill’ (beans’)*
  - one approach of dealing with this in the grammar is to have an idiomatic lexical entry associated with each of the decomposed components, and constrain idiomatic parses to have all of the idiom parts in particular relational configurations
Talk Outline

1. Introduction

2. MWEs at the Grammar–Lexicon Interface

3. A Case Study: English Determinerless PPs
   - Problem Statement
   - Analyses

4. Summary
Contents

1 Introduction

2 MWEs at the Grammar–Lexicon Interface

3 A Case Study: English Determinerless PPs
   - Problem Statement
   - Analyses

4 Summary
Definition

Determinerless PPs ("PP–Ds") are MWEs comprising a preposition (P) and a singular noun ($N_{Sing}$) without a determiner:

<table>
<thead>
<tr>
<th>Institution</th>
<th>Media</th>
<th>Metaphor</th>
<th>Temporal</th>
<th>Means/Manner</th>
</tr>
</thead>
<tbody>
<tr>
<td>at school</td>
<td>on film</td>
<td>on ice</td>
<td>at breakfast</td>
<td>by car</td>
</tr>
<tr>
<td>in church</td>
<td>on TV</td>
<td>at large</td>
<td>at lunch</td>
<td>by train</td>
</tr>
<tr>
<td>in gaol</td>
<td>to video</td>
<td>at hand</td>
<td>on break</td>
<td>by hammer</td>
</tr>
<tr>
<td>on campus</td>
<td>off screen</td>
<td>at leave</td>
<td>by night</td>
<td>by computer</td>
</tr>
<tr>
<td>at temple</td>
<td>in radio</td>
<td>at liberty</td>
<td>by day</td>
<td>via radio</td>
</tr>
</tbody>
</table>

Source(s): Baldwin et al. [2006], Stvan [1998]
The Syntax of PP–Ds

- Variability in syntactic markedness, productivity and nominal modifiability for different PP–D constructions
- Non-productive, non-modifiable PP–Ds: *ex cathedra, ad hominem, ad nauseum*
- Fully-productive, highly-modifiable PP–Ds: *per recruited student that finishes the project*
- Most PP–Ds lie between these two extremes

*Source(s):* Ross [1995]
Syntactic Markedness

- Syntactically-unmarked PP–Ds: \( N_{Sing} \) is uncountable
  
  E.g. Institutions: *in school, in gaol*, but *in library* (cf. *school finished vs. *library finished*)

- Syntactically-marked PP–Ds: \( N_{Sing} \) is strictly countable
  
  E.g. PPs headed by *per*: *per person*, but *per information* (c.f. *by bus/public transport*)
Nominal Modifiability

- No modification: *mental/*small hospital
- Idiosyncratic modification: *long/*lengthy/*short last
- Relatively free modification: *great/*considerable/*tedious length
- Modification seldom unrestricted, except for fully productive prepositions (e.g. by)
Types of Modification in PP–Ds

Modification can be:

- impossible, optional or obligatory
- nominal, adjectival or both (or none)

<table>
<thead>
<tr>
<th></th>
<th>Obligatory</th>
<th>Optional</th>
<th>Impossible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>at <em>(eye)</em> level</td>
<td>on <em>(summer)</em> vacation</td>
<td></td>
</tr>
<tr>
<td>Adjective</td>
<td>at <em>(long)</em> range</td>
<td>in <em>(sharp)</em> contrast</td>
<td>on <em>(very)</em> top</td>
</tr>
<tr>
<td>Either</td>
<td>at <em>(company)</em> expense</td>
<td>in <em>(family)</em> court</td>
<td></td>
</tr>
<tr>
<td></td>
<td>at <em>(considerable)</em> expense</td>
<td>in <em>(open)</em> court</td>
<td></td>
</tr>
</tbody>
</table>
Contents

1 Introduction

2 MWEs at the Grammar–Lexicon Interface

3 A Case Study: English Determinerless PPs
   - Problem Statement
   - Analyses

4 Summary
Analysis 1: Lexical Listing

**Analysis:** all PP–Ds fully lexicalised

**Pros:**
- effective at capturing syntactically- and semantically-marked PP–Ds

**Cons:**
- unable to handle nominal modification
- data sparseness effects with productive classes
- inability to capture semantic compositionality

_Source(s):_ Baldwin et al. [2006]
Analysis 2: Simple Combination

- **Analysis**: use head complement rule to combine simplex P and N lexical entries

- **Pros**:
  - Effective at capturing syntactically-unmarked PP–Ds
  - Licenses unconstrained nominal modification

- **Cons**:
  - Inability to capture non-compositional semantics
  - Overgeneration (*by/on/... school*)
At school
Analysis 3: Idiosyncratic Selection

**Analysis:** introduce idiosyncratic lexical types for head nouns which specify constraints such that:

1. phrases that it projects will only appear as complements of (certain) Ps;
2. its specifier (determiner) will never be expressed;
3. it must combine with a (pre-head) modifier before it can combine with the P; and
4. it can only appear with particular modifier types (e.g. only nouns)
Analysis 3: Idiosyncratic Selection

Example for \textit{at eye/street/... level}:

\[
\text{level\textsubscript{i,n1}} := \text{n---c-brn_le} & \\
\quad [ \text{STEM < "level" >}, \\
\quad \text{SYNSEM [ LKEYS.KEYREL.PRED "\_level\_n\_1\_rel",} \\
\quad \text{PHON.ONSET con ] } ] .
\]

\[
\text{at+level} := \text{detless\_pp\_idiom\_mtr} & \\
\quad [ \text{INPUT.RELS <! [ PRED \_at\_p\_rel ],} \\
\quad \text{[ PRED "\_level\_n\_1\_rel" ],} \\
\quad \text{[ PRED idiom\_q\_i\_rel ] !> ]}. 
\]
Analysis 3: Idiosyncratic Selection

- **Pros:**
  - captures idiosyncracies of modifiability, P-N collocation

- **Cons:**
  - lexical redundancy
  - overgeneration
Analysis 4: \( \bar{N} \) Selection

- **Analysis:** allow particular P senses (e.g. *by MEANS*) to select for unsaturated NPs (\( \bar{N}s \))

- **Pros:**
  - allows for nominal modification and full productivity

- **Cons:**
  - semantic restriction of the noun?

(Dutch) *in CLOTHING* vs. *by MEANS*

\[
\begin{array}{c}
\text{SYN} \\
\text{CAT} \\
\text{HEAD} \\
\text{prep} \\
\text{VAL} \\
\text{COMPS} \langle \text{SPR} \langle \text{Det} \rangle \rangle \\
\end{array}
\]
Analysis 4:  $\bar{N}$ Selection

```
PP
   /\  \bar{N}
P by
   \  \N
car
```
Determining the Appropriate Analysis

Consider:

- degree and type of nominal modifiability
- productivity relative to a given P
- semantic markedness
- NP saturation (syntactic markedness)
Talk Outline

1. Introduction

2. MWEs at the Grammar–Lexicon Interface

3. A Case Study: English Determinerless PPs
   - Problem Statement
   - Analyses

4. Summary
Summary

- MWEs populate a broad spectrum, in terms of productivity, constructional composition, morphosyntactic flexibility, ...
- Important to understand the full complexities of the data, to be able to customise the syntactic analysis to the MWE
References


References


