Morphology 1

- Introduction
- Morphology
- Morphological Analysis (MA)
- Using FS techniques in MA
- Automatic learning of the morphology of a language

Morphology 2

- Morphology
  - Structure of a word as a composition of morphemes
  - Related to word formation rules
  - Functions
    - Flexion
    - Derivation
    - Composition
- Result of morphologic analysis
  - Morphosyntactic categorization (POS)
    - e.g. Parole tagset (VMIP1S0), more than 150 categories for Spanish
    - e.g. Penn Treebank tagset (VBD), about 30 categories for English
  - Morphological features
    - Number, case, gender, lexical functions

Morphology 3

- Morphologic analysis
  - Decompose a word into a concatenation of morphemes
  - Usually some of the morphemes contain the meaning
    - One (root or stem) in flexion and derivation
    - More than one in composition
  - The other (affixes) provide morphological features
- Problems
  - Phonological alterations in morpheme concatenation
  - Morphotactics
    - Which morphemes can be concatenated with which others

Morphology 4

- Problems
  - Affixes
    - Suffixes, prefixes, infixes, interfixes
  - flexive Affixes ≠ derivative Affixes
  - Derivation implies sometimes a semantic change not always predictible
    - Meaning extensions
    - Lexical rules
  - A derivative suffix can be followed by a flexive suffix
    - love => lover => lovers
  - Flexion does not change POS, sometimes derivation does
  - Flexion affects other words in the sentence
    - agreement

Morphology 5

- Morphotactics
  - Word formation rules
  - Valid combinations between morphemes
    - Simple concatenation
    - Complex models root/pattern
    - Regularity language dependent
- Phonological alterations (Morphophonology)
  - Changes when concatenating morphemes
  - Source: Phonology, morphology, orthography
  - variable in number and complexity
  - e.g. vocalic harmony

Morphology 6

Morphemes

- 1 morpheme:
  - evitar
- 2 morphemes:
  - evitable = evitar + able
- 3 morphemes:
  - inevitable = in + evitar + able
- 4 morphemes:
  - inevitabilidad = in + evitar + able + idad

Morphology 7

Flexive Morphology

- number
  - house houses
  - cheval chevaux
  - casa casas
- verbal form
  - walk walkes walked walking
  - amo amas aman ...
- gender
  - niño niña

Morphology 8

Derivative Morphology

- Form
  - Without change — barcelonés
  - Prefix — inevitable
  - Suffix — importantísimo
- Source
  - verb => adjective — tardar => tardío
  - verb => noun — sufrir => sufrimiento
  - noun => noun — actor => actorazo
  - noun => adjective — atleta => atlético
  - adjective => adjective — rojo => rojizo
  - adjective => adverb — alegre => alegremente

Morphological Analysis 1

Types of morphological analyzers

Formaries

- Dictionaries of word forms
- \+ efficiency
- \+ Languages with few variants (e.g. English)
- \+ extensibility
- \+ Possibility of building and maintenance from a morphological generator
- – Languages with high flexive variation
- – derivation, composition

- FS techniques
  - FSA
    - 1 level analyzers
  - FST
    - > 1 level analyzers

Morphological Analysis 2

2 levels morphological analyzers

- General model for languages with morpheme concatenation
- Independence between lingware and analyzer
- Valid for analysis and generation
- Distinction between lexical and superficial levels
- Parallel rules for morphophonology
- Simple implementation

Morphological Analysis 3

- Morphological rules
  - Define the relations betweens characters (surface) and morphemes and map strings of characters and the morphemic structure of the word.
- Spelling rules
  - Perform at the level of the letters forming the word. Can be used to define the valid phomological alterations.
- Ritchie, Pulman, Black, Russell, 1987

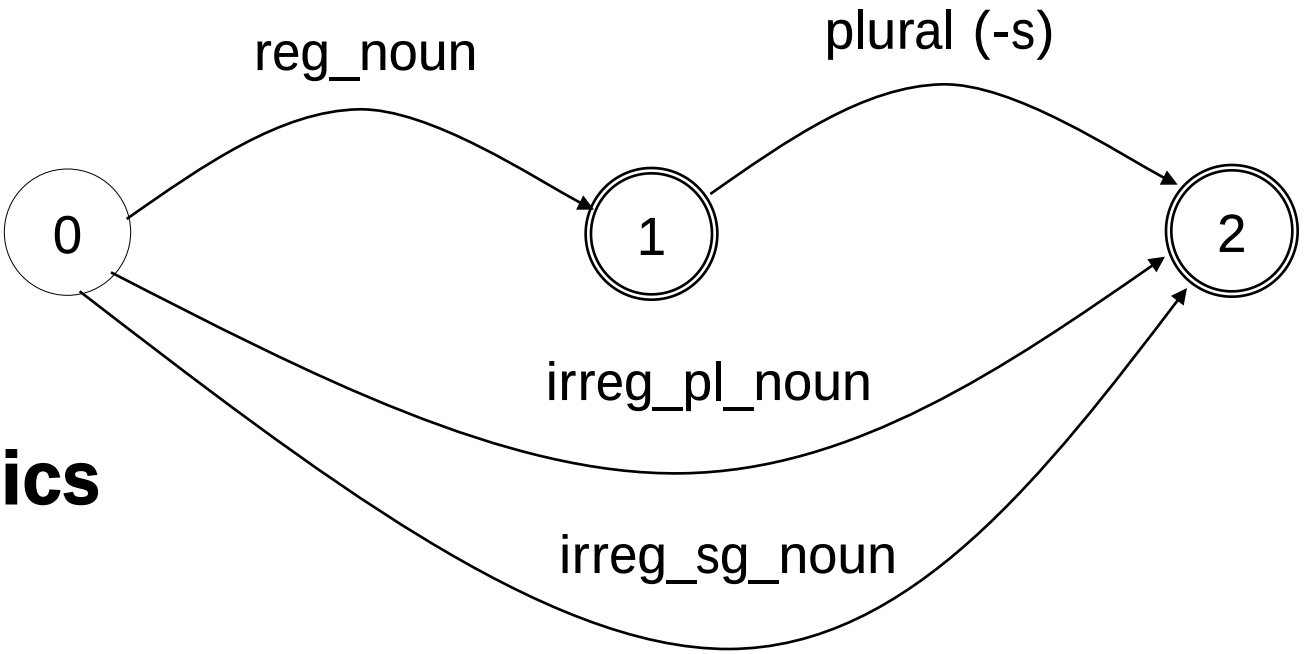
Morphological Analysis 4

- input:
  - form
- output
  - lemma + morphological features

Input	Output
cat	cat + N + sg
cats	cat + N + pl
cities	city + N + pl
merging	merge + V + pres_part
caught	(catch + V + past) or (catch + V + past_part)

Morphological Analysis 5

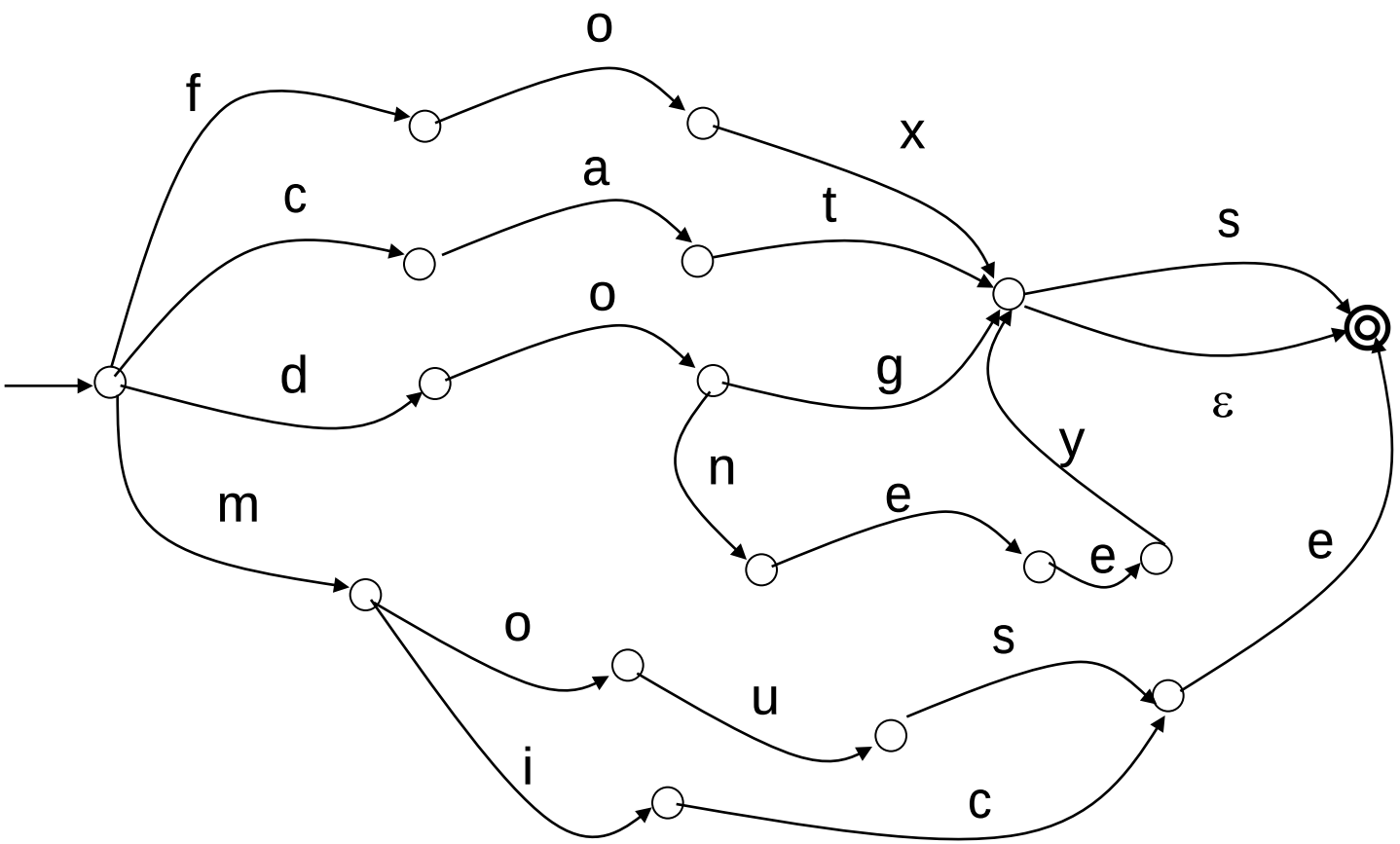
reg_noun	irreg_pl_noun	irreg_sg_noun	plural
fox	sheep	sheep	-s
cat	mice	mouse	
dog			



Morphotactics

Morphological Analysis 6

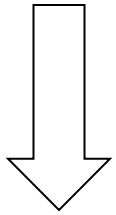
fog
cat
dog
donkey
mouse
mice



Letter Transducers

Morphological Analysis 7

upper level	lexic	cat + N	cat + N + pl
lower level	surface	cat	cats



c:c	a:a	t:t	+N:ε	+pl:s

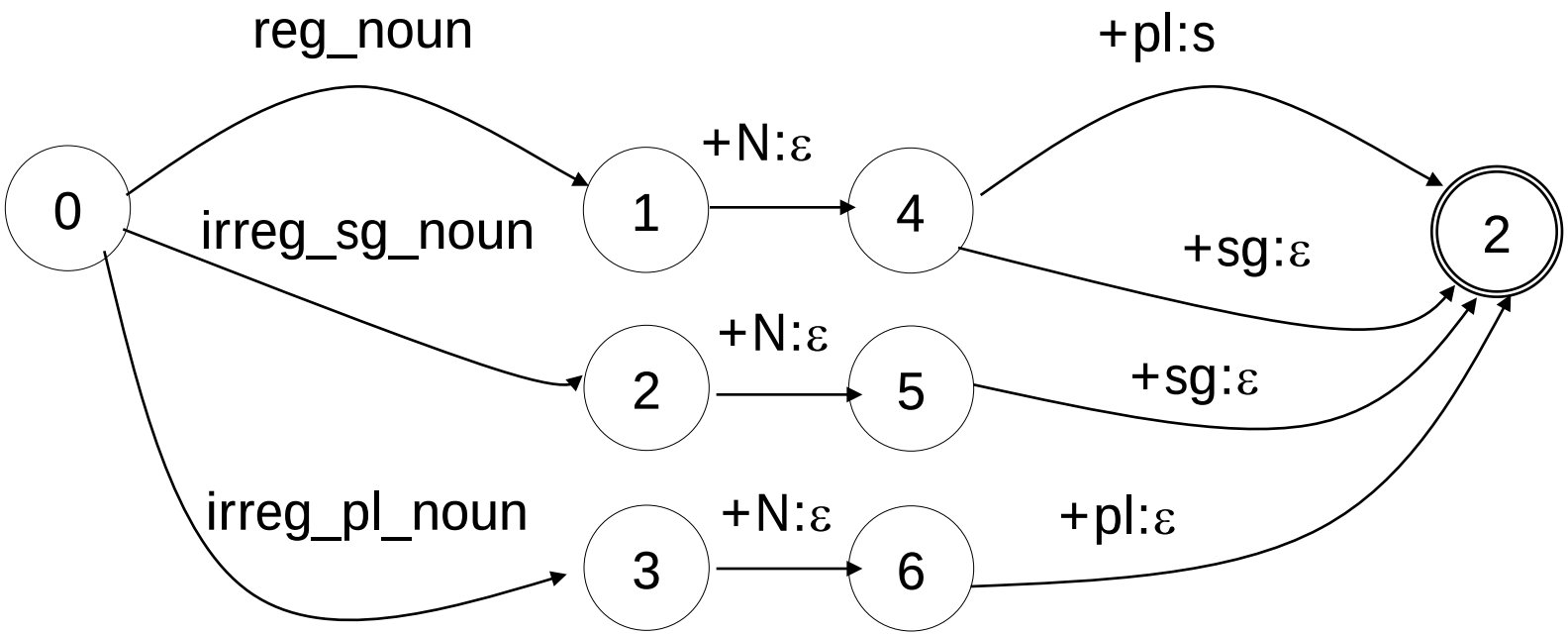
Morphological Analysis 8

Using FST

- As a recognizer
  - From a pair of input strings (one lexical and the other superficial) and answers if one is transduction of the other.
- As a generator
  - generated pairs of strings
- As a translator
  - From a superficial string generates its lexical transduction

Morphological Analysis 9

reg_noun	irreg_pl_noun	irreg_sg_noun	plural
fox	sheep	sheep	s
cat	m o:i u:ε ce	mouse	
dog	g o:e o:e se	goose	



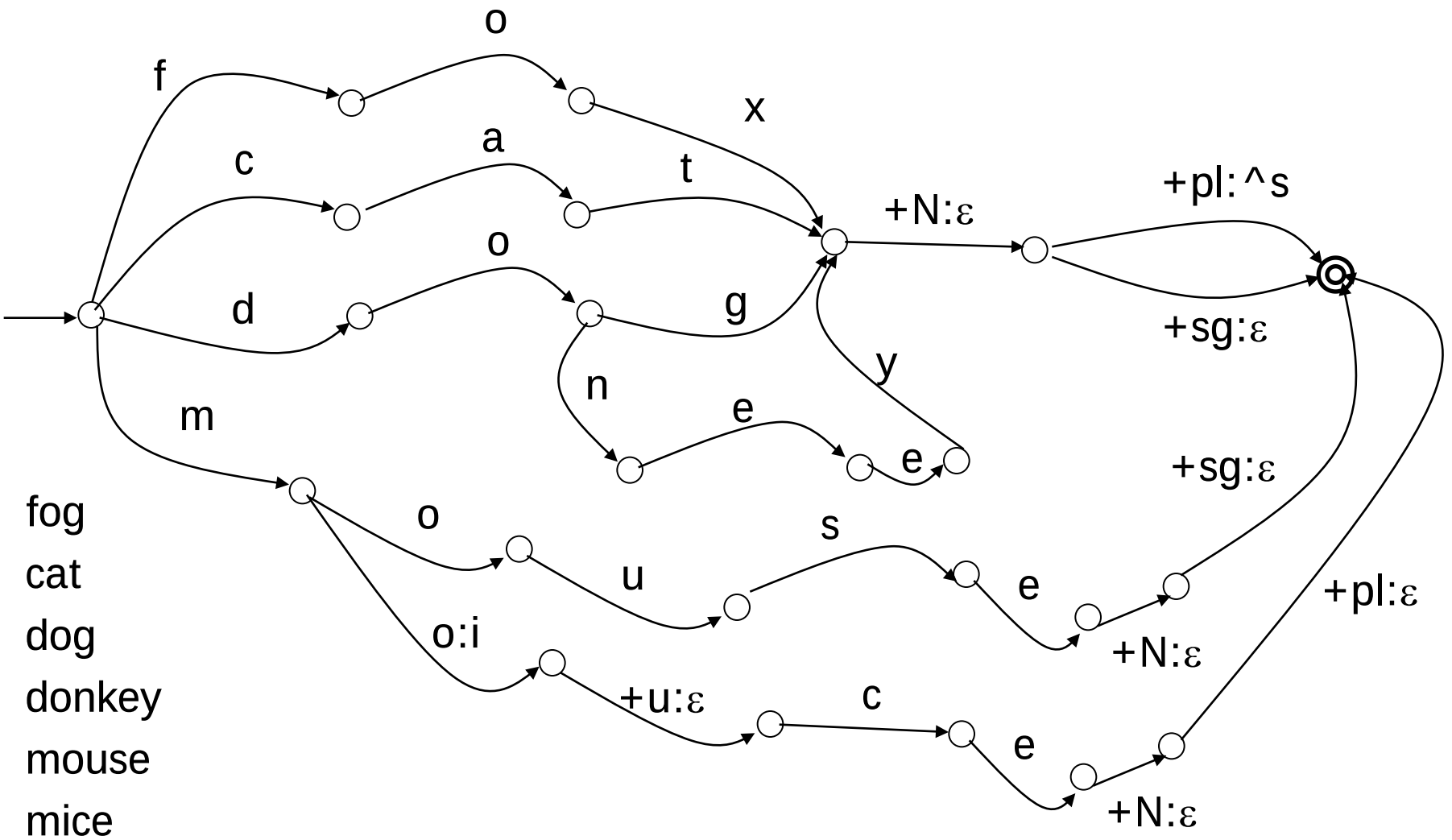
Morphological Analysis 10

morphotactics

spelling rules

lexical level	f	o	x	+N	+pl
intermediate level	f	o	x	^	s
superficial level	f	o	x	e	s

Morphological Analysis 11



Morphological Analysis 12

Spelling rules

name	description	example
consonant doubling	single letter consonant doubled before -ing/-ed	beg/begging
e deletion	silent e dropped before -ing/-ed	make/making
e insertion	e added after -s,-z,-x,-ch,-sh before -s	watch/watches
y replacement	-y changes to -ie before -s, to i before -ed	try/tries
k insertion	verbs ending with voyel +c add -k	panic/panicked

Morphological Analysis 13

Spelling rules: e-insertion

ε:e ⇔ [xsz]^:ε ___ s#

⇒ (down) **decomposition** (down) */⇐*

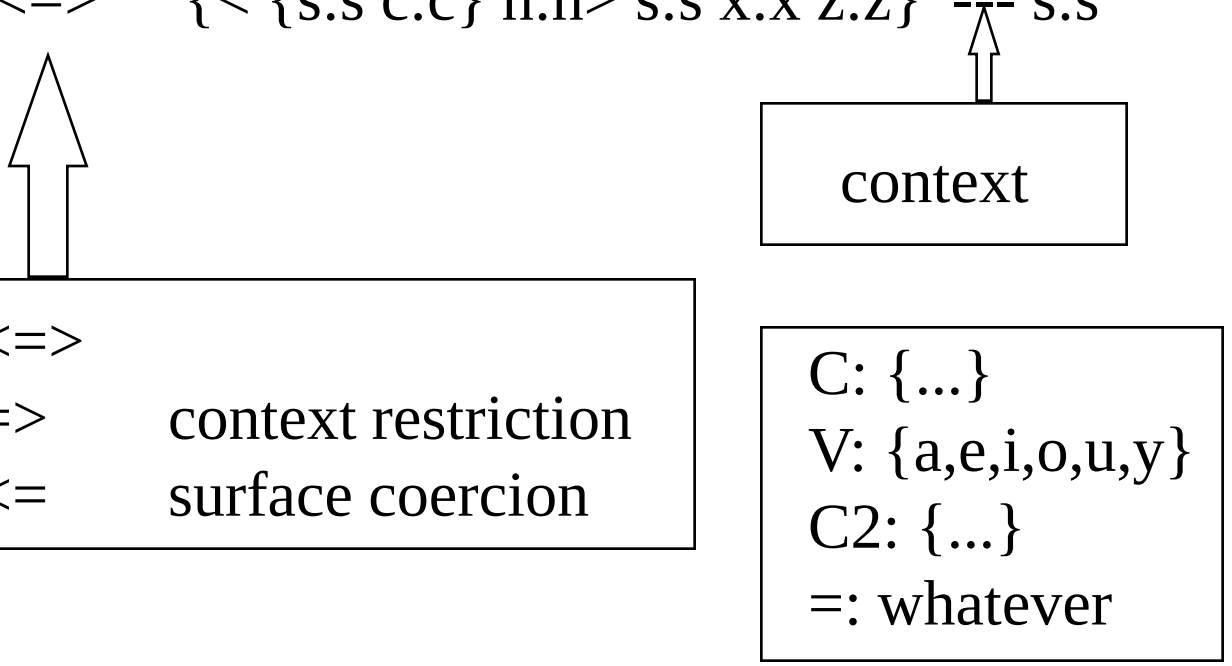
ε:e ⇒[xsz]^:ε ___ s#

ε:ε */⇐* [xsz]^:ε ___ s#

Morphological Analysis 14

<u>epenthesis</u>

+ : e <=> {< {s:s c:c} h:h> s:s x:x z:z} --- s:s



example:

box	+	s
box	e	s

Morphological Analysis 15

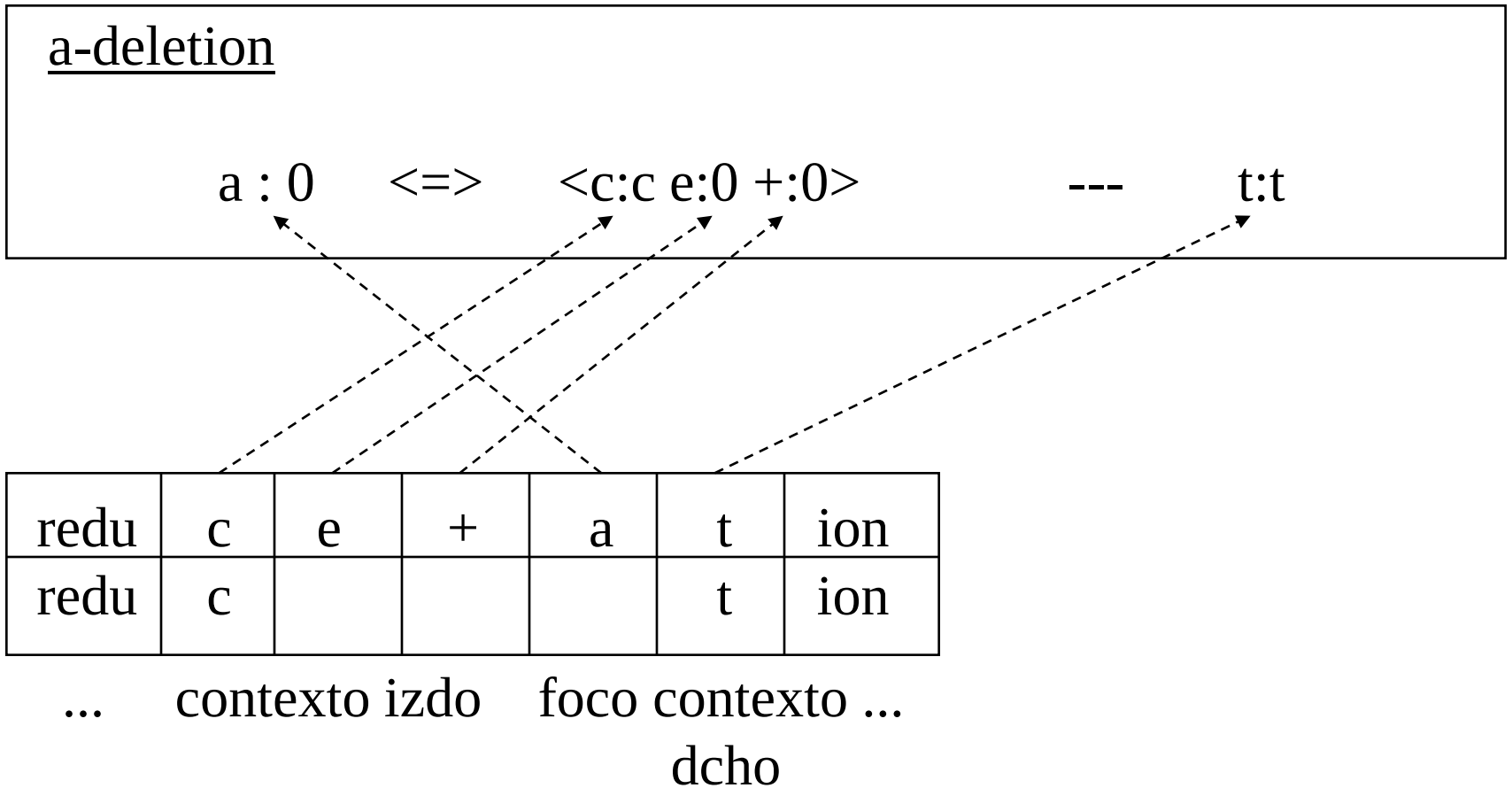
<u>e-deletion</u>

e : 0	<=>	= :C2	---	<+:0 V:= >
	or	<C:C V:V>	---	< +:0 e:e >
	or	<c:c g:g>	---	< +:0 {e:e i:i} >
	or	l:0	---	+:0
	or	c:c	---	< +:0 a:0 t:t b:b>

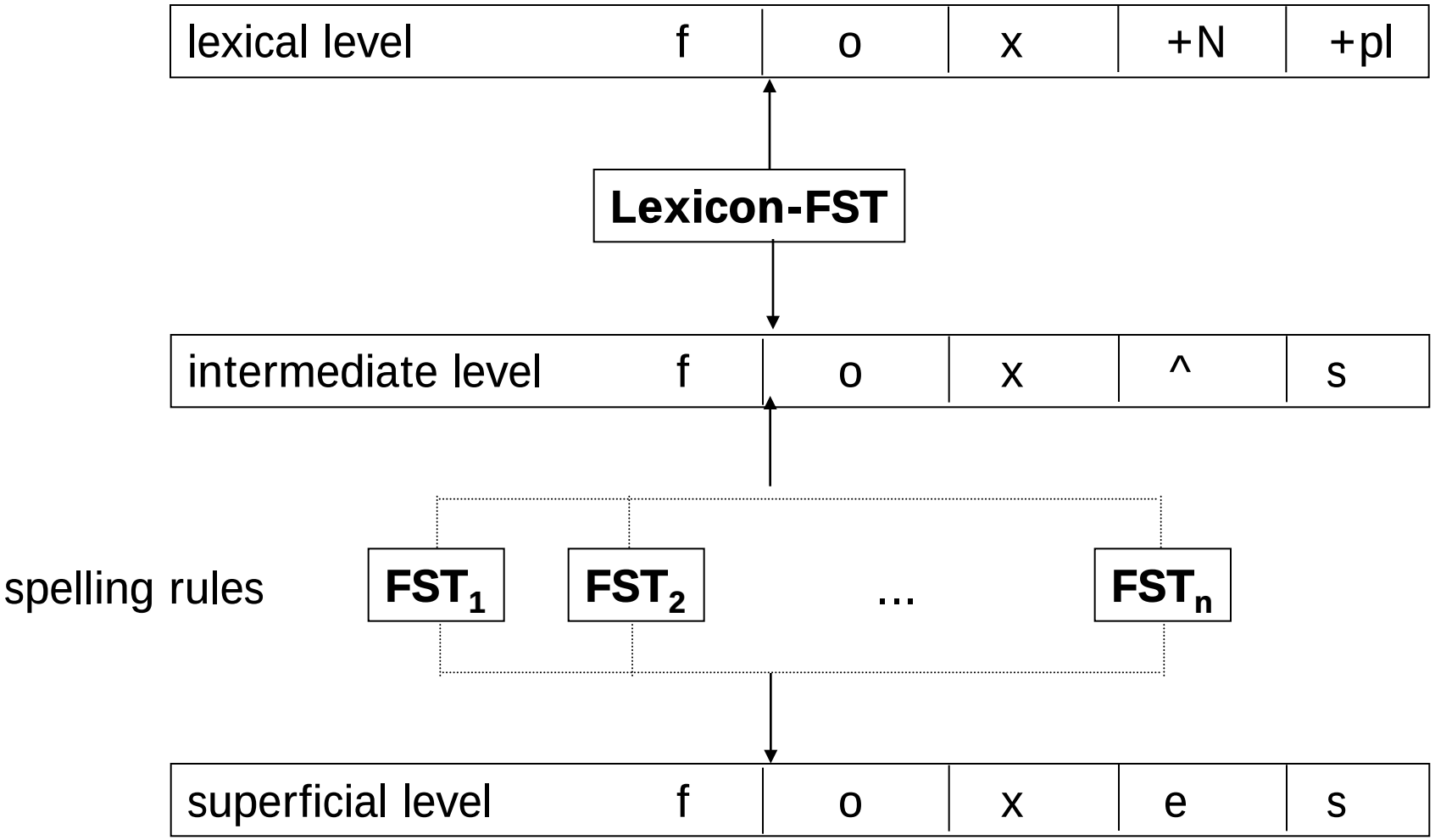
mov	e	+	ed
mov			ed

agre	e	+	ed
agre			ed

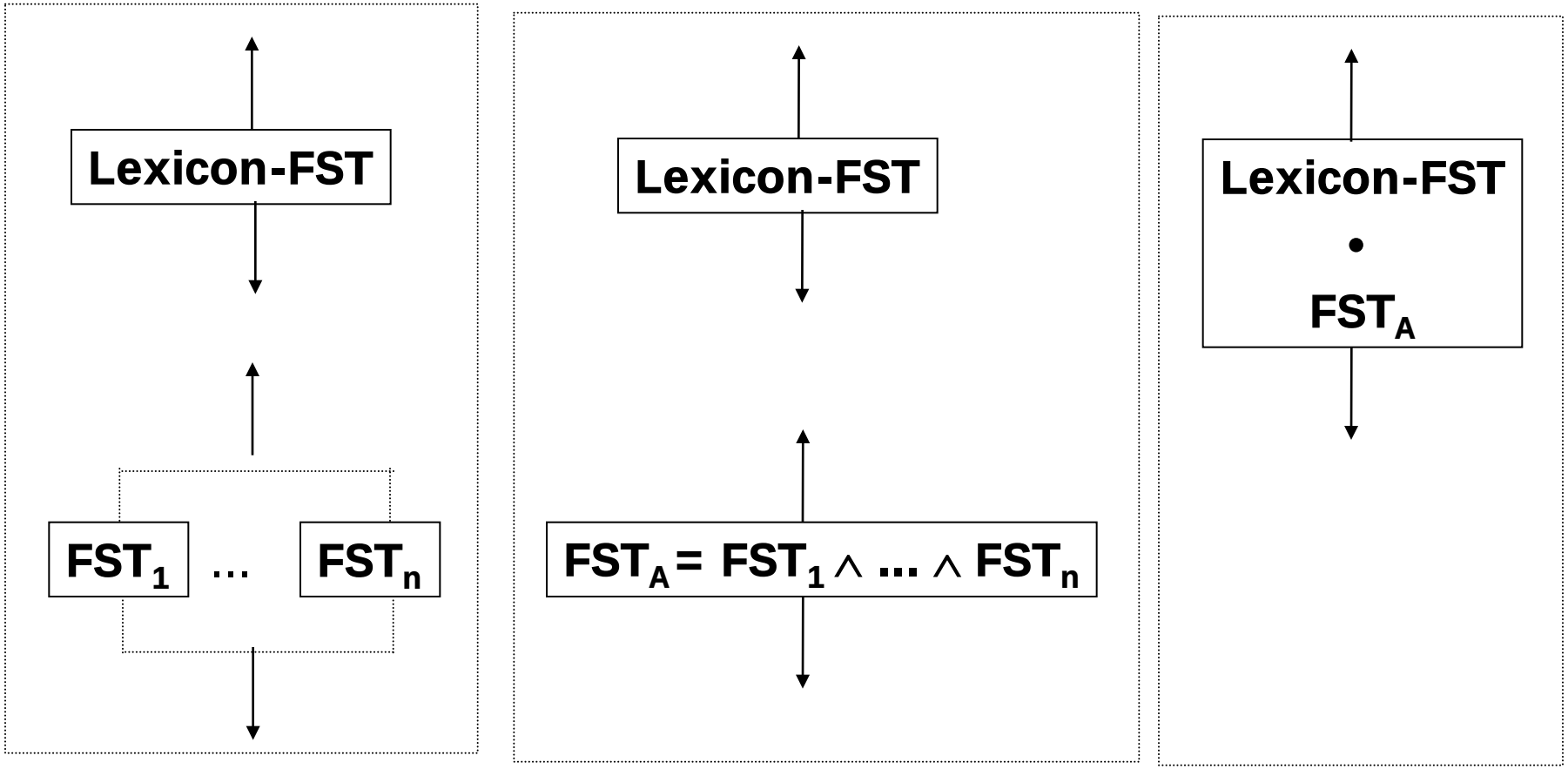
Morphological Analysis 16



Morphological Analysis 17



Morphological Analysis 18



Automatic morphology learning 1

- Problem
  - Paradigm stem + affixea
  - Obtaining the stems
  - Classification of stems into models
  - Learning part of the morphology (e.g. derivational)
- Two approaches
  - No previous morphologic knowledge is available
    - Goldsmith, 2001
    - Brent, 1999
    - Snover, Brent, 2001, 2002
  - Morphologic knowledge can be used
    - Oliver at al, 2002

Automatic morphology learning 2

- Automatic morphological analysis
  - Identification of borders betwen morphemes
    - Zellig Harris
      - {prefix, suffix} conditional entropy
  - bigrams and trigrams with high probability of forming a morpheme
  - Learning of patterns or rules of mapping between pairs of words
  - Global approach (top-down)
    - Golsdmith, Brent, de Marcken

Automatic morphology learning 3

- Goldsmith's system based on MDL (Minimum Description Length)
  - Initial Partition: word -> stem + suffix
    - split-all-words
      - A good candidate to {stem, suffix} splitting in a word has to be a good candidate in many other words
    - MI (mutual information) strategy
      - Faster convergence
  - Learning Signatures
    - {signatures, stem, suffixes}
  - MDL

Automatic morphology learning 4

- Semi-automatic morphological analysis
  - Oliver, 2004
  - Starts with a set of manually written morphological rules
    - TL:TF:Desc
      - lemma ending
      - form ending
      - POS
  - Lists of non flexive classes , closed classes and irregular words
  - Corpora
    - Serbo-Croatian 9 Mw
    - Russian 16 Mw