Table of Contents

1 Introduction

2 Variation in fixed MWEs

3 Alternatives to fixed

4 Conclusions
A problem: *fixed* as a syntactic relation

In UD, *fixed* is treated as a headless dependency relation, where the leftmost word is always regarded as the head. This makes a visible syntactic structure invisible [2, 6].

![Diagram showing two mutually excluding dependency analyses in UD.](image)

**Figure:** Two mutually excluding dependency analyses.
More on the UD relation $fixed$

- it links elements of grammaticalized expressions that behave as function words or short adverbials
More on the UD relation *fixed*

- it links elements of grammaticalized expressions that behave as function words or short adverbials
- it should be restricted to the most grammaticalized cases and be treated as a closed class.

Source: UD-page on *fixed* for v2.13
MWEs, including fixed MWEs, are generally quite common in languages
Other problems

• MWEs, including fixed MWEs, are generally quite common in languages
• The Swedish treebanks Talbanken and LinES (version 2.13) together hold 439 MWEs types annotated with and without *fixed*
Contributions

- A schema for analysis of fixed MWEs with a focus on kinds of variation, that can be applied to any language
- Identification of subtypes of fixed MWEs in Swedish, based on the schema that can support decisions on what to annotate with fixed in future versions
- Proposals for alternative syntactic annotations for the subtypes and for the treatment of MWEs in UD
Common properties of MWEs

- lexical units
- non-compositionality
- morpho-syntactic irregularity
A characterisation of fixed MWEs

Baldwin and Kim (2010) [1] defines fixed multiword expressions as expressions "that undergo neither morphosyntactic variation nor internal modification, often due to fossilisation of what was once a compositional phrase."
Introduction

Variation in fixed MWEs

Alternatives to fixed

Conclusions

Source: Baldwin and Kim, 2010:19 [1]
The space to explore

Introduction
Variation in fixed MWEs
Alternatives to fixed
Conclusions

Figure: Caption
## Introduction

Variation in fixed MWEs

Alternatives to fixed

Conclusions

---

### Analysis schema

<table>
<thead>
<tr>
<th>Expression</th>
<th>Freq</th>
<th>Talb.</th>
<th>Lin.</th>
<th>UPOS</th>
<th>Modif.</th>
<th>example</th>
<th>Infls</th>
<th>Synyms</th>
<th>Order</th>
<th>Abbr.</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>som om</td>
<td>73</td>
<td>0</td>
<td>73</td>
<td>SCONJ</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>som SC</td>
</tr>
<tr>
<td>så att</td>
<td>70</td>
<td>34</td>
<td>36</td>
<td>SCONJ</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>ADV SC</td>
</tr>
<tr>
<td>det här</td>
<td>61</td>
<td>14</td>
<td>47</td>
<td>DET/PRON</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>DET ADV</td>
</tr>
<tr>
<td>även om</td>
<td>43</td>
<td>31</td>
<td>12</td>
<td>SCONJ</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>ADV SC</td>
</tr>
<tr>
<td>i dag</td>
<td>43</td>
<td>39</td>
<td>4</td>
<td>ADV</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>P NN</td>
</tr>
<tr>
<td>därför att</td>
<td>41</td>
<td>25</td>
<td>16</td>
<td>SCONJ</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>ADV SC</td>
</tr>
<tr>
<td>den här</td>
<td>40</td>
<td>21</td>
<td>19</td>
<td>DET/PRON</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>DET ADV</td>
</tr>
<tr>
<td>på grund av</td>
<td>35</td>
<td>24</td>
<td>11</td>
<td>ADP</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>P NN P</td>
</tr>
<tr>
<td>för att</td>
<td>35</td>
<td>34</td>
<td>1</td>
<td>SCONJ</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>P SC</td>
</tr>
<tr>
<td>i stället</td>
<td>34</td>
<td>22</td>
<td>12</td>
<td>ADV</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>P NND</td>
</tr>
<tr>
<td>till exempel</td>
<td>31</td>
<td>10</td>
<td>21</td>
<td>ADV</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>P NN</td>
</tr>
<tr>
<td>till och med</td>
<td>28</td>
<td>3</td>
<td>25</td>
<td>ADV</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>P CC P</td>
</tr>
<tr>
<td>i regel</td>
<td>28</td>
<td>26</td>
<td>0</td>
<td>ADV</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>P NN</td>
</tr>
<tr>
<td>med hjälp av</td>
<td>27</td>
<td>6</td>
<td>21</td>
<td>ADP</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>P NN P</td>
</tr>
<tr>
<td>in i</td>
<td>26</td>
<td>0</td>
<td>26</td>
<td>ADP</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>ADV P</td>
</tr>
</tbody>
</table>
Aspects of variation, I

Each aspect corresponds to a property which an MWE may have, or not. The properties are framed as questions. Each property is illustrated with a positive example.

- **Internal modification**: Does any part allow a modifier? Yes or No.
  - Example: *mer än*, 'more than'
  - Variation: *mycket mer än*, 'much more than'

- **Optionality**: Is any part optional, or may an optional part be added? Yes or No.
  - Example: *efter det att*, 'after it that'
  - Variation: *efter att*, 'after that'

- **Inflections**: Does any part of the expression allow inflectional variants? Yes or No.
  - Example: *när det gäller*, 'when it concerns'
  - Variation: *när det gällde*, 'when it concerned'
Aspects of variation, II

- **Synonymy**: Is it possible to replace any part with a synonym? Yes or No.
  - Example: *med hjälp av*, 'with the aid of'
  - Variation: *med hjälp från*, 'with the aid from'
- **Order change**: Can the order of the parts be different? Yes or No.
  - Example: *i stort sett*, 'in large seen', 'largely'
  - Variation: *sett i stort*, 'seen in large'
- **Iterability**: Can some part be repeated? Yes or No.
  - Example: *om och om igen*, 'again and again',
  - Variation: *om och om och om igen*, 'again and again and again'
Aspects of variation, III

- **Separability**: Can, or must, some part be separated from the rest by other material? No, Optionally, Obligatory
  - Example: *vem som helst*, 'anyone',
  - Variation: *vem av dem som helst*, 'anyone of them'

- **Abbreviation**: Does an abbreviated form exist? Yes or No.
  - Example: *till exempel*, 'for example',
  - Variation: *t. ex.*, 'i. e.'

- **Collapsibility**: Does an equivalent single form exist? Yes or No.
  - Example: *i dag*, 'today',
  - Variation: *idag*, 'today'
Types of fixed MWEs

• **Rigid**: No aspect of variation can be found except possibly for Collapsibility and Abbreviation.
• There are 139 such expressions in the dataset
• **Non-rigid**: Some aspect of variation concerning a part is present. This means that the part is recognizable as a part and that the expression is structurally transparent.
Collapsible MWEs

Collapsible MWEs are numerous in the data. The different spellings invoke no differences in pronunciation or meaning. For this reason, collapsed versions are suggested to be treated as multiword UD tokens.

(a) De kan vara här när som helst
(b) De kan vara här närsomhelst
   ’They can be here at any moment’

<table>
<thead>
<tr>
<th></th>
<th>vara</th>
<th>vara</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>här</td>
<td>here</td>
</tr>
<tr>
<td>4</td>
<td>när</td>
<td>när</td>
</tr>
<tr>
<td>5</td>
<td>som</td>
<td>som</td>
</tr>
<tr>
<td>6</td>
<td>helst</td>
<td>helst</td>
</tr>
</tbody>
</table>
Alternatives for coordinations

**Figure:** Alternative dependency analyses for lexicalized coordinations.
PPs without determiners

Figure: Alternative annotations for determinerless PPs.
Graded syntactic regularity

A phrase consisting of a preposition and a countable noun is often not grammatical:

i trädgården  in the garden
i en trädgård  in a garden
*i trädgård  *in garden

However, it is a common pattern in fixed MWEs. Baldwin and Kim (2010) has a name for them, determinerless prepositional phrases, PP-Ds.
Instances of the same pattern?

The PP-Ds are not the only fixed MWEs introduced by a preposition:

- ADP + NOUN[Def]  i stället  ’instead’
- ADP + NOUN[Ind]  i dag  ’today’
- ADP + NOUN[SpecInfl]  i vårås  ’this (past) spring’
- ADP + ADJ/ADV  i stort  ’largely’
- ADP + cranberry morph  i fjol  ’last year’
- ADP + ??  till dels  ’partly’
Instances of the same pattern?

The PP-Ds are not the only fixed MWEs introduced by a preposition:

<table>
<thead>
<tr>
<th>ADP + NOUN[Def]</th>
<th>i stället</th>
<th>'instead'</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADP + NOUN[Ind]</td>
<td>i dag</td>
<td>'today'</td>
</tr>
<tr>
<td>ADP + NOUN[SpecInfl]</td>
<td>i våras</td>
<td>'this (past) spring'</td>
</tr>
<tr>
<td>ADP + ADJ/ADV</td>
<td>i stort</td>
<td>'largely'</td>
</tr>
<tr>
<td>ADP + cranberry morph</td>
<td>i fjol</td>
<td>'last year'</td>
</tr>
<tr>
<td>ADP + ??</td>
<td>till dels</td>
<td>'partly'</td>
</tr>
</tbody>
</table>

Common structure: ADP + ANY
A view from Construction Grammar

• Language is organised in terms of constructions, or subsystems [3, 4]
• Many of these subsystems concern MWEs
• Subsystems are related via hierarchical and other types of relations

Note that UD provides only one subsystem, Names, with special annotation, using the relation *flat*. 
Reference to weekdays, seasons, holidays, and parts of the day, of the immediate past form a subsystem in Swedish, which, in principle could be given a separate UD annotation.

\[
i \text{lördags} \quad \text{this past Saturday} \\
i \text{våras} \quad \text{this past spring} \\
i \text{julas} \quad \text{this past Christmas} \\
i \text{förmiddags} \quad \text{this past (late) morning}
\]
Syntactic annotation with features

For syntactically transparent expressions that form a unit, the property of fixedness can be represented at the level of features.

Figure: A template for ADP + ANY structures
Syntactic annotation with features

Figure: A template for ANY + CC + ANY
When the MWE is not a syntactic unit

When a fixed MWE is not a syntactic unit the proposed solution does not work. The syntactic structure can be captured by treating the two parts separately, but the feature needs an explicit reference.

Figure: Analysis for som om, ’as if’
A no solution

One proposal here is to give up, as UD has done with flexible MWEs, and turn to a more expressive format such as the .cupt formalism used by the PARSEME-MWE project [5, 6].
Conclusions

• 439 expressions currently annotated with *fixed* in Swedish UD treebanks have been analysed

• A language-independent schema, focused on variation, has been developed for this purpose

• Two properties, rigidness (i.e no variation) and syntactic non-transparency, have been identified as important for the future annotation with the *fixed* dependency

• I have proposed that a subset of fixed MWEs can adequately be annotated with a combination of syntactic dependencies and a feature

• I also argue that annotation of the full set of fixed MWEs need a more expressive feature formalism such as the .cupt format used in PARSEME-MWE.
Thanks for listening!
References are in the paper and in the following slides.
Fitting Fixed Expressions into the UD Mould: Swedish as a Use Case

Lars Ahrenberg

Introduction
Variation in fixed MWEs
Alternatives to fixed
Conclusions

Timothy Baldwin and Su Nam Kim.
Multiword expressions.

Kim Gerdes and Sylvain Kahane.
Dependency annotation choices: Assessing theoretical and practical issues of Universal Dependencies.

Thomas Hoffmann.
Construction Grammar: The Structure of English.

Benjamin Lyngfelt, Linnéa Bäckström, Lars Borin, Anna Ehrlemark, and Rudolf Rydstedt.
Constructicography at work: Theory meets practice in the Swedish constructicon.

Agata Savary, Cherifa Ben Khelil, Carlos Ramisch, and et al.
Parseme corpus release 1.3.

Agata Savary, Sara Stymne, Verginica Barbu Mititelu, Nathan Schneider, Carlos Ramisch, and Joakim Nivre.
Parseme meets universal dependencies: Getting on the same page in representing multiword expressions.