Derivation

- Basic derivational operations

  A taxonomy of typical lexeme-formation operations

- Productivity

  What does it mean to be a productive word-formation operation?

  How does one calculate productivity?

- Order of affixes

  Are there tendencies for affixes with certain meanings to be ordered relative to one another?

3 main types of morphological relations

- Inflection, derivation, and compounding
  - Inflectional morphology modifies properties of LEXEMES, while maintaining the basic meaning of the LEXEME.

mor-iksel'-i-ń	mor-iksel'-i-ť
sing-DES-PAST-1SG	sing-DES-PAST-2SG
`I wanted to sing'	`you wanted to sing'

(Erzya Mordvin)

3 main types of morphological relations

- Inflection, derivation, and compounding
  - Derivation relates lexemes in a word family

    eŕa-ms$_V$ ⇒ eŕa-ma$_N$
    live-INF live-NR = `life' (Erzya Mordvin)

  - Compounding combines LEXEMES

    repül-ő-gép-gyart-ás (Hungarian)
    fly-er-machine-produce-NR
    `airplane production'

3 main types of morphological relations

- Derivation relates LEXEMEs in a word family

  eŕa-ms$_V$ ⇒ eŕa-ma$_N$
  live-INF live-NR = `life' (Erzya Mordvin)

- Compounding combines LEXEMEs to create a new LEXEME

  repül-ő-gép-gyart-ás (Hungarian)
  fly-er-machine-produce-NR
  `airplane production'

Prototypical differences between inflection and derivation

	Derivation	Inflection
1.	Encodes lexical meaning	Encodes grammatical meaning
2.	Not syntactically relevant	Syntactically relevant
3.	Occurs close to the root & inside other derivation	Occurs outside all derivation
4.	Often changes lexical category	Does not change lexical category
5.	Often semantically opaque	Usually semantically obvious
6.	Often shows restricted productivity	Fully productive
7.	Optional	Obligatory

Basic derivational (lexme-formation) operations

- Permits the expansion of the lexicon of a language

Category-maintaining operations:

V ⇒ V_{CAUS}:	olvas	`read'	⇒	olvas-tat	`make read'
	legel	`graze'	⇒	legel-tet	`make graze'
V ⇒ N:	énekel	`sing'	⇒	énekl-ő	`a singer'
	sír	`weep'	⇒	sír-ó	`a weeper'
N/V ⇒ A:	felhő	`cloud'	⇒	felhő-tlen	`cloudless'
	mos	`wash'	⇒	mos-atlan	`unwashed'

Basic derivational (lexme-formation) operations

- Permits the expansion of the lexicon of a language

Category-changing operations:

A ⇒ Adv:	meleg	`warm'	⇒	meleg-en	`warmly'
	csunya	`ugly'	⇒	csunyá-n	`in an ugly way'
V ⇒ N:	énekel	`sing'	⇒	énekl-ő	`a singer'
	sír	`weep'	⇒	sír-ó	`a weeper'
N/V ⇒ A:	felhő	`cloud'	⇒	felhő-tlen	`cloudless'
	mos	`wash'	⇒	mos-atlan	`unwashed'
A ⇒ V	szép	`beautiful	⇒	szép-ül	`become beautiful'

External organization: words as participants in networks of relations

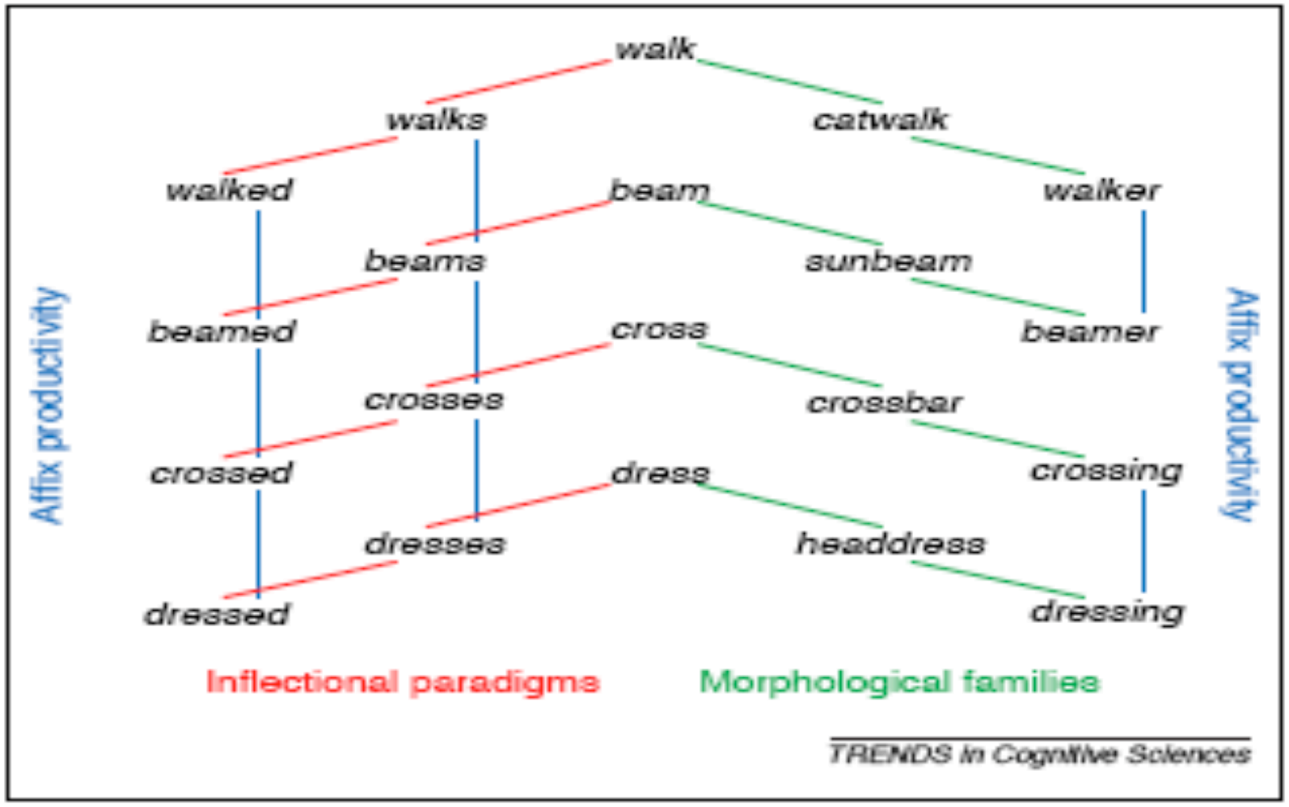


Figure 1. Examples of paradigmatic lexical relations in English. Relations between inflected variants (inflectional paradigms) are shown in red, relations between morphologically related compounds and derived words (morphological families) are shown in green, and relations between words sharing the same affix are shown in blue. Affixes that occur across many words are described as productive.

-th affixation

broad + th ⇒ breadth

deep + th ⇒ depth

long + th ⇒ length

strong + th ⇒ strengh

true + th ⇒ truth

warm + th ⇒ warmth

wide + th ⇒ width

phonology: X-/θ/, with various different base alternations

category of based: X = adjective

semantics: `state or property of being X'

What's the structure?

unactualizability

Derivation

- Basic derivational operations

  A taxonomy of typical lexeme-formation operations

- Productivity

  What does it mean to be a productive word-formation operation?

  How does one calculate productivity?

- Order of affixes

  Are there tendencies for affixes with certain meanings to be ordered relative to one another?

What's going on here?

- For some words, we can predict that -ity won't apply

glorious	*gloriosity	gloriousness
furious	*furiosity	furiousness
gracious	*graciosity	graciousness
fallacious	*fallaciousity	fallaciousness
acrimonious	*acrimoniosity	acrimoniousness

-able affixation

abominable
absorbable
abstractable
abusable
acceptable
accountable
accruable
achievable
acid-extactable
actable
actionable
actualizable
adaptable
addressable
adjustable
admirable
admissable
adorable
advisable
affable

Productivity

"Property of a morphological process: a process is productive if it can be applied to new (forms of) words." [Booij in glossary]

"The statistical readiness with which an element enters into new combinations (Bolinger 1948:18)

- Productivity isn't really an all-or-nothing concept

Productivity

- Some observations
  - Though many things are possible in morphology, some things are more likely than others (cf. walked and ran)
  - Though there are infinitely many potential words in a language, some are more likely to become actual words than others (cf. mini-burger, burgerlet, burgerette)
  - We need to consider actual words and potential words and what the relation is between them.

-th affixation

broad + th ⇒ breadth

deep + th ⇒ depth

long + th ⇒ length

strong + th ⇒ strengh

true + th ⇒ truth

warm + th ⇒ warmth

wide + th ⇒ width

phonology: X-/θ/, with various different base alternations

category of base: X = adjective

semantics: `state or property of being X'

Productivity

- So, the suffix -th is generally considered unproductive
- But, WWW searches turn up many citations:

> Coolth, once a nonce word made up on analogy with warmth, is now tiresomely jocular. (1923)
>
> Increase the capacity of your house to store coolth. (Yes, it is a real word.) Using the mass in your house...
>
> The team developed a strategy to capture night-time coolth and store it for release during the following day.
>
> Do we see the whiteness of the snow, but only believe in its coolth.

Productivity

- The suffix -th was once productive:
  - filth, health, length, mirth, strength, truth, dearth, depth, breadth, sloth, wealth
- Coolth (coined after warmth) goes back at least to 1547
- Width [wɪdθ] comes from widness (influenced by length) in 1627, not wide [waid]
- Later coinages: illth (opposite of wealth = well-being), greenth, loweth

And, what's going on here?

approve	approval	approbation
recite	recital	recitation
propose	proposal	proposition
arrive	arrival	* arrivation
refuse	refusal	* refusation
derive	* derival	derivation
describe	* describal	description

How can we describe constraints on the use of specific derivational operations?
Alternatively, how can account for degrees of productivity?

Productivity

- Words in -ness have three meanings
  - 'the fact that Y is X' — His callousness surprised me.
  - 'the extent to which Y is X' — His callousness surprised me.
  - 'the quality or state of being X' — Callousness is not a virtue.
- Words in -ity can have many specialized meanings

  The are several varieties of fish in the lake.
  They admired his dress, but only as a curiosity.

Blocking

- For some words, we can predict that -ity won't apply

glorious	*gloriosity	gloriousness
furious	*furiosity	furiousness
gracious	*graciosity	graciousness
fallacious	*fallaciousity	fallaciousness
acrimonious	*acrimoniosity	acrimoniousness

- The existence of a noun (glory, fury, ...) blocks the formation of a synonym
- Panini's Principle (aka Elsewhere Condition): A more specific rule trumps a more general rule

Blocking

- Let's assume that the operation that forms nouns with -ity is more restricted (applies to fewer stems with more conditions on its application) than the operation that creates nouns with -ness, thus, formation with -ity is more specific than formation with -ness

  glory blocks gloriousity and if glorioiusness is formed, then it doesn't mean was glory means.

  electricity blocks electricness

- Completely predictable forms aren't listed in the dictionary, so aren't subject to blocking effects; this makes claims about what we store in our mental lexicons and how "rules" interact with stored items.

- Blocking seems general (*this night / tonight), and somewhat mysterious

Storage and Rules: The role of history

- The impulse toward generative rules, i.e. operations that produce complex forms from smaller pieces, thus limiting storage:

  "The computers of that era (1950s) had comparatively reasonable computational capacity but very limited memory. For a program to work efficiently, it had to minimize storage...Linguistic data in electronic form did not exist. Not surprisingly, the linguistic theories of the time took formal languages as models, emphasizing the generative capacity of language, denying any role of importance to probability and statistics, elevating economy of storage in memory to a central theorem." (Baayen 2003:230)

Disparagement of Frequency

- Generative grammarians distinguish between I(-nternal) Language and E(xternal)-language, where the former is our mental representation of grammar and of theoretical interest, but the latter is just how this grammar is used in real time.
- Frequency considered to be part of performance/E-language and, hence, not of primary theoretical interest:
  - "It seems that probabilistic considerations have nothing to do with grammar, e.g. surely it is not a matter of concern for the grammar of English that 'New York' is more probable than 'Nevada' in the context 'I come from–'." (Chomsky 1962)
  - "But it must be recognized that the notion 'probability of a sentence' is an entirely useless one, under any known interpretation of this term." (Chomsky 1969)

Storage versus Rules

- What do people store in their mental lexicons?

  Hypothesis 1: Just the irregular formations, since listed elements and rules can give you the regular ones. (elegant, since non-redundant)

  Since spell-able is producible by the application of a productive -able affixation rule, it is not stored in the lexicon.

- It would be redundant to store both the complex word and the rule that could generate it.
- Thus, it is both elegant and economical to just store each of necessary and irreducible pieces just once.
- The only elements that must be stored are those cannot be generated by rule.

Word formation rules (WFRs): IP

- The bases of WFRs are themselves words, i.e. words are built from words, i.e. lexeme-formation operations.
  - Bases must be existing words; a possible but non-existent word (according to the hypothesis) cannot be the base of WFR, since only underived words are in the lexicon.
  - WFRs can take as a base only a single word, no more (e.g., phrases) and no less (e.g., morphs)
  - Both the input and the output of WFRs must be members of a major lexical category (noun, verb, adjective, preposition)

Storage versus Rules

- Let's say that the goal of morphology is create a lexicon that it is as non-redundant as possible
- It contains all of the morphemes of the language, i.,e, lexical representations for roots and affixes:

  store$_V$ <SUBJ, OBJ> `put way in order to have for a future date'

  able$_N$ `such that X can be V-ed'
- An -able word-formation rule:

  phonology: X/əbl/

  semantics: `X can be V-d'

  base: X = V(erb)

WFRs

- Speakers seem to be aware of WFRs:
- a WFR of English produces adjectives from verbs:

  phonology: X-/əbl/

  category of base: X = $V_{transitive}$

  semantics: `capable of being V-ed'

  spellable, traceable, singable, imaginable

- But, what about:

  workable solution, makeable mistake, perceptible error, saleable items, remarkable tenacity...

Word formation rules

- We also have words with similar form and meaning that are not formed by WFR

  possible, tangible, legible, edible (versus eatable) unflappable

- Given the rule, should speakers posit meanings and lexical categories for poss- and tang- and -unflap might be, though they aren't listed anywhere.

- Indeed, sometime WFRs can be used to motivate new formations.

Word formation rules

- WFRs can also explain back-formations
- The word babysitter is formed by the rule:

  $[X_N Y_N]_N$ 'an Y-er of X'

  like anteater and cardholder
- But, it superficially has the form of a word produced by the rule:

  $[X_V$ -er$]_N$ 'one who Xs habitually, professionally, etc.'
- If one assumes that -er in babysitter is the agentive affix, the stripping it off will yield babysit, a coinage that is motivated, but not predicted

Storage versus Rules

- Is there any reason to believe that human minds reflect the scientific goals of parsimony and non-redundancy?
- Is the standard IA and IP organization of morphology, in terms of minimizing storage and optimizing rules, simply a reflection of the historical fact that methods and metaphors for the analysis of morphology were developed at a time that favored and encouraged certain assumptions and tools and disfavored others?

Storage versus Rules

- What do people store in their mental lexicons?

  Hypothesis 2: Both the regular and the "irregular" ones, though patterns/rules also exist. (less elegant, since redundant)

  Though spell-able is producible by the application of a productive -able affixation rule, if frequent enough, it may still be stored in the lexicon.

- Though this is redundant, it may be psychologically accurate and, so, theories should reflect the storage of complex words.

- Do human minds reflect the scientific goals of parsimony and non-redundancy?

Interim summary

- There are regularities evident in lexeme-formation, i.e. the derived form is simply a product of the information in the rule that produces it.
- There are differences in productivity with respect lexeme-formation operations
- Some operations compete with one another, partitioning domains of lexeme-formation (-ity v. -ness as both deriving nouns)
- Some operations seem to apply straightforwardly and regularly (spellable), while others apply less straightforwardly and regularly (saleable).

Interim summary

- The task is to identify the nature of productivity and regularity and to understand what the consequences of this investigation are for the design of our morphological theories and for assumptions about the mental lexicon.
- To explore productivity, we have to examine frequency more carefully.

Frequency

(following discussion based on Plag 2003)

- Frequency effects are very important in morphology, perhaps more than any other subfield of linguistics
- What is frequency?
  - Absolute and relative frequency
  - Type and token frequency
  - Word, stem, and morph frequency
- Speakers are very aware of relative word frequencies
- More frequent items are (generally) processed faster
- What is linguistic theory supposed to reflect?

Productivity and frequency

- If productivity is a gradient concept, how can we define or measure it?
- Absolute type frequency
  - There are 3,604 words ending in -able in Webster's 2nd
  - Some, though, are French loans: acceptable, changeable, desireable
  - Some might be French loans: payable, regrettable
  - Calques: understandable

-able derivatives in the British National Corpus: Types and Tokens

	Frequency		Frequency
abominable	84	actionable	87
absorbable	1	actualizable	1
abstractable	2	adaptable	230
abusable	1	addressable	12
acceptable	3416	adjustable	369
accountable	611	admirable	468
accruable	1	admissable	2
achievable	176	adorable	66
acid-extactable	1	advisable	516
actable	1	affable	111

Measuring productivity in types

- Given access to historical data (dictionaries, corpora), we can compare the rate of additions licensed by word formation rules
- Many words with a particular affix, e.g. government, may reflect the former productivity of the affix in creating neologisms, i.e., at some time the affix may have been productive.
- Many dictionaries do not list the most productive affixes, i.e., smartness, so this type may be under-represented precisely because it is synchronically productive.
  - Dictionaries don't always list new words
  - Dictionaries don't always list fully predictable words

Measuring productivity

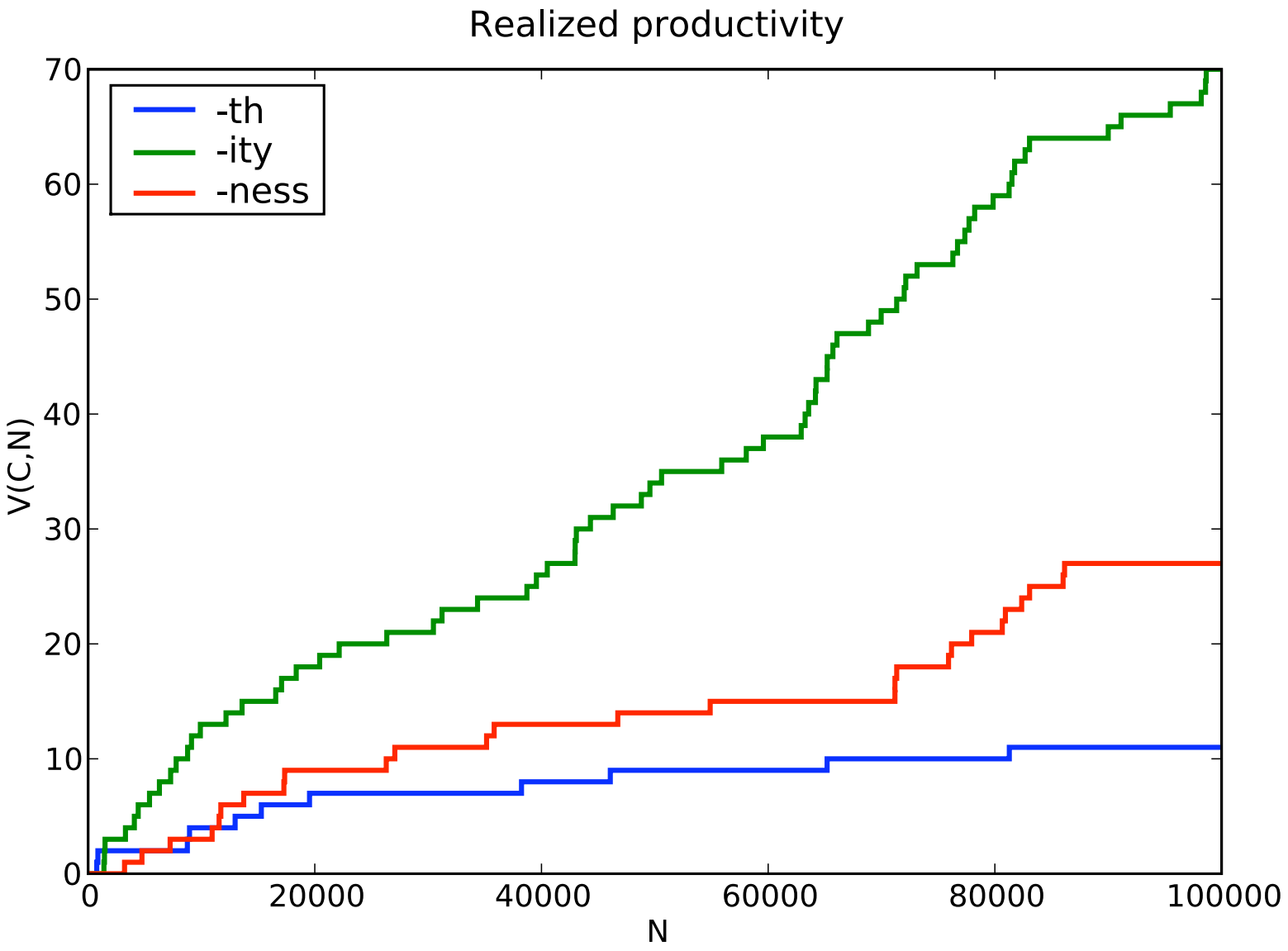
- We could look at the ratio of actual to possible words to get an index of productivity (Aronoff 1976, Baayen and Lieber 1991)

$$I = \frac{V}{S}$$

- This depends on being able to identify S, the number of words which a WFR ought to apply to (infinity?)
- And, we need to identify V, the number of words which it does apply to.
- Even if we can get past this, at best we get an index of past productivity, not current productivity

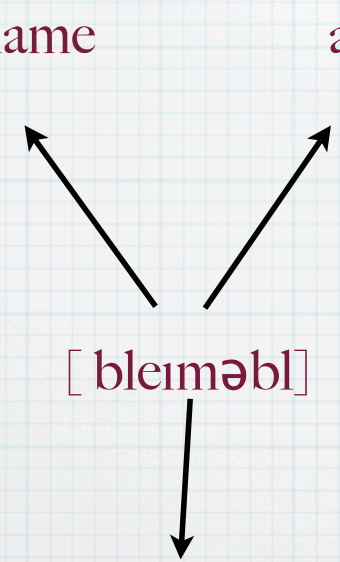
Measuring productivity

- Realized productivity is the success of a morphological pattern in the linguistic marketplace (cf. market share)
- It can be estimated by V(C,N), the number of word types following pattern C in a corpus of N tokens
- Brown corpus
  - 49,815 types, 1,161,192 tokens
  - 402 types in -ity
  - 322 types in -ness
  - 38 types in -th


Realized productivity
-th
-ity
-ness
V(C,N)
N
0
10
20
30
40
50
60
70
0
20000
40000
60000
80000
100000

Token frequency and the mental lexicon

Dual Route Model (Baayen): each word processed in 2 ways



Decomposition route

(economical wrt storage, costly wrt processing)

Whole word route

(costly wrt storage, economical wrt processing)

Frequency

- Words that occur more frequently are more easily stored and accessed than less frequent words.
- Higher frequency words have a higher resting activation than lower frequency words: it is this that makes them more accessible more quickly.
- All words are amenable to analysis in terms of the decomposition route and the whole word route, with the choice of which route applying in each case being determined by the level of resting activation.
- The storage of complex, highly frequent words (high resting levels) renders them more amenable to the whole word route.
- The low frequency words are decomposed, since there is no whole word.

Frequency

- The decomposition of low frequency words leads to the greater accessibility of affixes and, consequently, these affixes are more accessible for creation of neologisms (new derivatives).
- If only high frequency words enter into the system, then these are accessed as whole words and their internal composition, i.e. affixes, are less accessible for the neologisms.
- Unproductive morphological categories associated with many high frequency words and few low frequency ones, while
- correlatively, productive morphological categories associated with many low frequency words and few high frequency ones.

Measuring productivity via tokens

- Words with the lowest frequency are those which appear only once: these are hapax legomena or hapaxes
- They are likely to be coinages, made up on an ad hoc basis and they can be used to quantify productivity.

-able derivative in BNC and Webster's 3rd

3 of 6 hapaxes not listed; 3 other low frequency words not listed in Webster 3rd

-able derivative	token frequency	listed in Wrd
absorbable	1	yes
abusable	1	no
accruable	1	no
acid-extractable	1	no
actable	1	yes
abstractable	2	no
admissible	2	no
addressable	12	no
adorable	66	yes
abominable	84	yes
actionable	87	yes
affable	111	yes
achievable	176	yes
adaptable	230	yes
adjustable	369	yes
admirable	468	yes
accountable	611	yes
acceptable	3416	yes

First twenty word alphabetically in BNC:

13 of 20 hapaxes not listed in Webster 3rd

BNC hapaxes and their entries in *Webster's Third*

-able derivative	Listed in *Webster's Third*	*-able* derivative	Listed in *Webster's Third*
absorbable	yes	amusable	no
abusable	no	annotatable	no
accruable	no	applaudable	yes
acid-extractable	no	approvable	no
actable	yes	arrangeable	no
actualizable	yes	assessionable	yes
affirmable	yes	auctionable	no
again-fashionable	no	biteable	yes
aidable	no	blackmailable	no
air-droppable	no	blameable	no

Calculating productivity?

- To find the probability of finding a likely neologism in a corpus, one calculates the ratio of the number of hapaxes with the specified affix to all of the tokens containing that affix.
- The formula is: $P = n_1^{aff}/N^{aff}$
- That is, if there are 2 hapaxes with a specific affix and 100 tokens with that affix, the probability of encountering a new word is 2 percent.
- Thus, a higher number of hapaxes yields a higher value for P and, hence, a higher degree of productivity associated with the affix.
- In contrast, a high number of high frequency items leads to a lower value for P and, consequently, a lower productivity associated with the affix.

Calculating productivity

(Plag et. al. 1999, Plag 2003)

Frequencies of affixes in the BNC (written corpus)

Affix	V	N	n_1	*P*
-able	933	140627	311	0.0022
-ful 'measure'	136	2615	60	0.023
-ful 'property'	154	77316	22	0.00028
-ize	658	100496	212	0.0021
-ness	2466	106957	943	0.0088
-wise	183	2091	128	0.061

V = type frequency/'extent of use',
N = token frequency, n_1 = hapax frequency,
$P = n_1/N$ 'productivity in the narrow sense'

- Some forms have large number of types (V): -ness
- Some forms have large number of tokens (N): -ize
- Some forms have relatively small number of hapaxes:

  -ful `measure': tankful, truckful,...

  -ful `property': peaceful, mindful...

Calculating productivity

(Plag et. al. 1999, Plag 2003)

Frequencies of affixes in the BNC (written corpus)

Affix	V	N	n_1	*P*
-able	933	140627	311	0.0022
-ful 'measure'	136	2615	60	0.023
-ful 'property'	154	77316	22	0.00028
-ize	658	100496	212	0.0021
-ness	2466	106957	943	0.0088
-wise	183	2091	128	0.061

V = type frequency/'extent of use',
N = token frequency, n_1 = hapax frequency,
$P = n_1/N$ 'productivity in the narrow sense'

- `measure' -ful and `property' -ful have similar type (V) frequencies
- they have very different token frequencies
- they have very different hapax frequencies
- the relation of hapaxes to tokens (0.023 `measure' v. 0.00028 `property' suggests that `measure' -ful is the productive one.

Calculating productivity?

Frequencies of the most frequent adjectival *-ful* derivatives (BNC, written corpus)

derivative	frequency
successful	10366
useful	9479
beautiful	7964
powerful	7064
careful	4546
wonderful	4202

- That `measure' -ful is the productive one is further corroborated by closer examination of tokens and types for `property' -ful:

  the high frequency of a small number of types accounts for the high number of tokens (N)

- This is expected given that if a morphological category (an affixal pattern) contains a large concentration of highly frequent words, it tends to be less productive (whole-word route)

- In contrast, patterns with fewer highly frequent words and more low frequency words are generally productive (decomposition route)

Looking at a different type of frequency: affix ordering

- Some basic questions

  1. Do affixes contributing derivational and inflectional inflectional information exhibit reliable sequencings relative to the lexical root and to each other?

- 2. If there do seem to be such orders are they categorical or tendential?

  3. What might be the causes of such orderings, if they exist?

  4. Since investigating orderings presupposes the existence of morpheme-like entities, how do approaches that don't depend on morphemes deal with such data? (Conversely, are there generalizations that can be made for complex words that aren't analyzable in terms of pieces with identifiable meanings?)