Fitting Fixed Expressions into the UD Mould: Swedish as a Use Case

Lars Ahrenberg

Department of Computer and Information Science
Linköping University

Workshop on MWE-UD@LREC-COLING
Turin, May 25th, 2024

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

Table of Contents

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

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].
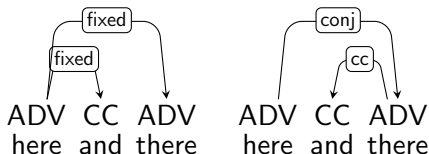

Figure: Two mutually excluding dependency analyses.

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

More on the UD relation *fixed*

- it links elements of grammaticalized expressions that behave as function words or short adverbials

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

Other problems

- MWEs, including fixed MWEs, are generally quite common in languages

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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*

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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
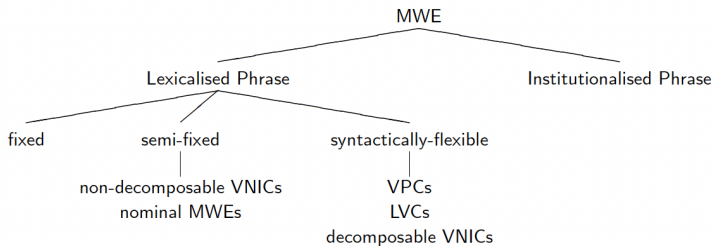
Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

Common properties of MWEs

- lexical units
- non-compositionality
- morpho-syntactic irregularity

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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."*

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

MWE

Lexicalised Phrase

Institutionalised Phrase

fixed   semi-fixed   syntactically-flexible

non-decomposable VNICs   VPCs
nominal MWEs   LVCs
decomposable VNICs

Source: Baldwin and Kim, 2010:19 [1]

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

The space to explore

Figure: Caption

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions
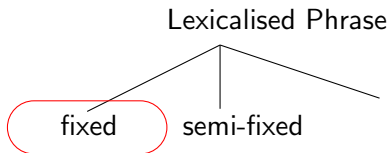
Analysis schema

fixedexpressions-2

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

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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'

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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'

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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'

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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.

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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'

3	vara	vara
4	här	here
5-7	närsomhelst	_
5	när	när
6	som	som
7	helst	helst

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

Alternatives for coordinations
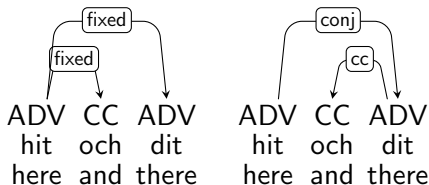
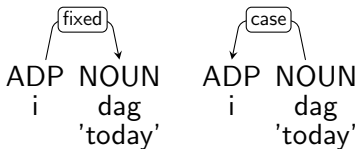Figure: Alternative dependency analyses for lexicalized coordinations.

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

PPs without determiners

Figure: Alternative annotations for determinerless PPs.

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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.

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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 våras	'this (past) spring'
ADP + ADJ/ADV	i stort	'largely'
ADP + cranberry morph	i fjol	'last year'
ADP + ??	till dels	'partly'

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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 våras	'this (past) spring'
ADP + ADJ/ADV	i stort	'largely'
ADP + cranberry morph	i fjol	'last year'
ADP + ??	till dels	'partly'

Common structure: ADP + ANY

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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*.

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

A Swedish subsystem

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 lördags	this past Saturday
i våras	this past spring
i julas	this past Christmas
i förmiddags	this past (late) morning

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

Syntactic annotation with features

For syntactically transparent expressions that form a unit, the
property of fixedness can be represented at the level of features.
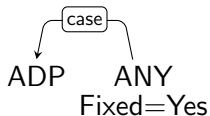
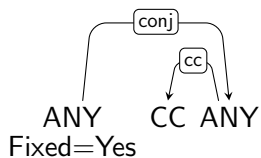Figure: A template for ADP + ANY structures

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Syntactic annotation with features

Figure: A template for ANY + CC + ANY

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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.
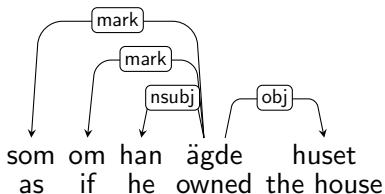

Figure: Analysis for *som om*, 'as if'

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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].

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Introduction

Variation in
fixed MWEs

Alternatives to
fixed

Conclusions

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.

Fitting Fixed
Expressions
into the UD
Mould:
Swedish as a
Use Case

Lars
Ahrenberg

Thanks for listening!
References are in the paper and in the
following slides.

📄 Timothy Baldwin and Su Nam Kim.
Multiword expressions.
In Nitin Indurkhya and Fred J. Damerau, editors,
*Handbook of Natural Language Processing, Second
Edition*. CRC Press, Boca Raton, USA,, 2010.

📄 Kim Gerdes and Sylvain Kahane.
Dependency annotation choices: Assessing theoretical and
practical issues of Universal Dependencies.
In Annemarie Friedrich and Katrin Tomanek, editors,
*Proceedings of the 10th Linguistic Annotation Workshop
held in conjunction with ACL 2016 (LAW-X 2016)*, pages
131–140, Berlin, Germany, August 2016. Association for
Computational Linguistics.

📄 Thomas Hoffmann.
Construction Grammar: The Structure of English.
Cambridge University Press, 2022.

📄 Benjamin Lyngfelt, Linnéa Bäckström, Lars Borin, Anna
Ehrlemark, and Rudolf Rydstedt.

Constructicography at work: Theory meets practice in the Swedish constructicon.
In B. Lyngfelt, L. Borin, K. Ohara, and T. T. Torrent, editors, *Constructicography: Constructicon development across languages*, pages 41–106. John Benjamins, Amsterdam, 2018.

Agata Savary, Cherifa Ben Khelil, Carlos Ramisch, and et al.
Parseme corpus release 1.3.
In *Proceedings of The 19th Workshop on Multiword Expressions (MWE 2023)*, pages 24–35, 2023.

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.
Northern European Journal of Language Technology, 9(1), 2023.

Fitting Fixed Expressions into the UD Mould: Swedish as a Use Case

Lars Ahrenberg

Introduction

Variation in fixed MWEs

Alternatives to *fixed*

Conclusions