

## IMPORTANT UNDERLYING ISSUES/INTERESTS

- Semantic/pragmatic vs. syntactic vs. discourse-related vs. ... taxonomies
- Overlap/similarities/differences with other MWE types/taxonomies
- Linguistic tests to justify classifications
- Syntax–semantics interface
- Role of pragmatics
- Representation
- How could this insight be used in the extractions of MWEs?

## REFERENCE:

R. Moon (1998) *Fixed Expressions and Idioms in English: A Corpus-based Approach*, Clarendon Press

## MAIN CLAIMS

- Fixed expressions and idioms (FEIs) can only be properly described and understood in situ, based on corpus evidence
- Frequency/distribution and contrastive corpus evidence are integral to MWE identification/determination
- Many idioms do not have a canonical form (*lick someone's shoes/boot*, *put one's foot/boot in it*, *put the boot in*, *to be in one's shoes*, *boots and all*, *put one's foot down*, ...)

## BOUNDS ON RESEARCH (1/2)

- Descriptive not theoretical research: focus on *what/how* not *why*
- MWE types not considered in this research (but recognised as MWEs)
  - compound nouns, adjectives and verbs (*civil servant, plain wrong, factory test*)
  - phrasal verbs (*put off*)
  - foreign phrases (*raison d'être*)
  - multiword inflectional forms of verbs, adjectives and adverbs (*had been read*)

## BOUNDS ON RESEARCH (2/2)

- What are claimed to not be MWEs:
  - indirect speech acts — idiomatic in nature but unbounded in number (impossible/undesirable to enumerate in lexicon, e.g. *do you have the time?*)
  - single words/hyphenated complexes derived through lexicogram-matical processes (*antidisestablishmentarianism, one-off*)

## SHORTFALLS/DISCUSSION POINTS

- Domain and MWE-hood
  - compositionality (*monosodium glutamate* – variation accessible from corpus?)
  - stylistic effects of domain on MWE choice/evidence
- What does lexical alternation between a MWE and single lexeme/hyphenated complex say about the MWE status of that word? (*stand alone/stand-alone, mailman/mail man*)
- Can this be applied in the evaluation of coverage of extraction methods?
- Moon excludes familial euphemisms from the scope of MWEs. Is she justified in doing so, or is this simply a highly specialised domain?

## DEFINING MWEs

*... there is no unified phenomenon to describe but rather a complex of features that interact in various, often untidy, ways and represent a broad continuum between non-compositional (or idiomatic) and compositional groups of words.*

## CONDITIONS ON MWE-HOOD

- **Institutionalisation** (process by which a string or formulation becomes recognised and accepted as a lexical item, determined quantitatively); diachrony and domain are important considerations here — **necessary but not sufficient**
- **Lexicogrammatical fixedness** (formal rigidity, preferred lexical realisation, restrictions on aspect/mood/voice) — **neither necessary nor sufficient**
- **(Semantic/pragmatic) non-compositionality** (incl. ungrammaticality and the case of the component lexical items having special meanings within the context of the MWE) — **not necessary but sufficient**
- Also: FEIs must be made up of two or more words, FEIs typically form syntactic or grammatical units in their own right, and intonation can sometimes be used to distinguish between compositional and non-compositional readings.

## CATEGORIES OF MWEs (1/3)

- **Anomalous collocations:** lexicogrammatically marked
  - *ill-formed collocations* (*at all, by and large*)
  - *cranberry collocations:* idiosyncratic lexical component (*in retrospect, kith and kin*)
  - *defective collocations:* idiosyncratic meaning component (*in effect, foot the bill*)
  - *phraseological collocations:* semi-productive constructions (*in/into/out of action, on show/display*)

## SHORTFALLS/DISCUSSION POINTS

- Cranberry collocations: what about lexemes which occur in multiple collocations but are not productive? (*ad* in *ad hoc*, *ad nauseum*, *ad infinitum*, *hoc* in *post hoc*, so *ad hoc* not cranberry collocation)

## CATEGORIES OF MWEs (2/3)

- **Formulae:** pragmatically marked
  - *simple formulae/sayings*: compositional strings with a special discourse function (*alive and well, a horse, a horse, my kingdom for a horse*)
  - *metaphorical/literal proverbs* (*you can't have your cake and eat it, enough is enough*)
  - *similes* (*as good as gold*)

## **SHORTFALLS/DISCUSSION POINTS**

- Metaphor/simile distinction meaningful? (Aren't similes a lexically bounded form of proverbs?)
- Very fuzzy boundary between literal proverbs and sayings

## CATEGORIES OF MWEs (3/3)

- **Metaphors:** semantically marked (non-compositional)
  - *transparent metaphors* (*behind someone's back, pack one's bags*)
  - *semi-transparent metaphors* (*on an even keel, pecking order*)
  - *opaque metaphors* (*bite the bullet, kick the bucket*)

## SHORTFALLS/DISCUSSION POINTS

- Overlap between metaphors and defective collocations (*admit defeat*); in fact, 60% of defective collocations are also classified as something else (formula, phraseological collocation, ...)
- Why three categories of metaphor, when effectively you have a cline of transparency?
- NOTE: Moon recognises that overlap between her categories exists (25% of her FEIs fall into 2 categories, 1% into 3), but claims that there is a preferable interpretation in most cases. BUT wouldn't an orthogonal set of taxonomies better capture MWEs?

## COLLOCATIONS

- ... *the occurrence of two or more words within a short space of each other in a text*
- Vanilla vs. anomalous collocations
- **Semantic collocations** (co-occurrence preferences – *jam* with **FOOD**, *ice cream* with **cone**<sub>RECEPTACLE</sub>)
- **Lexico-semantic collocations:** (*rancid butter/fat*, *face the truth/facts/problem*)
- **Syntactic collocations:** (*too ... to*)

## SHORTFALLS/DISCUSSION POINTS

- Despite being a corpus-based approach, Moon doesn't really capture fully-compositional but highly marked collocations such as *rolling blackouts*
- Moon discusses syntagmatic paradigms but not anti-collocations