

## 2007 Paper 12 Question 1

### Natural Language Processing

- (a) In German, the third person singular present inflection of weak verbs is generally formed by adding the 't' to the stem. Exceptions to this rule include verbs with stems that end in 't' or 'd' which are formed by adding 'et' instead of 't'. The following table gives some examples (^ is used as the affix marker in the underlying form).

stem	surface	underlying
kauf	kauft	kauf <sup>t</sup>
arbeit	arbeitet	arbeit <sup>t</sup>

Draw a finite state transducer (FST) that relates surface and underlying forms according to this pattern. (Only the inflected forms should be accepted by the transducer since the stems by themselves do not correspond to words.) Explain the notation that you use and outline how the FST could be used in morphological analysis and generation. [14 marks]

- (b) The past participle of these verbs is the same as the third person singular present, but with 'ge' before the stem. Assume that the underlying form of the past participle is treated as having the artificial suffix '^P', as indicated below.

stem	surface	underlying
kauf	gekauft	kauf <sup>P</sup>
arbeit	gearbeitet	arbeit <sup>P</sup>

Modify the FST that you gave in answer to part (a) above to allow for the past participle as well as the third person singular present. What does this example illustrate about limitations of the FST approach? [6 marks]