Magn(idea) = big:

Collocations for Lexicon Building

William S. Annis

The Digital Language Creation Conference
On the Online!
March 6-7, 2021
Lexical Collocations are not idioms

- A *red herring* is neither red, nor a herring.
Lexical Collocations are not idioms

- A red herring is neither red, nor a herring.
- But strong tea is still tea.
Lexical Collocations are not idioms

- A red herring is neither red, nor a herring.

- But strong tea is still tea.

  - Sp., té cargado ("loaded")
  - Fr., thé fort ("strong")
  - Ger., starker Tee ("strong")
  - Rus., krepkiy chay ("firm")
  - Shona, tea hobvu ("fat")
  - Ar. šay ṭaqiil("heavy")
  - etc.
Meaning-Text Theory

- Started in 60s, with ongoing work to this day
- A very formal theory of language
- Wants to do machine translation
- Needed dictionaries that captured collocations and other associations (the Explanatory Combinatorial Dictionary)
- Lexical Functions
Lexical Functions

- A notation to describe semantic relationships, for derivations and collocations
- They have to be general, within a language and cross-linguistically
- Latinate names: **Magn**(um), **Bon**(um), **Centr**(um), **Oper**(are), **Incep**(< *incipere*), etc.
- There are about 60 of them
**Magn: "very, intensely, to a high degree"**

- Lots of collocations for this (across different word classes):
  - **Magn**(tea) = strong
  - **Magn**(patience) = infinite
  - **Magn**(smoker) = heavy
  - **Magn**(easy) = very, as pie
  - **Magn**(skinny) = as a rake
  - **Magn**(prepared) = carefully
  - **Magn**(condemn) = strongly
  - **Magn**(laugh) = heartily, one’s head(etc.) off
  - Japanese: **Magn**(uso, "lie") = makka "crimson"
Magn: "very, intensely, to a high degree"

- The magnitude might apply to different domains:
  - $\text{Magn}_{\text{temp}}(\text{experience}) = \text{long}$
  - $\text{Magn}_{\text{quant}}(\text{experience}) = \text{considerable}$
  - $\text{Magn}_{\text{temp}}(\text{shortage}) = \text{chronic}$

- Can be negated (less rich area in the lexicon, $\sim 10:1$, $n=2$):
  - $\text{AntiMagn}(\text{applause}) = \text{scattered}$
  - $\text{Magn}(\text{temperature}) = \text{high}$
  - $\text{AntiMagn}(\text{temperature}) = \text{low}$
**Magn**: "very, intensely, to a high degree"

- This LF and a few other admit degrees:
  - \( \text{Magn}(\text{cost}) = \text{high, significant} < \text{huge} < \text{exorbitant} < \text{astronomical} \)
  - \( \text{Magn}(\text{toll}) = \text{heavy} < \text{devastating} \)

- Some words have many \textbf{Magn} options:
  - \( \text{Magn}(\text{conviction}) = \text{strong, firm, deep, thorough} \)
Polysemy and Synonymy: "Sick"

- \( \text{Magn}(sick\#1) = \text{very} < \text{as a dog} \)
Polysemy and Synonymy: "Sick"

- $\text{Magn}(sick\#1) = \text{very} < \text{as a dog}$
- $\text{Magn}(sick\#2) = \text{completely} < \text{to death}$
Polysemy and Synonymy: "Sick"

- \textbf{Magn}(sick\#1) = very < as a dog
  \textbf{Magn}(sick\#2) = completely < to death

- \textbf{Magn}(ill) = deathly, gravely
  \textbf{Magn}(ill) = *to death, *as a dog
  \textbf{Magn}(sick\#1) = ?deathly, gravely?
Polysemy and Synonymy: "Sick"

- \( \text{Magn}(sick\#1) = \text{very } < \text{ as a dog} \)
  \( \text{Magn}(sick\#2) = \text{completely } < \text{ to death} \)

- \( \text{Magn}(ill) = \text{deathly, gravely} \)
  \( \text{Magn}(ill) = \text{*to death} \)
  \( \text{Magn}(sick\#1) = \text{?deathly, gravely?} \)

- \( \text{IncepCopul}(sick\#1) = \text{become, get, fall, take} \)
  \( \text{IncepCopul}(ill) = \text{become, get(?), fall, take(?)} \)
  \( \text{IncepCopul}(sick\#2) = \text{become, get} \)
**Magn(Kílta)**

- **Os dust** has associations with entropy, time, mortality:
  - **Magn(kár, "sickness") = osarin "dusty"** (= deathly, grave)
  - **Magn[psych](kár) = luikin "heavy"** (= difficult)
  - But also: **osarin tirchoma "lifeless" desert**

- **Magn(ákin, "humble") = tokui "low"
  - **Magn(hottin, "trite, sentimental") = kummë "fatty, greasy"
  - **Magn(vós, "plague, epidemic") = mahin "hot"**
Magn(Mwanele) (by Miacomet)

- **Magn(ŋolu, "tea") = genó ("wide")**
  - **AntiMagn(ŋolu, "tea") = edaŋ ("thin")**

- Serial verb constructions can be used:
  - **Magn(gepwu, "stop") = ŋwelok ("fall down") (= "come to a crashing halt")**
  - **Magn(eméŋi, "run") = mebi ("be awake") (= quickly)**
Magn(Karyol) (by Zeke Fordsmender)

- **Magn(hargya, "coffee") = koñoñah ("piquant, spicy")**
- **Magn(tāo, "pain") = koñoñah ("piquant, spicy")**
- **AntiMagn(tāo, "pain") = bāge ("bland, insipid")**
- **AntiMagn(ecyōam, "surprise") = bāge ("bland, insipid")**

- **Magn(kaha, "liquor") = tohba ("heavy")**
- **Magn(myāle, "sound") = tohba ("heavy")**
Light Verb Constructions

- I walked to the store vs. I took a walk to the store.

- Cross-linguistically common:
  Eng. make a decision
  Fr. prendre une décision (take)
  Grm. eine Entscheidung treffen (meet)
  Turk. vermek (give)
  Ko. haerida/naerida (do, put down)
  etc.
Light Verb Constructions

- *I walked* to the store vs. *I took a walk* to the store.

- Cross-linguistically common:
  Eng. *make a decision*
  Fr. *prendre une décision* (take)
  Grm. *eine Entscheidung treffen* (meet)
  Turk. *vermek* (give)
  Ko. *haerida/naerida* (do, put down)
  etc.

- more potential for modification, dances with topicality and focus
MTT: "pseudo-predicates"

- Many non-verb words have implicit arguments.

- "Investigation": (= "investigate")
  1st = the person investigating ("subject")
  2nd = the person or thing being investigated ("object")

- These argument assignments are arbitrary but identified in a MTT dictionary.

- Some complex LFs reference different arguments.
\textbf{Oper}_n: "to perform, do, make, or have"

- The subscript indicates which argument is the subject of the LVC

- \textbf{Oper}_1(\textit{decision}) = make
- \textbf{Oper}_1(\textit{lecture}) = give, deliver
- \textbf{Oper}_1(\textit{search}) = carry out, conduct, do, make
- \textbf{Oper}_1(\textit{idea}) = have
- \textbf{Oper}_1(\textit{resistance}) = offer, put up
- \textbf{Oper}_2(\textit{resistance}) = meet, run into
- \textbf{Oper}_1(\textit{risk}) = pose, present
- \textbf{Oper}_2(\textit{risk}) = run
- \textbf{Oper}_1(\textit{control}) = have [\sim \text{ over N}]
- \textbf{Oper}_2(\textit{control}) = be [\text{under N's } \sim]
IncepOper\(_n\): "start to do, make, or have"

- \(\text{IncepOper}_1(fame) = \text{achieve}\)
- \(\text{IncepOper}_1(command) = \text{take}\)
- \(\text{IncepOper}_1(power) = \text{gain}\)
- \(\text{IncepOper}_1(victory) = \text{achieve, gain, score, win}\)
- \(\text{IncepOper}_2(injury) = \text{sustain}\)
- \(\text{IncepOper}_2(setback) = \text{suffer}\)
- \(\text{IncepOper}_2(difficulty) = \text{run into}\)
**IncepOper** \(n\): "start to do, make, or have"

- **IncepOper** \(_1\)(fame) = achieve
- **IncepOper** \(_1\)(command) = take
- **IncepOper** \(_1\)(power) = gain
- **IncepOper** \(_1\)(victory) = achieve, gain, score, win
- **IncepOper** \(_2\)(injury) = sustain
- **IncepOper** \(_2\)(setback) = suffer
- **IncepOper** \(_2\)(difficulty) = run into

- Kílta doesn’t need any light verbs just for **IncepOper**. It has an affix to mark this sense on the normal **Oper** verb.
**Ver**: "as it should be, meeting requirements"

- **Ver**(instrument) = precise
  **Ver**(argument) = convincing, valid, sound
  **Ver**(argue) = convincingly, strongly
  **Ver**(lie) = plausible
- **AntiVer**(fear) = unfounded
- **AntiVer**(lie) = obvious, transparent
- **Ver**(walk) = steadily

- **Kílta:**
  - **Ver**(ësikëroma, "PR synthetic reality") = ~ **lairo** ("grow")
    (the -om- element means earth, land, world)
  - **Ver**(alúta, "model, example") = **kemin** ("neatly fit together, unobstructed")
**Bon**: conventional praise

- **Bon**(different) = refreshingly
  **Bon**(future) = bright
  **Bon**(idea) = bright
  **Bon**(smile) = dazzling

- **AntiBon**(car) = //lemon
  **AntiBon**(start) = rocky
**Bon:** conventional praise

- **Bon**(different) = refreshingly
  **Bon**(future) = bright
  **Bon**(idea) = bright
  **Bon**(smile) = dazzling

- **AntiBon**(car) = //lemon
  **AntiBon**(start) = rocky

- It can sometimes be difficult to distinguish **Magn**, **Ver**, and **Bon**. For a conlanger, this doesn’t matter a great deal.
Any conlang aiming at "natural" should involve a hefty dose of collocations.

After ~150-200 content words it should be a normal part of vocabulary process.

Lexical Functions are very useful shorthand in conlanging notes, not just for collocations: derivations, word relations, exploring concept domains.
Links

- This talk: https://lingweenie.org/conlang/LCC2021/