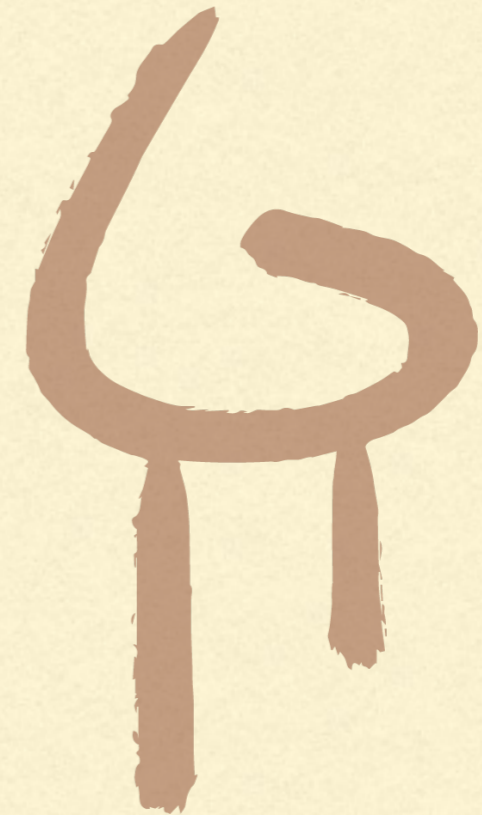


---

**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., *krepkiiy chay* ("firm")
    - Shona, *tea hobvu* ("fat")
    - Ar. *šay taqiil* ("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:
    - **Magn**[temp](*experience*) = long
    - **Magn**[quant](*experience*) = considerable
    - **Magn**[temp](*shortage*) = chronic
  - Can be negated (less rich area in the lexicon, ~10:1,  $n=2$ ):
    - **AntiMagn**(*applause*) = scattered
    - **Magn**(*temperature*) = high
    - **AntiMagn**(*temperature*) = low
-



---

# **Magn:** "very, intensely, to a high degree"

---

- This LF and a few other admit degrees:
    - **Magn**(*cost*) = high, significant < huge < exorbitant  
< astronomical
    - **Magn**(*toll*) = heavy < devastating
  - Some words have many **Magn** options:
    - **Magn**(*conviction*) = strong, firm, deep, thorough
-

---

# Polysemy and Synonymy: "Sick"

---

- **Magn**(*sick#1*) = very < as a dog

---

# Polysemy and Synonymy: "Sick"

---

- **Magn**(*sick#1*) = very < as a dog  
**Magn**(*sick#2*) = completely < to death

---

# Polysemy and Synonymy: "Sick"

---

- **Magn**(*sick#1*) = very < as a dog  
**Magn**(*sick#2*) = completely < to death
  
  - **Magn**(*ill*) = deathly, gravely  
**Magn**(*ill*) = \*to death, \*as a dog  
**Magn**(*sick#1*) = ?deathly, gravely?
-

---

# Polysemy and Synonymy: "Sick"

---

- **Magn**(*sick#1*) = very < as a dog  
**Magn**(*sick#2*) = completely < to death
  - **Magn**(*ill*) = deathly, gravely  
**Magn**(*ill*) = \*to death  
**Magn**(*sick#1*) = ?deathly, gravely?
  - **IncepCopul**(*sick#1*) = become, get, fall, take  
**IncepCopul**(*ill*) = become, get(?), fall, take(?)  
**IncepCopul**(*sick#2*) = become, get
-

---

# Magn(Kíлта)

---

- **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(Mwaneḷe) (by Miacomet)

---

- **Magn(ḡolu, "tea") = genó ("wide")**  
**AntiMagn(ḡolu, "tea") = edaḡ ("thin")**
  - Serial verb constructions can be used:  
**Magn(gepwu, "stop") = ḡwelok ("fall down") (= "come to a crashing halt")**  
**Magn(eméḡi, "run") = mebi ("be awake") (= quickly)**
-

---

# Magn(Karyol) (by Zeke Fordsmender)

---

- **Magn(hargya, "coffee") = koñoᅇah ("piquant, spicy")**  
**Magn(tāo, "pain") = koñoᅇah ("piquant, spicy")**  
**AntiMagn(tāo, "pain") = bāge ("bland, insipid")**  
**AntiMagn(ecyōam, "surprise") = bāge ("bland, insipid")**
  - **Magn(kaha, "liquor") = tohba ("heavy")**  
**Magn(myāᅇe, "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.
-

---

# **Oper<sub>n</sub>**: "to perform, do, make, or have"

---

- The subscript indicates which argument is the subject of the LVC
  - **Oper<sub>1</sub>**(*decision*) = make
  - **Oper<sub>1</sub>**(*lecture*) = give, deliver
  - **Oper<sub>1</sub>**(*search*) = carry out, conduct, do, make
  - **Oper<sub>1</sub>**(*idea*) = have
  - **Oper<sub>1</sub>**(*resistance*) = offer, put up
  - **Oper<sub>2</sub>**(*resistance*) = meet, run into
  - **Oper<sub>1</sub>**(*risk*) = pose, present
  - **Oper<sub>2</sub>**(*risk*) = run
  - **Oper<sub>1</sub>**(*control*) = have [~ over N]
  - **Oper<sub>2</sub>**(*control*) = be [under N's ~]
-

---

# **IncepOper<sub>n</sub>**: "start to do, make, or have"

---

- **IncepOper<sub>1</sub>**(*fame*) = achieve
  - IncepOper<sub>1</sub>**(*command*) = take
  - IncepOper<sub>1</sub>**(*power*) = gain
  - IncepOper<sub>1</sub>**(*victory*) = achieve, gain, score, win
  - IncepOper<sub>2</sub>**(*injury*) = sustain
  - IncepOper<sub>2</sub>**(*setback*) = suffer
  - IncepOper<sub>2</sub>**(*difficulty*) = run into
-

---

# **IncepOper<sub>n</sub>**: "start to do, make, or have"

---

- **IncepOper<sub>1</sub>**(*fame*) = achieve  
**IncepOper<sub>1</sub>**(*command*) = take  
**IncepOper<sub>1</sub>**(*power*) = gain  
**IncepOper<sub>1</sub>**(*victory*) = achieve, gain, score, win  
**IncepOper<sub>2</sub>**(*injury*) = sustain  
**IncepOper<sub>2</sub>**(*setback*) = suffer  
**IncepOper<sub>2</sub>**(*difficulty*) = run into
  - Kíлта 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
  - Kíлта:  
**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.
-

---

# Summary

---

- 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/>
-