# Finite State Transducers

Note: Most examples are taken from Jurafsky & Martin (2008)

# Eliza (Weizenbaum, 1966)

User: Men are all alike

ELIZA: IN WHAT WAY

User: They’re always bugging us about something or other

ELIZA: CAN YOU THINK OF A SPECIFIC EXAMPLE?

User: Well, my boyfriend made me come here

ELIZA: YOUR BOYFRIEND MADE YOU COME HERE

User: He says I’m depressed much of the time

ELIZA: I AM SORRY TO HEAR THAT YOU ARE DEPRESSED

# Recognition is not enough

Step 1: replace first person with second person references

I am depressed -> you are depressed
my boyfriend made me come here -> your boyfriend made me come here

Step 2: use regular expressions to generate replies

```
YOU ARE (depressed + sad) -> I AM SORRY TO HEAR YOU ARE \1
YOU ARE (depressed + sad) -> WHY DO YOU THINK YOU ARE \1
* all * -> IN WHAT WAY
* always * -> CAN YOU THINK OF A SPECIFIC EXAMPLE
```

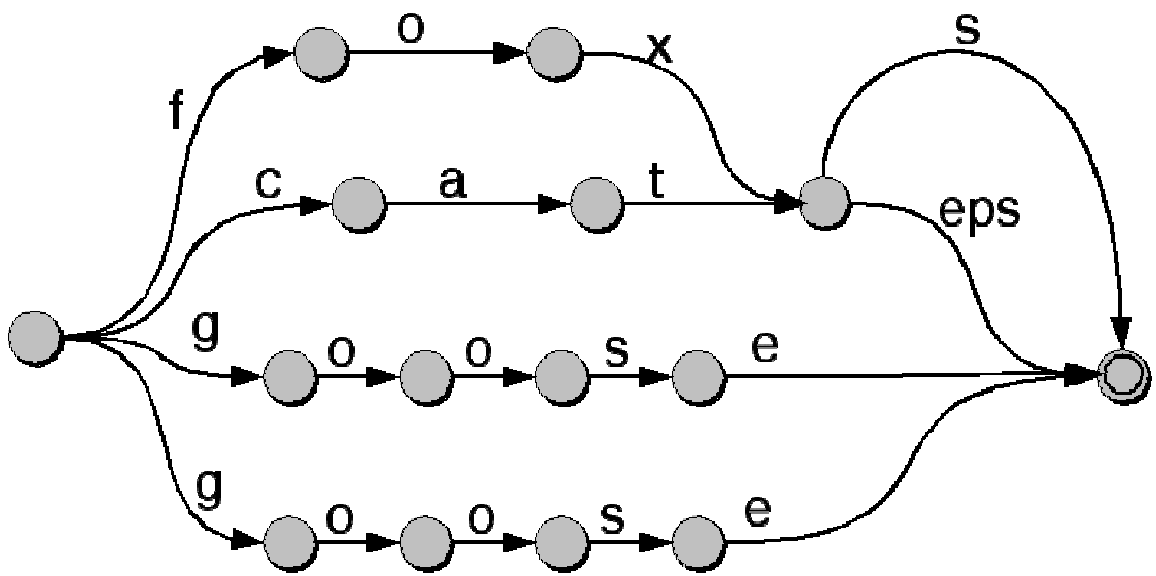
Step 3: possible transformations are ranked by scores

# Recognition

| | Regular Nouns | | Irregular Nouns | |
|---|---|---|---|---|
| Singular | cat | thrush | mouse | ox |
| Plural | cats | thrushes | mice | oxen |

-s ibis/ibises
-z waltz/waltzes
-sh thrush/thrushes
-ch finch/finches
-x box/boxes

-y butterfly/butterflies



# Many languages are morphologically much more complex than English

Turkish:

uygarlaştıramadıklarımızdanmışsınızcasına

*uygar +laş +tır +ama +dık +lar +ımız +dan +mış +sınız +casına*

civilized +BEC +CAUS +NABL +PART +PL +P1PL +ABL +PAST +2PL +AsIf

"(behaving) as if you are among those whom we could not civilize"

| | |
|---|---|
| +BEC | "become" |
| +CAUS | the causative verb marker ('cause to X') |
| +NABL | "not able" |
| +PART | past participle form |
| +P1PL | 1st person pl possessive agreement |
| +2PL | 2nd person pl |
| +ABL | ablative (from/among) case marker |
| +AsIf | derivationally forms an adverb from a finite verb |

# Another famous example


St Mary's church in the hollow of the white hazel near to the rapid whirlpool and the church of St Tysilio of the red cave

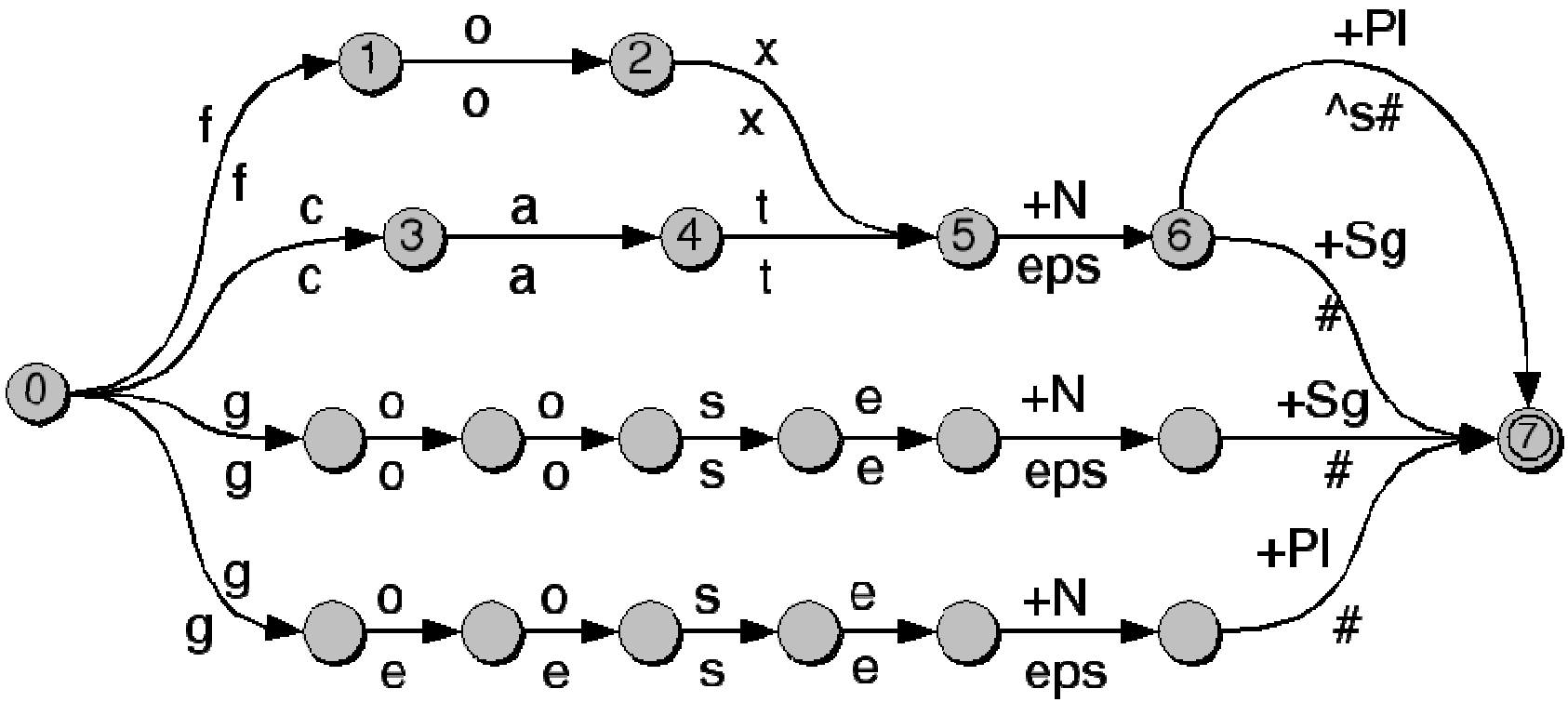
# Parsing/Generation vs. Recognition

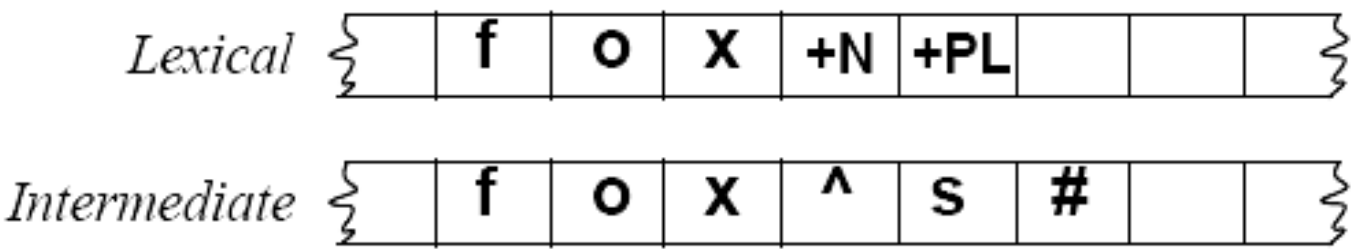
- We can only recognize words
- But this isn't the same as parsing
  - Parsing: building structure
  - Usually if we find some string in the language we need to find the structure in it (parsing)
  - Or we have some structure and we want to produce a surface form (production/generation)
- Example
  - From "`cats`" to "`cat +N +PL`"

# Finite State Transducers (FST)

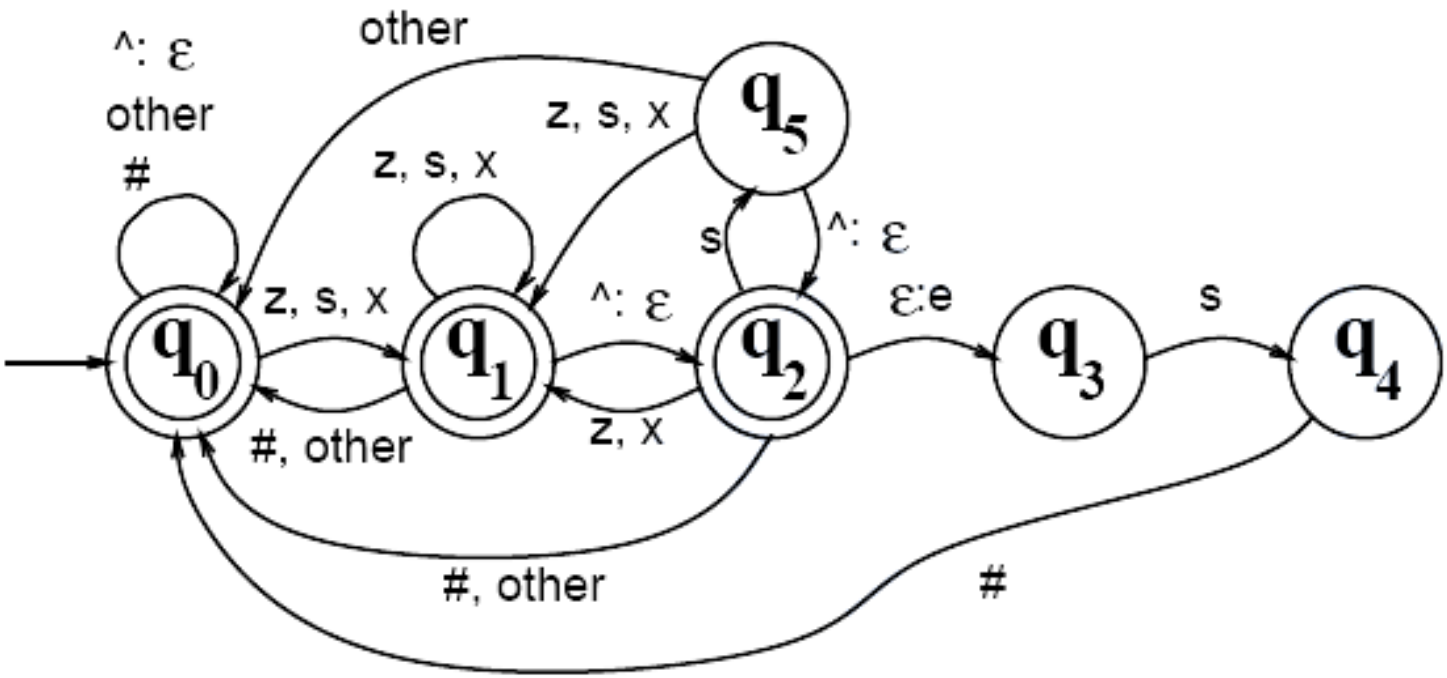
- The simple story
  - Add another tape
  - Add extra symbols to the transitions
  - On one tape we read "`cats`", on the other we write "`cat +N +PL`"

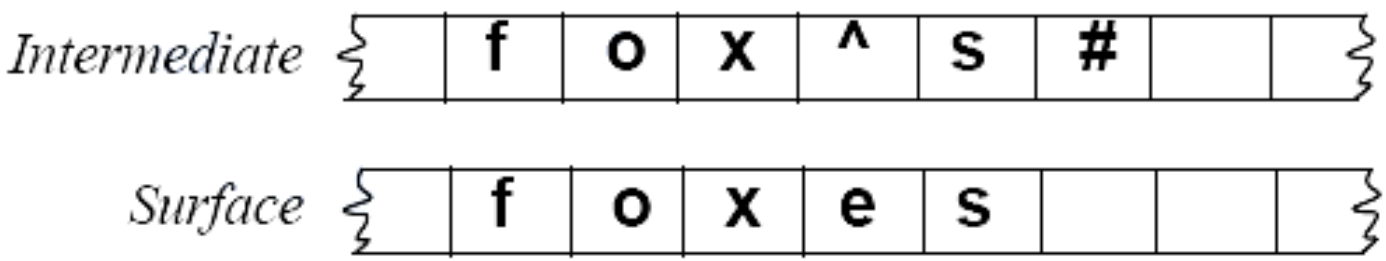
# FST: English Plural (1)



# English Plural: FST (2)



# Applications of Finite State Technologies in CL

- Phonology: models for speech recognition
- Representing lexicons and dictionaries
- Morphology: analysis and generation
- Stemming
- Spell checking
- Shallow parsing
- Named entity recognition
- Sentence boundary detection: segmentation
- Translation...

# Exercise: FST

- Write a finite state transducer for regular English verbs.
- Write a finite state transducers for regular comparative forms (late/later/latest).