A typology of semilexicality and the locus of grammatical variation

Chenchen Song

School of International Studies
Zhejiang University

The 9th International Conference of Formal Linguistics
Shanghai, Nov 5–7, 2021

Introduction

Vocabulary items are usually divided into two big classes based on their grammatical status—lexical or functional—but many have argued that this dichotomy is too coarse and that there are also semilexical items.

> *Although the distinction between content words and function words is a useful one, it seems too coarse. There are content words with a degree of "functionalness" and there are function words having a degree of "lexicalness." ...[F]or a better understanding of semilexicality a great deal of in-depth research is required. (Corver et al. 2001:10)*

Research in the past 20 years has accumulated some valuable results, but three critical questions remain unanswered.

- The WHAT question: What types of semilexicality are out there?

- The WHY question: Why are there such different types?

- The HOW question: How to formally derive/reconcile those types?

Introduction

My goal in this pilot study is quite humble. I will focus on the WHAT question and only slightly touch on the WHY/HOW questions.

Plan

1. A preliminary typology of semilexicality (main)
2. Correlation between semilexicality and morphological types
3. Derivation of semilexical elements
4. Implication for theory of grammatical variation

1. Introduction

2. Chinese (and other isolating languages)

3. Previous studies on semilexicality (mainly on European languages)

4. An unexplored area: semilexicality and polysynthesis

5. A little bit of theory

6. Summary

Chinese: A quintessential case

In Song (2019), I did a comprehensive survey of semilexical items in Chinese, which I termed *semifunctional* due to their stably functional status. Chinese is extremely abundant in such items:

- V-domain: light verbs, auxiliaries, sentence-final particles

- N-domain: classifiers, pronouns

- P-domain: "prepositions"

- Others: logical operators (AND, OR)

```
primary status:  functional
idiosyncrasies:  encyclopedic, stylistic
lexical usage?:  yes
```

Examples

(1) a. *Fànrén zuótiān bèi/?ràng/?gěi/?jiào jǐngchá dàibǔ le.* (passive)
    criminal yesterday PASS   police arrest PRF
    "The criminal was arrested by the police yesterday." `formal`

   b. *Nǐ-de   jiāqiǎor bèi/?ràng/gěi/??jiào māo chī le.*
      you-POSS sparrow PASS   cat eat PRF
      "Your sparrow was eaten by a cat!" `colloquial`

(2) a. *yì   zhī zhū* (classifier)
       one CL pig
       "a pig" `neutral`

   b. *yì   tóu zhū*
      one CL pig
      "a pig" `pejorative`

(3) a. *hālì bōtè yǔ/?hé/?gēn mófǎ-shí* (conjunction)
       Harry Potter and   magic-stone
       "*Harry Potter and the Philosopher's Stone*" `literary`

   b. *Xiǎomíng hé/gēn/?yǔ xiǎohóng dōu zài wàitou wánr.*
      Xiaoming and   Xiaohong all be.at outside play
      "Both Xiaoming and Xiaohong are playing outside." `colloquial`

High analyticity

Such omnipresent and flexible semifunctionality is a typical feature of highly analytic (aka isolating) languages. Another example is Vietnamese.

Example: Vietnamese negators (Wiktionary; Li Nguyen, p.c.)

không (default), *chẳng* (emphatic), *chả* (emphatic, informal), *đâu* (emphatic, colloquial), *nào* (colloquial but elevated), *đếch* (mildly vulgar), *đéo* (very vulgar), *cóc* (very informal)

High analyticity

Such omnipresent and flexible semifunctionality is a typical feature of highly analytic (aka isolating) languages. Another example is Vietnamese.

Example: Vietnamese negators (Wiktionary; Li Nguyen, p.c.)

không (default), *chẳng* (emphatic), *chả* (emphatic, informal), *đâu* (emphatic, colloquial), *nào* (colloquial but elevated), *đếch* (mildly vulgar), *đéo* (very vulgar), *cóc* (very informal)

In addition, many semilexical items are multifunctional; e.g., *không*

1. Original lexical meaning: empty, nothing, none (空); e.g., *tay không* 'empty handed'
2. Adj. only; e.g., *cơm không* 'only rice'
3. Interj. no; e.g., *Không, tôi không hút thuốc.* 'No, I do not smoke.'
4. Yes-no particle; e.g., *Đúng không?* 'Is that true?'
5. Num. zero; e.g., *hai nghìn không trăm lẻ một* '2001 (lit. 2 thousand 0 hundred 0 one)'

(Wiktionary, Tra Tu)

Variation

Not all highly analytic languages show the kind of massive semifunctionality in Chinese and Vietnamese. For example, Yoruba is far less abundant in semifunctional items, nor does it show noticeable stylistic variation.

Example: Yoruba conjuctions (Bowen 1858)

ti (for personal pronouns and verbs), *ati* (for nouns, pronouns, adverbs, and prepositions, but not for verbs), *oŋ* (preferably for nouns), *kpęlu* (for nouns and occasionally pronouns, ≺'with'), *sì* (for verbs, often with a pronoun, ≺*si ęhiŋ* 'go backwards')

The variation here is conditioned by categories rather than by pragmatics.

Two types of isolating language

A quick survey of reference grammars reveals that isolating languages worldwide are either more like Chinese (mostly languages in Southeast Asia) or more like Yoruba in terms of their semifunctional items.

- The Chinese type (e.g., Burmese, Lao, and especially Vietnamese)

    - Huge inventory
    - Systematic polysemy (often including original meaning)
    - Highly flexible usage
    - Alternation mainly conditioned by register or speaker attitude

- The Yoruba type (e.g., Samoan, Kaingáng)

    - Much smaller inventory
    - Less conspicuous polysemy
    - Less flexible usage
    - Alternation (if any) mainly subject to grammatical constraints

Examples

- **Burmese**: about 36 pre-head and 42 post-head "versatile" auxiliary verbs, which all have synchronic lexical uses (Soe 1999:118ff.)

- **Lao**: over 30 aspectual-modal markers, most of which also function as open-class items (mostly verbs); further modality distinctions are made by sentence-final particles (at least 30) (Enfield 2007:5,174)

- **Samoan**: no sharp-cut distinction between grammatical and lexical morphemes, which frequently show etymological relationship (Mosel & Hovdhaugen 1992:140ff.), but except for the adverbial, adpositional, and emphatic particles (which have more lexical content), other particles (e.g., 11 TAM, 2 Neg) mostly only have grammatically conditioned usage

- **Kaingáng**: altogether 12 or so particles, about 6 of which are "predicative elements" used to form predicates, mostly with relatable lexical meanings and lexically conditioned usage; some have developed tense/modal uses (Henry 1948:200ff.), but there is no stylistic variation

Outline

Literature on semilexicality

Previous studies using the term "semilexicality" are mostly on familiar Indo-European languages, which are typically synthetic (except for English and Afrikaans, which are more analytic). To get an impression of what researchers usually mean by "semilexicality," let's do a quick literature review.

Corver et al. (2001) is the only book-length publication (an edited volume) on semilexicality. Apart from familiar cases like classifiers and light verbs, it also covers some other reportedly semilexical categories, mainly from Germanic and Romance languages.

I. Special uses of motion verbs

Motion verbs in Romance/Germanic (Cardinaletti & Giusti 2001):

- As VP-shell elements ("semilexical")
- As auxiliaries ("functional")

(4) Examples of VP-shell usage:

    a.   *Vaju a piggiu u pani.*   [Marsalese]
        go.1SG to fetch.1SG the bread
        "I go to fetch bread."

    b.   *Jag går   och gör   mig en grogg.*   [Swedish]
        I   go.PRES and make.PRES myself a grogg
        "I go to make myself a grogg."

    c.   I go buy bread. / John will go visit Harry tomorrow.   [English]

I. Special uses of motion verbs

Motion verbs in Romance/Germanic (Cardinaletti & Giusti 2001):

- As VP-shell elements ("semilexical")
- As auxiliaries ("functional")

(5) Examples of auxiliary usage:

    a.   *La pasta va   / viene mangiata subito.*   [Italian]
         the pasta goes comes eaten   immediately
         "Pasta must be / is eaten immediately."

    b.   *Il va partir. / Il vient d'arriver.*   [French]
         he goes leave   he comes of.arrive
         "He will leave. / He has just arrived."

    c.   He is going to leave.   [English]

> This is closer to the Chinese style, though R/G languages often require extra elements around semilexical verbs (e.g., French *de*, English *to*).

I. Special uses of motion verbs

Motion verbs in Romance/Germanic (Cardinaletti & Giusti 2001):

- As VP-shell elements ("semilexical")
- As auxiliaries ("functional")

The AUX usage of motion/posture verbs is also studied in Pots (2020).

(6) a. *Ik heb de hele dag zitten te lezen.*   [Dutch]
          I   have the entire day sit   to read
          "I have been reading the entire day."

     b. *Ek het gister baie (ge-)loop (en) praat.*   [Afrikaans]
          I   have yesterday a.lot PRF-walk and talk
          "I have been (walking and) talking a lot yesterday."

Here too the semilexical verbs require extra grammatical elements (e.g.,
infinitive/participle markers, conjunctions), though the "bare" version
of the Afrikaans sentence is highly analytic (category/word = 1).

II. Lexical realizations of Pred

Predicative operators in Norwegian (Eide & Åfarli 2001):

- Serve to form the subject-predicate relation (via PredP)
- May be lexicalized by several types of elements

(7) a. *anse   Jon som gal*   [Norwegian]
         regard Jon as   crazy
         "regard Jon as crazy"

     b. *gjøre Jon til forbryter*
         make Jon to criminal
         "make Jon into a criminal"

     c. *ta   Jon for kelner*
         take Jon for waiter
         "take Jon for being a waiter"

> E & Å: These prepositions lexicalize Pred in small clauses, they are like
> nonverbal copulas (*som* 'as'–*være* 'be', *til* 'to'–*bli* 'become').

Old Norse complementizers

A similar case in Old Norse (C in relative clauses, Åfarli 1995):

(8) a. *kringla heimsins sú er mannfólkit byggir.*   [Old Norse]
         "the world that men live in."

  b. *þau helgu orð en í bókinni váru.*
      "those holy words that were written in the book."

  c. *gera hús þar sem eigi hafi fyrr verit.*
      "build a house where there had been no house before."

  d. *þeir allir, at flau tíðindi heyrðu.*
      "everybody that heard these news."

> E & Å: Despite their differences in frequency and geographical or historical distribution, these C items often cooccur in the same text and have no clear semantic difference. They all lexicalize a higher Pred.

III. Postpositions qua little p heads

Postpositions in German and Dutch (Zeller 2001):

- "Inflected" prepositions consisting a lexical element with a functional suffix (little p in Van Riemsdijk 1990)
- "Prepositional proforms" when there is no overt PP

(9) a. *Ein Tourist steigt (auf) den Berg   hin-auf.*   [German]
         a   tourist climbs up/on the mountain thither-up/on
         "A tourist climbs up the mountain."

     b. *Ein Tourist steigt   hin-auf.*
         a   tourist climbs thither-up/on
         "A tourist climbs up."

     c. *Het vliegtuig is onder de   brug   door   gevlogen.*   [Dutch]
         the airplane is under the bridge through flown
         "The airplane has flown through under the bridge."

     d. *omdat Jan de berg   op reed.*
         because Jan the mountain up drove
         "because Jan drove up the mountain."

IV. Acedo-Matellán & Real-Puigdollers (2019)

Derivational affixes behave differently from inflectional ones, since they carry conceptual meanings in addition to their categorizing functionality and have a variety of nonallomorphic exponents.

- English: *-ity*, *-al*, *-ic*, *-ness*...

- Catalan:
    - *mols-ós* 'abundant in moss' vs. *mols-ut* '(soft or fleshy) like moss'
    - *ferr-ós* 'containing or abundant in iron' vs. *ferr-ís* 'strong as iron'
    - *cendr-aire* 'person who buys and sells ash' vs. *cendr-er* 'ashtray'

Interim summary I

Except for the Old Norse case, phenomena labeled "semilexical" in previous studies usually concern categories near the lexical end:

- categorizers, light/shell categories, classifiers, Pred, Asp, Voice

They usually do not concern categories higher up in the tree:

- D (including pronouns), T, C, logical operators ($\wedge$, $\vee$, $\neg$)

Recall that these typical functional categories are also often semilexical in isolating languages (especially in those of the Chinese type).

> I call semilexicality near the lexical end *prototypical semilexicality*.

Interim summary II

A key feature that distinguishes Chinese-type isolating languages (especially Chinese and Vietnamese) on the one hand and Yoruba-type isolating languages and the familiar European fusional languages on the other hand is stylistic conditioning on the choice of function word.

> Feng (2010, 2012): Register plays a vital role in Chinese grammar.

Outline

Polysynthetic languages

The polar opposite of isolating languages:

- Isolating: category/word $= 1$
- Polysynthetic: category/word $= n$ (could be very large)

Yet strikingly, they also have many semilexical elements, which go by the name *field* or *lexical affixes* in the literature.

> **WE ARE ENTERING UNEXPLORED TERRITORY!** 🚀
>
> To my knowledge, no previous study on semilexicality has touched on polysynthetic languages, nor have studies on lexical affixes considered the phenomenon from the angle of semilexicality.

Lexical affixes

[B]ased on Mithun (1988), Fortescue (1994), and Drossard (1997), a language is acknowledged as polysynthetic here if it fulfills the following formal and semantic criteria:

(i) There are ... polymorphemic verb forms ... formed with the help of nonroot bound morphemes with rather "lexical" meaning and, optionally, through concatenation of more than one lexical root (i.e., allowing noun incorporation and verb serialization), and

(ii) these forms allow for components representing the following categories, at least one of which by nonroot bound morphemes: event or participant classification and quantification, setting..., location or direction, motion, instrument..., manner..., modality..., degree, scale ... and focus, chronology..., as well as ... the usual categories such as valence, voice, central participants, [TAM], and polarity.

(Mattissen 2002:385)

(NB: Mattissen uses the term "root" in a traditional morphological sense.)

Example I: Verbal classifiers

Verbal classifiers are similar in function to numeral classifiers except that they are manifested on verbs. They exist in many polysynthetic languages, such as Innu (North America), Palikur (South America), Waris (Papua New Guinea), and Gumuz (Ethiopia) (Aikhenvald 2017).

(10) a. *Mâk ni-tâpiškâkan mišta-miš-eci-ší-pan.*   [Innu]
          and 1SG-scarf very-big-CL$_{sheetlike}$-ANIM.INTR-PRT
          "And my scarf was very large."

      b. *Putitê-y-ašku-mutâ-w   ukutuškwêw-a anitêhêmicwâp-it.*
          inside-LK-CL$_{long.rigid}$-install-TI2–3 pipe-PL   there
          "He installs the pipes there inside the house."   (Aikhenvald 2017:377)

(11) a. *Wonda ka-m   mwan-vra-ho-o.*   [Waris]
          net.bag 1SG-DAT CL$_{soft.pliable}$-get-BEN-IMP
          "Give me a net bag!"

      b. *Nenas   ka-m   li-ra-ho-o.*
          pineapple 2SG-DAT CL$_{compact}$-get-BEN-IMP
          "Give me a pineapple!"   (Brown 1981:95)

Example II: Enriched derivational affixes

Derivational affixes in polysynthetic languages tend to be enriched with diversified lexical semantic meanings, which are a lot more concrete than those in derivational affixes of familiar European languages.

Example: Inuktitut nominalizers (Beach 2011)

-ji/ti	'one that does'	*-nirsaq*	'one that is more'
-siti	'one that does well'	*-nirpaaq*	'one that is the most'
-suuq	'one that habitually does'	*-vik*	'place or time of'
-lik	'one that has'	*-utik*	'device or reason for'
-gunaq	'one that seems to'	*-usiq*	'way or means of'
-gialik	'one that needs to'	...	

---

e.g.,   *nalli-gi-ji-ga*   'love-TR-NML$_{one.that.does}$-my; one that loves me'
       aangajaa-suuq   'be.drunk-NML$_{one.that.hab.does}$; one that is habitually drunk'
       illu-lik   'house-NML$_{one.that.has}$; one that has houses'

Example II: Enriched derivational affixes

Derivational affixes in polysynthetic languages tend to be enriched with diversified lexical semantic meanings, which are a lot more concrete than those in derivational affixes of familiar European languages.

Example: West Greenlandic verbalizers (Fortescue 1984)

-u	'be'	*-ssaaliqi*	'lack'
-nngur	'become'	*-liri*	'have pain in'
-miit/niit	'be in'	*-nngu*	'have pain (weaker)'
-gar	'have'	*-niar*	'hunt'
-gasag	'have lots of'	*-siur*	'look for'
-gig	'have (a) good'	*-liar*	'go to'
-lug	'have (a) bad'	*-miir*	'come from'
-it	'not have any'	...	

e.g., *atisa-ssaaliqi-vunga* 'clothes-VBL$_{lack}$-INFL; I lack clothes'
isi-liri-vuq 'eye-VBL$_{have.pain.in}$-INFL; he has a pain in his eye'
Qaqurtu-liar-pugut 'Qaqortoq-VBL$_{go.to}$-INFL; we went to Qaqortoq'

Example III: Quasi-IN affixes

Noun incorporation is highly common in polysynthetic languages, but there are also languages (mainly in northwestern North America) that have *quasi–incorporated noun (quasi-IN) affixes* instead of or in addition to true INs. Quasi INs have no free-standing (i.e., root) usage.

Example: Bella Coola (Mithun 1997:358–359)

-us 'face'	*-ak* 'hand'	*-anł* 'covering'
-an 'ear'	*-łp* 'tree'	*-ank* 'side'
-uc 'mouth'	*-lst* 'rock'	*-ulmx* 'earth, floor'
-alic 'tooth'	*-alus* 'piece'	*-mx* 'member of a tribe'
-ał 'foot'	*-lt* 'child'	...

Example III: Quasi-IN affixes

Quasi INs often have more general or diffuse meanings compared to their (etymologically related or unrelated) root counterparts.

[E]ach lexical suffix can probably best be viewed as representing a complex network of associations rather than a concrete or abstract base from which metaphorical extensions are made. (Montler 1986:66; via Mithun 1997:359)

(This is reminiscent of the distributed morphology notion *root*.)

Example: Bella Coola (Mithun 1997:360)

-*uc*   'mouth, food, opening, edge...'
        sqal-uc 'fruit', *squp-uc* 'beard', *kuɫh-uc* 'beach'...
        ʔulx̣-uc 'steal one's food', *namilc-uc* 'channel opens up'...
-*als*   'cheek, walls of containers, hollow structures...'
        stp-als 'freckle on cheek', *sɬs-als* 'bottle', *six-als* 'new pot'...

Example III: Quasi-IN affixes

Some quasi-IN affixes (or true-INs) have become classificatory stems.

(12) a. *cp-uɬ-ic*   *ti-yalk-uɬ-tx.*   [Bella Coola]
    wipe-CL_{round}-1/3   PROX-ball-CL_{round}-ART
    "I am wiping the ball."   (Davis & Saunders 1973; via Mithun 1997:361)

    b. *Ne-hra-taskw-ahkw-ha'*   *ha'*   *tsi:r.*   [Tuscarora]
    du-M-CL_{animal}-pick.up-SERIAL EMPH dog
    "He is a dog catcher."   (Williams 1976:60; via Rosen 1989:303)

    c. *Kassi' háh-'ič'á-sswí'-sa'.*   [Caddo]
    bead   PROG-CL_{eye}-string-PROG
    "She is stringing beads."   (Mithun 1984:865; via Rosen 1989:303)

Example III: Quasi-IN affixes

And some languages have further developed numeral classifiers.

(13) a. *łíxʷ-əqən   lisék*   [Halkomelem]
        three-CL_container sack
        "three sacks"

     b. *Te'cs-élə   kʷθə nə   mémənə.*
        eight-CL_people DET 1POS children
        "I have eight children."   (Gerdts & Hinkson 1996:10)

(14) a. *dikwh-okwł bołak*   [Yurok]
        three-CL_salmon salmon
        "three salmon"

     b. *nahks-oh   ha'aag*
        three-CL_round rock
        "three rocks"   (Conathan 2004:26–27)

Example IV: TAM markers

Tense-aspect-modality markers in polysynthetic languages are often semilexical too. Again, take West Greenlandic for example.

> *There are around fifty productive affixes concerned with aspect [in West Greenlandic]. It is possible to regard them, together with affixes of manner and degree, as a subcategory of derivational Aktionsart. (Fortescue 1984:277)*

(See also Song 2016 for the notion of nontemporal Aktionsart.)

Example: West Greenlandic aspect markers (Fortescue 1980:276–277)

-gajug 'often', *-juaar* 'continuously', *-llatsiar* 'for a short while', *-llattaar* 'from time to time', *-mmirsur* 'for some time', *-qattaar* 'again and again', *-ussar* 'keep on', *-arsug* 'half-heartedly', *-gasuar* 'quickly', *-jummir* 'all at once', *-liqqissaar* 'just about to', *-junnaar* 'no longer', *-sima/nikuu* 'perfective', *-(r)sari* 'be in middle of', *-kannir* 'more or less', *-laar* 'a little', *-piar* 'exactly'...

Example V: SA markers

And if speaker attitude is a functional category (as many has argued in recent years), then items in this category are mostly semilexical too.

> *There is a sizeable class of sentential affixes [in West Greenlandic] ... whose principal common effect is to modify the verbal base according the* the speaker's attitude *towards the action or state he is describing. (Fortescue 1984:295)*

(These are a bit like sentence-final particles in Asian languages.)

Example: WG "subjective coloration" markers (Fortescue 1980:277)

-ataar 'exclamation', *-gallar/niar* 'polite imperative', *-(s)innar* 'just', *-kasig* 'disdain', *-llariaa* 'negative imperative strengthener', *-kulug* 'bad', *-(ria)llar* 'surprising action', *-(r)luinnar* 'really', *-nnguar* 'affection/comfort', *-qi* 'intensity', *-ratar* 'surprise', *-riannguar* 'discovery', *-vallaar* 'so very much'...

Suffixation vs. prefixation

So far, we've only seen lexical affixes in the form of suffixation, but there are also polysynthetic languages that mainly rely on prefixation, and those languages have plenty of semilexical elements too.

Example: Slave verb structure (Rice 1989:425)

1	2	3	4	5	6	7	8	9	10	11	12	13
Adv	Dist	Cus	Inc	Num	DO	Dei	The	Asp	Conj	Mod	Sub	Stem

Positions 1, 4, 8, and 9 tend to be semilexical. The basic verb meaning is codetermined by 1-4-5-6-8, which together constitute the "verb theme."

e.g., (Adv, 1) *'a-*   'home'   (Inc, 4)   *fi-*   'head'
      'ejih-   'run away, undo'   *nah-*   'eye'
      'eshíh-   'pretend'   *da-*   'mouth'
      'ets'ǫ-   'lose, misplace'   *la-*   'work'

(Like quasi-IN suffixes, these "incorporated stems" have generalized meanings.)

Suffixation vs. prefixation

So far, we've only seen lexical affixes in the form of suffixation, but there are also polysynthetic languages that mainly rely on prefixation, and those languages have plenty of semilexical elements too.

Example: Lavrung verb structure (Lai 2017:293)

-11	-10	-9	-8	-7	-6 ~ -2	-1	0	1	2
Prog/Sup	Dir$_{TAM}$	Inv/Irre	Neg/Q	Cond	Voice/Appl	Inc	V	Agr	Red

Positions -1 and -10 are clearly semilexical. The verb root and an optional quasi IN together constitute the verb stem. Directional prefixes (-10) also mark TAM.

e.g., (Dir, -12)	*æ-*	'upward$_{PST}$'	(Inc, -1)	*xjár/sêm-*	'heart'
	næ-	'downward$_{IPFV.PST}$'		*fɕî-*	'tooth'
	kə-	'upstream$_{PFV}$'		*gúŋ-*	'price'
	rə-	'toward$_{NPST}$'		*tsʰâ-*	'good'

(Lavrung—and Gyalrongic languages [嘉絨語] in general—also has numeral classifiers and even Chinese-style sentence-final particles; see Huang 2007 and Jacques 2008)

Interim summary III

Semilexicality is a general feature of polysynthetic languages regardless of the direction of affixation. Characteristics:

- Many categories (pretty much omnipresent)
- Huge inventory (tens and even hundreds)
- Highly concrete lexical content (almost like content words)
- Varied degrees of grammaticalization (but no free-standing usage)
- Occasional stylistic variation (but far less common than in Chinese)

In a sense, the kind of semilexicality in polysynthetic languages is quite like that in Chinese-type isolating languages.

One big difference*: morphological status (affix vs. free morpheme)*

Interim summary IV

We can distinguish three types of semilexicality based on two dimensions.

Table: A two-dimensional view of semilexicality

	purely lex. sem.	purely fun. sem.	lex. & fun. sem.
free	canonically lex.	analytic	semilexical (analytic)
bound	semilexical (mor.)	canonically fun.	semilexical (synthetic)

Most previous studies on semilexicality (e.g., Corver et al. 2001, Song 2019, Pots 2020) have studied *analytic semilexicality*, fewer (e.g., Acedo-Matellán & Real-Puigdollers 2019) have touched on *synthetic semilexicality*, and none has considered *morphological semilexicality* a type of semilexicality.

Outline

Deriving semilexicality

In current generative syntax, purely lexical or idiosyncratic information is popularly encoded in categoryless *roots*. Two ways to enrich a functional category X with root information have been proposed:

- Root support (Song 2019)
  $[_X \ X \ \sqrt{\ }\ ]$ (syntactic Merge)

- Root insertion (Acedo-Matellán & Real-Puigdollers 2019)
  $X \Leftarrow \sqrt{\ }$ (postsyntactic insertion)

Song's (2019) theory is a direct, minimal extension of classical distributed morphology (with only a redefinition of "categorizer"), while A-M & R-P's (2019) theory assumes a greater deviation therefrom (with across-the-board late insertion and a modified grammatical architecture).

Deriving semilexicality

In current generative syntax, purely lexical or idiosyncratic information is popularly encoded in categoryless *roots*. Two ways to enrich a functional category X with root information have been proposed:

- Root support (Song 2019)
  $[_X \; X \; \sqrt{\;}]$ (syntactic Merge) ☞ analytic semilexical head

- Root insertion (Acedo-Matellán & Real-Puigdollers 2019)
  $X \Leftarrow \sqrt{\;}$ (postsyntactic insertion) ☞ synthetic semilexical head

Song's (2019) theory is a direct, minimal extension of classical distributed morphology (with only a redefinition of "categorizer"), while A-M & R-P's (2019) theory assumes a greater deviation therefrom (with across-the-board late insertion and a modified grammatical architecture).

> Both ideas may turn out to be useful in modeling semilexicality.

Deriving semilexicality

But there's arguably more than one way to derive synthetic semilexical heads. Two obvious alternatives are:

1. Root support + head movement
2. Root support + affixal spell-out

❶ is possible because head movement creates complex heads, which are synthetic by definition, while ❷ is possible due to the acategorial, amorphous nature of DM roots, which makes it possible to designate a particular spell-out form for each categorization. Thus, a root that spells out as a free morpheme when merged with a traditional lexical category (i.e., a DM little-*x*) may well spell out as an affix when merged with a non-little-*x* functional category.

> A detailed comparison of the three ways must be left to future research.

Deriving semilexicality

As for bound purely lexical morphemes (i.e., morphologically semilexical elements), maybe they are better treated as compound components. Thus, at least one common type of such lexical affixes—the modificational type—can be given a straightforward treatment in the root support theory, since the X in $[_X \text{ X } \sqrt{}]$ may be an underspecified category Cat, which creates adjuncts in a Set Merge-only fashion. In other words, being a modifier is also a function.

So, root support can potentially help us derive most of the semilexical elements we have seen. The only remaining case is that of nonmodifier, nonclassifier quasi INs, which must be given an ordinary compounding analysis and not treated as a type of semilexicality.

Example

- Analytical semilexical element:
  e.g., Chinese classifier: *tóu* ↔ [$_{Cl}$ Cl $\sqrt{\text{TÓU}}$]

- Synthetic semilexical element:
  e.g., WG aspect marker: *-gajug* ↔ [$_{Asp}$ Asp $\sqrt{\text{GAJUG}}$] + HM

- Morphologically semilexical element:
  e.g., Slave adverbial prefix: *'a-* ↔ [$_{Cat}$ Cat $\sqrt{\text{'A}}$] ([$_V$ Cat$_\sqrt{}$ V])

- Nonmodifier, nonclassifier quasi IN:
  e.g., *sjar-cʰǽ* 'heart-be.big; be.brave' ↔ [$_V$ v [$_N$ n $\sqrt{\text{SJAR}}$] [$_V$ v $\sqrt{\text{Cʰ Æ}}$]]

Locus of grammatical variation

> **Borer-Chomsky Conjecture (BCC; Baker 2008:156)**
>
> All parameters of variation are attributable to differences in the features of particular items (e.g., the functional heads) in the lexicon.

To the extent that the crosslinguistic variation in semilexicality type is predictable from certain regular factors (i.e., not entirely random), the BCC must be modified. Three factors may be relevant here:

- the structural nature of syntactic heads (bare or root-supported)
- the morphological nature of head exponents (free or bound)
- the language-specific situation of head movement (systematic or mixed)

Locus of grammatical variation

Only the last one falls under classical BCC, whereas the first two cannot be formulated in terms of formal features. They have more to do with lexical tendencies and hence are still variation *in the lexicon*, though.

Extended Borer-Chomsky Conjecture (EBCC; adapted from Song 2019:139)

All factors of variation are attributable to the lexicon, either to differences in the formal features of functional heads or to the ways those heads are realized.

This converges with a remark from Borer:

[A]ll variation, both within a language and across languages, is reducible not only to the properties of range assigners ... but [also] to their morphophonological properties. (Borer 2005:264)

Outline

1. Introduction

2. Chinese (and other isolating languages)

3. Previous studies on semilexicality (mainly on European languages)

4. An unexplored area: semilexicality and polysynthesis

5. A little bit of theory

6. **Summary**

Summary

Semilexical elements are found in a wide range of languages:

- Highly analytic languages: semifunctional items
- Familiar IE languages: (prototypically) semilexical items
- Polysynthetic languages: lexical/field affixes

Thus, semilexicality may turn out to be a universal in human language. I divided semilexical elements in human language into three types:

- Analytically semilexical element
- Synthetically semilexical element
- Morphologically semilexical element

Previous studies on semilexicality have mainly focused on the first type, but a complete theory of semilexicality should take all three types into account. Root support is a potential component in such a theory.

Remaining questions

For instance,

- Why are highly analytic and polysynthetic languages more abundant in semilexical elements?
— Perhaps due to their highly consistent syntactic setting.
- Why do Chinese-type isolating languages alone show conspicuous stylistic conditioning in their semilexical items?
— Perhaps an areal phenomenon.

I leave these and other remaining questions to future research.

Thank you!

Selected references I

📄 Acedo-Matellán, V. & C. Real-Puigdollers
Roots into functional nodes: Exploring locality and semi-lexicality
The Linguistic Review 36(3), 411–436

📄 Aikhenvald, A.
A typology of noun categorization devices
The Cambridge handbook of linguistic typology, 361–404
CUP, 2017

📄 Beach, M.
Studies in Inuktitut grammar
State University of New York at Buffalo dissertation, 2011

📕 Borer, H.
Structuring sense: In name only (Vol. 1)
OUP, 2005

Selected references II

📖 Corver, N., & H. van Riemsdijk
Semi-lexical categories: The content of function words and the function of content words
Mouton de Gruyter, 2001

📄 Feng, S.
语体语法: "形式-功能对应律" 的语言探索 [Register grammar: A linguistic exploration of the form-function mapping]
当代修辞学 [*Contemporary Rhetoric*] 6, 3–12

📖 Fortescue, M.
West Greenlandic
Croom Helm, 1984

Selected references III

📄 Lai, Y.
Grammaire du khroskyabs de Wobzi
Université Sorbonne Paris Cité dissertation, 2017

📄 Mattissen, J.
The missing link between different types of polysynthetic
languages
CLS 38 Proceedings, 385–399

📄 Mithun, M.
Lexical affixes and morphological typology
Essays on language function and language type, 357–371
John Benjamins, 1997

Selected references IV

📄 Pots, C.
Roots in Progress: Semi-lexicality in the Dutch and Afrikaans verbal domain
KU Leuven dissertation, 2020

📄 Song, C.
A minimalist study of complex verb formation: Crosslinguistic patterns and variation
University of Cambridge PhD first-year report, 2016

📄 Song, C.
On the formal flexibility of syntactic categories
University of Cambridge dissertation, 2019