Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 860–867
Marseille, 11–16 May 2020

Material Philology Meets Digital Onomastic Lexicography: The NordiCon Database of Medieval Nordic Personal Names in Continental Sources

Michelle Waldispühl • Dana Dannélls, Lars Borin
Department of Languages and Literatures • Språkbanken Text/Department of Swedish
University of Gothenburg, Sweden
{michelle.waldispuhl, dana.dannells, lars.borin}@gu.se

Abstract

We present NordiCon, a database containing medieval Nordic personal names attested in Continental sources. The database combines formally interpreted and richly interlinked onomastic data with digitized versions of the medieval manuscripts from which the data originate and information on the tokens' context. The structure of NordiCon is inspired by other online historical given name dictionaries. It takes up challenges reported on in previous works, such as how to cover material properties of a name token and how to define lemmatization principles, and elaborates on possible solutions. The lemmatization principles for NordiCon are further developed in order to facilitate the linking to other name dictionaries and corpora, and the integration of the database into Språkbanken Text, an infrastructure containing modern and historical written data.

Keywords: Digital Humanities, Lexicon/Lexical database, LR Infrastructures and Architecture

1. Introduction

Recent years have seen a development in computational processing and storage technology enabling the creation of new kinds of language resources in support of research in many fields in the humanities and social sciences.

Even though digital language resources have been around for a long time – almost as long as computers (Winter, 1999) – they have typically been subject to many limitations. In the case of language corpora, many aspects of the original texts are lost when they are included in a corpus. Modern texts may lose typographical information and illustrations such as drawings and photographs accompanying the text often do not make it into the corpus. In the case of older, handwritten sources, both palaeographic information and aspects of the spatial organization of texts may be discarded in the digitization process.

In the field of linguistic typology with its long tradition of building linguistic databases, a useful distinction is often made between linguistic data at varying degrees of remove from natural linguistic interaction (see, e.g. Himmelmann 2012). The terminology varies among authors, and here we will talk about *primary*, *secondary*, and *tertiary* language data. Primary or raw data are recordings of naturalistic linguistic interaction, original texts in a language, etc., whereas secondary data are lexicons and grammars, i.e. heavily curated and interpreted as well as formally structured data, higher in generalization but with an inevitable concomitant loss of – sometimes relevant – details. Tertiary data are found in linguistic databases such as the *World Atlas of Language Structures* (WALS; Dryer and Haspelmath 2013),[1] at considerable remove from the raw data of linguistic interaction, and typically collected from grammars and lexicons, rather than from primary sources.

However, this way of building and organizing language resources is not a logical necessity, and vigorous digitization efforts together with enormous advances in storage and networking technologies now allow us to both have our cake and eat it, as it were. Linguistic databases containing highly curated secondary and tertiary data can be combined with digital versions of the original raw language data on which the generalizations have been made, which makes for greater transparency and reproducibility of research.

In this paper, we describe such a database which combines formally structured, interpreted and richly interlinked onomastic data with digitized versions of the medieval manuscripts from which the data originate. Such a resource offers opportunities for research in different disciplines such as onomastics, historical linguistics, history in general and prosopography in particular, cultural and literary studies. Including material properties of the original source provides greater transparency of research and it moreover opens up for research questions regarding the production and reception of writing at the time.

The database is currently under development.[2] In the following, we present the design and first version of the structure of the database and its online availability and use. We also address philological problems and other challenges reported on in previous works on historical personal name dictionaries and how these will be met in NordiCon. The main issues we deal with in this paper are the database's particular focus on covering material properties of the name tokens, the lemmatization principles that were developed in order to facilitate the connection to other name dictionaries and corpora, and the integration of the database into Språkbanken Text, an infrastructure containing modern and historical written data.

2. The Project

The NordiCon database is part of the project *Variation and contact in medieval personal names* and covers medieval Nordic personal names in Continental sources outside Scandinavia.[3] One particular feature of this name cor-

[1] https://wals.info

[2] It will be available by May 1st 2020 at https://spraakbanken.gu.se/karp/tng/?mode=nordicon

[3] "Nordic" is here used as basically synonymous with "Scandinavian" in the linguistic sense, i.e. the bearers of the names in

pus is that many instances were recorded by German and French speaking scribes under the conditions of a presumably oral language contact situation.

The significance of this digital resource is manifold. The names serve as historical record of Nordic mobility, religious practices, usage of Latin script, naming customs, text arrangement etc. Not least, they provide highly interesting material for linguistic analysis and language contact studies due to their multilingual scripting context. Given the interdisciplinary significance of the name corpus, interest groups from different fields with their respective research traditions have to be considered in the conception of the database.

Nordic names in Continental sources outside Scandinavia have not been documented comprehensively yet and older editions of parts of the data have several shortcomings as they partly lack lemmatization and do not meet philological standards of modern editions. The NordiCon database aims at filling this gap and overcoming limitations of previous research. The edition makes use of digital methods and will be published as an open source online platform. In this regard, the project aims to improve digital edition principles for historical personal names and contribute to the broader fields of historical onomastic lexicography and the Digital Humanities.

The approach presented here is new to historical onomastic lexicography. To date, other digital onomastic resources (cf. §3.2 for examples) do not provide a combination of formally interpreted and highly interlinked data with original sources. However, the requirement to integrate single textual witnesses is for example met in recent editions of medieval works, which make use of the potential of multimediality in digital editions. They include both a diplomatic transcription of the texts and a photographic reproduction of the actual manuscript sources, cf. e.g. the *Parzival Projekt*,[4] or the *Homer Multitext project*.[5]

3. Methods in Historical Onomastic Lexicography

3.1. Historical Lexicography and Material Philology

In the compilation of a name dictionary, several methodological questions arise, such as how the relevant linguistic data is delimited, selected and arranged. This concerns theoretical considerations of how a proper name and/or a certain name type is defined, the selection of data material to excerpt from, which interest groups are addressed, and what retrieval purposes shall be possible (cf. Möller 1995). NordiCon includes only given names (forenames), i.e. "obligatory" names that individualize a person – in contrast to bynames that are optional and family names that assign an individual to a small social group. Family names had not been established yet in the Northern countries by the time covered in NordiCon.

A clear challenge in the lexicography of historical given names has been the principle of arrangement. The main interest groups are historical linguists on the one hand who are interested mainly in the name forms, and historians on the other hand who want to get information about particular persons. Since the writing of names was not standardized in historical languages, spelling variation occurs not only for the same name (such as *Catharina* and *Katharina*), but also for the name of the same person. Following an alphabetical order in the arrangement of names, two name forms of the same person might be listed at two different places which makes the retrieval for a historian difficult. An ordering principle with a primary focus on persons, however, tears linguistically related name forms apart (on the dilemma cf. e.g. Autenrieth et al. 1979, XLV f.; Patzold 2012/2013, 40–42). Many recent editions of historical given names include therefore both an index of name forms and an index of persons (cf. most recently Geuenich et al. 2019).

For historical data it is furthermore particularly relevant to thoroughly document the single records and give exact references to the historical source where the single name entry can be found. Older name dictionaries pose difficulties in this respect since they refer to data in editions that followed older editorial practices of medieval manuscripts. These are not necessarily correspondent with current practices where the single manuscript and its materiality are central to the edition of word (and name) forms (cf. e.g. Haugen 2017), but may have applied normalizing principles to facilitate the reception of medieval works. Consequently, these edited forms are of little use for the linguists on the one hand who might be interested in contemporary spelling practices and the original phonology of names, for example. On the other hand, with respect to the "material turn" in philology in general and in literary and historical studies in particular, many research questions regarding the visual presentation of writing and other features in manuscripts cannot be addressed when the philological context is not documented in a database.

3.2. Historical Given Name Dictionaries Online

With the advent of digital editions, a wider spectrum of possiblities for arrangement and presentation of the data emerges and some of the challenges mentioned in the previous subsection can be addressed more easily than in printed editions. The following examples show different structures and uses of historical given name dictionaries online.

The Nomen et Gens database covers names and persons in Continental sources from 400 to 800 CE.[6] It was established jointly by historical linguists and historians and makes searches for name lemmas and historical persons possible. Further search entries include name gender, dating of when a person was mentioned, ecclesiastical or lay status of a person, attributes (such as ecclesiastical status) mentioned together with a name token, "ethnicity" of a name, sources and dating of the sources. These data types are provided for each name token, and additionally, an excerpt of the sentence the name token occurs in is given. The sources appear as a link where information on the original source

question belong to the North Germanic speech communities in continental and insular Scandinavia.

[4] http://www.parzival.unibe.ch/

[5] http://www.homermultitext.org/

[6] http://www.neg.uni-tuebingen.de/

(shelf mark, dating, script origin) and a reference for the edition of the respective source are given. The data structure is well-designed and addresses an interdisciplinary group of scholars and therefore, it serves as a model for NordiCon (§5.1). It provides, however, no access to original sources.

The "Dictionary of Medieval Names from European Sources" is an ambitious project that aims to cover all given names written in European sources between 500 and 1600 CE.[7] The big advantage of the dictionary is that name forms in different European languages are given for a name (lemma). The sources in which a single name token occurs appear as a link, where information on the edition of the source can be found. Named persons are however only named occasionally and it is not possible to browse for them. The website moreover does not provide background information on compilation or lemmatization methods. This dictionary is most probably addressed at a general audience and does not intend to mainly serve linguistic or historical research purposes.

Whereas the just-mentioned two dictionaries were intended and structured as online databases, the online dictionary of "Sveriges medeltida personnamn" (Medieval personal names of Sweden) is a digitization of its printed version.[8] It is possible to search for a name (or name string) to access the digitized entries for the single name lemmas as a PDF, where variant spelling forms, their dating and source are listed. The sources are given as an abbreviation and the reference for the source can be found in a list via a separate link on the homepage. Obviously, there is potential in linking the data in this resource and making it accessible more easily and in more diverse ways.

4. Data and Source Materials

The names of interest here are, as mentioned above, Nordic names from medieval Continental sources outside Scandinavia. The names belong most probably to Nordic pilgrims traveling to Rome and further to Jerusalem (cf. Jørgensen and Jónsson 1923; Naumann 2009). On their way, they passed by Continental monasteries and let their names be registered in the monasteries' *Libri vitae*. These are books that consist mainly of name lists. They became popular on the Continent in Carolingian 800 CE and were often in use for decades or even centuries. The manuscripts had a liturgical and memorial function: if your name was in the book, you would be included in the convent's prayers and remembered on Judgment Day.

To date, five medieval Continental manuscripts are known that contain Nordic personal names (cf. Naumann 2009): the Reichenau *Liber vitae* (Autenrieth et al., 1979) with approximately 740 tokens, the obituary of S. Martin-de-Champs / Paris (Heim and Wollasch, 1982) with around 50 tokens, the obituaries of Merseburg and Lüneburg (Althoff and Wollasch, 1983) and the *Liber vitae* of Remiremont (Hlawitschka et al., 1970). In the latter two sources, the number of tokens is still unclear. Only the Nordic names in the Reichenau manuscript have been documented and examined before. Naumann (2009) mentions the other sources, but does not give the exact records. This is why the NordiCon database, in its current state, only includes the Nordic names in the Reichenau *Liber vitae*. The manuscript is available both in a facsimile edition with a lemma index (Autenrieth et al. 1979, see below §5.3) and in a digitized version on *e-codices – Virtual Manuscript Library of Switzerland*.[9] The other sources are planned to be added at a later stage.

5. Structure of the Database

5.1. General Structure

The relational database is structured to cover data of philological, linguistic and historical interest, and its central part is the name token table (see figure 1). Every name token will be edited and described separately, then assigned to a name lemma (see below, §5.3) and, if a name bearer can be identified, to a historical person. Hence, the basic structure consists of the tables name token, sublemma/name lemma and person.

Furthermore, other tables are included to describe the philological context. The name tokens normally occur in name lists written by one single scribe (see figure 2). Single lists might also date from different times. Since it might be of relevance both to linguists who want to analyse name forms and name inventories and to historians who are interested in the social affiliation of a person, this list arrangement of the name tokens is accounted for in the database. Hence, it is possible to search for a certain list and get all the name tokens occurring in it. On the other hand, when searching for a name token, one can obtain its list context. A list is assigned to a scribal hand for which the main palaeographic characteristics are named.

The general database model is inspired by the Nomen et Gens database (cf. Kettemann 2006; Kettemann and Jochum-Godglück 2009) on names and persons in Continental sources from 400 to 800 CE.[10] In contrast to their database, however, NordiCon lays more stress on material and graphic details of the name forms.

At the time of writing, a prototype of the user interface is being developed. In the course of this process, the design of the database possibly may be affected.

5.2. Material Philology as an Integral Component

At the centre of the database structure is the single name token as it is attested in the historical source. As argued above (cf. §3.1), the material characteristics of the token entry and its material context are of central importance in modern philology. The database meets these requirements by giving information on material details on different levels.

First, the single token entry is not only transcribed diplomatically and as close to the original spelling as possible, but also its conservation status, its physical properties such as ink colour or ink blots, erasure or other amendments,

[7] http://dmnes.org/

[8] http://www.isof.se/sprak/namn/personnamn/sveriges-medeltida-personnamn/smp---natutgava-och-fullstandigt-register.html

[9] https://www.e-codices.unifr.ch

[10] http://www.neg.uni-tuebingen.de/

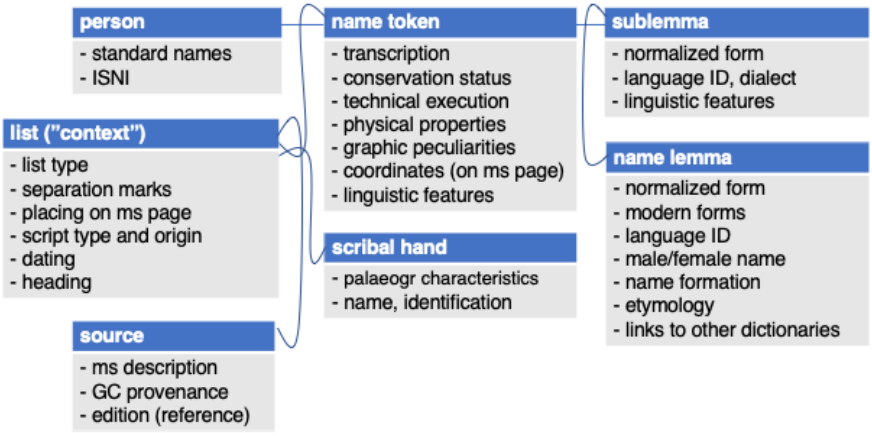

Figure 1: General structure of the database

and characteristics of the technical execution and writing method are described.

Second, the original record will be located and marked on the digitized manuscript page which will be displayed in the online database together with the token entry (cf. §6.1).

Third, layout characteristics of the list a token occurs in will be given together with other visual properties such as separation marks between the names and codicological details such as ruling of a manuscript page and the list's placement in relation to the ruling. The list marked in the top left corner of the manuscript page shown in figure 2, for instance, includes thirteen name tokens and a heading *hislant terra*. It is a vertical list (i.e. arranged in a column) with basically one name on each line, but with two exceptions of two names on the same line. Another property of the list is the consistent use of dots after every name token. The list can palaeographically be dated to the middle of the 1100s and probably originates from Southwestern Germany. The other marked list in the top right corner on the same page, on the other hand, contains of 27 name tokens and is also arranged vertically, but with three to four tokens per line and one token on the last line. It also uses dots as separation marks after the tokens, however inconsistently. Dating and origin of the script are the same as for the list in the top left corner, but on the basis of palaeographic characteristics the list is assigned to another scribal hand.

These descriptive details, together with the digitized manuscript page are intended, on the one hand, to provide the scholar with a transparent basis for their own judgment of the data. On the other hand, they facilitate systematic research on material issues relating to production and reception of the manuscript using the database.

5.3. Lemmatization of the Names

Lemmatization will be manually provided for all the single tokens. In NordiCon, lemmatization follows two different principles: Concordance of lemma forms in previous research and structural hierarchisation on two levels (etymological lemma and sublemma).

The purpose of the first principle, concordance, is to link and accumulate previous research on the names. The Nordic names in the Reichenau *Liber vitae* were first lemmatized in Jørgensen and Jónsson (1923) and in the facsimile edition (Autenrieth et al., 1979), a different lemmatization principle was applied. The name token *Durchgetil*, for example, is assigned the Old Icelandic form *Þórketil* in the former and the reconstructed Germanic form *thōr ketil* in the latter. Moreover, some of the Reichenau names are included in the Nordic name dictionaries of Lind (1905–1915) and DGP (Knudsen et al., 1936–48), where they are sometimes assigned again other lemma forms (e.g. *Thorkil* for *Durchgetil* in DGP). To achieve better comparability between previous research but also between the Nordic name corpus in Continental sources treated here and the Nordic homeland records, all these previously applied lemma forms will be included in the database. Moreover, modern forms of the respective names are given whenever possible to facilitate the integration of NordiCon into Språkbanken Text's lexical infrastructure through SALDO (cf. §6).

The second principle, structural hierarchization, follows an approach suggested by Alhaug (1992). On the linguistically most abstract level, an etymological lemma will be given. Moreover, the single etymological lemmas are, whenever appropriate, divided into sublemmas. For the etymological level, the reconstructed Old West Norse forms proposed by Lena Peterson in her *Dictionary of Proper Names in Scandinavian Viking Age Runic Inscriptions* (Peterson, 2007) will be used whenever possible. She based her lemmatization principles both on the actually recorded forms, on the names' etymology, and on language change processes during the Viking Age (Peterson, 2007, 11). On the sublemma level, morphological and phonological variation occurring in the contemporary Nordic records are aimed to be

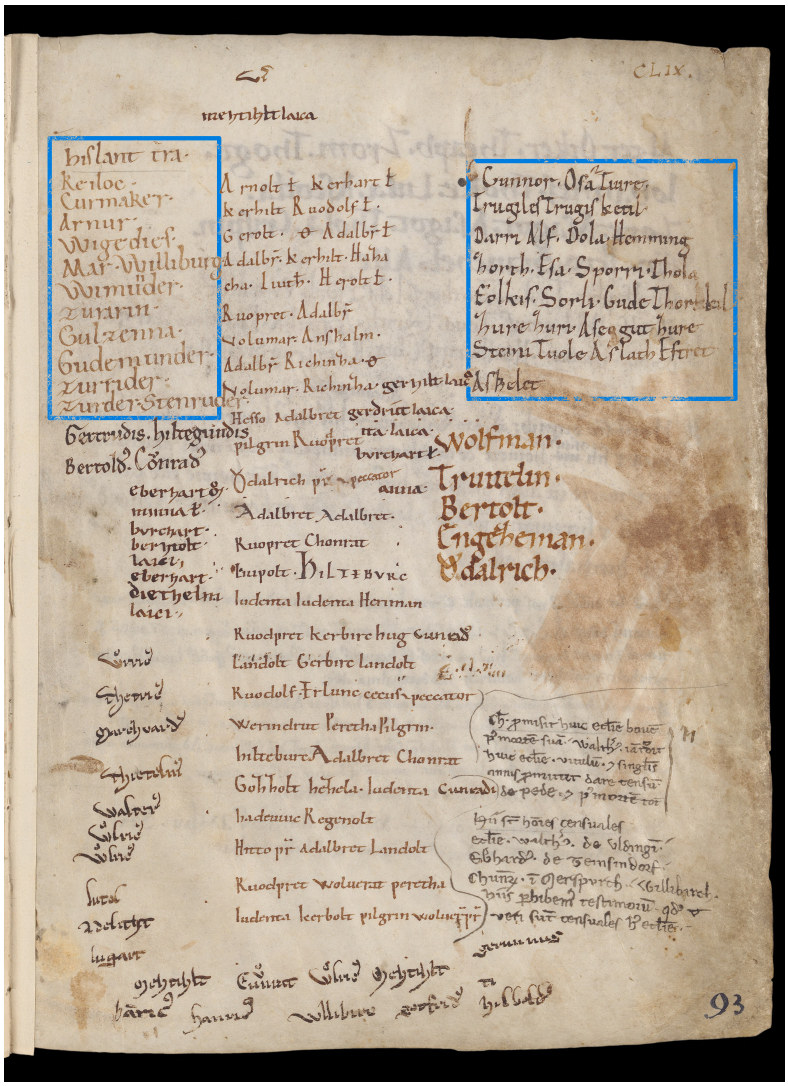

Figure 2: Example of a manuscript page from the Reichenau *Liber vitae*, Zentralbibliothek Zürich, Ms. Rh. hist. 27. fol. 93r, (www.e-codices.unifr.ch) (boxes at top left and right added for emphasis of elements described in the text)

reflected. For the etymological lemma *Þōrkætill*, for example, three different sublemmas are included: (1) *Thorketil* with the full form in the second name element; (2) *Thorkil* with a reduced second element; and (3) *Throkil* with a reduced second element and metathesis in the first element. Here, existing lemma forms from previous research named above are used whenever possible, and the linguistic processes that have led to the different sublemma forms are described. The sublemma fields are free text fields. The purpose of this second principle is to facilitate comparability to the Nordic material on the one hand, and to reflect structural variability on the other hand.

Single names that cannot be interpreted or where an interpretation and assignment to a lemma is uncertain, are highlighted and the uncertainty is commented on. Comments will also be added where assignment of name forms to two different lemmas is plausible. The name token *VVilliburg*, for example, can be assigned to both German *Williburg* or to North Germanic *Vilbiǫrg*. In the latter case, the name token has to be interpreted as a result of a linguistic adaptation from North Germanic to German. This is a plausible case here since the name is written by a German scribe and occurring as one of thirteen names in a list with otherwise unambiguously North Germanic names and titled *hislant terra* 'Iceland' (cf. Waldispühl 2018, 145f.). In all cases of lemma ambiguity, a ranking of the different assignments and its motivation will be given.

Following the practice of the Nomen et Gens database, historical persons will also be indicated by their "standard name(s)", i.e. their modern names that are used in historical literature in different languages (cf. Patzold 2012/2013, 36 f.). These names will be registered in the table of the respective historical person, which will be linked to a certain name token, and via the token to a name lemma. The user will be able to search the database not only for the philological lemmas, but also for these standard names. This practice is necessary to make the database accessible for historians and the interested public. So far, there are only very few instances in the NordiCon database where a named historical person could be possibly identified. One example are the name tokens *Hestrit* and *Osthein* that could belong to

Estrid and her husband Östen of the Jarlabanke clan named on three runestones in Swedish Uppland (cf. Edberg 2006).

6. Integrating NordiCon into Språkbanken Text's Lexical Infrastructure

Språkbanken (the Swedish Language Bank) is a national research infrastructure funded for the period 2018–2024 by the Swedish Research Council and the ten collaborating partners, seven universities and three national cultural-heritage institutions. Its division Språkbanken Text at the University of Gothenburg is committed to making NordiCon widely available to researchers. Språkbanken Text is also a CLARIN B center, which will ensure that the dataset will be available as part of the European research infrastructure provided through CLARIN ERIC.

NordiCon will be part of Karp, Språkbanken Text's open lexical infrastructure (Borin et al., 2012a). Karp has a backend and a frontend side. The backend side gathers and organizes contemporary and historical lexicons in a large network.[11] Lexical entries in the different lexicons are interlinked in the network either on the lexical sense level or on the lemma level. The majority of the contemporary lexicons in the network have been constructed for natural language processing (NLP) according to the ISO Standard LMF (Francopoulo, 2013) and connected through SALDO's lexical sense identifier (Borin et al., 2013). Historical resources that originally were not structured for NLP applications were converted to machine readable formats and were interlinked to SALDO through a diachronic pivot (Borin and Forsberg, 2011). Their base forms and morphological inflections were either extracted from the lexicon or inferred through language technology analysis (Adesam et al., 2014).

6.1. Linking Historical Lexical Resources

Linking historical resources to a lexical infrastructure requires specific techniques that not only are tailored to the material but also to the technical solutions of the infrastructure. An example of a technique that is implemented in Karp is the ability to access the source document, and the possibility of viewing it in the interface. When a user performs an entry search he/she will be presented with a link to the facsimile, given it is available, as a part of the search results. The user can further click the link and access the facsimile which is enlarged at the center of the page. This functionality has been proven useful for accessing entries in Hellquist's Swedish etymology database.[12]

Several lexical resources for Old Swedish have been added to the lexical infrastructure over the years, including Schlyter (1884), Söderwall (1884), just to name a few. Each required specific solutions to link it to the lexical network. For a database such as NordiCon where each lexical entry is provided in a separate field and is linked to the lemma, linking involves matching the entries against the correct sense in SALDO, for example the entry *Inga* appearing in the NordiCon database will be linked to SALDO sense id Inga..1 which in turn is paired with its part-of-speech proper noun. Out of the 729 entries in the current version of NordiCon we have linked 28 to SALDO senses. An obstacle here could arise in cases where there are more than one occurrence of the same name which will require disambiguation, but since occurrences of ambiguous personal names in SALDO is very low, this might not be a problematic issue.

An advantage of linking the NordiCon database to Karp is access to name spelling variations, that potentially could enrich the other lexical resources such as Hellquist's Swedish etymology database.

6.2. Online Access through Karp

All lexicons that are available in Karp are also browsable and accessible via Karp's search interface. Users can search in each lexicon separately or in all lexicons simultaneously. Technically, the interface offers solutions that are contingent on the annotation schema and the structure of different lexical databases. It allows for simple browsing possibilities on different levels, which makes it conventionally easy to search for content, not only lexical content, in different fields in the database. Users can browse for the lemma, base form, part-of-speech category, paradigm, conservation status, material properties, print year, etc.

One prominent functionality of Karp that more often is appreciated by researchers coming from the digital humanities and social sciences is direct access from Karp to Språkbanken's corpus infrastructure, Korp (Borin et al., 2012b). The user can click on the lexical entry and get directed to a set of corpus examples where the entry appears.

Karp also provides a function for accessing external databases directly from the interface such as Nomen et Gens or for linking to personal names in other relevant databases such as the Biographical Dictionary of Swedish Women (SKBL),[13] and the International Standard Name Identifier (ISNI).[14] Once the NordiCon database is available in Karp we will add links to these external resources.

7. Conclusion

NordiCon is a an example of a database where formally interpreted and richly interlinked onomastic data is combined with information on material properties and digitized versions of the medieval manuscripts from which the data originate. This architecture not only provides data required in studies with a focus on materiality, but also guarantees transparency for the structural interpretations of the names. The lemmatization principles developed for NordiCon, (1) *concordance* of lemma forms in previous research on Nordic names and in SALDO; and (2) *hierarchization* on two structural levels facilitate the integration of the database into Karp, Språkbanken Text's lexical infrastructure, and show at the same time structural variability in the name forms. Entries in NordiCon will be accessable via Karp's search interface and linked both to lexica that are already provided there and to Korp, where suitable corpus data is available. Moreover the possibility will be given to

[11] All lexicons are freely available under CC-BY license.

[12] https://spraakbanken.gu.se/eng/resource/hellqvist

[13] https://skbl.se/en

[14] http://www.isni.org/

access relevant external databases for further exploration and exploitation of its content.

8. Acknowledgements

This project has been supported by the Swedish Research Council, grant number 2018-01556 – Variation and Contact in Historical Personal Names. The national research infrastructure *Nationella språkbanken* is jointly funded for the period 2018–2024 by the Swedish Research Council (grant number 2017-00626) and its 10 partner institutions.

9. Bibliographical References

Adesam, Y., Ahlberg, M., Andersson, P., Bouma, G., Forsberg, M., and Hulden, M. (2014). Computer-aided morphology expansion for Old Swedish. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14) May 26-31, 2014 Reykjavik, Iceland*, pages 1102–1105.

Alhaug, G. (1992). Lemmatisering av namnevariantar – ein nivåmodell basert på norsk materiale. *Studia anthroponymica Scandinavica. Tidskrift för nordisk personnamnsforskning*, 10:115–150.

Althoff, G. and Wollasch, J. (1983). *Die Totenbücher von Merseburg, Magdeburg und Lüneburg*. Monumenta Germaniae historica. [Antiquitates]. Libri memoriales et necrologia. Nova series. Hahnsche Buchhandlung, Hannover, [faksimile-ausgabe] edition.

Autenrieth, J., Geuenich, D., and Schmid, K. (1979). *Das Verbrüderungsbuch der Abtei Reichenau*. MGH. Libri memoriales et Necrologia. N. S. Hahnsche Buchhandlung, Hannover.

Borin, L. and Forsberg, M. (2011). A diachronic computational lexical resource for 800 years of Swedish. In *Language technology for cultural heritage*, pages 41–61. Springer.

Borin, L., Forsberg, M., Olsson, L.-J., and Uppström, J. (2012a). The open lexical infrastructure of Språkbanken. In *Proceedings of LREC 2012*, pages 3598–3602, Istanbul. ELRA.

Borin, L., Forsberg, M., and Roxendal, J. (2012b). Korp — the corpus infrastructure of Språkbanken. In *Proceedings of LREC 2012*, pages 474–478, Istanbul, May. ELRA.

Borin, L., Forsberg, M., and Lönngren, L. (2013). SALDO: a touch of yin to WordNet's yang. *Language Resources and Evaluation*, 47(4):1191–1211.

Matthew S. Dryer et al., editors. (2013). *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Edberg, R. (2006). Spår efter en tidig jerusalemsfärd. *Fornvännen*, 101:342–347.

Gil Francopoulo, editor. (2013). *LMF: Lexical markup framework*. ISTE/Wiley, London/Hoboken.

Geuenich, D., Ludwig, U., Crivello, F., Erhart, P., and Zettler, A. (2019). *Die St. Galler Verbrüderungsbücher*. Monumenta Germaniae Historica. Libri memoriales et necrologia. Nova series. Harrassowitz Verlag, Wiesbaden.

Haugen, O. E. (2017). Levels of granularity: Balancing literary and linguistic interests in the editing of medieval texts. In Harry Lönnroth, editor, *Philology Matters!*, pages 118–135. Brill, Leiden.

Heim, W.-D. and Wollasch, J. (1982). *Synopse der cluniacensischen Necrologien*. Münstersche Mittelalter-Schriften. W. Fink, München.

Himmelmann, N. P. (2012). Linguistic data types and the interface between language documentation and description. *Language Documentation & Conservation*, 6:187–207.

Hlawitschka, E., Schmid, K., and Tellenbach, G. (1970). *Liber memorialis von Remiremont*. Weidmann, Dublin/Zürich.

Jørgensen, E. and Jónsson, F. (1923). Nordiske Pilegrimsnavne i Broderskabsbogen fra Reichenau. *Aarbøger for nordisk Oldkyndighed og Historie*, 13:1–36.

Kettemann, W. and Jochum-Godglück, C. (2009). Namenüberlieferung und Personennennung im Kontext. Bedingungen und Möglichkeiten von Erfassung und Auswertung in der Datenbank "Nomen et gens". In Albrecht Greule, editor, *Namen des Frühmittelalters als sprachliche Zeugnisse und als Geschichtsquellen*, Ergänzungsbände zum Reallexikon der germanischen Altertumskunde, pages 267–277. De Gruyter, Berlin.

Kettemann, W. (2006). Vom Überlieferungsbefund zur Ergebnispräsentation. Konzeption und Organisation der datenbankbasierten Projektarbeit. In Dieter Geuenich et al., editors, *Name und Gesellschaft im Frühmittelalter. Personennamen als Indikatoren für sprachliche, ethnische, soziale und kulturelle Gruppenzugehörigkeiten ihrer Träger*, Deutsche Namensforschung auf sprachgeschichtlicher Grundlage, pages 41–53. Georg Olms Verlag, Hildesheim.

Knudsen, G., Kristensen, M., and Hornby, R. (1936–48). *Danmarks gamle Personnavne. I. Fornavne*. I Kommission hos G.E.C. Gad, København.

Lind, E. H. (1905–1915). *Norsk-isländska dopnamn ock fingerade namn från medeltiden*. Lundequist / Harrassowitz, Uppsala / Leipzig.

Möller, L. A. (1995). Methods and problems in proper name lexicography. In Ernst Eichler, et al., editors, *Namenforschung: ein internationales Handbuch zur Onomastik*, volume 1 of *Handbücher zur Sprach- und Kommunikationswissenschaft*, pages 324–328. De Gruyter, Berlin / New York.

Naumann, H.-P. (2009). Die nordischen Pilgernamen von der Reichenau im Kontext der Runennamenüberlieferung. In W. Heizmann et al., editors, *Analecta Septentrionalia*, pages 776–800. De Gruyter, Berlin.

Patzold, S. (2012/2013). Personen, Namen, Lemmatisierung: Namenforschung zwischen Geschichts- und Sprachwissenschaft am Beispiel des Projekts "Nomen et Gens". *Namenkundliche Informationen*, 101/102:22–48.

Peterson, L. (2007). *Nordiskt runnamnslexikon*. Institutet för språk och folkminnen, Uppsala, 5 rev. utg. edition.

Schlyter, C. (1884). *Glossarium ad Corpus iuris Sueo-Gotorum antiqui, Ordbok till Samlingen af Sweriges gamla lagar*. Berlingska boktryckeriet, Lund, Sweden.

Söderwall, K. (1884). *Ordbok Öfver svenska medeltidsspråket*, volume Vol I–III. Svenska Fornskriftsällskapet, Lund, Sweden.

Waldispühl, M. (2018). ‚Deutsch' oder ‚nordgermanisch'? Sprachliche Bestimmung von Namen im Reichenauer Verbrüderungsbuch vor dem Hintergrund von Sprachkontakt und Mehrsprachigkeit. In Christoph Hoffarth et al., editors, *Ambiguität und die Ordnungen des Sozialen im Mittelalter*, Das Mittelalter. Perspektiven mediävistischer Forschung. Beihefte, pages 129–150. De Gruyter, Berlin.

Winter, T. N. (1999). Roberto Busa, S.J., and the invention of the machine-generated concordance. *The Classical Bulletin*, 75(1):3–20.