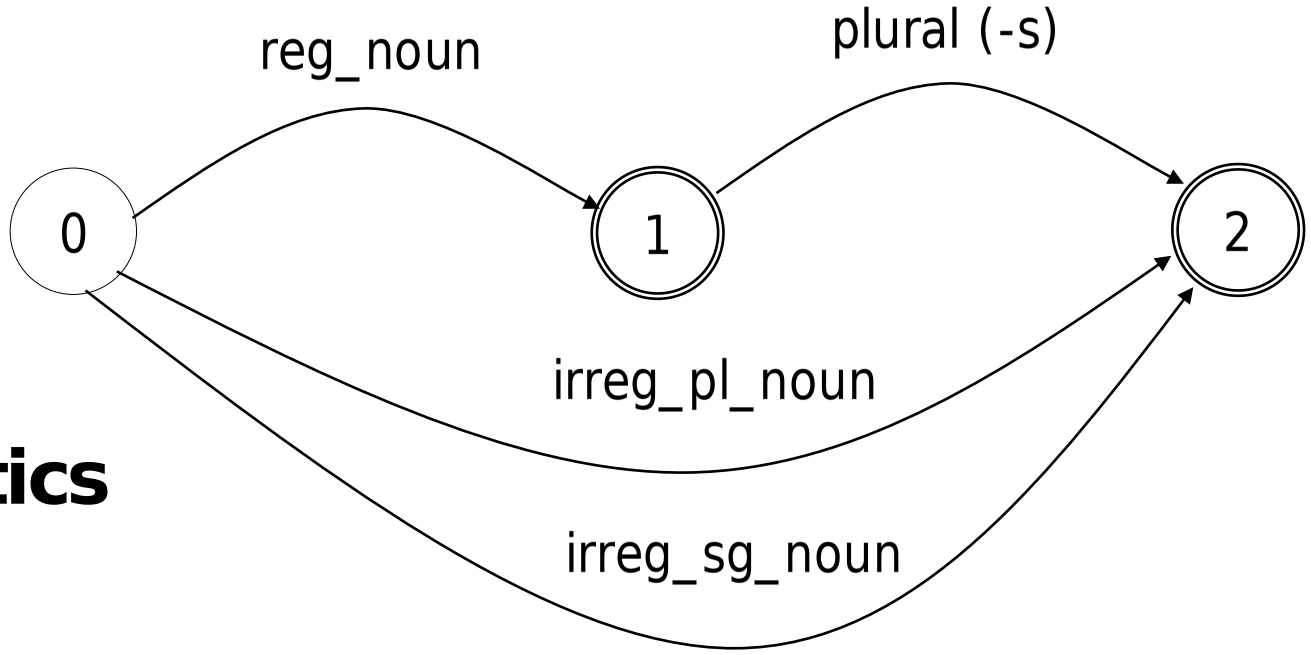
Morphological Analysis 5

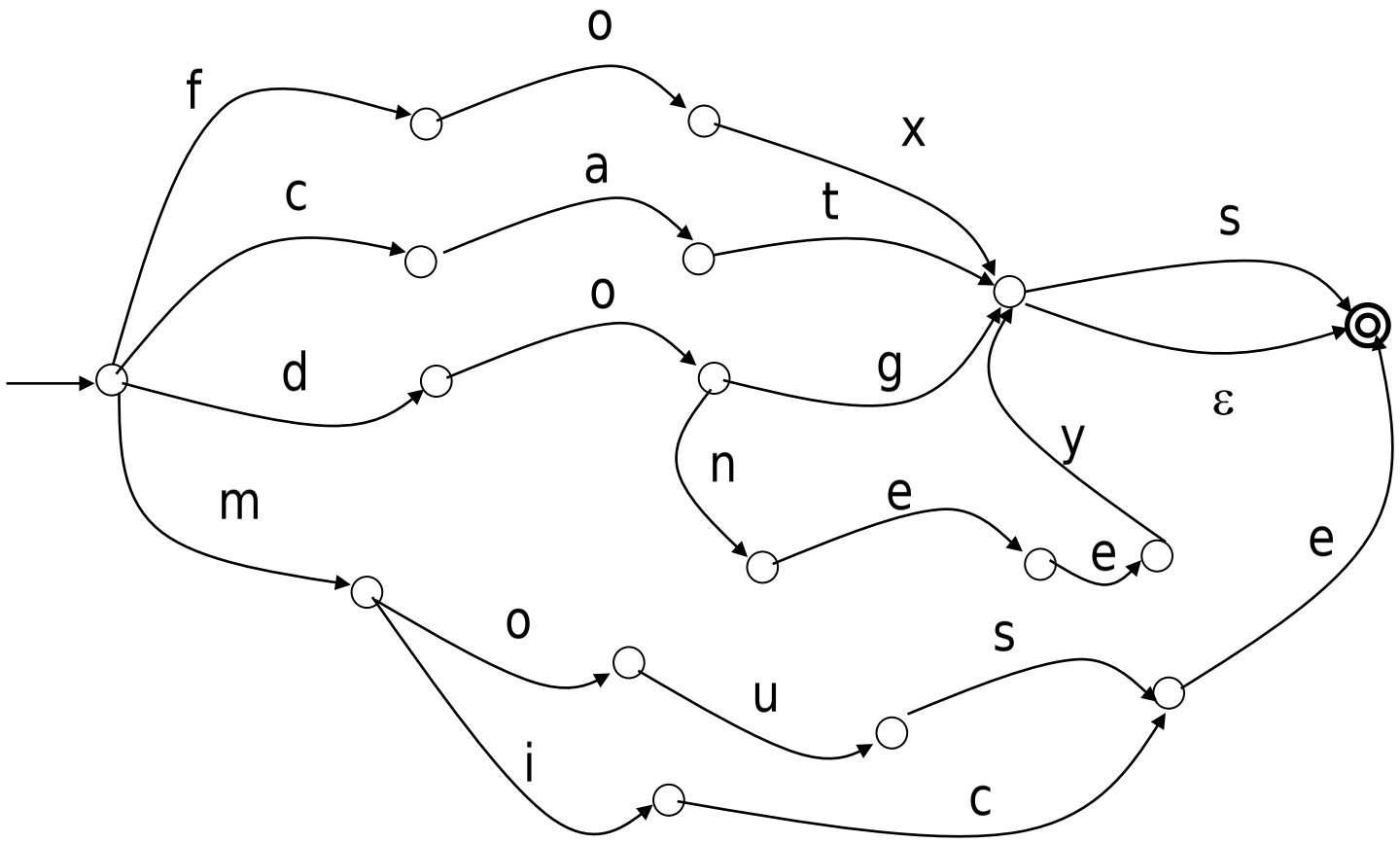
reg_noun	irreg_pl_noun	irreg_sg_noun	plural
fox	sheep	sheep	-s
cat	mice	mouse	
dog			



Morphotactics

Morphological Analysis 6

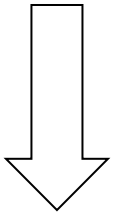
fog
cat
dog
donkey
mouse
mice



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

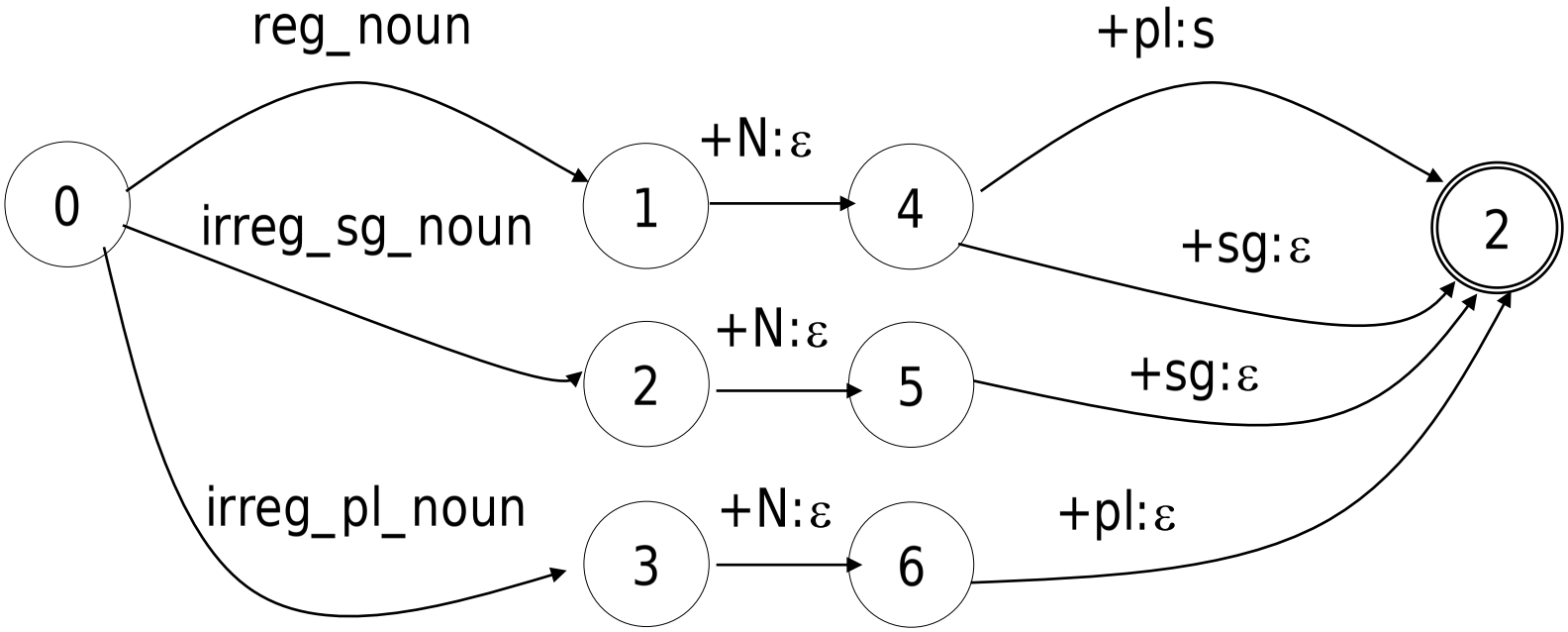
Morphological Analysis 8

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	



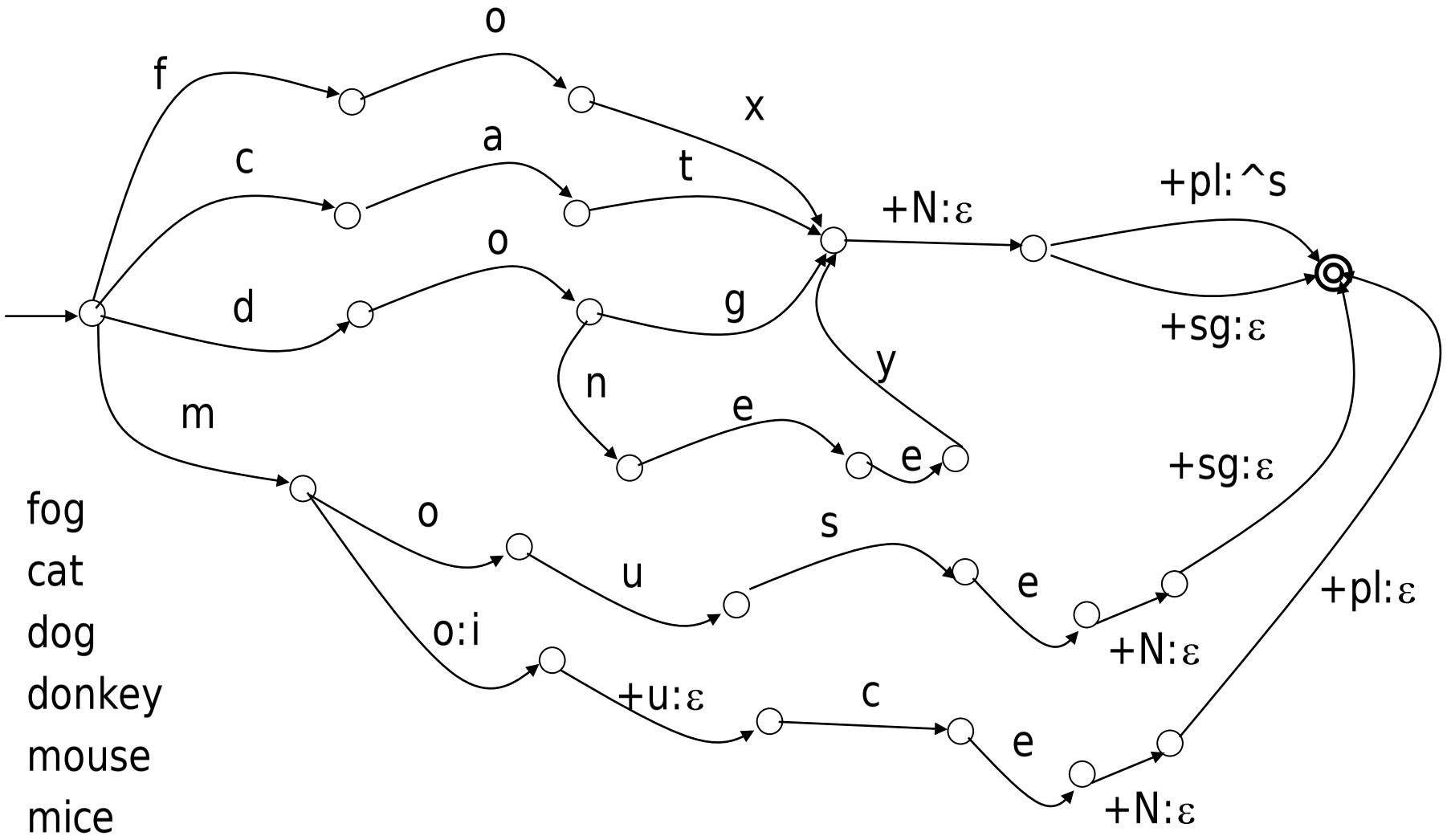
Morphological Analysis 10

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



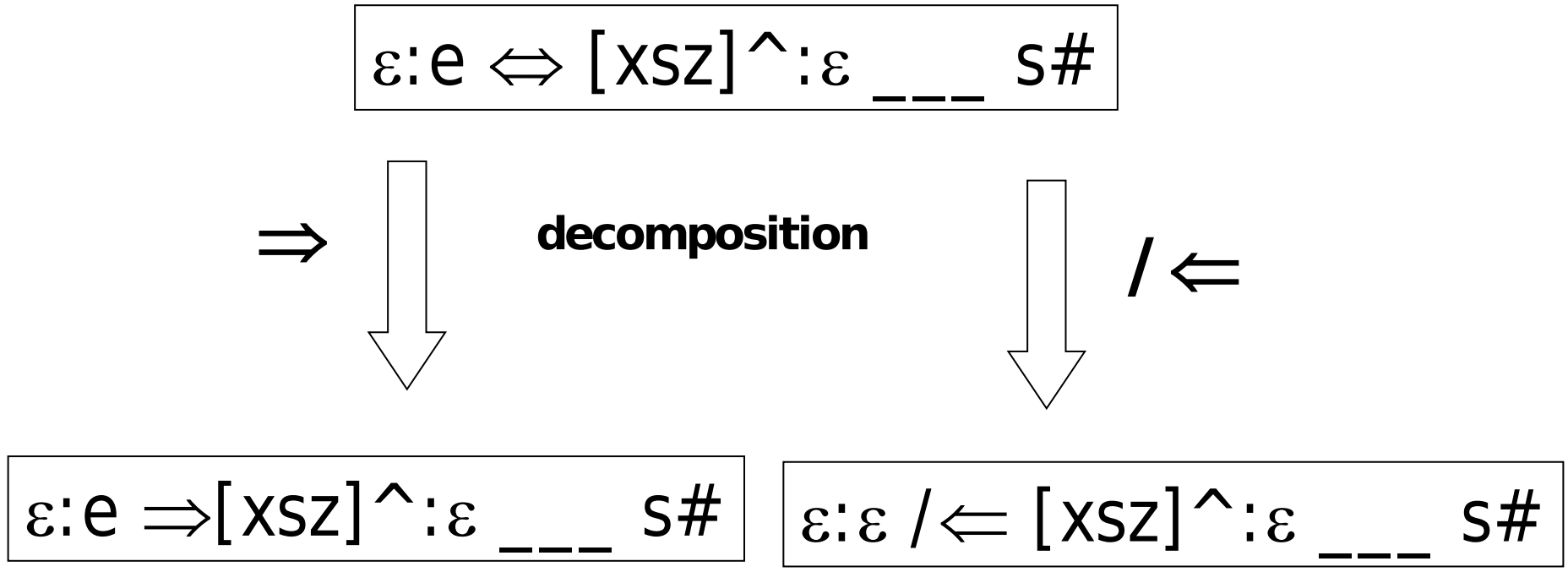
Morphological Analysis 12

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

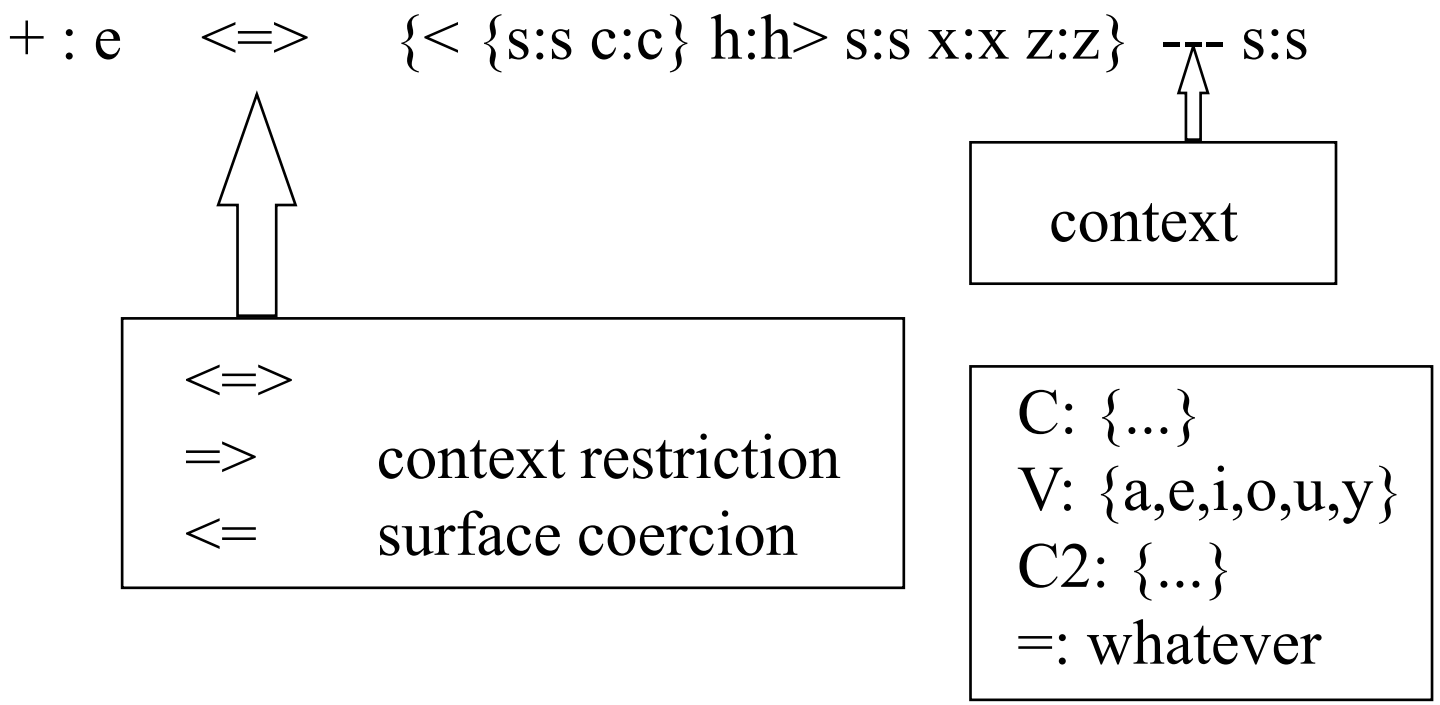
Morphological Analysis 13

Spelling rules: e-insertion



Morphological Analysis 14

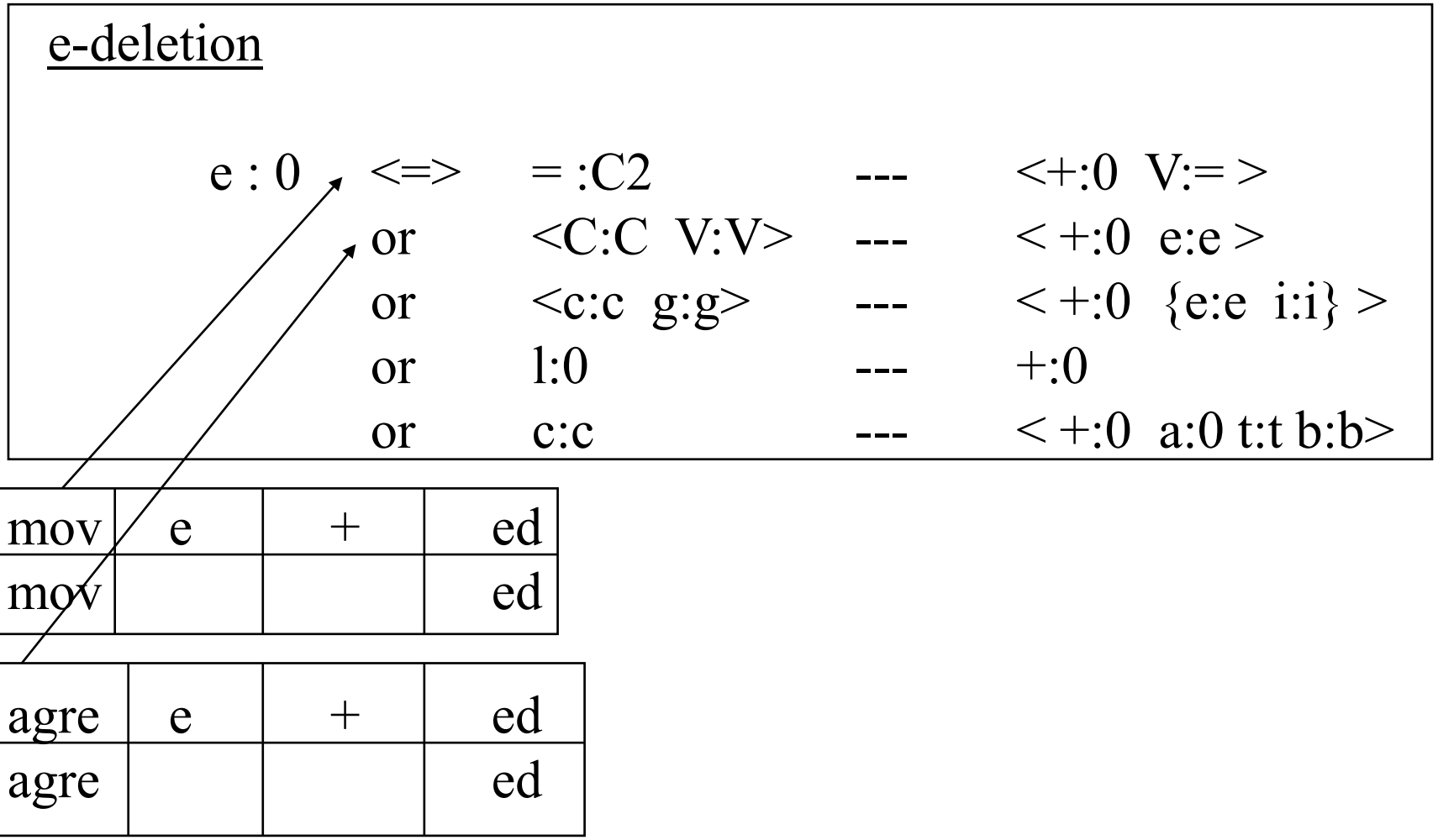
epenthesis



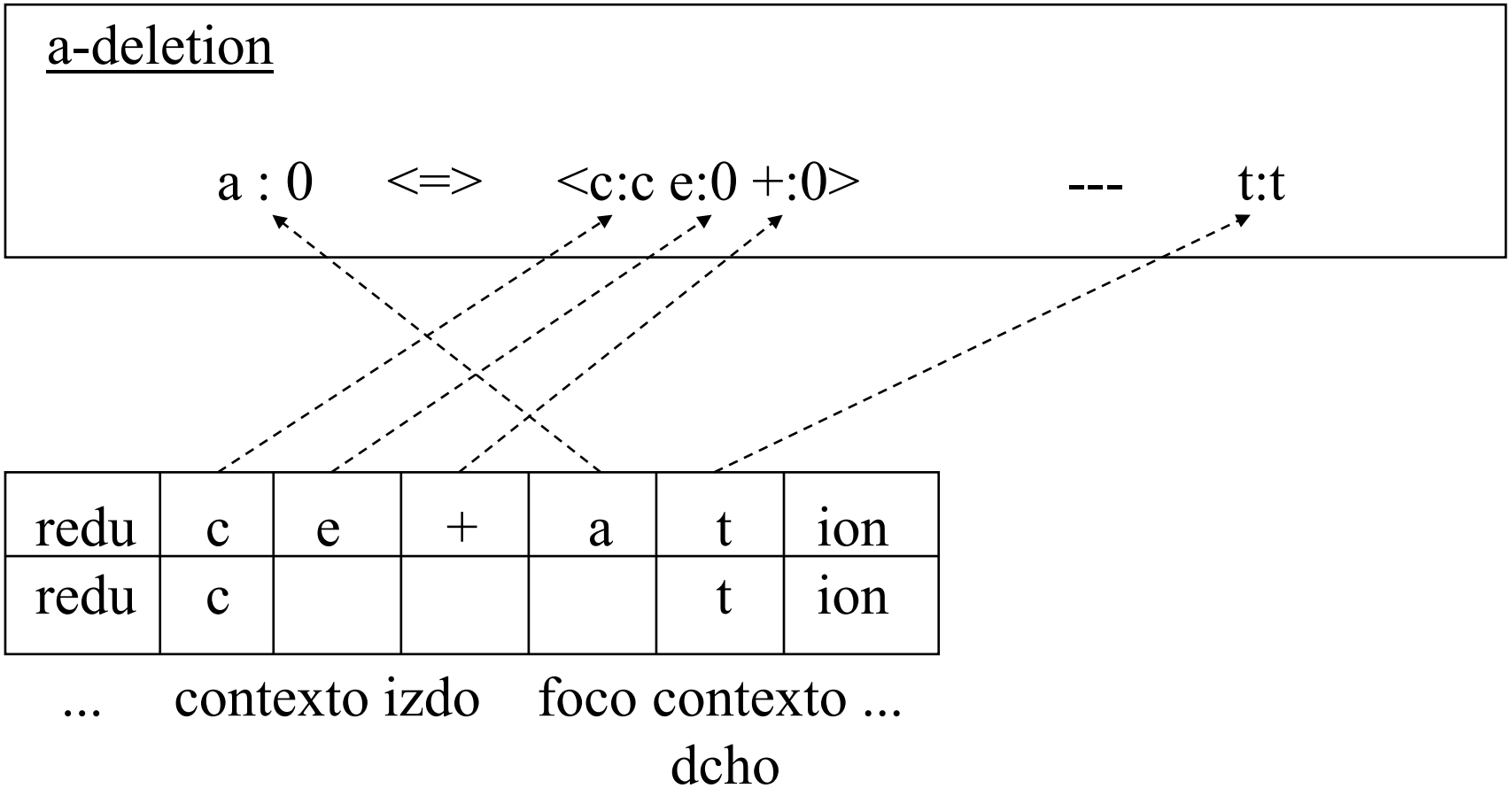
example:

box	+	s
box	e	s

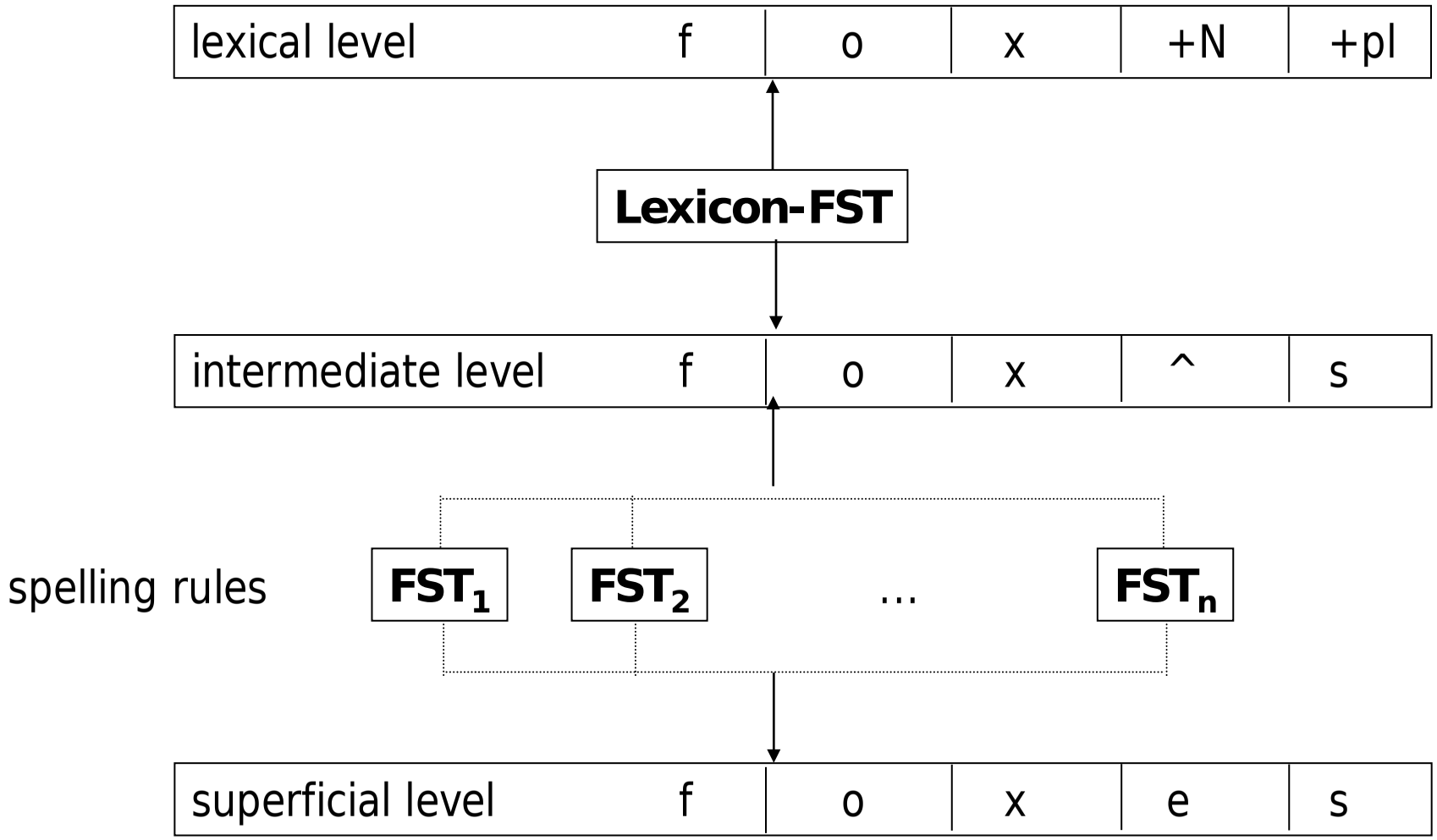
Morphological Analysis 15



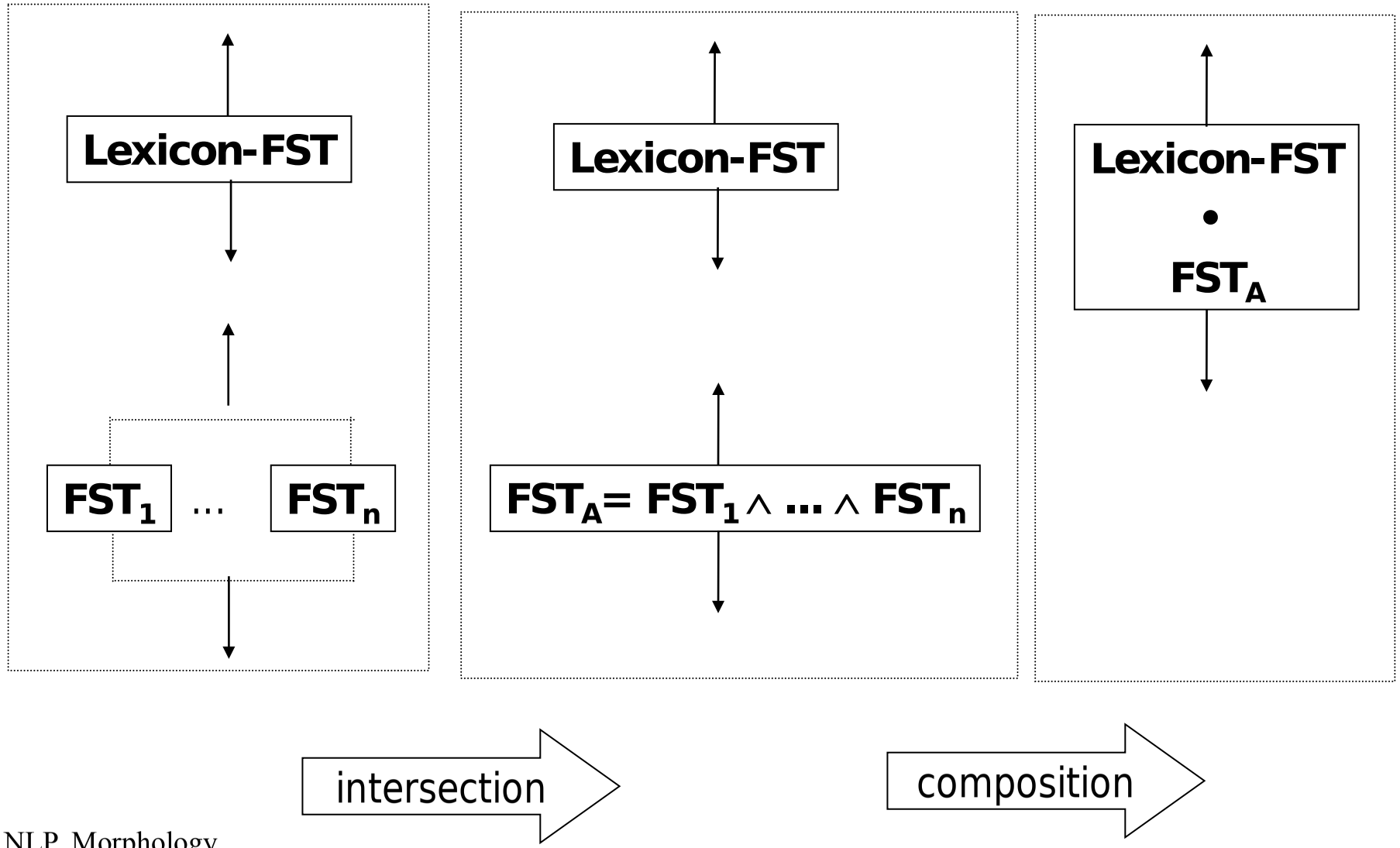
Morphological Analysis 16



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