Natural Language Processing Question 1   2003
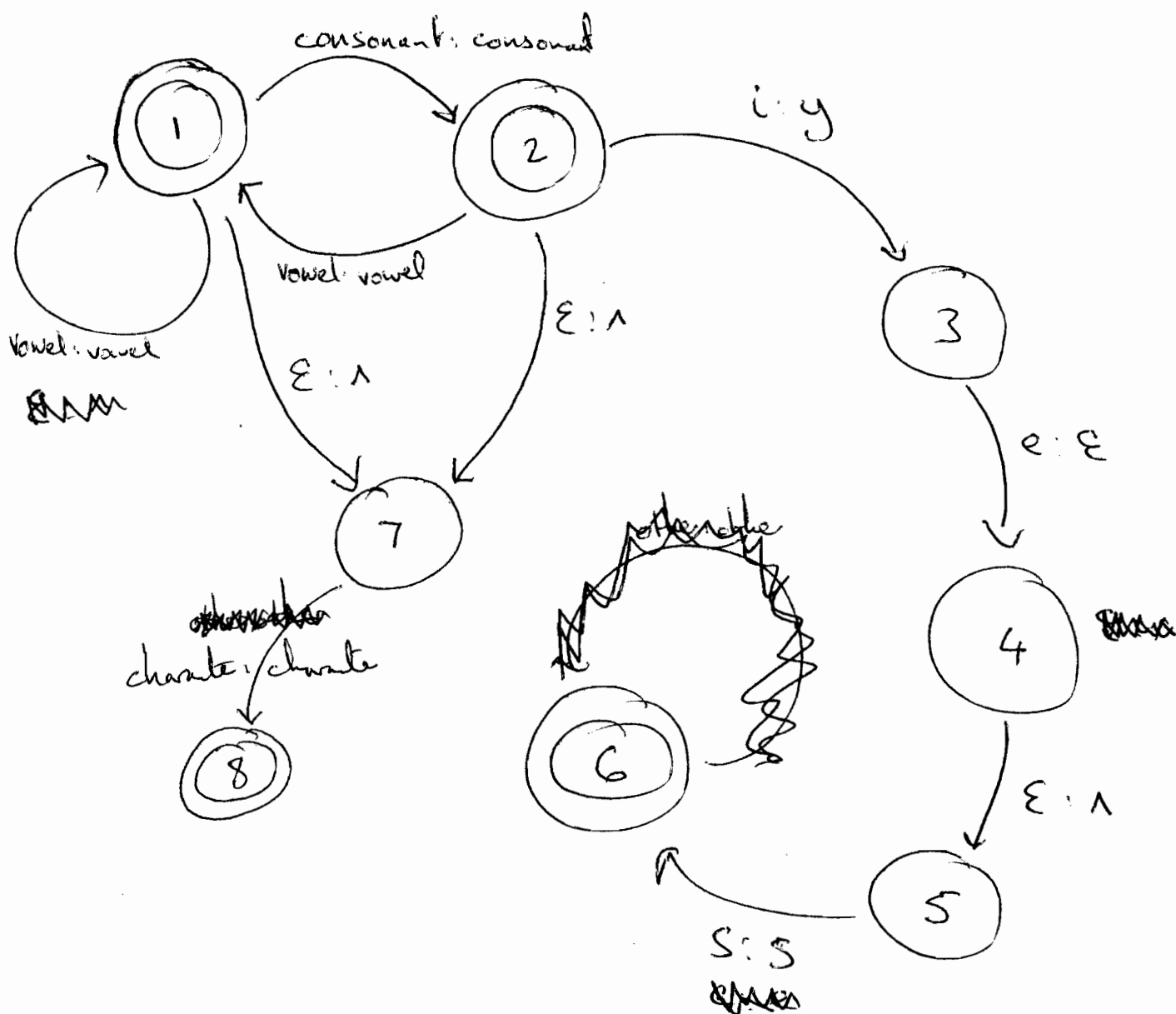--

Corresponds to lecture 2 on the syllabus

(a)

(i) morpheme is the minimal information carrying unit
ii) affix is a morpheme that cannot occur alone i.e., without a stem

(b) see hand-drawn FST

(c) the FST might be simply extended by adding states after 5 to
correspond to the affix 'ed' (or, more generally, by adding e:e to the
5 to 6 transition and puuting a loop on 6). However, the correct formulation
depends on the other FSTs in the system: there should be a separate
FST for 'e-deletion' operating in parallel with this FST.

b)

The finite-state transducer diagram shows:

- State 1 (accept state) with a self-loop labelled **vowel : vowel**
- Transition from state 1 to state 2 labelled **consonant : consonant**
- Transition from state 2 to state 1 labelled **vowel : vowel**
- Transition from state 2 to state 3 labelled **i : y**
- Transition from state 1 to state 7 labelled **ε : ∧**
- Transition from state 2 to state 7 labelled **ε : ∧**
- Transition from state 7 to state 8 (accept state) labelled **character : character**
- Transition from state 3 to state 4 labelled **e : ε**
- Self-loop on state 4 labelled **other : other** (struck through)
- Transition from state 4 to state 5 labelled **ε : ∧**
- Transition from state 5 to state 6 (accept state) labelled **s : s**
- Self-loop on state 6 labelled **other : other**

surface : underlying

ε is empty character

∧ is affix boundary

◎ accept state

Convention is that no characters can be "left over"