

Kieler Arbeiten zur skandinavistischen Linguistik 5

Iben Nyholm Debess

HOSTCol 1.0

Guide to the
Hamburg Older Scandinavian Text Collection
(Version 1.0)

Kiel 2019

Christian-Albrechts-Universität zu Kiel
Institut für Skandinavistik, Frisistik und Allgemeine Sprachwissenschaft

© Iben Nyholm Debess, Kiel 2019

Iben Nyholm Debess
Kiel University
Institute of Scandinavian Studies, Frisian Studies and General Linguistics

Published 18 March 2019

DOI: 10.21941/publ/kas15

Die *Kieler Arbeiten zur skandinavistischen Linguistik* publizieren Arbeitspapiere und Forschungsergebnisse aus der skandinavistischen Linguistik am Institut für Skandinavistik, Frisistik und Allgemeine Sprachwissenschaft der Christian-Albrechts-Universität zu Kiel.

Contents

1	Introduction	5
2	Text collection overview	7
3	Document structure	9
3.1	File description	9
3.2	Encoding description	10
3.3	Profile description	11
3.4	Revision description	11
4	Text structure and annotation	13
4.1	Division of the text	13
4.2	Other annotation	14
	References	15

1 Introduction

This paper provides a short description of the *Hamburg Old Scandinavian Text Collection* (HOSTCol), Version 1.0, which can be accessed online via the permanent identifier (PID) <http://hdl.handle.net/11022/0000-0007-DoF1-D>.

The texts in the collection originate from the project *Niederdeutsch und Skandinavien* at the University of Hamburg.¹ The project was investigated how Low German affected the Scandinavian languages in the Middle Ages. More information about the project and its results can be found in Braunmüller & Diercks (1993), Braunmüller (1995) and Jahr (1998).

Before the editing of the texts in 2018, all the original texts were in an old Word-document-format with limited information about the texts. Most of the texts have been transcribed to a digital format manually, but some seem to have been automatically scanned. The text were originally transcribed with all diacritics and other information, but only some of this information remains in the texts available today. The current XML text collection is the result of reorganizing and editing of the original texts, as well as a conversion to the XML format. All the XML documents are compliant with TEI and Menota guidelines. The current version was prepared as part of a project at Kiel University, which aimed at converting legacy corpus data into sustainable formats and ensuring internal consistency.²

¹ The project *Niederdeutsch und Skandinavien* (1990–1995) was led by Kurt Braunmüller and funded by the German Research Foundation.

² The project (2017–2018) was founded by the German Research Foundation (DFG), hosted by the Specialist Information Service Northern Europe at Kiel University Library, led by Ruth Sindt and supervised by Steffen Höder, in cooperation with the Hamburg Center for Language Corpora. Iben Nyholm Debess was responsible for carrying out the technical work.

2 Text collection overview

The texts in HOSTCol are in Old Swedish or in Old Danish. The language of each text is indicated by the last letter in the text code, example: FLEND (the last letter *D* indicates Danish), BONDS (the last letter *s* indicates Swedish). The first four letters of the text code is inspired by the title of the text.

The texts are grouped in three groups: RECHT (law texts), VOLKSB (short for German *Volksbuch*, includes texts from chapbooks) and SONSTIGE (miscellaneous).

List of texts in the collection:

Text code	Title
AVGLS	Äldre Västgöotalagen
CHRID	Kong Christian den Fjerdes egenhændige breve, 1589–1625; et udvalg
FLEND	Flensborg stadsret. Text II; Prolog & den oprindelige lov 1–133
JYSLD	Den Jyske Lov – Chapter 1
BONDS	Peder Månssons Bondakonst
ERIKS	Erikskrönikan
FLOSS	Flores och Blanzeflor – Her sigx aff Flores oc Blanzaflor
LAEKS	Läke- och Örtre-Böcker
COBUS	Codex Bureanus
SPATS	S. Patriks-Sagan: S. Patrik och hans Järtecken, Legend 1-5
NAVAS	Namnlös och Valentin.
REYND	En Ræffue Bog: Herman Weigeres Oversættelse af Reinke de Vos
BAESS	Bältare-svennernes skrå
SJAES	Sjællinna thrøst / Første delin
GRISD	Griseldis. Tuende deylige oc Nyttelige Historier at laese. Den förste om Griseldis.
GRISS	En skön och märkelig historia om Grisilla
KOPLD	En skön Historie Om tuende Kiöbmend / oc om en ærlig oc dydelig Quinde. Oc huorledis det gick den ene Kiöbmands Hustru vdi hans frauærelse / Oc huad falsk oc suig dem baade vederfarit er.
KOPLS	Äktenskapskärleks Ärekrona
RAUSD	Broder Russes Historie
RAUSS	Broder Rus, det är Broder Ruses historia eller chronica
REYNS	Reyncke Foß
DODED	Den gamle danske dödedans

3 Document structure

All the documents in the text collection contain an XML declaration (`<?xml>`), a document type declaration (`<!DOCTYPE>`) and a root element (`<TEI>`), which in turn consists of the header (`<teiHeader>`) and the actual text (`<text>`).

The header (`<teiHeader>`) contains the relevant metadata. It consists of the file description (`<fileDesc>`), the encoding description (`<encodingDesc>`), the profile description (`<profileDesc>`), and the revision description (`<revisionDesc>`).

3.1 File description

The first element of the header, the file description, contains information about the document, which is presented as a digital edition of an original text. The title statement (`<titleStmt>`) includes bibliographic information about the digital text (title, editor, specific responsibilities of each editor). The edition statement (`<editionStmt>`) specifies the version of the digital edition. The `<extent>` element gives the number of words in the text. The publication statement (`<publicationStmt>`) names the institution that has made the digital text available, specifies the accessibility of the text, defines a short (unambiguous) identifier of the text and dates the publication. The source description (`<sourceDesc>`) contains a bibliographic description of the printed edition from which the digital text was derived.

```

<fileDesc>
  <titleStmt>
    <title>Äldre Västgötalagen (Cod. Holm. B 59): digital version</title>
    <editor>
      <name xml:lang="ger">
        <persName>Iben Nyholm Debess</persName>
        <orgName type="affiliation">Christian-Albrechts-Universität zu Kiel</orgName>
      </name>
    </editor>
    <respStmt>
      <resp>Transcription of the printed edition. Transcription does not include comments,
        footnotes and information about expansions and abbreviations. </resp>
      <name>
        <persName>Unknown</persName>
      </name>
    </respStmt>
    <resp>Converting the document to xml-version compliant with TEI and Menota
      guidelines</resp>
      <name xml:lang="ger">
        <persName>Iben Nyholm Debess</persName>
        <orgName type="affiliation">Christan-Albrechts-Universität zu Kiel</orgName>
      </name>
    </resp>
  </titleStmt>
  <editionStmt>
    <edition n="1.0">First published version <date>2018-06-30</date>
  </edition>
</editionStmt>
<extent n="15008">15.008 words</extent>
<publicationStmt>
  <istributor/>
  <idno>AVGLS</idno>
  <date when="2018-06-30">30 June 2018</date>
  <availability status="restricted">
    <p>This text is available for purposes of academic research and teaching only.
      Re-distribution in any form without prior permission is prohibited. Short extracts may
      be cited with full acknowledgement of the source.</p>
  </availability>
</publicationStmt>
<sourceDesc>
  <bibl>Samlingar av Sveriges gamla lagar (SSGL 1), utg. H.C. Collin/ C.J. Schlyter, I-XIII,
    Lund 1827-77, 1-48.</bibl>
</sourceDesc>
</fileDesc>

```

3.2 Encoding description

The encoding description documents the relationship between the digital text and its printed source. The project description (`<projectDesc>`) contains information about the text encoding standard that has been applied. The editorial practices declaration (`<editorialDecl>`) specifies how and to what degree the text has been proofread, normalised and annotated.

```

<encodingDesc>
<projectDesc>
<p>This text has been transcribed into a digital version in a project at Hamburg University
1990-1995, investigating how low German affected the Nordic languages ("Lågtyska och
Skandinavien" / "Niederdeutsch und Skandinavien", Kurt Braummüller. See - Jahr, Ernst
Håkon, 1998: Språkkontakt i Norden i middelalderen, særlig i hansatiden :
forskningsprogrammet Norden och Europa, - Braummüller, Kurt and Diercks, Willy, 1993-1995,
Niederdeutsch und die skandinavischen Sprachen I and II).
The text has in 2018 been converted and edited into an xml-format compliant with TEI (P5) and the Menota handbook (2.0).</p>
</projectDesc>
<editorialDecl>
<correction status="unknown">
<p>The text has not been fully proof-read. </p>
</correction>
<normalization level="dipl">
<p>This text follows the printed edition. </p>
</normalization>
<interpretation lemmatized="none" morphAnalyzed="none">
<p>This text has been divided into paragraphs, chapters, sections and headings. It
indicates words and punctuation marks, and, if present, corrected and supplied elements.
It also indicates page breaks and line breaks, if the information has been available.
</p>
</interpretation>
</editorialDecl>
</encodingDesc>

```

3.3 Profile description

The profile description is used to define the three-letter language tags used to identify different languages both in the text and in the header. The Menota recommendation is that the Old Nordic languages should be referred to by the abbreviations **oda** (Old Danish), **oic** (Old Icelandic), **onw** (Old Norwegian), and **osw** (Old Swedish).

```

<profileDesc>
<langUsage>
<language ident="eng">English</language>
<language ident="ger">German</language>
<language ident="lat">Latin</language>
<language ident="osw">Old Swedish</language>
<language ident="swe">Swedish</language>
<language ident="oda">Old Danish</language>
<language ident="dan">Danish</language>
</langUsage>
</profileDesc>

```

3.4 Revision description

The final element of the header, the revision description, contains the revision history of the document in reverse chronological order, beginning with the initial release of the published edition. Each change (<change>) is dated, attributed to an individual co-worker and described in detail. The numbers (@n) correspond to the version of the document.

Iben Nyholm Debess

```
<revisionDesc>  
  <change n="1.0">  
    <date>2018-06-30</date>  
    <name>Iben Nyholm Debess</name> First published version. </change>  
</revisionDesc>
```

4 Text structure and annotation

4.1 Division of the text

The actual text is represented by the `<text>` element, which also specifies the main language of the source (**osw** or **oda**).

The default element for text in Menota documents is the paragraph (`<p>`). All higher-level elements (such as `<body>`) must include one or more paragraphs.

After the `<body>`-element, all texts start with a `<head>`-element containing the original head or title of the text.

The text consists primarily of words (`<w>`) and punctuation marks (`<punct>`).

All texts are divided into different parts with the element `<div>`. The division elements used are (in hierarchical order with the **@type**): **chapter**, **paragraph** and **section**. The chapter, paragraph and section elements are numbered (**@n**) when appropriate according to the original text. The individual sections may contain headings as well (`<head>`), in accordance with the original text.

The level of division is different in the various texts and depends on the structure of the original text. The minimum level of division is into chapters.

```
<text xml:lang="osw">
  <body>
    <head>Äldre Västgötalagen</head>
    <div n="1" type="chapter">
      <head><w>1</w>
        <w>Kirkiu</w>
        <w>Bolkær</w>
      </head>
      <div n="1" type="paragraph">
        <p>
          <w>1</w>
          <punct>.</punct>
          <w>Her</w>
          <w>byriar</w>
          <w>laghbok</w>
          <w>væsgöta</w>
          <w>Krister</w>
          <w>ær</w>
          <w>fyrst</w>
          <w>i</w>
          <w>laghum</w>
          <w>warum</w>
          <w>tha</w>
          <w>ær</w>
          <w>cristna</w>
          <w>var</w>
          <w>oc</w>
          <w>allir</w>
          <w>cristnir</w>
        </p>
      </div>
    </div>
  </body>
</text>
```

4.2 Other annotation

Page breaks and line breaks are indicated by an empty `<pb/>`-element or `<lb/>`-element, respectively, possibly with an attribute `@n` specifying the page or line number. Page and line breaks and numbers are indicated in those texts, where the information was available from the original document.

Two philological tags are included to indicate text corrected (`<corr>`) or supplied by the editor (`<supplied>`).

References

- Braunmüller, Kurt (ed.). 1995. *Niederdeutsch und die skandinavischen Sprachen*. Vol. 2 (Sprachgeschichte 4). Heidelberg: Winter.
- Braunmüller, Kurt & Willy Diercks (eds.). 1993. *Niederdeutsch und die skandinavischen Sprachen*. Vol. 1 (Sprachgeschichte 3). Heidelberg: Winter.
- Jahr, Ernst Håkon (ed.). 1998. *Språkkontakt i Norden i middelalderen, særlig i Hansatiden. Forskningsprogrammet Norden og Europa* (Nord 1998:4). København: Nordisk Ministerråd.