

12. Identifying context-dependent meaning components of figurative expressions

Supervisors: Anke Holler, Caroline Sporleder

PhD Student: Irene Pagliai

I. The form-meaning mismatch

- Figurative expressions can also have a literal interpretation, depending on the context.
- Case in point: verb phrase idioms as in example (1).
 - Dad had to break the ice on the chicken troughs so that they could get water.
[Gigaword Corpus NYT200008]
- As (2) to (4) illustrate, both interpretations can even hold simultaneously:
 - (EN): Left holding the baby, single mothers find it hard to fend for themselves. [BNC CRA]
 - (EN): You're like a restless bird in a cage. When you get out of the cage, you'll fly very high. [BNC FR6]
 - (DT): Wenn das Kind gerade einen fieberhaften Infekt hat, ist den meisten Eltern klar, dass das Planschen buchstäblich erst einmal ins Wasser fallen muss.
[Web, www.praxisvita.de]
- Such "mixed usages" are not a homogeneous class:
 - Example (2): Both the literal and the figurative interpretation would be felicitous given the context. (Typically the idiomatic one might be more likely.)
 - Example (3): An idiom is embedded in a metaphor and takes on a literal meaning within the figurative meaning of the surrounding metaphorical context.
 - Example (4): The figurative reading is dominant, but the idiom is chosen in such a way that its constituents exhibit lexical cohesion (Halliday & Hasan 1976) with the literal context. Such mixed usages are often chosen for humorous effects.
- "Mixed usages" constitute a 1:many form-meaning mismatch in that one form can take on literal and figurative meanings, components of which may even be accessible simultaneously.

II. Motivation

- Figurative expressions generally pose a challenge for natural language processing (Sag et al. 2002).
- A robust system for natural language understanding needs to be able to detect potentially figurative expressions and to determine their interpretation in the given context.
- This is typically done by comparing the constituents of the target expression to the context and computing for example the semantic 'fit' (e.g. in terms of lexical cohesion) between the context and the expression (Li & Sporleder 2009 *i.a.*).
- Under the assumption that the semantics of figurative expressions is not (or not entirely) compositional, a good 'fit' with the context indicates a literal interpretation.
- However, this approach is bound to fail for mixed usages, which exhibit cohesion under both interpretations. While mixed usages are rare, their frequency is not negligible (Sporleder et al. 2010).

III. Research questions

- Aim: Explore linguistically-informed algorithmic approaches to identifying and interpreting occurrences of figurative expressions which have mixed interpretations in a given context.
- Assuming that multi-word figurative expressions are generally taken to be morpho-syntactically 'frozen' (Fraser 1970) and semantically non-compositional to varying degrees (Nunberg et al. 1994; Fellbaum 2014):
 - **Under which circumstances are mixed usages possible?**
- E.g. one might hypothesize that expressions which are semantically more transparent and/or syntactically less fixed are more likely to allow mixed interpretations.
- It is also possible that certain properties of the context license such readings.
- Current computational models of figurative expression interpretation tacitly assume figurativeness to be an all-or-nothing category whilst e.g. (1-4) suggest:
 - that different interpretations hold simultaneously and
 - the possibility of nested figurative and literal interpretations.

IV. Hypotheses and method

The PhD student will

- carry out a corpus study in order to categorize mixed usages and identify subclasses
 - good starting point: the IDIX corpus (Sporleder et al. 2010)
 - contains annotations of literal, figurative and mixed interpretations for 5.836 occurrences of 78 different idioms in the British National Corpus.
- hypothesize about linguistic markers/licensors for mixed usage, with respect to the target expression itself and the context.
- develop a computational model of mixed usage interpretations, building on recent advances in distributional semantics and neural language processing (Salton et al. 2016; King & Cook 2018).

V. Connections to other research projects

- Type of form-meaning mismatch: [10](#), [11](#)
- Empirical domain: [3](#), [6](#), [9](#) (language processing and acquisition)
- Content: [3](#) and [9](#), since it will benefit from cross-talks on word-form and word-meaning mismatches in language acquisition and on the impact of the semantic vs. pragmatic access from form to meaning on sentence processing.

VI. Possible follow-up studies

- Mixed usage of figurative expressions across domains (a corpus study).
- Sensitivity of processing of figurative language to internal factors (degree of predictability) and external factors (context sensitivity) using eye-tracking in reading.
- Modelling of mixed usages at the grammar-pragmatics interface.