

Introduction

- ▶ Preliminary investigations on how to model syntactic properties of different types of MWEs within LFG.
- ▶ **Motivation:** which types of MWEs can be described at the level of the lexicon and which (if any) require corresponding descriptions at the level of LFG phrase-structure rules?

Structurally regular MWEs

Constructions which can be successfully **handled by productive grammar rules**.

Agreement

F-description for the expression *NP vider DET sac* 'to express NP's secret thoughts', lit. 'NP empty DET bag', **assigned to the head** *vider*:

(↑ OBJ PRED FN) =_c 'sac'
 (↑ OBJ SPEC POSS) = *f*
 (*f* PERS) = *g* (*f* NUM) = *h*
 (↑ SUBJ PERS) = *g* (↑ SUBJ NUM) = *h*

Modifiers

- ▶ **Syntactically flexible** *to spill the beans* 'to reveal a secret': verb entry extended with (↑ OBJ NUM) =_c *pl* and lexical constraints (and nothing more).
- ▶ French **fixed expression** *NP casser sa pipe* 'to die, lit. NP break his/her pipe': ¬(↑ OBJ ADJUNCT) ensures that object is not modified.

Passivization

- ▶ An expression may **lose its idiomatic meaning when passivized**, e.g. *a bucket kicked by him*.
- ▶ LFG posits **separate active and passive verb entries**.

References

- Abeillé, A. and Schabes, Y. (1989). Parsing Idioms in Lexicalized TAGs. In *Proceedings of the Fourth Conference on European Chapter of the Association for Computational Linguistics*, EACL '89, pages 1–9, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Asudeh, A., Dalrymple, M., and Toivonen, I. (2013). Constructions with lexical integrity. *Journal of Language Modelling*, 1(1):1–54.
- Attia, M. A. (2006). Accommodating Multiword Expressions in an Arabic LFG Grammar. In *Proceedings of the 5th International Conference on Advances in Natural Language Processing*, FinTAL'06, pages 87–98, Berlin, Heidelberg. Springer-Verlag.
- Mahowald, K. (2011). An LFG Account of Word Order Freezing. In Butt, M. and King, T. H., editors, *Proceedings of the LFG11 Conference, Hong Kong*, pages 381–400, Stanford, Ca. CSLI Publications.
- Patejuk, A. (2014). Handling MWEs in walenty, a new valence dictionary for Polish. Poster presented at the 2nd PARSEME general meeting, 10–11 March, Athens, Greece.

Structurally idiosyncratic MWEs

Constructions which **require** corresponding **descriptions at the level of phrase-structure rules**.

Verb-particle constructions

- ▶ Should not be handled by productive grammar rules – only **specific pairs**, e.g. *look up* 'search for a reference' but not *look at*, can be **interpreted as verb-particles**.
- ▶ English ParGram: a standard, transitive VP rule accepts a particle either before or after the direct object of the verb.
- ▶ Verbs linked with particles through **unification** (= and =_c) **over a dedicated PRT_FORM feature**.

Correlative conjunctions

- ▶ Constructions such as *either _ or _* and *both _ and _*
- ▶ Handled similarly as verb-particles: unification over a dedicated COORD-FORM feature.

Phrasal configuration

Swedish traversal construction, e.g. *Sarah armbågade sig genom mängden* 'lit. S. elbowed SELF through crowd.DEF':

- ▶ Consists of a verb, a weak reflexive (coindexed with the subject), and a directional PP.
- ▶ The particle follows the direct object of the verb, while normally it would adjoin to the verb.
- ▶ Most elegantly modeled by a dedicated c-structure rule (Asudeh et al., 2013).

Word-order preferences

- ▶ Polish *doręczyć NP do rąk własnych* 'deliver NP as hand delivery, lit. deliver NP to hands own' **vs.** *doręczyć NP do własnych rąk* 'lit. deliver NP to own hands'.
- ▶ A particular **word order** can be **blocked** with the help of an *f-precedence* operator.
- ▶ **No satisfying solution** for modeling soft, quantitative word-order preferences seems to exist.

'Ungrammatical' constructions

- ▶ Irregular structure **not recognized by productive grammar rules**.
- ▶ Such as *all of a sudden* 'suddenly' idiomatic expression.
- ▶ 'Ungrammatical' constructions are **typically fixed** and thus can be handled as **words-with-spaces**. This will sometimes lead to input **segmentation ambiguities**.
- ▶ Idioms can lead to **phrasal template**. Polish *nie wszystko złoto co się świeci* 'what seems ideal is not necessarily so, lit. not everything gold which shines' in which the **main verb is missing** could be handled by S → *nie wszystko NP co się VP*.