

# Morphology 1: Introduction

Doug Arnold  
doug@essex.ac.uk

## 1 Doing Without

```
det --> [the].  
n --> [child].  
n --> [children].
```

- words as atoms
- words in the grammar

Problems:

- All forms have to be explicitly listed – even if this is possible, it is unattractive.
- Lexical lookup is inefficient (e.g. top down).
- The grammar is repetitive.

Grammar and lexicon can be separated:

```
det(Word) --> [Word], {dict(Word,det)}.  
n(Word) --> [Word], {dict(Word,n)}.
```

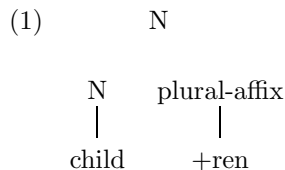
```
dict(the,det).  
dict(child,n).  
dict(children,n).  
etc.
```

But this is still unattractive.

## 2 Morphological Processing

It would be better to find some form of morphological processing: replace words (word forms) by stems plus structure and information, in the form of:

- features (e.g. *children*  $\rightsquigarrow$   $\text{child}_{[+\text{plural}]}$ ); and/or
- a sequence of morphemes:  $\text{child}+\text{ren}$ , or  $\text{child}+\text{plur}$ ; and/or
- a morphological structure (e.g. a tree):



## 3 Morphology: A Word of Warning

Morphology is a complicated business, without a consensus view (syntax and semantics are much more cohesive). Not only is there no consensus about formal machinery, morphologists don't agree on terminology, on the nature of the subject (e.g. whether morphemes exist), or even on whether the subject really exists separate from syntax and phonology (Chomsky and Halle (1968)).

Some terminology. I will say that *words* (lexemes, lexical items) are realized as different *word forms*: so *take, takes, taken*, etc. are different forms of the word/lexeme **take**.

Many people use terms like *lemma, stem, root* in similar, but often distinct ways.

## 4 Types of Morphology

It is useful to distinguish three basic kinds of morphological process:

- inflection;
- derivation (word-formation);
- compounding

### 4.1 Inflection

Intuitively, this involves creating different forms of a word, typically expressing grammatical information (e.g. tense, aspect, person, number, case, agreement, etc):

- (2) a.  $N_{\text{plural}} \rightarrow N+s$   
b.  $\text{baby} \rightarrow \text{babies}$

$$(3) \text{ break}_V \rightarrow \left\{ \begin{array}{l} \text{breaks} \\ \text{breaking} \\ \text{broke} \\ \text{broken} \end{array} \right\}$$

### 4.2 Derivation

Intuitively, this involves combining a bound morpheme (affix) with a word to create a new word:

- (4) a.  $A+ize \rightsquigarrow V$   
b.  $[_V [_A \text{ legal} ] \text{ize} ]$

- (5) a.  $V + er \rightsquigarrow N$   
b.  $[_N [_V \text{ destroy} ] \text{er} ]$

- (6)  $[_N [_V [_N [_V \text{ compute} ] \text{+er} ] \text{+ize} ] \text{+ation} ]$

Cf. also: nominalization (*destroy/destruction*), passivization, causativization, etc. etc).

### 4.3 Compounding

Intuitively, combining two words to make a third:

- (7) a.  $N+N \rightsquigarrow N$   
b.  $[_N [_N \text{ table} ] [_N \text{ top} ]]$

- (8) tube  
 feed-tube  
 in-line feed tube  
 in-line feed tube adaptor  
 in-line feed tube adaptor hose  
 in-line feed tube adaptor hose cover  
 in-line feed tube adaptor hose cover cleaning  
 in-line feed tube adaptor hose cover cleaning instruction  
 in-line feed tube adaptor hose cover cleaning instruction sheet
- (9) a.  $[_N [_N \text{ adult film } ] \text{ society } ]$   
 b.  $[_N \text{ student } [_N \text{ film society } ]]$

“When tweetle beetles battle, it is called a tweetle beetle battle. When tweetle beetles battle with paddles, it is called a tweetle beetle paddle battle, and when tweetle beetles do it in a puddle, it is called a a tweetle beetle paddle puddle battle. . .”

## 5 Discussion/Comparison

### 5.1 Inflection

E.g. plural formation:

- semantically regular (non-aggregate  $\rightarrow$  aggregate)
- productive (applies to any semantically appropriate N)
- preserves category (N  $\rightarrow$  N)
- produces different forms of one word/lexeme (*baby/babies*)

(Similarly for tense/aspect inflection, case marking, etc.)

Often on inflectional morpheme realizes several pieces of information, e.g. *+o* in Latin *amo* expresses 1st person, singular, present, indicative and active.

### 5.2 Derivation/Compounding

Derivation:

- semantically unpredictable (*computer, typewriter*; most derived nominals can refer to events (*destruction*), but *suction* cannot: \**Sam's suction of the coffee*.)
- semi-productive (*rare:rarity, red:\*redity*)
- category changing (*destroy<sub>V</sub>: destroyer<sub>N</sub>*)
- words/stems  $\leadsto$  new words

Compounding:

- *alligator shoes* vs *horse shoes* vs *brake shoes*; *blackboard*; *pumpkin bus* . . .
- (but N+N compounding is rather productive)
- *under-write* (P+V  $\leadsto$  V)

**Derivation:** stem + bound morpheme

**Compounding:** free morpheme + free morpheme

### 5.3 Borderline Cases

- Comparative/superlative formation:

- (10) a. big, bigger, biggest  
 b. tall, taller, tallest
- Adverb formation:
 

(11) a. quick, quickly  
 b. happy, happily
  - Prefixation:
 

(12) a. pre-; pseudo-: pseudo-scientific, pseudo-computational, pseudo-intellectual, ...  
 b. un-: unhappy, unintelligent, unfortunate, ..., (\*unsad, \*unpathetic)  
 c. -less: windowless, deskless, sugarless, ...

## 6 Forms of Morphological Combination/Operation

1. Concatenative, ‘beads on string’: (English: see above; Turkish)
2. Infixation: (Bontoc)
 

(13) fikas f-**um**-ikas  
 ‘strong’ ‘be strong’
3. Circumfixion: (Indonesian)
 

(14) besar : **ke**-besar-**an**  
 big : bigness
4. Reduplication: (Indonesian)
 

(15) orang : **orang**+ orang  
 man : men
5. Sound change (e.g. tone, vowel/consonant change)
6. Zero morphology:
 

(16) a. The *poor* are always with us.  
 b. He always *examples* his way out of giving an explanation.
7. Templatic: (Arabic) verb stems are made up of:
  - a root (usually three consonants), e.g. *ktb* ‘write’
  - a vowel pattern (e.g. active is A, passive UI)
  - a template (e.g. CVCVC for ‘write’, CVCCVC for ‘cause to write’)

(17) Active(A)    Passive(UI)    Template  
 kAtAb        kUtIb        CVCVC    ‘write’  
 kAttAb       kUttIb       CVCCVC   ‘cause to write’

## 7 Some Conceptual Problems

### 7.1 Bracketing Paradoxes: *unhappier*

- *unhappier* means ‘more unhappy’. So the structure should be (18a).
  - *+er* only attaches to adjectives with two or fewer syllables: *sad+er:sadder*, *clever+er:cleverer*, but *\*elegant+er*. So the structure must be (18b).
- (18) a. [[un+happy]+er]  
 b. [un+[happy+er]]

Maybe we need two structures: one for morpho-phonology one for morpho-syntax.

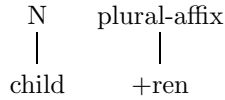
### 7.2 Morphology and Syntax: Clitics

Clitics are phonological affixes to their hosts, but are syntactically separate.

### 7.3 Do Morphemes Exist?

- Word grammar: morphology describes the way morphemes can be combined together; morphemes are lexical items ('item and arrangement'):

(19) N



Maybe the plural-affix subcategorizes for the N...

- Word formation rules: there are no morphemes, just lexical items (stems), on which 'word formation rules' operate to produce other items ('item and process'):

(20)  $f_{\text{plural}}(\text{child}) = \text{children}$

(21) a.  $[_A X] \rightarrow [_N X + \text{ity}]$   
 b.  $\text{scarce} \rightarrow \text{scarcity}$

Many other, and intermediate, positions are possible.

## 8 A First Approximation

A first approximation is to distinguish:

- rules/principles of spelling/pronunciation:
- rules/principles of word-formation (at least: compounding, derivation):
- rules/principles of word-formation (at least: compounding, derivation):
  - *red+ness* is a possible word, \**red+ity* is not;
  - *+ize* can be added to an adjective to make a verb;
- rules/principles of spelling/pronunciation:
  - nasals assimilate to the following consonant:
 

(22) a. *in+possible*  $\rightsquigarrow$  *impossible*  
 b. *in+tolerable*  $\rightsquigarrow$  *intolerable*
  - *+s* (3rd singular/plural marker) is realized as
    - \* *s* (phonologically /s/ or /z/), normally, but
    - \* *es* (phonologically /Iz/) if the word ends in *s*, *z*, *ch*, *sh*, *o*, *x*:
 

<i>bus</i> : <i>buses</i>	<i>jazz</i> : <i>jazzes</i>
<i>church</i> : <i>churches</i>	<i>fish</i> : <i>fishes</i>
<i>tomato</i> : <i>tomatoes</i>	<i>box</i> : <i>boxes</i>
    - \* *es* with words ending in *y* (making *y* into *i*):  
*fly* : *flies*
    - \* unless the penultimate letter is a vowel (in which case, *s*):  
*play* : *plays*

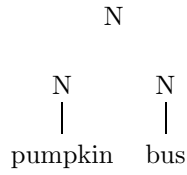
Different techniques seem appropriate: 'structural' or 'tree-based' approaches look right for word formation:

Compound Nouns:

$N \rightarrow N \ N$

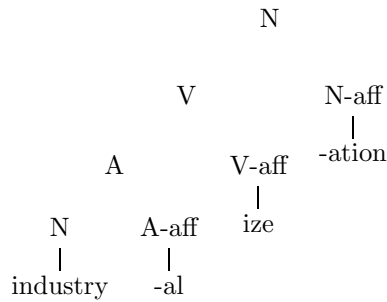
$N \rightarrow \text{pumpkin}$

$N \rightarrow \text{bus}$

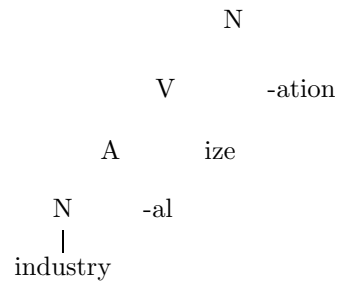


Derivation:

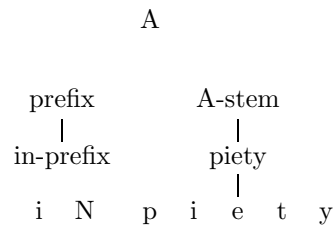
$N \rightarrow V \quad N\text{-aff}$   
 $N\text{-aff} \rightarrow \text{ation}$   
 $V \rightarrow A \quad V\text{-aff}$   
 $V\text{-aff} \rightarrow \text{ize}$



or ('morphemeless'):



This does not look so attractive for the 'spelling' rules, where adjacency is at least as important as structure (the presence of the p in *impiety* must be signalled to the N, but this has to go via several nodes of the tree structure):



## 9 Reading

The best introduction to computational morphology is Sproat (1992) which contains an excellent chapter overviewing morphology in general. There is a very nice discussion of how to handle some aspects of inflectional morphology in Covington (1994, Ch9). Ritchie et al. (1992) is also useful.

Spencer (1991) provides a more detailed general introduction to morphology. On word-formation (derivation and compounding), see Bauer (1983). Another useful general introduction is Matthews

(1974).

## References

- L. Bauer. *English Word Formation*. CUP, Cambridge, 1983.
- Noam Chomsky and Morris Halle. *The sound pattern of English*. Harper and Row, New York, 1968.
- Michael A Covington. *Natural Language Processing for Prolog Programmers*. Prentice Hall, Englewood Cliffs, NJ, 1994.
- P H Matthews. *Morphology: An introduction to the theory of word-structure*. CUP, Cambridge, Ma., 1974.
- G.D. Ritchie, G.J. Russell, A.W. Black, and S.G. Pulman. *Computational Morphological: Practical Mechanisms for the English Lexicon*. MIT Press, Cambridge, Mass., 1992.
- A. Spencer. *Morphological Theory*. Basil Blackwell, Oxford, 1991.
- Richard Sproat. *Morphology and Computation*. ACL-MIT Series in Natural Language Processing. MIT Press, Cambridge, Ma., 1992.