

Übungsaufgaben 5
(Abgabetermin: 21.1.2003)

5.1

- a. Lesen Sie aus dem CL-Handbuch (Carstensen et al.) den Artikel 4.2 (Textkorpora). Relevant sind vor allem die Abschnitte 4.2.2 und 4.2.4.
- b. Zeigen Sie anhand von Beispielen aus dem folgenden Text (Saarbrücker Zeitung vom 7.1.2003) Probleme bei der Tokenisierung auf:

Zum zweiten Mal nach 1998/99 triumphierte der 25-jährige Finne bei der deutsch-österreichischen Tour und sorgte für den insgesamt 13. Sieg des finnischen Teams. Für seinen Erfolg erhielt Janne Ahonen ein Luxus-Auto im Wert von 70 000 Euro.

- 5.2 Versuchen Sie, den folgenden Satz mit den Wortart-Tags des Stuttgart-Tübinger Tagsets (STTS) zu annotieren. Eine Übersicht findet sich unter <http://www.ims.uni-stuttgart.de/projekte/corplex/german-tagsets.html> bzw. <http://www.ims.uni-stuttgart.de/projekte/corplex/TagSets/stts-table.html>. Eine Kopie ist in der Bibliothek im Ordner.

Jedoch lässt sich auch nach Feststellung der Wortart die Grundform oft nicht eindeutig bestimmen.

Wenn Sie im Einzelfall Probleme bei der Wortart-Zuordnung haben, bitte kurz charakterisieren.