

HiNTS: A Tagset for Middle Low German

Fabian Barteld^{*,†,‡}, Sarah Ihden^{*,†}, Katharina Dreessen[†], Ingrid Schröder[†]

* these authors contributed equally

[†]Department of Language, Literature and Media, Institute for German Studies, Universität Hamburg, Germany

[‡]Language Technology Group, Department of Informatics, Universität Hamburg, Germany
firstname.lastname@uni-hamburg.de

Abstract

In this paper, we describe the “Historisches Niederdeutsch Tagset” (HiNTS). This tagset has been developed for annotating parts-of-speech and morphology in Middle Low German texts, a group of historical (1200–1650) dialects of German. A non-standardized language such as Middle Low German has special conditions and requirements which have to be considered when designing a tagset for part of speech and morphology. We explain these requirements, i.e. the need to encode ambiguities while allowing the annotator to be as specific as possible, and our approach for dealing with them in the tagset. We then describe two special features of the tagset. In order to prove the benefit of these tags and corresponding annotation rules, we present example searches and the possible analyses arising from the results of such searches. Besides the usefulness of our tagset, we also considered its reliability in annotation using inter-annotator agreement experiments. The results of these experiments are presented and explained.

Keywords: non-standard language, part-of-speech tagset, morphological tagset

1. Introduction

In this paper, we describe a tagset called “Historisches Niederdeutsch Tagset” (HiNTS). The tagset was developed at the University of Hamburg in the context of the project ‘Reference Corpus Middle Low German / Low Rhenish (1200–1650)’ (ReN). The aim of the project is the creation of an annotated corpus of Middle Low German (GML) texts, a group of historical dialects of German that were used between 1200 and 1650 in the northern part of the German language area and in the Baltic regions. The corpus will be part of the ‘Corpus of Historical German Texts’, together with the corpora ‘Altdeutsch’ (Old German), ‘Mittelhochdeutsch’ (Middle High German), and ‘Frühneuhochdeutsch’ (Early New High German).¹ ReN provides diplomatically transcribed Middle Low German and Low Rhenish texts from 1200 to 1650 that are lemmatised and annotated with part of speech (POS) and morphology using HiNTS, the tagset described in this paper. Since GML is a historical language, which is not well described and hardly any resources for annotating GML texts exist, there are special requirements for the tagset which we will focus on in this paper.

When annotating a historical language, the annotators lack the intuition of a native speaker. Therefore, decisions about a category can only be based on the comparison and statistical analysis of texts (Dipper et al., 2013, 2). This however, should be avoided while annotating a text and left for later analysis using the annotated data. We present design decisions for our POS tagset that result from this requirement. HiNTS is also used to annotate morphology. Here, many tokens are ambiguous, for example, as regards their inflectional ending. To allow the annotators to encode such an ambiguity, HiNTS allows the combination of multiple features of one inflectional category.

¹The project is described in Peters and Nagel (2014) and on the website www.referenzkorpus-mnd-nrh.de. For information on the annotation used in ReN and possible grammatical analyses, see Schröder (2014).

In the following section, we present related work and detail the special needs for a tagset designed to annotate a historical language. We subsequently describe special features concerning the POS (Section 3) and the morphological tags (Section 4) of HiNTS dealing with the relevant requirements. To evaluate our design choices, we show example queries that illustrate the usefulness of the tagset. Furthermore, we present inter-annotator agreement experiments which show that GML can be annotated reliably using HiNTS.

2. Related Work and Requirements for a Tagset

The Stuttgart-Tübingen Tagset (STTS, Schiller et al. (1999)) is the de-facto standard for the annotation of German texts with parts of speech and morphology. In recent years, it has been adapted to non-standard, for example historical, texts (Zinsmeister et al., 2014). One of these adaptations is the Historical Tagset (HiTS, (Dipper et al., 2013)) that has been developed for the annotation of Old German (750–1050), Middle High German (1050–1350) and Early New High German (1350–1650) texts in the context of a group of projects that aim to create a ‘Corpus of Historical German Texts’.

One of the additions that HiTS introduced is the distinction between a type-based and a token-based tag. This allows the tagset to encode differences in the function of a word as a result of a grammatical change (Dipper et al., 2013, 15). However, during the annotation process other ambiguities may appear (Barteld et al., 2014; Seemann et al., 2017). These might have different reasons and can only be resolved using an aforementioned statistical analysis in certain cases. However, even in these cases, such an analysis might best be left till after the corpus is annotated. Furthermore, the annotators might not be able to distinguish these cases from other cases, where the ambiguity cannot be resolved.

Consequently, HiNTS – which is derived from HiTS for the annotation of GML texts – is designed following two

principles: on the one hand, it aims to allow the annotator to assign tags that are not fully specified while on the other hand it allows the annotator to be as specific as possible. We will illustrate this on the level of POS and of morphology in this paper.

3. Special POS Tags in HiNTS

According to our requirements for a tagset, we developed some specific tags and annotation rules for HiNTS. For POS, one of the main differences between HiNTS – our tagset for GML – and HiTS – the tagset used for Old German, Middle High German and Early New High German – lies in the category of determiners and pronouns. According to HiTS, determiners are lexemes that can appear either determining a noun – such as *this* in *this speech* – or substituting for a noun and thus, constituting a noun phrase on its own such as *this* in *this is the beginning*. In contrast, pronouns are lexemes that never determine but always substitute for a noun – hence, a pronoun is always the head of a noun phrase, e.g. the indefinite pronoun *somebody*. In such a lexeme-based approach, it is not possible to decide between pronoun and determiner based on a specific instance where the word appears as the head of a noun phrase. Only if another instance of the same word exists where it is used in the determiner slot, can it be said that the word in question is a determiner. In the case of pronouns on the other hand, it cannot be decided definitely that a word is a pronoun based on corpus data. Still, lexemes can be distinguished that are pronouns with a high probability, if they appear frequently in a corpus and only as heads of noun phrases. The problem of a lexeme-based approach is that you have to know while annotating which words can only appear as pronouns. Especially for less widely researched historical languages – where no native speaker intuition is available –, this is not possible without any doubt or any exception. Moreover, given the current state of knowledge about GML, even scholars of this language will not be able to resolve the ambiguity between pronoun and determiner. This holds for many other aspects of GML as well.

Therefore, our tagset HiNTS avoids lexeme-based assumptions: the categories should be assignable given a single instance. In the case of determiners and pronouns, HiNTS simply distinguishes between annotation units determining a noun, such as *dyt* ‘this’ in *dyt ghut kanstu allene nyth ghewynnen* ‘you cannot get this property by your own’ and tokens providing the head of a noun phrase such as *nemant* ‘nobody’ in *dat my nemant kunne lyken* ‘that nobody could resemble me’.

After the annotation of a corpus using HiNTS, this allows one to identify all lexemes which only appear as heads of noun phrases – substituting for a noun – and thus are likely to be pronouns in the basic word meaning.

A further difference between HiNTS and HiTS concerns the distinction between coordinating and subordinating conjunctions. In Modern German main clauses and subordinate clauses can be easily distinguished: in main clauses the finite verb usually stands in the second position whereas in subordinate clauses the verb stands in the final position. In GML word order is more open to variation. Due to this, it is not always possible to distinguish between

main and subordinate clauses and consequently between subordinating and coordinating conjunctions. This is especially true if the conjunction is ambiguous. According to Schiller and Lübben (1875–1880), the word *wente* for example can mean ‘denn’ (‘because’; coordinating) or ‘weil’ (‘because’; subordinating), ‘dass’ (‘that’; subordinating) or ‘aber’ (‘but’; coordinating). This ambiguity is exemplified in (1). Härd (2008, 1461) states that *wente* can be a coordinating as well as a subordinating conjunction.

- (1) vnde ik sach et · vnde betugede et · **wente** dit is godes sone
 ‘and I saw it and attested it **because/that/but** this is god’s son’
 (Buxteh. Ev.)²

Often, in sentences with ambiguous conjunctions such as *wente*, it is very difficult and sometimes even impossible to decide whether it is a coordinating or a subordinating conjunction. In order to avoid a potentially wrong interpretation by the annotator and to provide a consistent annotation, structure-based rules are necessary. In HiNTS the tag ‘KON’ is used for conjunctions in verb-second sentences, e.g. *wente dit is godes sone*, while ‘KOUS’ is the tag for the conjunction in a sentence with the finite verb in a later position than verb second (including verb final, the position where verbs with subordinating conjunctions appear in Modern German), e.g. *wente dit godes sone is*. In ambiguous sentences such as *wente he kam* (‘because/but/that he came’) where the finite verb could stand in the second as well as in the final position the tag ‘KO*’³ is used. The advantage of such an annotation principle is that the user of the corpus can search for specific structural contexts of a conjunction. In that way, one can investigate which conjunction is used how often in combination with the respective position of the finite verb. This helps historical linguists to research change processes concerning the position of the finite verb in coordinated and subordinate clauses in historical German.

In the following, we will illustrate this by the results of a search for the prototypical subordinating conjunction *dat* in the third pre-release of the “Reference Corpus Middle Low German / Low Rhenish (1200–1650)” (ReN) (ReN-Team, 2017) containing 32 texts. In Modern German, a subordinator usually appears only in sentences with verb final position. In GML there is much more variation. As the results in Table 1 show, in 1,775 sentences with the subordinator *dat* the finite verb appears in a later position than verb second. This includes not only sentences like (2) with a verb final position like in Modern German but also sentences like (3) where the finite verb stands in a later position than verb second but not in the last position. As Dreessen and Ihden (2015) have shown the reason for this divergent word order lies in the structure ‘finite verb before infinite verb’ within the verbal complex as in (3) as well as in the

²The examples are taken from the Reference Corpus Middle Low German / Low Rhenish (1200–1650). More information on the particular texts can be found in the metadata of the corpus.

³The asterisk is not part of a regular expression here but simply marks the ambiguity.

POS tag	Frequency
KOUS	1,775
KON	260
KO*	68
total	2,103

Table 1: POS tags of the conjunction ‘dat’ in ReN (ReN-Team, 2017)

post-field. Moreover, Table 1 shows that there are even 260 *dat*-sentences with a verb-second structure (the word order of coordinating sentences in Modern German) such as in (4) (annotated as ‘KON’). Here, the reason for the divergent word order again is the post-field (*eyn droghenere*). Both the structure ‘finite before infinite verb’ in the verbal complex and the post-field are frequently used in GML texts. In a lexeme-based approach one would have annotated the same tag for the conjunction *dat* in all of the sentences. Thus, no information on divergent word order in the sentences beginning with a conjunction – a special feature of historical German – would become visible through the tags.

- (2) weret szake · **dat** hgy dar eynd jar effte sossze blyuen *worden*
‘[if] it was the case that you would stay there one year or six’
(Agneta Willeken)
- (3) O wy vnde we · **dat** ik gy *wart* gheboren
‘oh alack and alack that I ever was born’
(Bord. Marien-Kl.)
- (4) Se seden · **dat** vnse here *were* eyn droghenere
‘they said that our lord was a cheat’
(Buxteh. Ev.)

As we pointed out, the structure-based annotation rule in HiNTS concerning conjunctions provides a consistent annotation without a potentially wrong interpretation by the annotator. Moreover, the annotation is absolutely transparent to users of ReN which is a crucial advantage for search queries and the further work with the results.

4. Special Features of the Morphological Tagset: Annotating Ambiguity

A great challenge in annotating non-standard languages such as historical languages is the annotation of ambiguity (Barteld et al., 2014). In GML, words that are morphologically ambiguous in the view of a researcher are rather frequent, for example concerning grammatical gender. Some of the nouns having one specific grammatical gender in Modern German could occur with different genders in historical German. In (5), according to the GML dictionary by Lasch et al. (1956ff), the noun *spegel* ‘mirror’ can be either a masculine or a neuter noun. The annotators should try to use contextual clues, e.g. the form of a determiner, to decide between possible genders for a specific instance. However, this is often not possible. In (5), for instance, the linguistic context does not allow a definite disambiguation for the

	Frequency (token)	Frequency (lemma)
Masc-Neut	1219	178
Masc-Fem	765	142
Neut-Fem	326	70
total	2310	390

Table 2: Ambiguity between two genders in ReN

gender of *speyghel*⁴, since there is no determiner of this noun giving a clue concerning its gender: The word *der* is part of a genitive construction determining the noun *Sassen*. Disambiguation could for example be enabled by the definite article for a masculine noun *de* (*dit is de spegel*) or the one for a neuter noun *dat* (*dit is dat spegel*). Hence, it is not possible to assign an unambiguous gender to *speyghel*. In such cases, annotators should have the option of encoding this ambiguity. STTS and HiTS use the asterisk (*) for this. According to this, the token *speyghel* in (5) would be annotated with ‘NA.*.Nom.Sg’. The problem with this mark-up is that it only encodes that there is a gender ambiguity but not that the female gender is not possible. According to our requirements for a tagset explained above, in HiNTS as much information as possible must be provided. Therefore, we use portmanteau tags (Leech et al., 1994), which allow the tagset to exclude the female gender in (5) and to represent the ambiguity between the masculine and neuter noun. Hence, in (5) the complete tag for the token *speyghel* in HiNTS is ‘NA.Masc-Neut.Nom.Sg’.

- (5) Dit is der sassen **speyghel**
‘this is the Saxon’s mirror’
(Oldb. Ssp., headline)

In search queries one could use the gender portmanteau tags to find out which ambiguity between two genders occurs most frequently. Table 2 shows that in ReN (ReN-Team, 2017) the most frequent ambiguity is the one between the masculine and the neuter gender; the ambiguity between the neuter and the feminine gender occurs least frequently. By having a closer look at one of the lemmas with gender ambiguity then, one could also examine which gender is used more often in the unambiguous cases. Table 3 shows the results for the two examples *lîf*¹ ‘life’ and *strî*¹ ‘fight’.⁵ Whereas concerning the lemma *lîf*¹, there are more unambiguous proofs for the neuter (34) than for the masculine gender (11), the lemma *strî*¹ provides no unambiguous proofs for the neuter but 24 for the masculine gender. This could lead to the assumption that in contrast to the lemma *lîf*¹, the lemma *strî*¹ is generally used as a masculine noun and that the neuter noun is on the fringes. When the corpus is completed, this hypothesis could be examined on a wider range of texts.

This shows how encoding all possibilities in the case of gender ambiguity as portmanteau tags can help us to understand the gender system of GML.

⁴This is a spelling variant of *spegel*.

⁵The numbers denote the different word senses according to the GML dictionary by Lasch et al. (1956ff).

	Frequency	
	lif ¹	strît ¹
Masc-Neut	85	60
Masc	11	24
Neut	34	0
total	130	84

Table 3: Gender of lif¹ and strît¹

However, one could suppose that such a detailed annotation as above could be more difficult for the annotator than only using the asterisk and thus could lead to more disagreement between annotators. We will address this issue – among others – in an inter-annotator agreement experiment described in the following section.

5. Inter-Annotator Agreement

In the previous sections, two principles guiding the creation of the tagset HINTS have been described: firstly, leaving decisions between categories open (as with pronoun and determiner) while annotating formal features that are correlated with these categories (concerning pronoun/determiner: determining or substituting a noun; concerning conjunctions: the position of the finite verb). Secondly, the tagset contains portmanteau tags for specifying morphological ambiguity.

For the assessment of inter-annotator agreement, a segment of a GML text has been annotated independently by two annotators. While annotating POS and morphology, the annotators are allowed to change the segmentation (and also to correct the transcription of the text). This has to be taken into account when calculating the inter-annotator agreement. The annotations are aligned using the method described in Barteld et al. (2016). For this study, only tokens that were aligned by the described method are considered. These are 988 tokens.

Looking only at the coarse-grained POS tags (i.e. without morphology), the overall percentage agreement is 94.33%. This is lower than what has been reported for Modern German with a comparable tagset, e.g. 98.57% for a newspaper corpus (Brants, 2000). But it is quite high, when taking the non-standard nature of the historical texts into account. Scheible et al. (2011) report 91.6% for Early Modern German (1650–1800).

Table 4 shows the agreement for different subsets of the part-of-speech tagset. The numbers for *conjunctions*, i.e. the tags *KOUS*, *KON* and *KO** is higher than the overall average and also higher than most of the other subsets of the non-inflected parts of speech. This is despite the fact that the distinction between subordinating and coordinating is ambiguous in GML and shows that the categorization scheme presented in Section 3 – along with its usefulness for users – can be applied with a rather high agreement. Table 5 shows the confusion matrix of the respective tags. This matrix shows that while the annotators mostly agree, there are deviations between the annotators in the seemingly simple case of distinguishing between subordinating and coordinating conjunction, despite the structural rule that should be applied by the annotators. This supports

Tags	Agreement (%)
Inflected	91.29
Non-inflected	96.06
Prepositions	100.00
Conjunctions	95.31
Particles	94.44
Adverbs	90.00
All	94.33

Table 4: Agreement

	KO*	KON	KOUS
KO*	0	0	0
KON	0	49	0
KOUS	1	2	12

Table 5: Confusion matrix for KON, KOUS and KO*

the assertion that the distinction between coordination and subordination is not straightforward for GML.

Allowing the annotators to include ambiguity in the annotation as described in Section 4 leads to a huge tagset with 42,752 individual tags. However, many of the possible ambiguous combinations in morphology will only be theoretically valid and never appear in the corpus. In the annotated segment, one annotator used 194 tags, the other 196. When simplifying the annotations based on the scheme used by STTS and HITS, i.e. encoding ambiguity with *, these numbers are reduced only marginally to 191 for both annotators. Using this tagset, an agreement (Cohen’s κ , Cohen (1960)) of 0.818 is reached. The agreement is similar (0.819) when only using “*” to encode ambiguity. This is partly due to the fact, that the annotators often disagreed about ambiguity itself: one annotator marked 54 tokens as ambiguous, the other annotator 95. This leads to another argument for the more specific annotation of ambiguity. Using only “*”, there is no information on the possible tags – when adding all possible genders, cases and so on, this information is kept.

6. Conclusion

We showed that a historical language such as GML brings special conditions and requirements having to be considered in a tagset for annotating POS and morphology of GML texts: GML is not well described and there are nearly no resources which could be used for the annotation. Moreover, in contrast to the current German language, GML cannot be annotated based on the intuition of a native speaker. Since the annotation of the texts should be used for a new grammar of GML, one of our most important requirements is that some decisions should not be made before annotating but be left for later studies in order not to anticipate the results coming from the analysis of the annotated texts. Therefore, in the POS part of the tagset, we avoid lexeme-based assumptions (as for the group of determiners and pronouns) and prefer structure-based rules. Concerning the distinction of coordinators and subordinators, such rules reduce the influence of potentially wrong interpretations of the annotators and thus ensure a consistent annotation. As we showed in an example search, these rules could

for example be used to analyse the position of the verb in sentences beginning with a conjunction.

The greatest challenge in annotating the morphology of GML are ambiguous forms, such as those concerning grammatical gender. In order to provide as much information as possible, we use portmanteau tags. An example query on nouns with gender ambiguity illustrated how our portmanteau-tag annotations could be used for studies on the frequency of specific gender ambiguities and the dominance of one gender in unambiguous forms.

According to the results of our inter-annotator agreement experiments concerning POS, the overall percentage agreement (94.33%) is quite high for non-standard texts. Moreover, within the group of conjunctions, agreement is even higher than the overall rate and higher than most of the other non-inflected parts of speech. This shows that our structure-based annotation rule concerning conjunctions is not only of benefit for the user but it can also be applied by the annotators reliably. Concerning morphology, the results of the inter-annotator agreement experiments have shown that the use of portmanteau tags in our corpus does not lead to a significantly higher rate of disagreement and therefore is not only useful but can also be annotated as reliably.

7. Acknowledgements

The authors are funded by the German Research Foundation (DFG) as part of the project “Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200–1650)”.

We would like to thank the anonymous reviewers for their helpful remarks and Piklu Gupta for improving our English.

8. Bibliographical References

- Barteld, F., Ihden, S., Schröder, I., and Zinsmeister, H. (2014). Annotating descriptively incomplete language phenomena. In *Proceedings of LAW VIII - The 8th Linguistic Annotation Workshop*, pages 99–104, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.
- Barteld, F., Schröder, I., and Zinsmeister, H. (2016). $t\gamma$ -Inter-annotator agreement for categorization with simultaneous segmentation and transcription-error correction. In Stefanie Dipper, et al., editors, *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)*, Bochumer Linguistische Arbeitsberichte, pages 27–37, Bochum, Germany, September.
- Brants, T. (2000). Inter-annotator agreement for a German newspaper corpus. In *Proceedings of the Second International Conference on Language Resources and Evaluation, LREC 2000*, Athens, Greece, May/June. European Language Resources Association.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.
- Dipper, S., Donhauser, K., Klein, T., Linde, S., Müller, S., and Wegera, K.-P. (2013). HiTS: ein Tagset für historische Sprachstufen des Deutschen [HiTS: A tagset for historical variants of German]. *Journal for Language Technology and Computational Linguistics*, 28(1):1–53.
- Dreessen, K. and Ihden, S. (2015). Korpuslinguistische Studien zur mittelniederdeutschen Syntax [Corpus studies on GML syntax]. In Markus Hundt et al., editors, *Deutsch im Norden*, volume 6 of *Jahrbuch für Germanistische Sprachgeschichte*, pages 249–275. Walter De Gruyter, Berlin, Boston.
- Härd, J. E. (2008). Syntax des Mittelniederdeutschen [GML syntax]. In Werner Besch, et al., editors, *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*, 2. Teilband, volume 2 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, pages 1456–1463. De Gruyter Mouton, Berlin, Boston, 2nd edition.
- Lasch, A., Borchling, C., Cordes, G., Möhn, D., and Schröder, I. (1956ff). *Mittelniederdeutsches Handwörterbuch [Concise dictionary of GML]*. Kiel/Hamburg.
- Leech, G., Garside, R., and Bryant, M. (1994). CLAWS4: The tagging of the British National Corpus. In *Proceedings of the 15th International Conference on Computational Linguistics, COLING 1994, Volume 1*, pages 622–628, Kyoto, Japan, August.
- Peters, R. and Nagel, N. (2014). Das digitale ‚Referenzkorpus Mittelniederdeutsch / Niederrheinisch (ReN)‘ [The digital ‘Reference Corpus Middle Low German/Low Rhenish (ReN)’]. In Vilmos Ágel et al., editors, *Paradigmen der Sprachgeschichtsschreibung*, volume 5 of *Jahrbuch für Germanistische Sprachgeschichte*, pages 165–175. Walter De Gruyter, Berlin, Boston.
- Scheible, S., Whitt, R. J., Durrell, M., and Bennett, P. (2011). A gold standard corpus of Early Modern German. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 124–128, Portland, Oregon, USA, June. Association for Computational Linguistics.
- Schiller, K. and Lübben, A. (1875–1880). *Mittelniederdeutsches Wörterbuch [GML dictionary]*. Aschendorff, Münster.
- Schiller, A., Teufel, S., Stöckert, C., and Thielen, C. (1999). Guidelines für das Tagging deutscher Textcorpora mit STTS [Guidelines for tagging German corpora with the STTS]. Technical report, University of Stuttgart and University of Tübingen.
- Schröder, I. (2014). Das Referenzkorpus: Neue Perspektiven für die mittelniederdeutsche Grammatikographie. [The Reference Corpus: New perspectives for GML grammaticography]. In Vilmos Ágel et al., editors, *Paradigmen der Sprachgeschichtsschreibung*, volume 5 of *Jahrbuch für Germanistische Sprachgeschichte*, pages 150–164. Walter De Gruyter, Berlin, Boston.
- Seemann, N., Merten, M.-L., Geierhos, M., Topfink, D., and Hüllermeier, E. (2017). Annotation challenges for reconstructing the structural elaboration of Middle Low German. In *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 40–45, Vancouver, Canada, August. Association for Computational Linguistics.
- Zinsmeister, H., Heid, U., and Beck, K. (2014). Adapting a part-of-speech tagset to non-standard text: the case of

STTS. In Nicoletta Calzolari, et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, pages 4097–4104, Reykjavik, Iceland, May. European Language Resources Association (ELRA).

9. Language Resource References

ReN-Team. (2017). *Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200-1650)*. Archived in Hamburger Zentrum für Sprachkorpora, Version 0.3. Publication date 2017-06-15. <http://hdl.handle.net/11022/0000-0006-473B-9>.