Genus

Wie handhaben wir die Genusangaben im XSD?

Laut TEI wärs so:

```
<trans>
  <tr>rabotin</tr>
  <gen>m</gen>
</trans>
```

Das reicht uns nicht, weil der Genus sich auf ein einzelnes Wort in einem ganzen TR beziehen kann.
Also sowas (fiktives) wie: ein Buch {n} mit Silberrand.

XML-Alternative

Das könnte man (angelehnt an die bisherige Syntax) wie folgt modellieren:

```
<trans>
  <tr>ein Buch {n} mit Silberrand</tr>
</trans>
```

oder eher XML-konform:

```
<trans>
  <tr>
    ein <n>Buch</n> mit Silberrand
  </tr>
</trans>
```

bzw. ausführlicher:

```
<trans>
  <tr>
    ein <main_noun genus="n">Buch</main_noun > mit Silberrand
  </tr>
</trans>
```

Für die Einarbeitung ist eine Version mit wenig Schreibarbeit praktischer. Das würde für die Version mit geschweifter Klammer sprechen. Echter XML-Code wäre wohl eleganter, wobei die letzte Version am ehsten auch von Menschen zu verstehen ist. "Nomen" als Name der Markierung ist wohl nicht optimal. Es geht nicht darum, dass es ein Nomen ist - das ist "Silberrand" bzw. "Silberschnitt", was wohl gemeint ist, auch. Es geht darum, dass es das wichtigste Nomen ist; also besser z.B. "main_noun"

> *Dan:*
> *Es ist eigentlich egal, was für ein Tag benutzt wird, man kann das alles wieder problemlos umwandeln, wie man es braucht.*
> *Für Editoren sollte es aber kurz sein, um den Schreibaufwand gering zu halten. Deswegen wäre wie bisher {n} oder <n></> ganz ok.*

Generische-Alternative

Eine andere Möglichkeit wäre noch, dass man den Genus ganz rausrechnet, also in einer seperaten Tabelle hält (Wort, Genus), und den Genus im XML nur dort pflegt, wo es halt vom Standard abweicht.

Das hätte folgende Vorteile:

- Vermeidung von unnötigen Redundanzen
- Erleichterung bei Neueinträgen
- Man kann sich ein allgemeines Hervorhebungstag überlegen, dass nicht nur für Nomen greift, z.B.

```
<trans>
  <tr>
    ein <main>Buch</main> mit Silberrand
  </tr>
</trans>
```

> *Ulrich:*
> *Markierung des wesentlichen Nomens eines Eintrages macht praktisch dieselbe Arbeit, wie gleich das Genus anzugeben, bzw. es ist umgekehrt, das wichtige Nomen wird bislang durch Genusangabe als solches gekennzeichnet. Wenn man sich klar macht, was das wesentlich Nomen ist, weiß man auch das Genus. Man muss sich auch nichts überlegen, um Spezialfälle wie der, die oder das Korpus abzufangen. Eine zweite Datei zur Überprüfung der Genera ist jeoch bestimmt sinnvoll.*

> *Tom:*
> *Eine manuelle Markierung des Genus macht in jedem Fall mehr Arbeit, aber vorallem auch Arbeit, die falsch sein kann, und dann an mehreren Stellen geändert werden müsste. Mit dem Reinrechnen des Genus hätte man diese Probleme nicht, sie könnte zentral gepflegt werden.*

> *Dan:*
> *Ich halte die direkte Angabe des Genus für am besten. Natürlich ist so eine automatische Zuweisung sehr praktisch. Doch man wird da immer auf Grenzen stoßen. Und die Arbeit für die Genusangabe sollte man nicht scheuen.*

Anmerkungen

Die generische Alternative ist bezüglich Redundanz und Fehlerminimierung sicherlich vorzuziehen.

Bezüglich des Terminus '*wesentliches Nomen*'

Dieser impliziert, dass nur ein wesentliches Nomen vorkommen kann, aber es Fälle wie z.B. '*Trauer {f} und Freude {f}*'.

XML-Darstellung der Genera

Vorschlag:

```
ein Buch <genus type="n" /> mit Silberrand
```

Dies hätte den Vorteil, dass man eine XLST-Transformation sehr einfach schreiben kann.

Natürlich stellt sich hier die Frage, ob es einen Mehrwert darstellen würde, wenn man z.B.

```
ein <nomen genus="n">Buch</nomen> mit Silberrand
```

schreibt? Ein Vorteil dieser Notation würde sein, wenn man mal sehr weit in die Zukunft schaut, dass man die deutsche Übersetzung "*Bäume mit eingeschlossenen Fossilien*" derart taggt:

fiktives Beispiel:

```
<token type="nomen" genus="mpf" citeform="Baum">Bäume</token><token type="prep">mit</token>

<token type="verb part" citeform="eingeschlossen" citeform="einschließen">eingeschlossenen</token><token type="
nomen" genus="npf" citeform="Fossil">Fossilien</token>
```

Daraus ließen sich z.B. Hyperlinks generieren, die nicht nur nach **Bäume** suchen, sondern nach allen Vorkommen von **citeform="Baum"**.

Erstellung einer Generatabelle

Die Erstellung gestaltet sich als recht '*einfach*', wenn man in der WaDoku-Präambel schreibt, dass sich die Genusangabe auf das 'Standard Deutsch' bezieht, ungeachtet standardisierter Varietäten wie z.B. Österreichisch.

Eine solche Tabelle könnte folgendes Aussehen haben:

...	
Mitteilen	n
Mitteilung	f
...	
See	m
See	f
...	

Eine derartige Tabelle "nomen" existiert bereits, und wird bei dem Tag2Xml-Konverter schon automatisch gefüllt. Derzeit hat diese über 90.000 Einträge.

> *Dan:*
> *Woran soll man erkennen können, welches Genus gerade passt? Den japanischen Deutschlernern nützt das nicht viel. Obwohl das zweifellos hilft, wenn man mit der Maus über ein Nomen fährt, welches kein "wesentliches Nomen" ist, und dessen Genus erfährt, was wiederum ein Punkt ist, der es den Deutschlernern einfacher machen könnte.*

Wie soll man dieses Tag beim Bearbeiten oder Neuerstellen eines Eintrages angeben?

Vorschlag:

```
ein Buch {} mit Silberrand
```

Hier ist es möglich, einen Parser zu schreiben, der das Wort, vorausgesetzt, man kann **Wort** eineindeutig definieren, vor {} in einer Tabelle nachschlägt, um den Genus zu erhalten.
Daraus könnte ein XML-Code derart errechnet werden:

```
ein Buch <genus type="n" /> mit Silberrand
```

Diese Schreibweise würde mit der derzeitigen Schreibweise konform laufen, man muss lediglich {n|m|f|kA|[nmf]kA|[nmf]pl} mit {} ersetzen, wenn das Nomen nur einmal in der Generatabelle vorkommt. Bei z.B. '**See {m} eines Vulkans**' muss der Genus erhalten bleiben. Dies setzt bei der Bearbeitung oder Neuerstellung voraus, dass bevor der Eintrag gespeichert wird, eine Prüfung stattfinden muss, um z.B. '**See {} eines Vulkans**' abzufangen.

Ein Problem könnte auftauchen, wenn das mit einem Genus zu versehendes Nomen aus mehreren Wörtern besteht, wie z.B. 'Compact Disc'. Ein Regel der deutschen Sprache lautet, dass sich der Genus eines zusammengesetzten Substantivs nach der letzten Komponente richtet, also in diesem Falle nach 'Disc' - weiblich.

```
Compact Disc <genus type="f" />
```

Auch hier stellt sich die Frage, ob eine Alternativschreibweise

```
Compact <nomen genus="f">Disc</nomen>
```

oder besser

```
<nomen genus="n">Compact Disc</nomen>
```

vorzuziehen ist? Letztere bedingt außerdem eine völlig andere Art der Eingabe!

Was passiert bei Nomen, die in der Genustabelle noch nicht aufgeführt sind?

Beispiel eines neuen Eintrages:

```
Segge {}
```

Dies muss man abfangen und den Benutzer auffordern, den Genus in die Tabelle einzutragen. Die Frage hier nur ist, ob es technisch einfach realisierbar ist?

Ein anderer Ansatz wäre, dies einfach zuzulassen. Der XML-Parser würde in diesem Falle {} ignorieren und keinen Genus-Tag ausgeben. Erst nachdem für *'Segge'* der Genus in die Generatabelle eingetragen wurde, würde die XML-Ausgabe aktualisiert.

Andere Vorschläge

Gibts andere Vorschläge bzw. Präferenzen?