

Transkription

Das Deutsche Textarchiv (DTA) erfasst strukturelle und linguistische Merkmale deutschsprachiger Texte des 17. bis 19. Jahrhunderts, mit dem Ziel „ein ausgewogenes historisches Referenzkorpus in deutscher Sprache zu schaffen“ (Deutsches Textarchiv (B), 2017) ([Quellen.html#ZielDTA](#)) Unter anderem stellen sie eine frei verfügbare Transkription von Davidis' Kochbuch in (Deutsches Textarchiv (A), 2008) ([Quellen.html#DavidisKochbuchDTA](#)) zur Verfügung. Abb. 1a zeigt den Scan eines Rezeptes und Abb. 1b die in TEI (<http://www.tei-c.org/index.xml>) transkribierte Version des Rezeptes. Wir sind allerdings nicht an linguistischen oder strukturellen Merkmalen interessiert, sondern bereiten eine kulinarische Analyse vor. Dementsprechend haben wir die Kodierung des DTA für unsere Zwecke überarbeitet. Dies ist in Abb. 1c zu sehen.

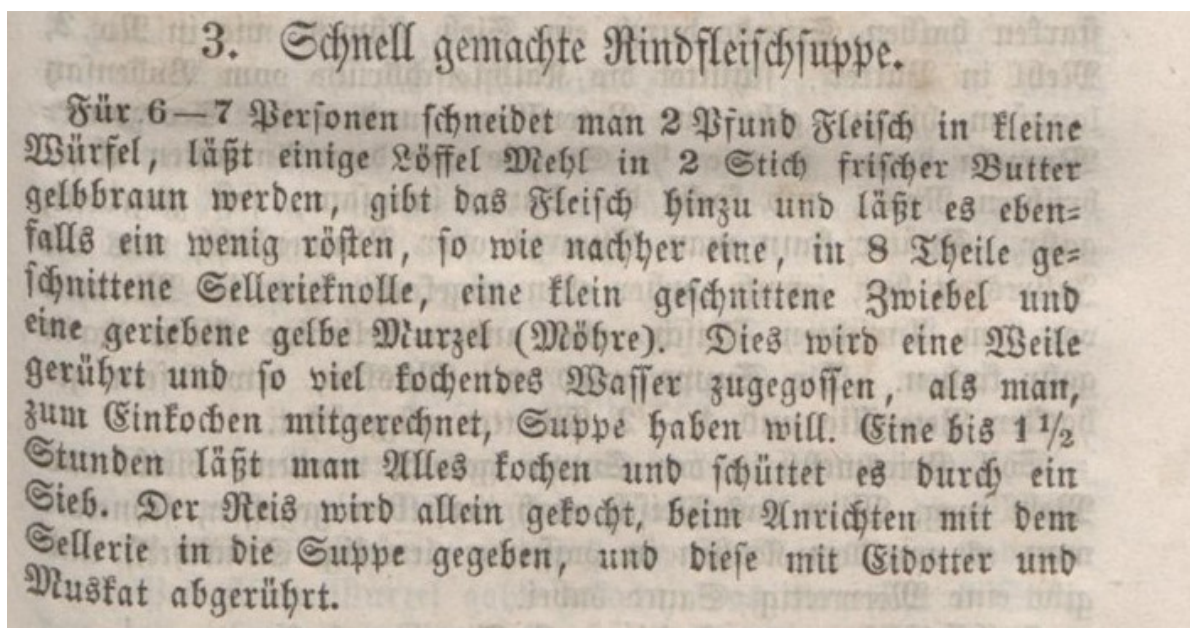


Abb. 1b (Deutsches Textarchiv (A), 2008) (Quellen.html#DavidisKochbuchDTA)

```

<div n="3">
  <head>3. Schnell gemachte
    Rindflei&#x017F;ch&#x017F;uppe.</head>
  <lb/>
  <p>Für 6&#x2014;7 Per&#x017F;onen &#x017F;chneidet
    man 2 Pfund Flei&#x017F;ch in kleine<lb/>
    Würfel, läßt einige Löffel Mehl in 2 Stich
    fri&#x017F;cher Butter<lb/> gelbbraun werden,
    gibt das Flei&#x017F;ch hinzu und läßt es
    eben-<lb/> falls ein wenig rö&#x017F;ten,
    &#x017F;o wie nachher eine, in 8 Theile
    ge-<lb/> &#x017F;chnittene Sellerieknolle, eine
    klein ge&#x017F;chnittene Zwiebel und<lb/> eine
    geriebene gelbe Murzel (Möhre). Dies wird eine
    Weile<lb/> gerührt und &#x017F;o viel kochendes
    Wa&#x017F;er zugego&#x017F;en,
    als man,<lb/> zum Einkochen mitgerechnet, Suppe
    haben will. Eine bis 1½<lb/> Stunden läßt man
    Alles kochen und &#x017F;chüttet es durch
    ein<lb/> Sieb. Der Reis wird allein gekocht,
    beim Anrichten mit dem<lb/> Sellerie in die
    Suppe gegeben, und die&#x017F;e mit Eidotter
    und<lb/> Muskat abgerührt.</p>
</div>

```

Abb. 1c

```
<cue:recipe type="Suppen." rcp-id="B-3">
  <head>Schnell gemachte Rindfleischsuppe.</head>

  <p>Für 6–7 Personen schneidet man 2 Pfund
    Fleisch in kleine Würfel, läßt einige Löffel
    Mehl in 2 Stich frischer Butter gelbbraun
    werden, gibt das Fleisch hinzu und läßt es
    ebenfalls ein wenig rösten, so wie nachher
    eine, in 8 Theile geschnittene
    Sellerieknolle, eine klein geschnittene
    Zwiebel und eine geriebene gelbe Murzel
    (Möhre). Dies wird eine Weile gerührt und so
    viel kochendes Wasser zugegossen, als man,
    zum Einkochen mitgerechnet, Suppe haben will.
    Eine bis 1½ Stunden läßt man Alles kochen und
    schüttet es durch ein Sieb. Der Reis wird
    allein gekocht, beim Anrichten mit dem
    Sellerie in die Suppe gegeben, und diese mit
    Eidotter und Muskat abgerührt. </p>
</cue:recipe>
```

Zum einen haben wir kodierte Sonderzeichen zu Gunsten der Lesbarkeit durch ihre entsprechende Standard-Pendants ersetzt; z.B das „f“ (lange S) durch ein heute übliches „s“ (rundes S). Dazu haben wir strukturelle Merkmale wie die Kodierung von Zeilenumbrüchen (<lb/>) oder die Verschachtelungstiefe innerhalb des Buches (<div n="3"/>) entfernt. Stattdessen haben wir jedes Rezept, welches zu unseren Zielobjekten gehört, in einem einheitlichen Tag (<cue:recipe>) gekapselt. Zusätzlich haben wir das Tag für Rezepte mit der Kapitelüberschrift als Typ angereichert (type="Suppen"), sowie eine eindeutige ID vergeben (rcp-id="B-3"). Des Weiteren haben wir die interne Kapitelnummerierung (3.) aus der Rezept-Überschrift entfernt, da diese nichts zur kulinarischen Analyse beiträgt, jedoch bei separater Betrachtung irreführend ist.

Der Vollständigkeit halber sei erwähnt, dass das DTA die Transkription von Nicht-Muttersprachlern im Double-Keying-Verfahren (<http://www.deutschestextarchiv.de/doku/basisformat/transkription.html>) hat durchführen lassen. Ihre Intention dabei ist, dass so unbewusste Modernisierungen, Korrekturen und Wertungen beim Transkribieren vermieden werden. Allerdings führt das zu Fehlern wie *Eßlöffel* statt *Eßlöffel*. Dem aufmerksamen Leser wird aufgefallen sein, dass Ersteres (*Eßlössel*) kein deutsches Wort ist. In diesem Beispiel wurde fälschlicherweise das „ff“ für zwei lange „s“ gehalten. Sofern uns solche Fehler aufgefallen sind, haben wir sie in unserer Transkription korrigiert.

Abschließend sei darauf hingewiesen, das sämtliche Anerkennung der Arbeit, die Kodierung des DTA zu unserer Kodierung zu transformieren, meinem betreuenden Prof. Herrn Luttenberger gebührt, wie in diesem Blog-Abschnitt (Blog.html#NL2) zu lesen ist.