



Carnegie Mellon University
Language
Technologies
Institute

11-324/11-624/11-724 Human Language for AI

Morphological Theories and Typology

David R. Mortensen

October 10, 2022

Language Technologies Institute
Carnegie Mellon University

Introduction

Learning Objectives

At the end of this lecture, students will have developed the following skills:

- Identifying cases where a push-down automaton would be more useful than a purely finite state approach to morphology
- Using LEXC flag diacritics (particularly of the unification type) to model long-distance morphological relationships

Students will be able to identify the following types of exponence:

- Direct/simple
- Extended
- Cumulative

- Null
- Vacuous

Students will understand the differences between, strengths, and weaknesses of the following general theories of morphology:

- Item-and-arrangement
- Item-and-process
- Word-and-paradigm

Students will be able to apply the traditional morphological typologies. In addition, they will understand the advantages of Sapir's two-dimensional typology.

Flag Diacritics and Long-Distance Morphological Dependencies

Sometimes Morphological Relationships or Non-Local

Take the following example from the Arabic word *kitaab* ‘book’. The following words are possible:

DEFINITE		INDEFINITE	
kitaab-u	nom def	kitaab-uN	nom indef
kitaab-a	acc def	kitaab-uN	acc indef
kitaab-i	gen def	kitaab-iN	gen indef

However, when the definite article *al-* is prefixed, only the indefinite forms are possible:

DEFINITE		INDEFINITE	
*al-kitaab-u	nom def	al-kitaab-uN	nom indef
*al-kitaab-a	acc def	al-kitaab-uN	acc indef
*al-kitaab-i	gen def	al-kitaab-iN	gen indef

Flag Diacritics Allow Us To Model Non-Local Dependencies

Multichar_Symbols

```
@U.ART.PRESENT@ @U.ART.ABSENT@  
uN aN iN
```

LEXICON Root

```
Article ;
```

LEXICON Article

```
al@U.ART.PRESENT@ Stems ; ! optional article prefix  
Stems ; ! empty string entry
```

LEXICON Stems

```
kitaab Case ; ! In reality, there would be many stems
```

LEXICON Case

```
u # ;  
a # ;  
i # ;  
@U.ART.ABSENT@ IndefCase ;
```

LEXICON IndefCase

```
uN # ;  
aN # ;  
iN # ;
```

@U.ART.PRESENT@ and @U.ART.ABSENT@ are FLAGS.

Flags have three parts:

1. Flag type (e.g., **U** for unification)
2. Feature name (e.g., **ART** for article; can be defined by the user)
3. Feature value (e.g., **PRESENT** for “article is present;” also user defined)

UNIFICATION features must unify—a path through the automaton is valid iff all flags for a given feature agree. If

@U.ART.PRESENT@ is present in a path, paths with @U.ART.ABSENT@ (etc.) are ruled out.

Declaring Multichar Symbols

All of the flags (`@type.feature.value@` combinations) must be declared as multichar symbols.

Multichar_Symbols

@U.ART.PRESENT@ @U.ART.ABSENT@

uN aN iN

Root, Article, and Stem Lexicons

In the **Article** lexicon, **ART** is set to **PRESENT** for paths beginning with **a1**:

LEXICON Root

Article ;

LEXICON Article

a1@U.ART.PRESENT@ Stems ; ! optional article prefix

Stems ; ! empty string entry

LEXICON Stems

kitaab Case ; ! In reality, there would be many stems

Any string that starts with the **a1** prefix must contain no flags with
@U.ART.ABSENT@.

Case Lexicons

To get to the **IndefCase** lexicon with the case markers that go with indefinite nouns, the **@U.ART.ABSENT@** is passed:

LEXICON Case

```
u          # ;  
a          # ;  
i          # ;  
@U.ART.ABSENT@ IndefCase ;
```

LEXICON IndefCase

```
uN         # ;  
aN         # ;  
iN         # ;
```

This means that words cannot have both the **a1** definite prefix and the indefinite **uN**, **aN**, and **iN** case suffixes.

Exponence

Cumulative Exponence in Latin

	SINGULAR	PLURAL
NOM	amīca	amīcae
VOC	amīca	amīcae
ACC	amīcam	amīcās
GEN	amīcae	amīcārum
DAT	amīcae	amīcīs
ABL	amīcā	amīcīs

Exponence Summarized

Direct



dogs (n.)

Vacuous

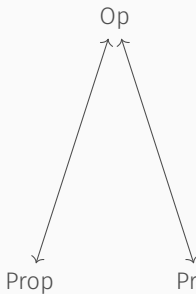


Sino-American friend (v.)

Null

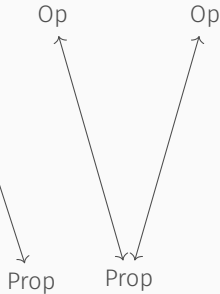


Cumulative



laughs (v.)

Extended



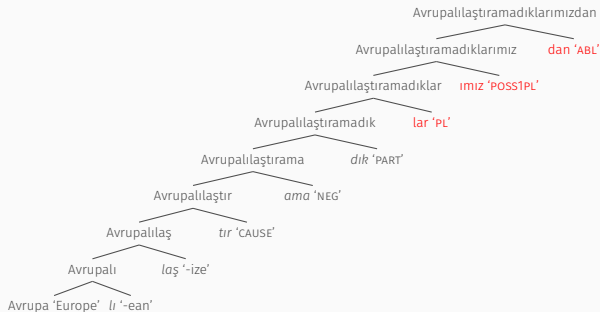
gemacht (v.)

Morphological Typology

- (1) 里斯对这个案件的调查 进行
Lisi dui zheige anjian de diaocha jinxing
Lisi to this CLF case DE investigation last
了一个小时
le yi ge xiaoshi
ASP one CLF hour
'Lisi's investigation of the case lasted an hour.'

Agglutinative Languages

An example from Turkish:



‘of ours that were unable to be Europeanized’

Fusional/Flexional Languages

	PRESENT	PERFECT	PRETERIT
1SG	make	gemacht	machte
2SG	machst	gemacht	machtest
3SG	macht	gemacht	machte
1PL	machen	gemacht	machten
2PL	macht	gemacht	machtet
3PL	machen	gemacht	machten

Templatic Languages

	PERFECT		IMPERFECT		PARTICIPLE	
	Active	Passive	Active	Passive	Active	Passive
I	katab	kutib	ktub	ktab	kaatib	ktuub
II	kattab	kuttib	kattib	kattab	kattib	kattab
III	kaatab	kuutib	kaatib	kaatab	kaatib	kaatab
IV	?aktab	?uktib	ktib	ktab	ktib	ktab
V	takattab	tukuttib	takattab	takattab	takattib	takattab
VI	takaatab	tukuutib	takaatab	takaatab	takaatib	takaatab
VII	nkatab	nkutib	nkatib	nkatab	nkatib	nkatab
VIII	ktatab	ktutib	ktatib	ktatab	ktatib	ktatab
IX	ktab(a)b	ktab(i)b	ktab(i)b			
X	staktab	stuktib	staktib	staktab	staktib	staktab

Polysynthetic Languages

Polysynthetic languages are languages in which noun arguments like objects can be expressed as part of a verb, meaning that full sentences can be expressed as a verb alone (not just through agreement with person and number, but through the “incorporation” of the noun into the verb). Take the following example from Nahuatl:

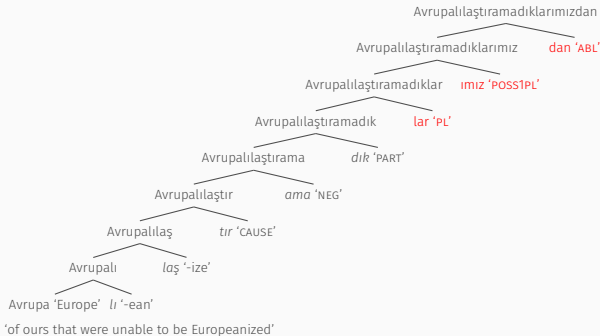
- ni-c-qua in nacatl
I-it-eat the flesh
'I eat the flesh.'
- ni-naca-qua
I-flesh-eat
'I eat flesh.'

- **Degree of synthesis.** How many properties are there per word?
- **Degree of fusion.** How many properties are there per operation?

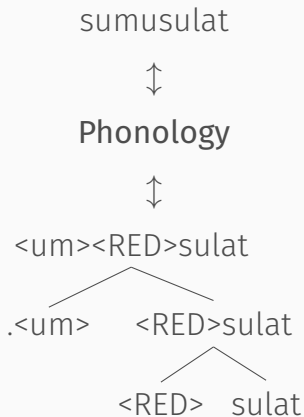
Classical Theories of Morphology

Item-and-Arrangement Morphology

- In item-and-arrangement (IA) morphology, the only operation is **concatenation**
 - Other apparent morphological properties are actually just phonology
- acting on concatenated morphemes
- IA morphology asserts that all languages, at their foundation, are agglutinative or isolating

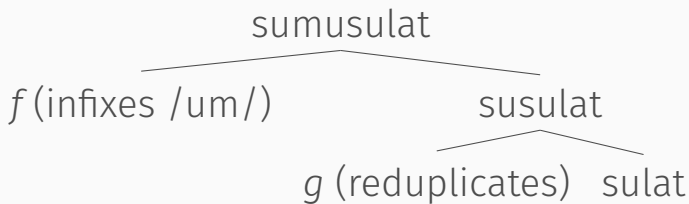


What about Infixation, Reduplication, and Apophony?



Item-and-Process Morphology

- In item-and-process (IP) morphology, only roots are morphemes in the classical sense (items) are processes—pairings of meaning and a function from one form to a transformed form
- Affixes and other operations



Word-and-Paradigm Morphology

- Word-and-paradigm (WP) morphology does away with morphemes altogether
- The basic units of morphology are words, which are grouped into paradigms and related by rules of realization and rules of referral
 - REALIZATION RULES specify

how a property or set of properties are realized in a word

- RULES OF REFERRAL specify how word forms must be formally related to one another

	SG	PL
NOM	amīca	amīcae
VOC	amīca	amīcae
ACC	amīcam	amīcās
GEN	amīcae	amīcārum
DAT	amīcae	amīcīs
ABL	amīcā	amīcīs

Tagalog Again

STEM	PERFECTIVE	CONTEMPLATI	IMPERFECTIVE	GLOSS
kain	kumain	kakain	kumakain	'eat'
sulat	sumulat	susulat	sumusulat	'write'
hanap	humanap	hahanap	humahanap	'seek'

Questions?