

# The lexicalization of morphosyntactically complex expressions

David Beck, University of Alberta

Definitions of parts of speech or lexical classes are often restricted to single words or “lexical items,” defined as morphological stem plus affixes. This paper argues that, for nouns, the notion “lexical item” can be extended to any conventionalized expression meeting the semantic and syntactic criteria for nounhood, irrespective of internal morphosyntactic complexity.

Semantically, nominal entries in the lexicon should

- (1) a. have a *conventionalized* meaning expressing a semantic KIND (Wierzbicka 1988)
- b. be *conceptually autonomous* (Langacker 1991)

Syntactically, nominal entries in the lexicon should

- (2) a. be unmarked as syntactic actants
- b. be syntactically *closed* (*i.e.* must have an unfilled syntactic valency of zero)

Clearly, these semantic and syntactic criteria are complementary, level-specific instantiations of commensurate principles—particularly (1b) and (2b)—and serve as potential measures of *Weak Iconicity* (Beck 2002) at the semantics  $\Leftrightarrow$  syntax interface. More importantly,

|| these formulations make no reference to the internal syntactic make-up of lexical items, meaning that clausal or multi-word items with these properties can qualify as nouns.

Leaving aside (2a), I will examine the criteria remaining criteria in (1) and (2), beginning with **closedness** and **conceptual autonomy** (Section 1) and then moving on to **conventionality** (Section 2). Section 3 concludes by relating these ideas to Wierzbicka’s (1988) notion of KIND.

## 1) Closedness and conceptual autonomy

Prototypical common nouns are **conceptually autonomous** entities (Langacker 1991)

- they exist on their own and can be conceptualized independently of other entities
- they have no connotational/intensional semantic arguments

This lack of semantic arguments is reflected in the syntax in that prototypical common nouns have no (unmarked) syntactic actants and so are **syntactically closed**.

This pairing of a semantic and a syntactic feature is a reflection of the Principle of Weak Iconicity (Beck 2002):

### (3) *The Principle of Weak Iconicity*

In the unmarked case, syntactic structure will be isomorphic with, or a direct reflection of, its underlying semantic structure

This principle also makes the prediction that *non-isomorphism* between the semantic and the syntactic levels of representation will result in either

- cross-linguistically variable treatment of the elements in question (1.1), or
- use of additional morphosyntactic devices creating marked uses of lexical items (1.2)

### 1.1) Conceptually non- (or less-) autonomous nouns

**Relational nouns** such as bodypart and kinship terms are non-prototypical semantically in that they entail the existence of some other discrete, potentially individuable entity in their semantic profile. Thus, they are conceptually autonomous, but less so than common nouns:

- ✓ referents of relational nouns (RNs) exist on their own and clearly designate KINDS
- ✗ they require another entity as a *classificatory landmark* (CLM) for their identification

In many languages, this reduced conceptual autonomy is recognized through special possessor-marking strategies (Beck 2003):

#### Inherent possession

Inherently possessed nouns are syntactically non-closed in that they can not be used without the expression of their CLM as a possessor, as in Upper Necaxa Totonac (words in citation form bear the third-person possessive prefix *if-*):

<u>Upper Necaxa Totonac</u>				
(4)	<i>ifna:ná</i>	'his/her grandmother'	<i>istampí:n</i>	'its base, lower part'
	<i>ifnapa:skín</i>	'her sister-in-law'	<i>istampún</i>	'its bottom (cup, etc.)'
	<i>ifa?alo?ót</i>	'its horn'	<i>ifli:má:n</i>	'oneself'
	<i>iftfe?én</i>	'his/her/its leg'	<i>iftapát</i>	'its price, value'
	<i>if?ósní</i>	'its point, tip'	<i>iflakamatsát</i>	'his/her salted tortilla'

- inherent possess nouns include kinship terms and bodyparts, but
- may also subsume expressions of part-whole relations, things that can not exist without a possessor, or things that are culturally salient as possessions

Such nouns are always realized by speakers with a possessive prefix and are generally rejected as ungrammatical if they are offered without one — and so would be considered syntactically non-closed, a direct result of reduced their conceptual autonomy.

#### Inalienable possession

Inalienable possession refers to a grammatical system that uses a special paradigm of possessive morphemes for the possessors of certain RNs:

<u>Tunica</u>					
(5)	?o:-siku	'his father'	: ?o:-rusa	'he knows'	
	?uhk-?iyut?eku	'his hog'	: ?eh-?uhk-i	'he kicked'	(Mithun 1996: 151)

- possessors of RNs take a different possessive marker than ordinary possessors
- inalienable possessive prefixes are homophonous with patient person-markers, ordinary possessives with agentive (though the opposite pattern exists in other languages)
- like inherent possession, inalienable possession applies to different nouns in different languages, depending on perceived conceptual autonomy

CLMs are part of the semantic constituency of RNs in all languages, although there is cross-linguistic variation as to how this is recognized in the syntax:

- languages that have *inherent possession* treat CLMs as individuable entities, obligatorily elaborated as possessors
- languages that don't have inherent possession treat CLMs as an ordinary sub-component of the word's meaning, optionally realized as a possessor
- some of these recognize the special status of CLMs using *inalienable possession*

Thus, RNs depart from prototypical nouns in that they have *reduced conceptual autonomy* and so are *syntactically non-closed*. However, they are consistently expressions of KINDs and function as unmarked actants and so are treated by the syntax of most languages as nouns (for some interesting exceptions to this, see Evans 2000).

## 1.2) Syntactically less-closed nominal expressions

Common nouns are prototypically semantically closed. All languages have morphosyntactic means of creating syntactically closed or partially-closed expressions through processes of **nominalization**. This is accomplished by suppressing the unmarked expression of one or more of the verb's actants (Beck 2000), most often the subject:

### Lushootseed

- (6) a.  $\text{ʔu-ʔabyid} \quad \text{tʃəd} \quad \text{ti} \quad \text{tʃ'atʃ'as}$   
 PNT-give 1SG the boy  
 'I gave something to the boy'
- b.  $\text{tiʔəʔ} \quad \underline{\text{d-s-ʔabyid}}-\emptyset$   
 this 1PO-NP-give-3SG/PL  
 'what I give to 3SG/PL' (Hess 1993)

- the subject of (6a), *tʃəd*, is realizable only as a possessor in the nominalized (b)
- the *s*-prefix shifts the profile of the expression from an act of giving to the object given
- the agent becomes an indexical element, used to identify the particular gift
- it is realized as are indexical non-actants of ordinary nouns (possessors)

Such processes are, in effect, methods of taking complex events and expressing them in their totality (or near-totality) as single syntactic units which are (near-)equivalents of words. The fact that they are syntactically more closed than clausal expressions accounts for many of their noun-like properties.

## 2) Conventuality

It is a commonplace for Americanists that many languages have expressions which are semantically the equivalent of English nouns but are (or appear to be) structurally the equivalents of finite clauses. Consider this data from Upper Necaxa Totonac:

### Upper Necaxa Totonac

- (7) a.  $\text{ti:} \quad \emptyset\text{-ma:}-\text{w-í:}$   
 HREL 3OBJ-CS-eat-CS  
 'his wife'  
 (lit. '3SG who feeds 3SG')
- b.  $\text{ti:} \quad \text{ki-ma:}-\text{w-í:}$   
 HREL 1OBJ-CS-eat-CS  
 'my wife'  
 (lit. '3SG who feeds me')
- c.  $\text{ti:} \quad \text{ik-ma:}-\text{w-í:}$   
 HREL 1SUBJ-CS-eat-CS  
 'my husband'  
 (lit. '3SG who I feed')

- these behave internally as finite clauses (cf. *kintʃitʃí* 'my dog', *mintʃitʃí* 'your dog', etc.)
- they are conventionalized (*ti: ikma:wí:* = 'my wife', even if I (MASC) do the cooking)
- they express KINDs and have the NP-external syntax of monomorphemic nouns

Syntactically, this is not an unfamiliar situation even in English, where finite clauses overlap in syntactic distribution with nouns, as they do in most other languages:

English

- (8) a. I know [that he is coming]  
 b. I wonder [what he will do]  
 c. [That he doesn't care] annoys her

As units, these finite subordinate clauses are,

- conceptually autonomous (they denote events that were realized independently of entities other than their participants)
- syntactically closed—that is, they have an unfilled syntactic valency of zero

Semantically, however, they differ from the examples in (7) in that they are not *conventionalized* (and so would not have an entry in the lexicon) and do not represent KINDS.

2.1) Complex expressions as lexical items

Nominalizations like those in (6) and (8) have a large number of nominal properties, but they are clearly not nominal “lexical” items in that they are the expressions of specific *instances* of things (6b) or events (8) and, as such, would not be listed in the lexicon. In some cases, conventionality alone distinguishes nouns (expressing KINDS) from clauses (expressing instances):

Tuscarora (Mithun 1976)

- |     |  |         |   |         |
|-----|--|---------|---|---------|
| (9) | a. ra-kwá:this<br>MASC-young<br>‘he is young’ or ‘boy’   | (p. 26) | b. ka-téskr-ahs<br>NEUT-stink-IMPF<br>‘it stinks’ or ‘goat’ | (p. 30) |
|     | c. ra-kwá:this wa-hr-Ø-atkáhto-?<br>MASC-young PST-MASC:SUBJ-OBJ-look:at-PNT<br>‘the boy looked at the goat’ |         | ka-téskr-ahs<br>NEU-stink-IMPF                              | (p. 32) |

On the second of their interpretations,

- *katéskrahs* ‘goat’ and *rakwá:this* ‘boy’ express KINDS and are conceptually autonomous
- they are unmarked actants of a verb and syntactically closed (9c)

Sasse (1993) and Hengeveld (1992) thus argue that Tuscarora does not distinguish nouns and verbs — that is, that there are no nominal entries in the lexicon, rather noun-equivalents are formed in the syntax as finite clauses based on verbs. However,

- not all nouns are analyzable, and not all finite clauses have readings as KINDS
- *katéskrahs* ‘goat’ and *rakwá:this* ‘boy’ have conventionalized meanings not predictable from their component parts (other things can stink or be young)
- in their nominal uses, they are conventionalized and morphosyntactically inert:

Tuscarora

- (10) ka-téskr-hs-hahk  
NEUT-stink-IMPF-PAST  
'it used to stink', \*'ex-goat, goat<sub>PAST</sub>' (glosses M. Mithun, p.c.)

Clearly, the Tuscarora lexicon has an entry for 'goat' which is superficially identical in form to a finite clause, but which nevertheless meets the criteria set out here for a prototypical noun. The same is true of multi-word nominal expressions in Upper Necaxa Totonac:

Upper Necaxa Totonac

- |         |                           |             |          |    |                           |               |          |
|---------|---------------------------|-------------|----------|----|---------------------------|---------------|----------|
| (11) a. | ik-l̥ʔtsí-t̥              | wa-kán      | tʃitʃiní | c. | ik-l̥ʔtsí-t̥              | wa-kán        | małkuyúx |
|         | 1SUBJ-see-PFV             | eat-IDF     | sun      |    | 1SUBJ-see-PFV             | (PST-)eat-IDF | moon     |
|         | 'I saw the solar eclipse' |             |          |    | 'I saw the lunar eclipse' |               |          |
| b.      | *ik-l̥ʔtsí-t̥             | if-wa-kán   | tʃitʃiní | d. | *ik-l̥ʔcí-t̥              | if-wa-kán     | małkuyúx |
|         | 1SUBJ-see-PFV             | PST-eat-IDF | sun      |    | 1SUBJ-see-PFV             | PST-eat-IDF   | moon     |
|         | *I saw the solar eclipse' |             |          |    | *I saw the lunar eclipse' |               |          |

- these expressions are syntactically inert in that they do not inflect for tense or aspect (or, more accurately, they are “frozen” in the present imperfective)
- they are treated by the syntax as nouns (they are unmarked syntactic actants)
- they have *conventionalized* meanings as KINDS, are *conceptually autonomous*, and *syntactically closed*

Because they are conventionalized, expressions like these would have to be listed in the lexicon in spite of internal morphosyntactic complexity.

### 3) The grammaticalization of KINDs

Prototypically, nouns denote KINDs (Wierzbicka 1988) or classificatory relations—that is, they identify their referents

as one might put a label on a jar of preserves. One might say that a noun [referring to a person] is comparable to an identifying construction: “that’s the kind of person that this person is”. (Wierzbicka 1988: 468)

- KINDs may have constellations of features associated with them, but they are not reducible to any given combination of those features
- instead, KINDs “can be identified by means of a certain positive image, or a certain positive stereotype, which transcends all enumerable features” (Wierzbicka 1988: 471)

Monomorphemic nouns are linguistic signs associated by convention to an identifiable, conceptually autonomous entity characterizable as a KIND. Expressions like those in (6) and (8)

- designate unique objects or one-off events or instances of their referents, rather than classes of things
- being unique, they are constructed, not prefabricated, and so are not conventionalized

Multi-word and clausal nominal expressions may begin this way, but over time become increasingly associated with specific classes of referents, becoming increasingly frozen and conventionalized. This is a process familiar from morphology, where stem-affix combinations follow a cline of grammaticalization:

transparent meaning	<i>interviewer</i>	<i>the one who tilled my garden</i>
transparent vs. conventionalized	<i>runner</i> : a person who runs a type of shoe	<i>katéskrahs</i> : 'it stinks' 'goat'
predominantly conventionalized	<i>stapler</i> : desktop stapling tool ?someone who staples	<i>ti: ma:wí::</i> : 'his wife' '?she who feeds him'
conventionalized meaning only	<i>computer</i> : 'computer' *one who computes	<i>wakán tʃítʃíní:</i> 'solar eclipse' *'the sun is eaten'

As the table shows, the same cline applies to morphosyntactically complex elements, strengthening the parallel with the root–affix combinations traditionally designated as lexical items. The same path leads to the creation of nominal lexical entries, irrespective of how many words the expression contains.

#### Abbreviations

CLM classificatory landmark	MASC masculine	PNT punctual
CS causative	NEU neuter	PO possessive
HREL human relative pronoun	NP nominalizing prefix	PST past
IDF indefinite agent	OBJ object	SG singular
IMPF imperfective	PFV perfective	SUBJ subject

#### References

- Bates, Dawn, Hess, Thomas M, & Hilbert, Vi. 1994. *Lushootseed dictionary*. Seattle: U of Washington Press.
- Beck, David. 2000. Nominalization as complementation in Bella Coola and Lushootseed. K. Horie (ed.), *Complementation: Cognitive and Functional Perspectives*, 121–47. Amsterdam: Benjamins.
2002. *The typology of parts of speech systems: The markedness of adjectives*. New York: Routledge.
2003. Conceptual autonomy and the typology of parts of speech in Upper Necaxa Totonac and other languages. In G. Casad & G. Palmer (eds.), *Bringing Non-Indo-European Languages into Focus: Studies in Cognitive Linguistics*, 135–56. Berlin: Mouton.
- Evans, Nicholas. 2000. Kinship verbs. In P.M. Vogel & B. Comrie, *Approaches to the typology of word classes*, 103–72. Berlin: Mouton.
- Hengeveld, Kees. 1992. *Non-verbal predication: Theory, typology, diachrony*. Berlin: Mouton.
- Hess, Thomas M. 1993. *Lushootseed reader with introductory grammar: Vol. I*. Victoria: Tulalip.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar, Volume 2*. Stanford: Stanford University Press.
- Mithun, Marianne. 1976. *A grammar of Tuscarora*. New York: Garland.
1996. Overview of general characteristics. In Ives Goddard (ed.), *Handbook of North American Indians, vol. 17: Languages*, 137–57. Washington D.C.: Smithsonian.
- Sasse, Hans-Jürgen. 1993. Das Nomen—eine universale Kategorie? *Sprachtypologie und Universalien-Forschung* 46, 187–221.
- Wierzbicka, Anna. 1988. *The semantics of grammar*. Amsterdam: Benjamins.

Department of Linguistics  
University of Alberta  
4-45 Assiniboia Hall  
Edmonton, AB T6E 2G7

dbeck@ualberta.ca  
<http://www.ualberta.ca/~dbeck>  
<http://www.arts.ualberta.ca/~totonaco>