Article

Towards A Taxonomy of Uncertainties: Analysing Sources of Spatio-Temporal Uncertainty on the Example of Non-Standard German Corpora

Renato Rocha Souza, Amelie Dorn *, Barbara Piringer and Eveline Wandl-Vogt

Austrian Centre for Digital Humanities, Austrian Academy of Sciences, 1010 Vienna, Austria
* Correspondence: Amelie.Dorn@oeaw.ac.at

check for updates

Abstract: Different types of uncertainties occur in almost all datasets and are an inherent property of data across different academic disciplines, including digital humanities (DH). In this paper, we address, demonstrate and analyse spatio-temporal uncertainties in a non-standard German legacy dataset in a DH context. Although the data collection is primarily a linguistic resource, it contains a wealth of additional, comprehensive information, such as location and temporal detail. The addressed uncertainties have manifested because of a variety of reasons, and partly also because of decades of data transformation processes. We here propose our own taxonomy for capturing and classifying the various uncertainties, and show with numerous examples how the remedying but also re-introduction of uncertainties affects DH practices.

Keywords: digital humanities; uncertainty; indigenous languages; spatial uncertainty; temporal uncertainty; lexical data uncertainty

1. Introduction

Uncertainty is an inherent aspect of humanity, of our daily activities, actions and interactions. It is typically associated with unknown or lacking information, imprecise or incomplete knowledge, inaccurate measurements and risk. As uncertainty permeates our lives in various forms and ways, it has also become a topic of discussion in policy making and entrepreneurship, and has equally been taken up in the scholarly and scientific discourses.

In the light of recent developments in the European policy landscape regarding science, research and innovation, the boosting of "openness" on several levels (i.e., innovation, data), has also given fresh impetus for uncertainty to resurface in the wider scientific discourse [1,2]. Nowotny [3] (p. 1602) states that "[...] science thrives on the cusp of uncertainty—the rare moments of intense creativity when completely new perspectives and visions open up and lead to new discoveries and insights." Scientific research and innovation processes are thus inherently uncertain, and the more so as their progress evolves towards ecosystem networks of actor groups with increased inclusion, collaboration and participation of different stakeholders, and the pressing necessities to meet human needs and face societal challenges. "It is therefore crucial to distinguish when and where science needs time and space when to engage with uncertainty." [3] (p. 1603). As a consequence, embracing uncertainty, creating a culture of learning from errors and allowing for the creation and acknowledgement of serendipitous discovery conditions are key; yet they lie at the centre of the ongoing discussion around scientific innovation and progress, not only at the policy level. Across academic and scholarly discourses, uncertainty—for decades—has been a fundamental topic of interest across scientific disciplines, including philosophy [4], psychology [5], physics [6], information science [7], economics [8], law [9] and statistics [10], just to name a few examples (see also [11]). There, uncertainty is typically dealt

with from the point of view of risk assessment, measurements of uncertainty or methods of removing uncertainty to attain higher degrees of certainty (cf. [12]).

In this context, a distinction can be made in how uncertainty is dealt with, for example, in the fields of natural sciences in contrast to the humanities. Whilst uncertainties in natural sciences are mostly related to the expected limits in the possibilities of making measurements and also inherent to the statistical properties of what can be inferred on the empirical samples, uncertainties in the humanities, however, can also involve subjective aspects related to perception, ambiguity, vagueness, incompleteness, credibility, etc.

The research purpose of the present paper thus evolves out of the uniqueness of the legacy collection dealt with here. The types of uncertainties encountered in humanities datasets already differ from those more typical in the natural sciences. Our specific non-standard dataset, and the focus on spatio-temporal aspects, makes it a necessity to devise our own taxonomy in order to fully capture the relevant uncertainties.

The paper is thus structured as follows: Section 1 (Introduction) presents an overview of existing and relevant taxonomies dealing with one or more of the aspects (temporal, spatial, uncertainties) central to the research purpose, and also an outline of the diverse contexts that uncertainty has been dealt with in the digital humanities field. In Section 2 (Materials and Methods), the specific dataset is illustrated, followed by a pointer to previous analyses of spatial and temporal aspects. In addition, the methods of devising our taxonomy are outlined. Section 3 (Results) presents the DBÖ (Datenbank der bairischen Mundarten in Österreich/Database of Bavarian Dialects in Austria) taxonomy of uncertainties illustrated with detailed examples of its specific categories. Finally, Section 4 (Discussion/Conclusion) discusses the newly composed taxonomy against the background of existing taxonomies that have been reviewed in Section 1. Aside from this, reflections on how the remedying or re-introduction of uncertainties affects digital humanities (DH) practice are provided.

1.1. Research Framework and Context

This study is realised in the context of exploration space. The exploration space was established in the working group 'Methods and Innovation' [13] at the Austrian Centre for Digital Humanities (ACDH) as a virtual and physical environment for fostering experimentation and innovation in the networked humanities, encouraging exchange at the interface of humanities, geography and environmental studies, as well as design and arts. Actors co-design innovation processes and experiment on questions of cultural, linguistic and biological diversity. Subsequently, exploration space has been listed as a best-practice example of open innovation by the Ministries of Transport, Innovation and Technology (bmvit) and Education, Science and Research (bmwfw) [14]. The particular framework and approach taken in the project has given rise to, and enabled, the establishment of the Open Innovation Research Infrastructure (OI-RI) and exploration space at the Austrian Centre for Digital Humanities (ACDH-OeAW) at the Austrian Academy of Sciences, in which exploreAT! is reified.

exploreAT!—exploring Austria's culture through the language glass (cf. [15])—has evolved as a cross-disciplinary project at ACDH-OeAW since 2015. It brings together expertise from different disciplines and collaboration partners in the fields of cultural lexicography and open innovation (ACDH-OeAW, Austria), semantic technologies (ADAPT Centre, Dublin City University, Dublin, Ireland) and human–machine interaction via visualization (VisUSAL, Universidad de Salamanca, Salamanca, Spain) (cf. [16–19]). The project more generally aims at making the implicit, but to-date still partially unlocked, cultural knowledge contained within a non-standard language legacy dataset accessible, connectable and reusable for different disciplines and actor groups. To achieve this, a variety of cross-disciplinary knowledge and agile research and design thinking methods are drawn upon. Following the principles of the Open Innovation Strategy for Austria [20], it also connects to different actor groups from society and industry.

The exploreAT! project evolves around a digitised non-standard language resource of the Bavarian Dialects in Austria (DBÖ - Datenbank der bairischen Mundarten in Österreich/Database of Bavarian Dialects in Austria) and related dbo@ema (Database of Bavarian Dialects Electronically Mapped; Wandl-Vogt, E. (2010; Ed.). Datenbank der bairischen Mundarten in Österreich electronically mapped [Database of the Bavarian Dialects in Austria electronically mapped] (dbo@ema). Wien. [Processing status: 2018.01.]) (cf. [21]). This highly heterogeneous collection captures the language and through it the culture of the local society in the area of the former Austro-Hungarian empire from the beginnings of the German language up until now. Besides capturing the local non-standard speech of the population, it contains a wealth of cultural information on detailed aspects of the former day-to-day life of the local population, including professions, customs, religious festivities, folklore medicine and food.

Inspiration also came from another DH project ProvideDH (Progressive Visual Decision Making in Digital Humanities) [22] (cf. [23]), who are developing their own set of metrics based on a distinction of epistemic uncertainty (systematic uncertainty; reducible) and aleatory (inherent uncertainty; irreducible) as proposed by [24] (cf. [25]).

The ProvideDH project ([23,25]) is implemented at the University of Salamanca (GrialUSAL) (ES) in collaboration with exploration space @ ACDH-OeAW Austrian Academy of Sciences (AT), Trinity College Dublin (IE) and the Supercomputing Centre Poznan (PL). The project is also run within a digital humanities environment and seeks to propose and develop innovative progressive visualisation tools that track and convey the degree of uncertainty of a given (humanities) dataset during its evolution, and how these data are affected by different computational models applied. The different sources of uncertainty introduced over the course of time and affecting the DH practice are made visible and can be assessed by the researcher, ultimately supporting the decision-making process (cf. [25]). Tools are developed for different use-cases based on humanities data (XML/TEI) (the 1641 depositions) and scenarios (according the Open Innovation Research Infrastructure Design © ewv 2017), and aim to enable collaborative and visual annotation in an uncertainty-aware context (cf. [26]) (Figure 1).
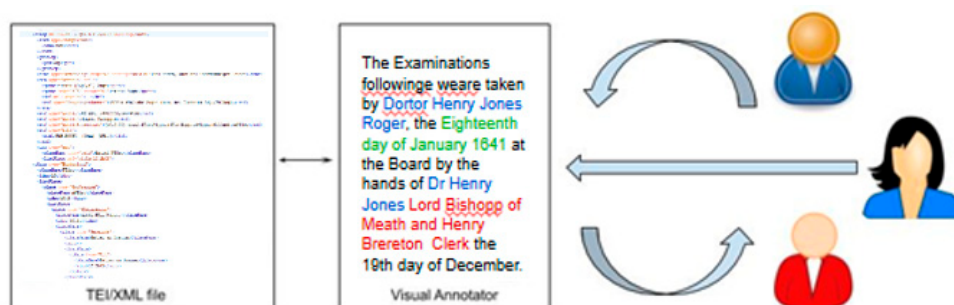
Figure 1. The collaborative visual annotation process supporting decision-making in uncertainty-aware contexts. Figure redrawn from [26].

1.2. Taxonomies of Uncertainty: A Concise Overview

This section presents a review of selected taxonomies dealing with one or all of the categories relevant for the present research purpose of capturing spatial and temporal uncertainties. The chosen taxonomies offer also a broad basis for evaluating whether existing classifications can be readily adopted, or whether categories more specific to lexicographic collections would need to be devised.

There are different kinds of taxonomies that have been proposed across disciplines to classify them [27]: Specific taxonomies of uncertainty can be found for various given areas, such as biology [28], health [29] and trading regulations [30]. A very general taxonomy is presented by the New World Encyclopaedia entry on uncertainty [31] (Figure 2). The difference between epistemological and ontological uncertainty was taken into account when devising our own taxonomy.
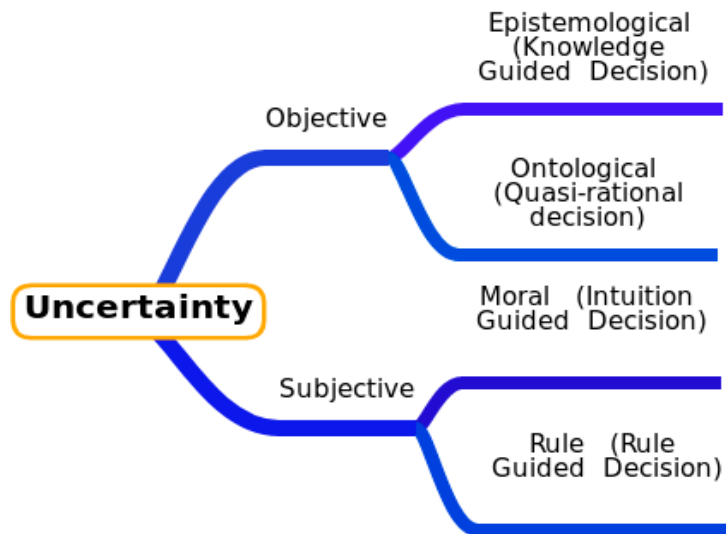
Figure 2. Taxonomy of uncertainty. Figure redrawn from [31].

Smithson [32] presents a comprehensible one (Figure 3), adapted from [33]. In this taxonomy, uncertainty appears as a specific kind of incompleteness, but not as an error.

Figure 3. Taxonomy of ignorance and uncertainty. Figure redrawn from [32].

Shattuck, Lewis Miller, and Kemmerer [34], on the other hand, make the distinction between uncertainty produced by the flow of information and by the individuals dealing with given information (Figure 4). This is the same approach as [35].
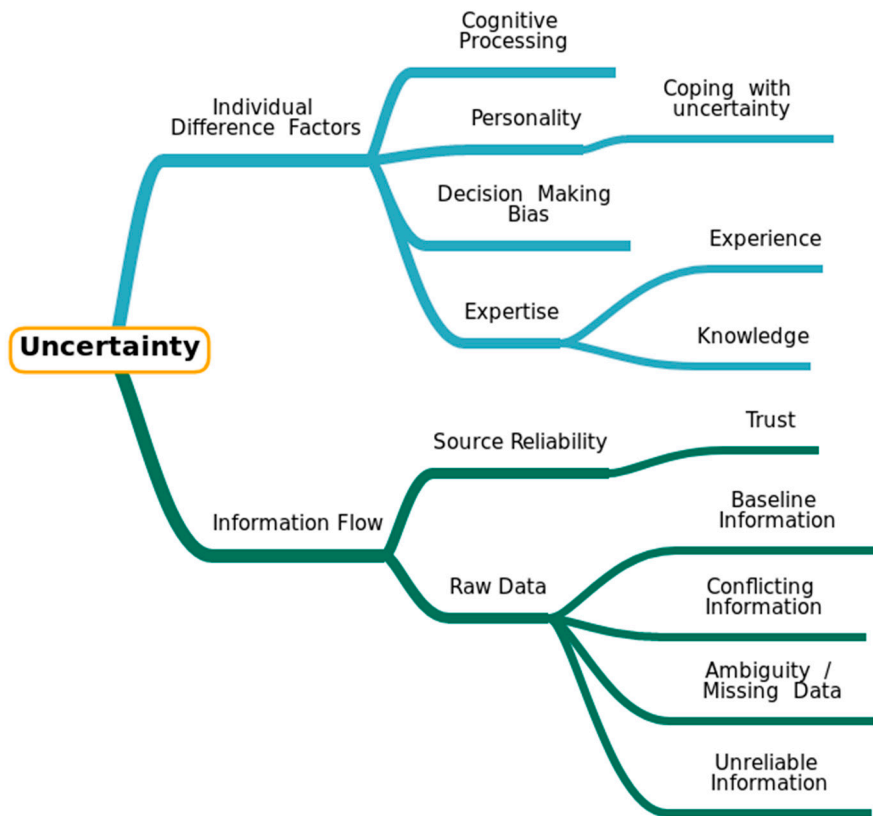
Figure 4. Integrated taxonomy of uncertainty. Figure redrawn from [34].

Lovell [36], in an extended digression on the topic, presents a detailed compilation of uncertainties of all sources (Figure 5). In this view, which adds another aspect to the previous one, uncertainties can be originated in (i) the world itself, (ii) the empirical evidence and (iii) the human subjects that interpret them (decision makers).
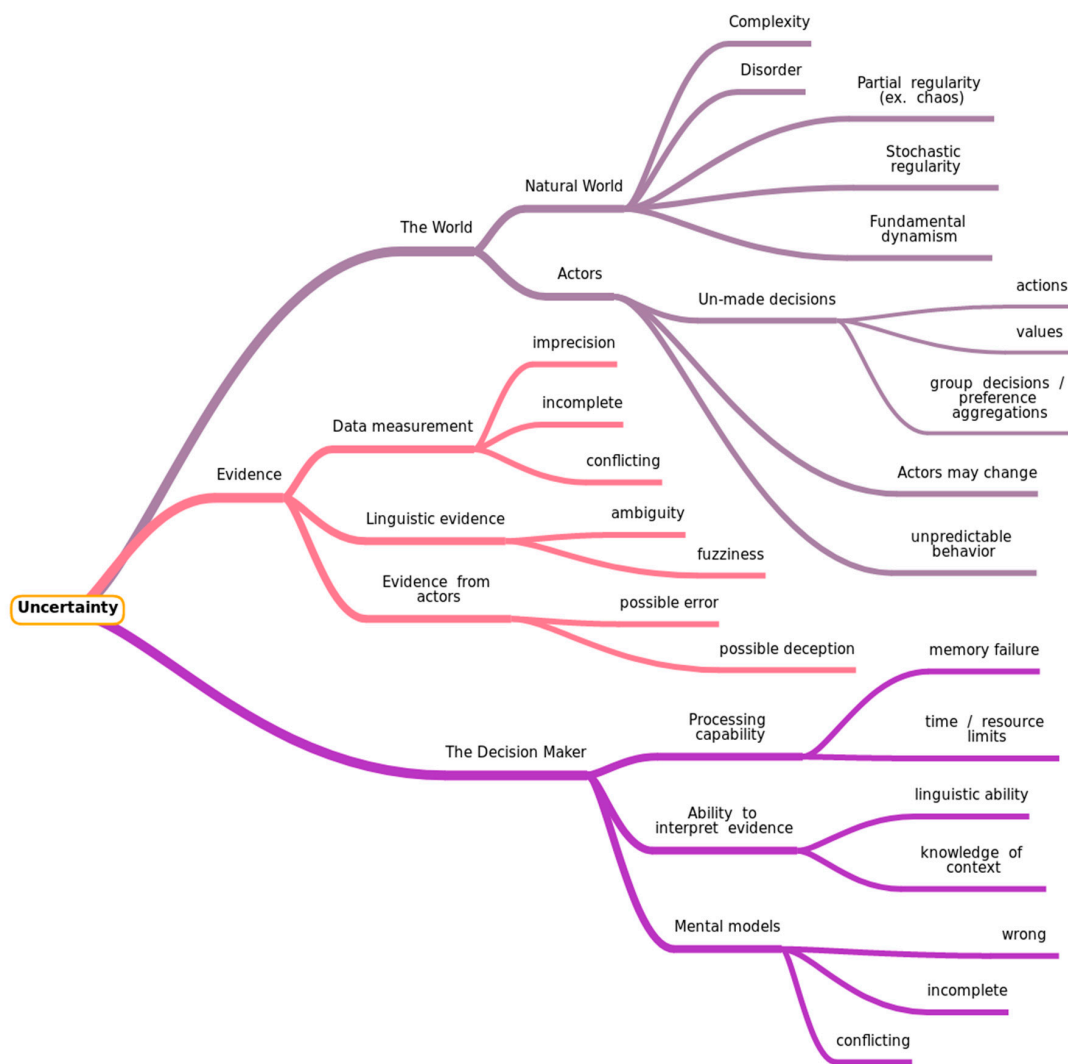
Figure 5. Uncertainties from all sources. Figure redrawn from [36].

Vullings, de Vries, and de Borman [37], based on [24], devised a fairly complete model for dealing with spatial uncertainties (Figure 6).

The most salient aspect is the distinguishment of "data (input/output)" and "planning process".

Temporal uncertainties often come associated with spatial data, as pointed out by [38]. Aigner et al. [39] distinguish time points and time intervals, and also draws attention to the kind of events that are being described when they involve other variables (as space). Kissling et al. [40] identify the variation of length of time series and the precision of time in the collection process as sources of temporal uncertainty.
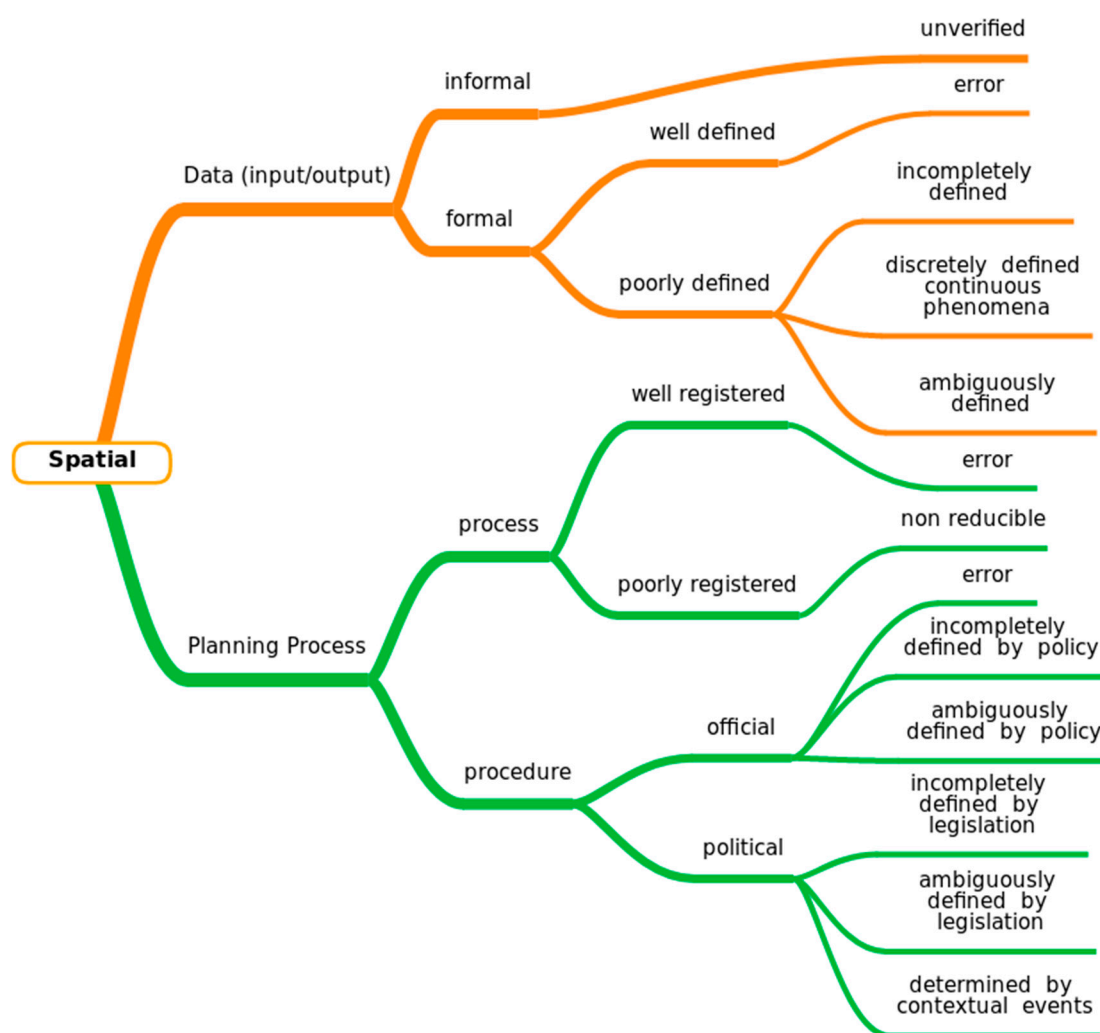
Figure 6. A taxonomy of uncertainty in spatial planning. Figure redrawn from [37].

1.3. Uncertainty in (Digital) Humanities

In the scope of the present paper, we specifically deal with the exploration of uncertainty in the field of humanities and, in particular, digital humanities (DH). There, uncertainty has in recent years also been in the spotlight of discussion and has generated an increased interest, particularly in relation to data and data treatment. Our understanding of uncertainty in this paper concerns the process of data transformation and evolution, and in understanding the sources of uncertainty that can affect the DH practice, which we illustrate with concrete examples from a case study dataset, focusing on spatio-temporal aspects.

While some have addressed the specific possibilities of encoding and modelling uncertainty in humanities data [41], others have discussed the differentiation between humanistic "fact" and interpretation that shapes the nature of humanistic research questions and attitudes towards sources [42]. Further issues relevant to the DH field were also discussed in different studies in a special track on "Uncertainty in Digital Humanities" at the recent Conference of Technological Ecosystems for Enhancing Multiculturality (TEEM) in 2018 [43] (see also [25,42,44,45], etc.). The concept of uncertainty in relation to data driven innovation (DDI)—the production of innovative outputs from data—has more-specifically been addressed in [46], who urge the re-thinking and re-organising of views on data collection in the light of open science and cross-organisational collaborations to enable new designs of DDI networks for dealing with aspects of heterogeneity in data.

We follow a data-driven research approach, and address aspects of uncertainty in a non-standard language legacy dataset (DBÖ); Österreichische Akademie der Wissenschaften. (1993–). Datenbank der bairischen Mundarten in Österreich [Database of Bavarian Dialects in Austria] (DBÖ). Wien. [Processing status: 2018.01.] in the context of the DH project exploreAT! [15]). Even though heterogeneous data offers a wide spectrum of different kinds of uncertainties to be addressed, we here more-specifically focus on the analysis of uncertainties in spatial and temporal aspects of our legacy language collection.

Uncertainty in data pertaining geographic information systems (GIS), and spatial information in general, is a frequently explored topic (cf. [47–50]) and also finds its own entry in the GIS dictionary [51].

In relation to our language dataset and spatio-temporal dimensions, uncertainties arisen in the process of data evolution include imprecise or erroneous information and knowledge, incomplete information, spelling mistakes, abbreviations, ambiguous information, missing information or uncertainties introduced in the process of digital data transformation and standardisation by tools or persons. In particular—in combination with language phenomena and changes in linguistic processes, such as shifts in language borders/boundaries—uncertainties in the spatio-temporal aspects play an important role and also give insights. To this end, we have grounded our analysis of the uncertainties over specific already-existent sets of categories—eventually modified to include novel aspects we have found in our data uncertainties. The model proposed by [37] based on [24], as introduced in Section 1.1, proved fairly successful when used to classify our examples of spatial uncertainties. We have adapted the taxonomies presented and incorporated new elements. Figure 7 depicts the classes of uncertainties used to classify the samples in our collection regarding the spatial, temporal and linguistic dimensions. We will provide some examples of each one in the results section of this paper.
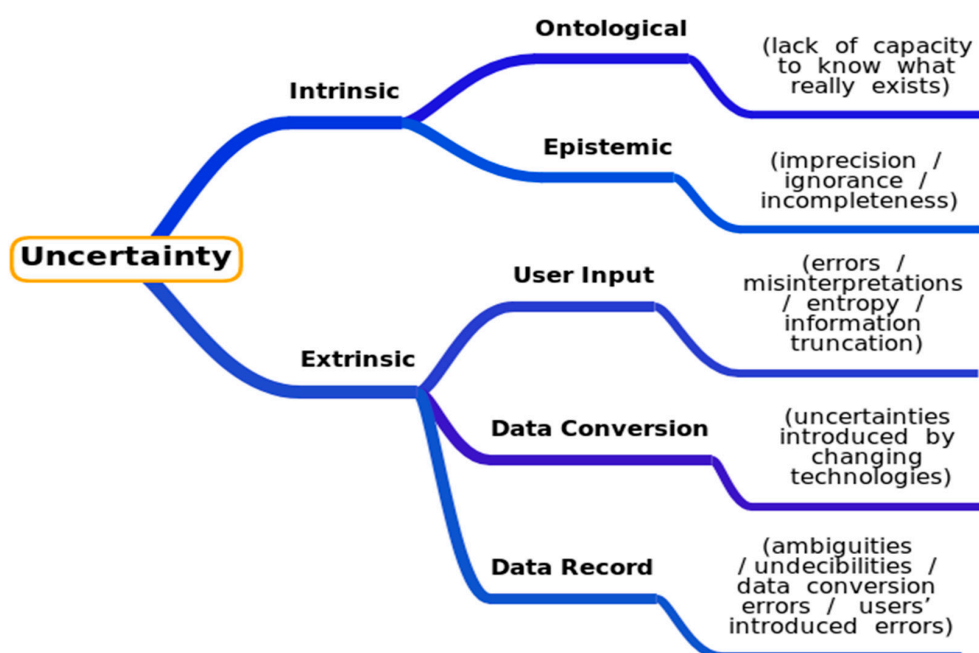

Figure 7. Uncertainty dimensions. Source: the authors.

Our selected focus for this study is relevant in the digital humanities field and to all related disciplines, in that spatial and temporal information are frequent aspects of data collections that makes our paper pertinent also to other fields.

2. Materials and Methods

This section presents an overview of the materials and methods central to the paper. We provide an overview of the exploreAT! project and exploration space as a general context and framework for the scenario described in this study, introducing the indigenous language data collection and discussing

sources of uncertainty, focusing on spatio-temporal dimensions in our linguistic dataset. The methods explain the composition of our specific taxonomy.

2.1. Materials and Data Description

A prominent part of the resource constitutes digitised data collection questionnaires and related answers from paper slips, which initially also pertained to a dictionary project (WBÖ [52], Wörterbuch der bairischen Mundarten in Österreich [Dictionary of Bavarian Dialects in Austria]), intending to capture the German language spoken by the local population, including also the compilation of a linguistic atlas of the local dialect geography [53]. Aside from this dataset, the DBÖ collection further contains digitised information from excerpts of folklore literature, vernacular dictionaries or plant names and mushroom catalogues. The data then follows a lexicographic structuring, consisting of lemmas, definitions, sources, time stamps, location information and a variety of other fields. In addition to this rich texture of linguistic, cultural and societal content captured, there is also detailed information available on persons (authors, collectors, editors) (cf. [54]) and spatio-temporal information (places, regions, GIS locations, etc.) of the collection (cf. [55]).

Table 1 presents a numerical overview of a relevant sub-set (linguistic, location, time-related) of the major entities contained in this non-standard language data collection in two of the main current digital sources (XML/TEI files and records from a relational MySQL database).

Table 1. Numerical overview of selected linguistic, spatial and temporal parameters of the collection. Source: the authors.

	XML/TEI Files		MySQL Records	
Number of Entries	**2,416,499**		**65,839**	
Number of lemmas	**Total**	**Unique**	**Total**	**Unique**
• Original mainlemmas	1,315,183	197,123	98,272	39,853
• Original additional lemma	112,309	42,432	14,842	9874
• Normalized mainlemmas	1,314,494	191,691	-	-
• Normalized additionallemma	112,236	40,069	-	-
• Entries with no lemmas	115,516		0	
Number of Locations	**Total**	**Unique**	**Unique**	
• Bundesland	1,316,889	9	-	
• Großregion	1,296,722	32	415	
• Kleinregion	1,286,463	323	415	
• Gemeinde	1,198,447	1146	3058	
• Ort	1,198,447	1145	19,946	
• Ort (without associated Gemeinde)	395,186	24,788	-	
	With Location	**Without Location**	**With Location**	**Without Location**
Entries vs location	1,712,705	703,794	7333	58,506
Time span of entries	**Oldest**	**Newest**	**Oldest**	**Newest**
	1010	2008	1196	2012

Looking first at the overall number of entries in Table 1, we noted a striking difference in size between the two data collections, with the MySQL database containing considerably fewer entries than the XML/TEI data. As regards different parameters contained in each entry, some, but not all, were relevant for our current context and analysis. Thus, we here concentrate on a sub-set of specific linguistic, spatial and temporal parameters only, and provide an overview of the total number of entries in each of the two data sets, as well as the number of unique (i.e., non-repeated) entries (Table 1, column 1). The specific linguistic parameters concern headwords/lemmas with the following distinction: original mainlemmas (i.e., headwords transcribed in its original form with special characters, e.g., (zu{o}t-aufen), normalized mainlemmas (i.e., headwords transcribed without special characters, e.g., zuotaufen), original additional lemmas (e.g., (Winkel)êe), normalized additional lemmas (i.e., additional headword transcribed without special characters; e.g., Winkelêe) and entries with no lemmas. The specific spatial parameters on the other hand are: Bundesland (county; e.g., Steiermark/St.), Großregion

(big region; e.g., mittelbairische Obersteiermark/mbair.Obst.), Kleinregion (small region; e.g., Erzberger Gegend/Erzbg.Geg.), Gemeinde (municipality; e.g., Radmer), Ort (location; e.g., Radmer) and entries without a given location. The distinction between the different types/sizes of regions were made according to an internally developed system of identifiers for regions, so-called sigles, consisting of a letter–number combination and denoting a hierarchical structure, as can be seen in Figure 8.

```xml
<listPlace xml:id="sigle:3.2q02">
  <place type="Bundesland">
    <placeName>St.</placeName>
    <idno>3</idno>
    <listPlace>
      <place type="Großregion">
        <placeName>mbair.ObSt.</placeName>
        <idno>3.2</idno>
        <listPlace>
          <place type="Kleinregion">
            <placeName>Erzbg.Geg.</placeName>
            <idno>3.2q</idno>
            <listPlace>
              <place type="Gemeinde">
                <placeName>Radmer</placeName>
                <idno/>
                <listPlace>
                  <place type="Ort">
                    <placeName>Radmer</placeName>
                    <idno>3.2q02</idno>
                  </place>
                </listPlace>
```

Figure 8. Example of the nested location codes in an entry of the XML files. Source: the authors.

Comparing next the numbers of total and unique lemmas in the XML/TEI and MySQL datasets, we again noted an overall higher number of both total and unique lemmas in the TEI/XML dataset. Overall the number of unique entries was significantly smaller than the overall number of entries. Whereas normalized lemmas (i.e., lemmas written without special characters) were included in the XML data, this parameter did not occur in the lemmas of the MySQL dataset. As for entries lacking lemmas, a considerably high number of such entries were found in the XML dataset, whereas this was not at all the case in the MySQL data.

Comparing next the total and unique numbers of locations, we first noted again considerably higher numbers of entries in the XML dataset compared with MySQL, but also striking structural differences between these two datasets. Whereas the majority of XML entries contained a hierarchical structure of location information (Bundesland > Großregion > Kleinregion > Gemeinde > Ort), some parameters (Bundesland, Großregion, Kleinregion) were not accessible in a structured way, but had been merged in a single column. A noticeable difference between the datasets, however, emerged, in that a higher number of unique location entries was contained in the MySQL dataset.

Looking finally at entries that were or were not linked to location parameters, again an overall higher number could be observed for the XML dataset. There, a higher number of entries was also linked to location information, whereas the opposite was the case in the MySQL dataset.

As for the time information of entries in both data sets, the oldest and most recent time information was determined, which most typically refers to the publication year of a source in the case of books, or the year of collection in the case of a questionnaire-related entry. In both datasets, we noted rather large time spans, which again highlights the heterogeneity of the material.

While this numerical overview offers an impression of the type of data contained, at the same time, it gives insights into the various levels at which uncertainties in this particular dataset can arise and the extent of heterogeneity. The records were not homogeneous, given differences in the details

from the myriad of sources, and also because of differences in the transformation and conversion processes from the legacy sources for the current ones.

2.2. Data Transformation and Process Description

The DBÖ collection has suffered many transformations since its beginning in the early 20th century (~1913) until today (2019). Originally, the collection was initiated with paper questionnaires and answers noted on individual paper slips. From these analogue forms, the cohort of the now digital data has undergone several stages of digitization and digital transformation through the years (Figures 9 and 10) until reaching its current state; partly in XML/TEI formats [56] and partly as a MySQL database (dbo@ema) (cf. [57]).
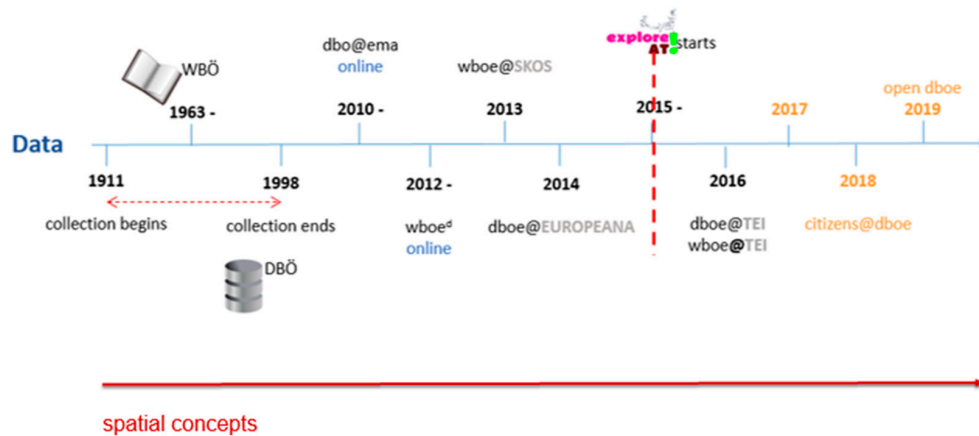

Figure 9. Timeline of the data transformation process relative to the beginning of the exploreAT! project. Image: Amelie Dorn, Eveline Wandl-Vogt, 2018.
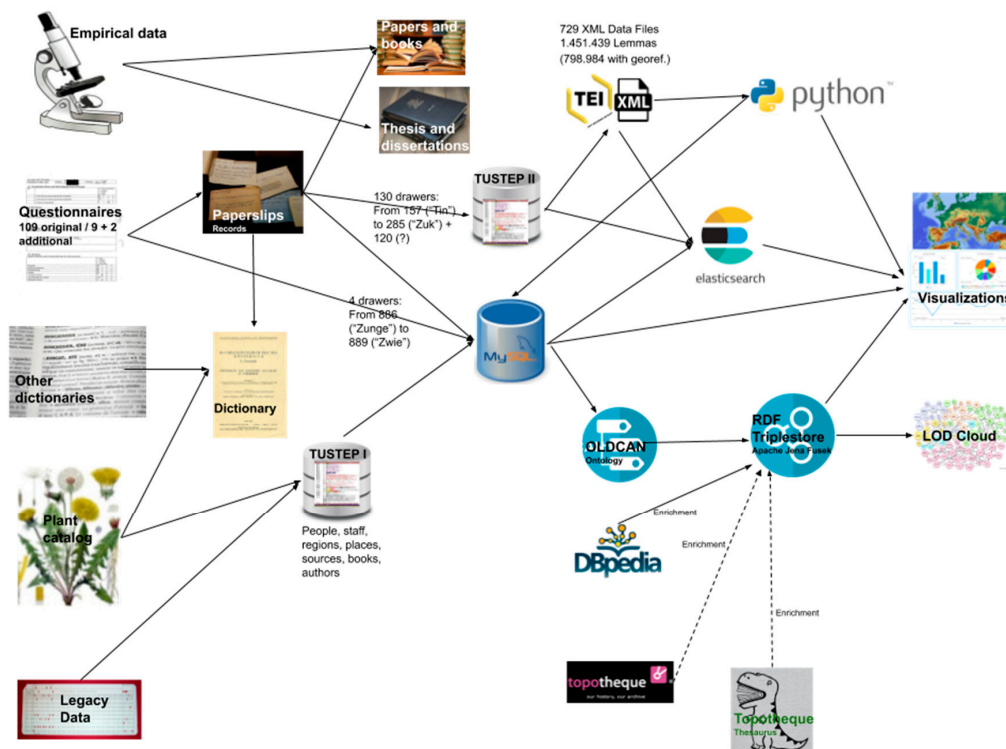

Figure 10. Overview of the data transformation process. Source: the authors.

During the turmoil of World War II, the collection suffered some notable losses, and thus the precise quantitative overview of the original material can only be speculated. In the first stage of digitization (1993–2011), all available information noted on the paper slips (including, for example, headword, meaning, pronunciation, location, date and collector name) was manually entered by several data typists into TUSTEP (Tuebinger System von Textverarbeitungs-Programmen/Tuebingen System of Text Processing tools), resulting in ~2.43 million entries. Additionally, auxiliary databases, persons databases, a literature database, a plant name database and location data were also created in TUSTEP. Towards the end of this first digitization process (2007), part of these TUSTEP data were transferred to a MySQL database as part of the dbo@ema project [57]. For the first time, different separate databases were joint; a geographic visualization interface (maps) and GIS locations were added, and information and data were made publicly accessible and visible on the Internet via a project website [56]. From then, the heterogeneity of the data increased again, with parts of the original data being still available in TUSTEP, another part having been converted to MySQL and additional newly digitized data being directly entered in MySQL. Following the dbo@ema project, the next step marked the transformation process towards the networked, open data realm. The conversion process of the remaining TUSTEP data to XML/TEI format was one of the first endeavours in exploreAT!, starting in 2015. Data entered in MySQL, however, remained unaltered at first. In the course of exploreAT!, the MySQL data and some of the XML/TEI data were converted to RDF and linked to the Linked Open Data (LOD) Cloud (2017–). In addition, the data is currently in the process of being enriched with lexical concepts and linked to DBpedia concepts. Figures 9 and 10 show an overview of the data transformation process.

On a concrete example, we can see in Figure 11 the partial steps of conversion, beginning with a paper slip (1), a TUSTEP entry (2) and an XML/TEI file excerpt (3).

Figure 11. Examples of the data conversion process on the word "Strützel". Source: the authors.

Regarding the MySQL database, data was added straight from the records, as depicted in Figure 10. An overview of the database schema can be seen in Figure 12.
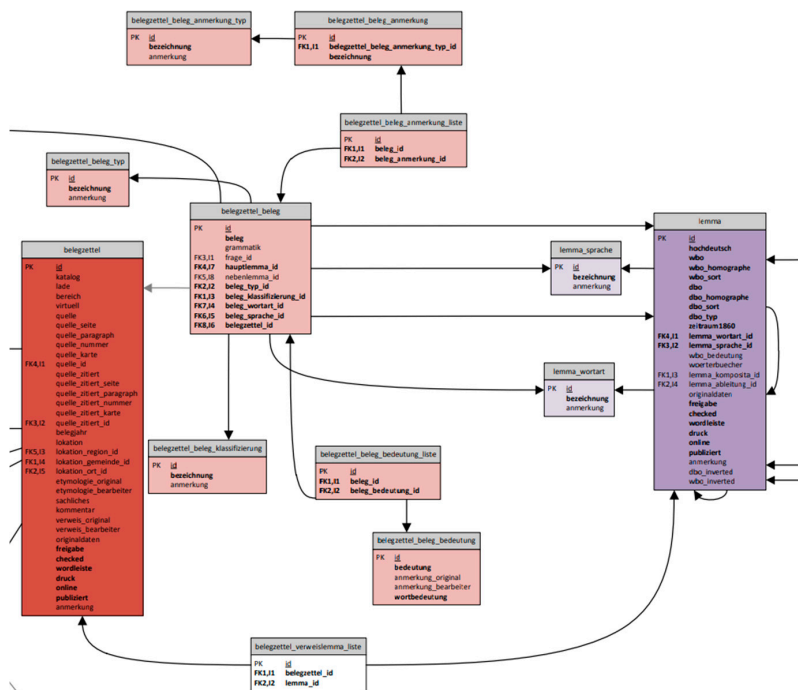

Figure 12. Except of the MySQL database schema. Source: the authors.

2.3. Spatial and Temporal Dimensions in the DBÖ: A Review

Since the very beginning of the data collection and in the course of the DBÖ data transformation, spatial and geographical concepts contained in dbo@ema have been a fundamental aspect of the data analysis and transformation [58–62].

With the existence of recent novel digital tools and methods, new ways of exploring these spatial dimensions were enabled, and they have thus served as the data basis for various studies, collaborations and theses in the context of data visualisation [16] (Figure 13), data modelling [60,63] or linked data representations [55,64] (Figure 14).
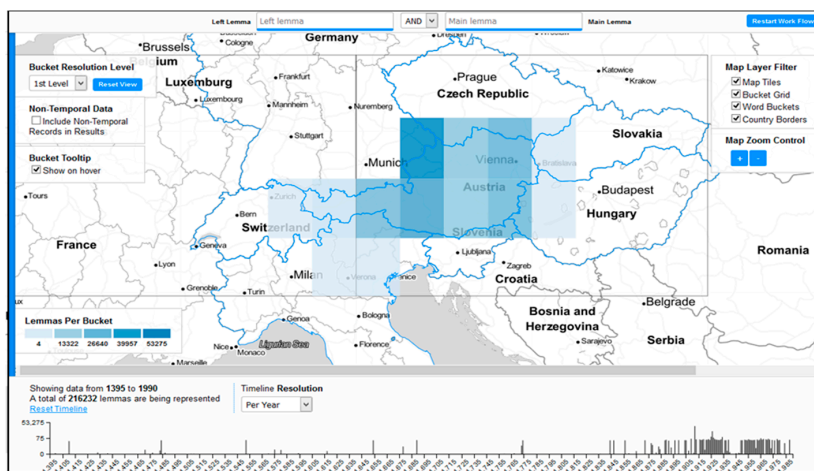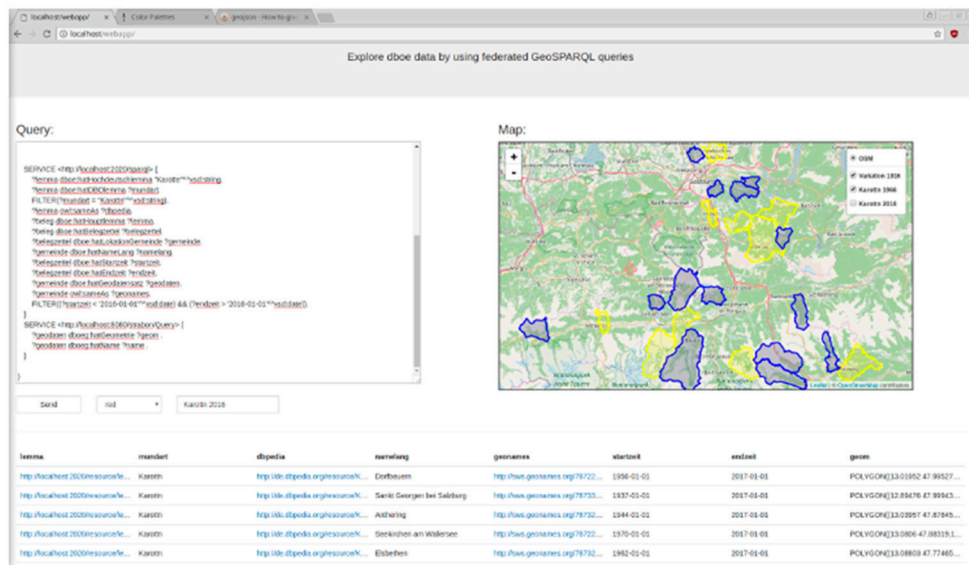

Figure 13. Web-browser based visual analysis of TEI-encoded data. Collection explorer (Benito et al., 2016, exploreAT!) Source: https://exploreat.acdh-dev.oeaw.ac.at/exploreAT-collectionexplorer/ [65]; https://github.com/acdh-oeaw/exploreAT-collectionexplorer [66].

Figure 14. Prototypical Linked Open Data (LOD) modelling of spatial data (Johannes Scholz, Emanuel Hrastnig, Eveline Wandl-Vogt 2017, Graz University of Technology).

The association of spatial and linguistic features also allowed for the analysis of the areas which yielded more concepts, as shown in the heatmap of Figure 15.
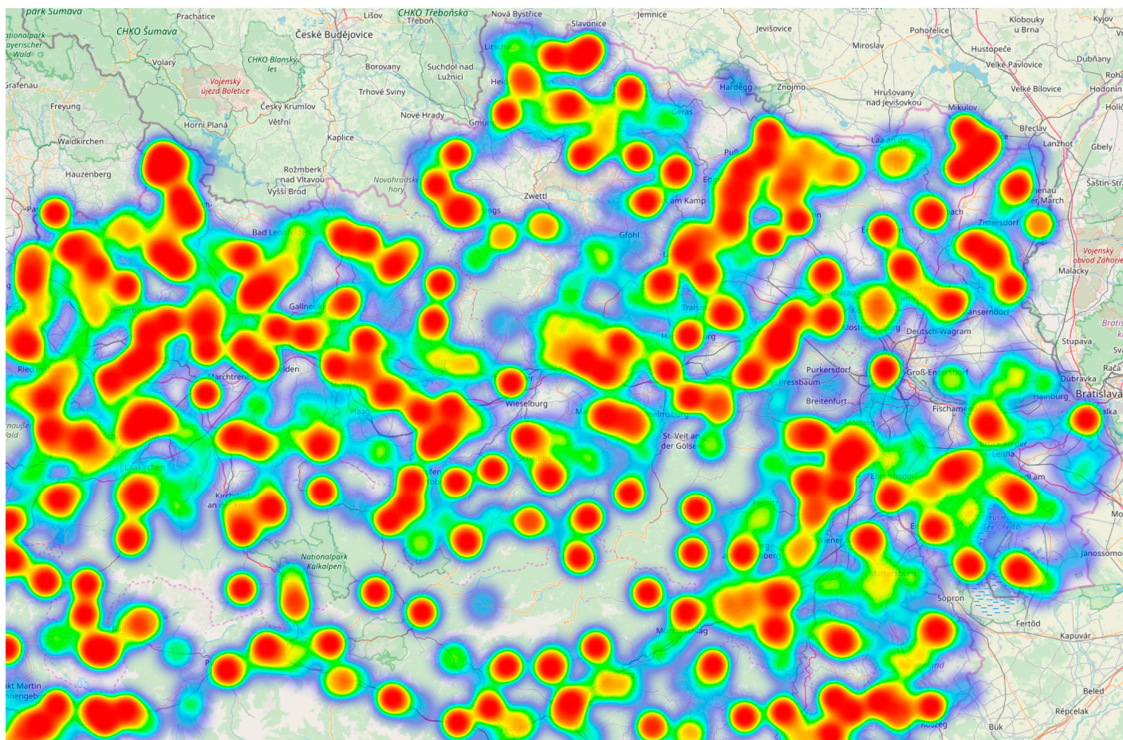
Figure 15. Part of the heatmap of the spatial concentration of collected concepts. Source: the authors.

2.4. Methods

The composition of our taxonomy of uncertainties was preceded by a review of carefully-selected existing models, aiming at covering the spatial, temporal and linguistic aspects across different fields. In spite of the abundance of taxonomies dealing with spatial and temporal aspects, we found almost no coverage of the linguistic aspects. On the basis of this review, entities for our taxonomy were selected

against the background of our dataset that were deemed most appropriate in capturing the underlying uncertainties. We tried to the maximum extent to reuse the spatial and temporal categories, presenting them in a simple but unified and coherent structure. Additionally, we introduced the linguistic aspects in a way in which the whole set of dimensions proposed could fit into an all-encompassing instrument that would be able to be applied to our ad hoc collection. Where no match was found, we established our own categories. The devised taxonomy based on uncertainties in the DBÖ dataset is discussed in further detail in Section 3.

3. Results

As commonly occurs in long data transformation and conversion processes, uncertainties have both been remedied as well as reintroduced through time, such as in the differences in DB schemas due to assignment of fields during database conversion and imperfect matches between lexical and LOD concepts in the enrichment process. Most of these uncertainties are common in a plethora of long-term, data-intensive projects. However, some are very particular to this collection.

To characterize the uncertainty sources that were found, a specific taxonomy was developed, which is now introduced.

3.1. Uncertainty in DBÖ—A Specific Taxonomy

In Section 1.2, a non-exhaustive review on attempts to offer taxonomies of uncertainties was presented. In this section, we present a taxonomy of our own, based on specific categories relevant to our dataset. Table 2 below, based on Figure 7 (Section 1.3), depicts the classes and sources of uncertainties found in our collection regarding spatial, temporal and linguistic dimensions, along with some examples of its sources.

While the *intrinsic/ontological* dimension deals with the problem of indeterminacy that arises from the limits on our capacity to know what exists in reality, the *intrinsic/epistemic* dimension deals with limits in the human capacity to measure, observe, record or determine precisely what is to be known. Conversely, uncertainties with origins in the extrinsic dimensions would all be avoidable, given a certain effort and control, but these are often unfeasible. *User input* is related to the role humans have in the transformation of data and information substrates; *data conversion* is related to the technological pitfalls when converting automatically from format to format and—last, but not least—*data record* is inherent to ambiguity in the data per se, and not directly related to the consequences of transformations and users' actions.

Table 2. Classes and sources of spatial, temporal and linguistic dimensions of uncertainties.

	Uncertainties				
	Intrinsic		**Extrinsic**		
	Ontological	**Epistemic**	**User Input**	**Data Conversion**	**Data Record**
Dimensions	(lack of capacity to know what really exists)	(imprecision/ignorance/ incompleteness)	(errors/misinterpretations/ entropy/information truncation)	(uncertainties introduced by changing technologies)	(ambiguities/undecibilities/ data conversion errors/users' introduced errors)
Spatial	- Places that ceased to exist	- Unknown places - Exact place vs approximate/region	- Typos - Abbreviations - Changing transcription guidelines - Assumptions on certain orthographies - Lack of precision on creating data records - Guessing - Prejudice and biases	- Language codification errors - Errors in the conversion of formats and databases - Heterogeneity of data sources	- Homographs (places) - Difference in details among records
Temporal	- Events with fiat limits [67] - Punctual events vs processes over time	- Events for which it is impossible to know the exact start, beginning and/or duration			- Differences in data formats - Differences in data standards (e.g., Roman numbers) - Differences in details among records
Linguistic	- Dialects that ceased to exist	- Lack of standard orthographies for a certain lemma			- Homographs (lemmas) - Difference in details among records - Different standards for representing (e.g., IPA vs. TUSTEP)

Source: the authors. TUSTEP = Tuebinger System von Textverarbeitungs-Programmen/Tuebingen System of Text Processing tools.

3.2. Uncertainty in DBÖ—Examples

Having first established our own taxonomy of uncertainties, we now present concrete examples for the most relevant cases from our data to exemplify and illustrate each category. The concrete examples in the following subsections correspond to the dimensions outlined in Table 2 above.

3.2.1. Spatial/Intrinsic/Ontological

A typical example of this kind of source of uncertainty is, for the DBÖ spatial data, a somewhat common historical region in the far north west of Bohemia in the Czech Republic at the border with Germany, called "Egerland". It was characterised by the German-speaking population until 1945 (cf. [68]), but now the names of the places in this region have all changed. Other common cases for this collection are regions that used to belong to Austria but are now in a different country, such as "Stilfs-Stelvio, Vinschgau" that are now "Val Venosta, BZ, TAA, 39,029 in Italy", or "Groß Olkowitz" that are now "Oleksovice" in the Czech Republic. Some of these examples are presented in Figure 16, together with their original, corresponding paper slips.

(a)	(b)
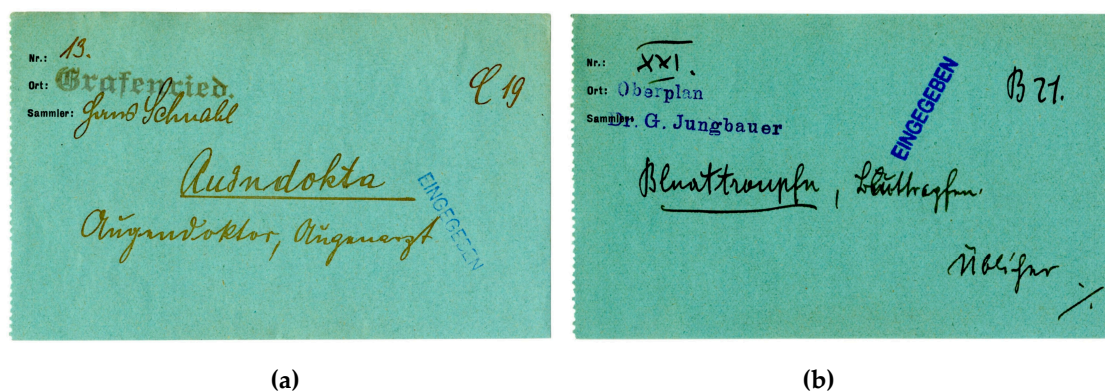
Figure 16. Examples of spatial/intrinsic/ontological uncertainties on the original paperslips: (**a**) Grafenried (= Lučina) in the Bohemian Forest has completely vanished as a geographical name or political region; (**b**) "Oberplan" (= Horní Planá) is an exemple of a region that used to belong to the Austrian empire, but now is in the Czech Republic. Source: the authors.

3.2.2. Spatial/Intrinsic/Epistemic

In the epistemic side, spatial examples are many, given the abundance of sources for imprecision, ignorance or incompleteness. Such examples can be differences in detail (a point/coordinate or a region/polygon), as in "Südmähren" (region) and "Obweg, Landeck, Tirol, 6571, Österreich" (a very well defined point).

3.2.3. Spatial/Extrinsic/User Input

Euphemistically speaking, there is no shortage of cases where the collectors' and data typists' "creativity" introduces uncertainties when putting data into systems. The lack of standards, guidelines or even their mutable dynamic over time, along with errors that arise from biases or different backgrounds, have the power to produce records that range from "ambiguous" to "data chimera". As an illustration, we can mention typos and abbreviations (e.g., "Kapellen BöW ^#^# ON/Gm.?", or "o.Ang.", that mix upper and lowercases, special characters and different types of information, such as the collector initials) in the same data field. There are, however, also other interesting cases, where biases and mistaken interpretations can lead to errors in the chain. Figure 17 shows an entry where the data typist read wrong twice, as explained:
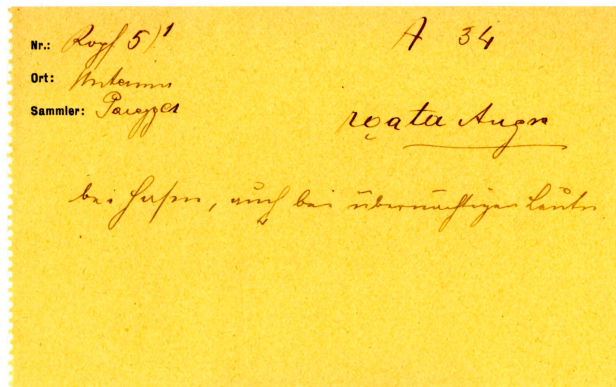
1. The correct location is "Unterinn" and not "Matrei". The error caused a wrong assignment, thus the standardised source in TUSTEP and later in the TEI/XML file (see below) is not correct. The

data typist was probably mistaken, as the collectors in Matrei and in Unterinn have similar names (Egger/Paregger).

2. There is also an error in the meaning—it is not "übermächtigen Leuten" (overpowering people), but "übernächtigen Leuten" (people that are tired out). As it almost sounds true and is not all that far-fetched, it is easy to overlook. These interpretations have introduced both spatial and linguistic errors in the collection.

3.2.4. Spatial/Extrinsic/Data Conversion

Data conversion and enrichment can be a source of augmenting information conveyed by the systems, but can introduce errors as well. In the specific case of spatial data, wrongly-matched geolocations have occurred often with our dataset. We have mainly used Geopy [69] (a Python package) for geolocation, as it offers access to many free and paid geocoders. The default and free option, OSM Nominatim [70], is quite good, but not as reliable as GoogleV3 or Bing APIs, which are paid. One example of mismatch is the place "Weikersdorf" that was identified as being "Vikýřovice, okres Šumperk, Olomoucký kraj, Střední Morava, 78813, Česko" by OSM geolocation, whilst Google would have pointed the correct location "Weikersdorf am Steinfelde", in Austria.
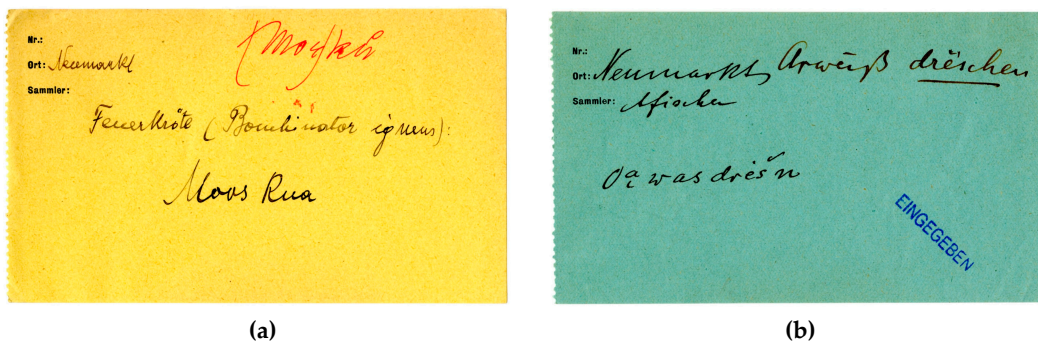
(a)

(b)

Figure 17. *Cont.*

```
<entry xml:id="r664_qdb-d1e124701" xml:lang="bar">
  <form type="hauptlemma">
    <orth>rôt</orth>
  </form>
  <gramGrp>
    <pos>Adj</pos>
  </gramGrp>
  <note type="anmerkung" resp="0" corresp="this:BD">bei Hasen, auch bei übermächtigen Leuten</note>
  <cit type="kontext" n="1">
    <quote>r;oata [P,pl,fl] Augn</quote>
    <quote resp="#JB" change="01">r;oata Augn</quote>
    <def xml:lang="de">--</def>
    <ref type="fragebogenNummer">13A34: rötliche Augen (bes. bei Tieren); Adj./Volksgl.</ref>
  </cit>
  <ref type="archiv">HK 664, r6640221.kro^#108</ref>
  <ref type="quelle">Matrei Paregger</ref>
  <ref type="quelleBearbeitet">{1C.2h02} mittl.NTir.Wippt.:NTir.Wippt.:Wippt.:Sillgeb.:mNTir.</ref>
  <ref type="bibl">
    <bibl>FbB.EGGER· (19xx) [SFb.]</bibl>
  </ref>
  <usg type="geo">
    <placeName type="orig">Matrei NTir.</placeName>
    <listPlace ref="sigle:1C.2h02">
<place type="Bundesland">
  <placeName>NTir.</placeName>
  <idno>1C</idno>
  <listPlace>
    <place type="Großregion">
      <placeName>mNTir.</placeName>
      <idno>1C.2</idno>
      <listPlace>
        <place type="Kleinregion">
          <placeName>mittl.NTir.Wippt.</placeName>
          <idno>1C.2h</idno>
          <listPlace>
            <place type="Gemeinde">
              <placeName>Matrei am Brenner</placeName>
              <idno/>
              <listPlace>
                <place type="Ort">
                  <placeName>Matrei am Brenner</placeName>
                  <idno>1C.2h02</idno>
                </place>
              </listPlace>
```

(c)

Figure 17. A chain of errors in user input: (**a**) Original record with the correct location "Unterinn"; (**b**) Location was changed to "Matrei" in TUSTEP; (**c**) TEI/XML file reproducing the error introduced previously. Source: the authors.

3.2.5. Spatial/Extrinsic/Data Record

Even if the process of inputting data is made in the correct way, it may happen that some information has ambiguities that may lead to uncertainties. In the case of places, we have as its main sources homographs and differences in details among records. As an example, we can present the case of "Neumarkt" where the location is not exactly specified and the information recorded can be interpreted as seven different places. We can only differentiate, then, referring to additional information, such as the colour of the paper slip, the name of the collector and/or the handwriting (for identification of the collector). Figure 18 shows two examples of paper slips for "Neumarkt".
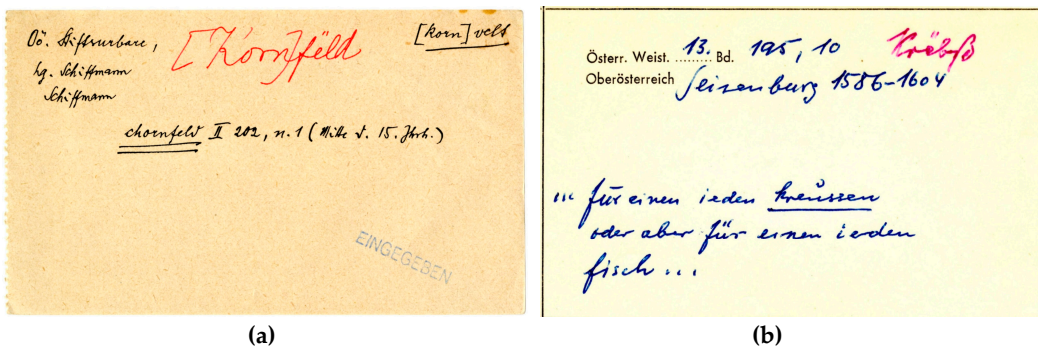
Figure 18. Two examples of paper slips referring to different homograph places: (**a**) Neumarkt (referring to a location in South Tyrol, now Alto Adige, Italy); (**b**) Neumarkt referring to Neumark (Czech.: Všeruby) in the Bohemian Forest). Source: the authors.

3.2.6. Temporal/Intrinsic/Ontological and Epistemic

Ontologically speaking, there are many aspects of time events that are uncertain. Examples are events with debatable fiat limits (e.g., the beginning and end of World Wars), and also the difference over punctual events (e.g., the day a person has answered a question) vs processes over time (e.g., when a word started to be used). Sometimes it may be the case that the event is ontologically well-defined, but there is no way to get access to the information about it (epistemic uncertainty), which can lead to imprecisions in the records. Both cases can give rise to differences in detail, such as:

- Reference to day (e.g., "11. 8. 1735"); month (e.g., "Juli 1927!"); year (e.g., "1465"; "a. 1630") or century (e.g., "15. Jh.").
- Imprecise reference (e.g., "Mitte d. 15. Jh.", "1608; 1651 [?]", "vor 2. Weltkrieg", "Biedermeierzeit", "Vorkriegszeit" or "nach 1. Weltkrieg") (Figure 19).

Figure 19. Examples of paper slips with imprecise time references: (**a**) Imprecise reference to "Mitte d. 15. Jhrh."; (**b**) Imprecise reference to "1586–1604". Source: the authors.

3.2.7. Temporal/Extrinsic/User Input

It is often difficult to tell whether a time record is ontologically/epistemic uncertain or if it is the case of erroneous user inputs. As an example of the latter, we have references to two (or more) different periods (e.g., "1560 (Kop. 18. Jh.)"). Other cases can be regarded as pure naivety of the data typist about the diachronic interpretations of statements (e.g., "heute noch"; "Allerheiligen" or "vor Jahrzehnten"). These were ontologically and epistemically correct in the time they were created, but almost useless in terms of information.

3.2.8. Temporal/Extrinsic/Data Conversion and Data Record

Data conversion can play a role in introducing uncertainties in time references. One of the most common cases are the differences in date formats [71], but can also be the case of idiosyncrasies in user

input that, in spite of not being errors, can also cause problems in the conversion, as the use of Roman notation for numbers, or even a mix of Arabic and Roman (e.g., "O 1670 8.VII"). These will introduce errors in the data records.

When trying to extract information from data, we have faced the need to deal with vague records, such as "17. Jh." and "16. Jh."—the two more common references of the collection, with an example in Figure 20.
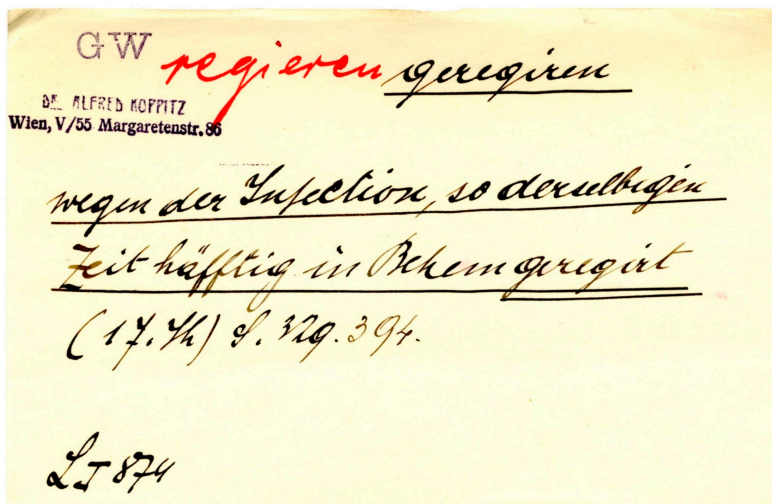

Figure 20. Vague reference to the "17. Jh.". Source: the authors.

For characterizing the collection, we decided to assign the year in the middle of the century for each imprecise reference such as these, which led to some mid-century peaks in the temporal depiction of dates, as can be seen in Figure 21. Whether a random assignment would convey more information is a matter of interpretation.
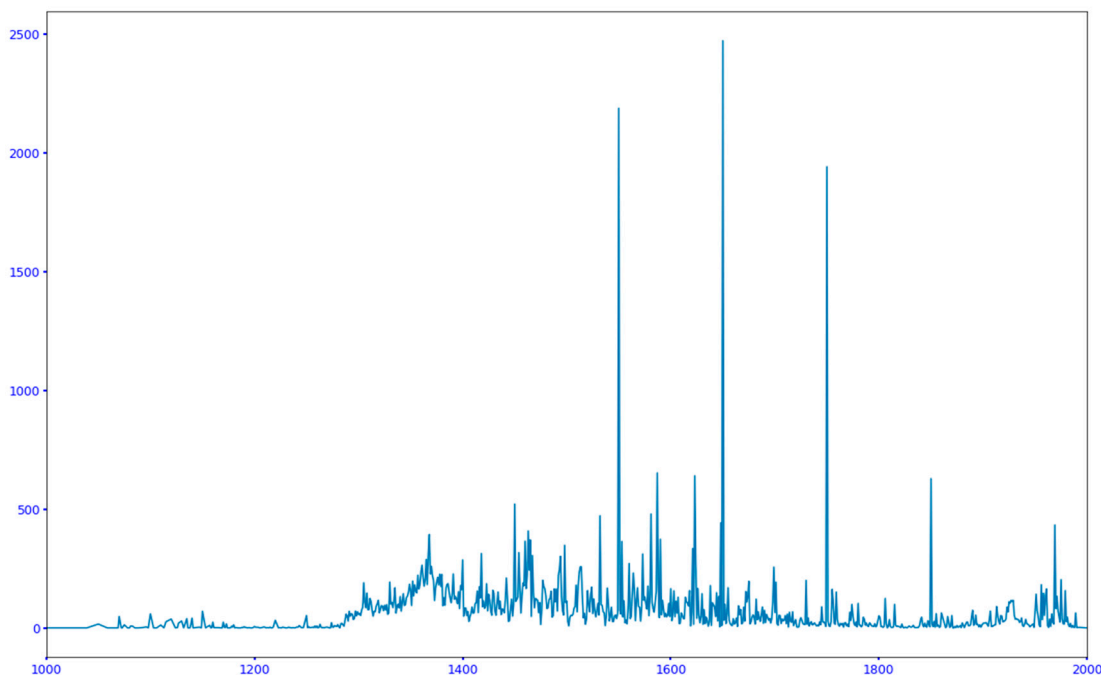
Figure 21. Temporal distribution of the collection records. Source: the authors.

3.2.9. Linguistic/Intrinsic/Ontological

Ontological uncertainties in the linguistic aspects are sometimes hard to tell apart from epistemic ones, but the phenomenon of dialects ceasing being used is undoubtedly an ontological problem, as a dialect exists as soon as there are at least two individuals using it. It brings uncertainties when the typists or researchers cannot distinguish them from other common errors and typos. They can also be interpreted as different words (homographs or semi-homographs) with different meanings. Figure 22 shows an example of a paperslip that was rejected for the dictionary because of the inherent uncertainty on the interpretation. The meaning is clear, but the word/pronunciation is unusual, so one cannot reliably deduce a lemma (Espetze?) for it or assign it to an already existing lemma. This is another argument for the importance of current research being done in endangered languages [72,73].
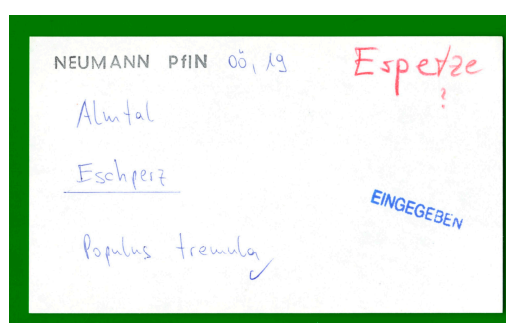
Figure 22. An example of an ontological uncertainty. Source: the authors.

3.2.10. Linguistic/Intrinsic/Epistemic

Epistemic uncertainties are common in linguistic records, mainly because of the process of collecting speech samples and choosing the most adequate spellings for the words unknown to the collectors. Even for those that they know, they may lack standard orthographies for certain vocables.

3.2.11. Linguistic/Extrinsic/User Input

Besides the type of errors in inputting information, as illustrated in Figure 17, Section 3.2.3, another emblematic example of uncertainty caused by user input is the observance of different standards for the phonetic representation of words. For the digitization and during the transformation processes of the collection, both the International Phonetic Alphabet (IPA) [74] and (coded) Teuthonista have been used (cf. [75]). Unfortunately, the documents referring to the latter are now available only through the Wayback Machine [76]. Figure 23 shows an example of entry in the TEI/XML database that shows both coded Teuthonista (referred as "tustep" notation) and IPA formats for representing the lemma "Alté".

```xml
- <body>
    - <entry xml:lang="bar" xml:id="e206_qdb-d1e2">
        - <form type="hauptlemma">
              <orth type="orig">Alt;é</orth>
              <orth type="normalized">Alté</orth>
          </form>
        - <gramGrp>
              <pos>Subst</pos>
          </gramGrp>
        - <form type="lautung" n="1">
              <pron notation="tustep">)aöt!e -E</pron>
              <pron notation="ipa" change="01" resp="#JB">)aöt!e -E</pron>
            - <gramGrp>
                  <gram>[f]</gram>
              </gramGrp>
          </form>
```

Figure 23. Teuthonista and IPA formats for representing the lemma "Alté". Source: the authors.

As the knowledge of the collectors about the standards were heterogeneous, we have many errors related to the correct representation of words, which led to uncertainties.

3.2.12. Linguistic/Extrinsic/Data Conversion

Perhaps the most common problem of all regarding data conversion from legacy systems is the difference in character encoding. Before Unicode Transformation Format encoding UTF-8 [77] and Unicode standards [78] being adopted as the preferred choice to represent characters, different systems used one of the many available character encodings [79], leading to undesirable special characters when the conversion is unaware of the previous formats. Figure 24 shows an example of such conversion errors of the lemmas in the MySQL database. Words that present the squared character are a result of a character encodings from the legacy system that could not be interpreted in the new one.


```
+?+ (Alben) © 2008-2080 jost nickel
+?+ (Alm)rausch © 2008-2080 jost nickel
+?+ (Alm)rôse © 2008-2080 jost nickel
+?+ (Átlas)pere © 2008-2080 jost nickel
+?+ (Átles)pere © 2008-2080 jost nickel
+?+ (Donner-zalten)pl□e © 2008-2080 jost nickel
+?+ (Dörner)- © 2008-2080 jost nickel
+?+ (Dörner)k□lein © 2008-2080 jost nickel
+?+ (Famílle)pl□me © 2008-2080 jost nickel
+?+ (Faul-per[en])pāum © 2008-2080 jost nickel
+?+ (Fisch-per)pāum © 2008-2080 jost nickel
+?+ (Frucht)edibeer © 2008-2080 jost nickel
+?+ (Fr□)sendach © 2008-2080 jost nickel
+?+ (Füg. [ge]sëlhtes- Fleisch-und Sch.)wurzel © ...
+?+ (Galmei)pl□me © 2008-2080 jost nickel
+?+ (Glän)pere © 2008-2080 jost nickel
+?+ (Hasel-h□mer)pere © 2008-2080 jost nickel
+?+ (Hexen)drëck © 2008-2080 jost nickel
+?+ (Hexen)k□se © 2008-2080 jost nickel
+?+ (Hexen)pësen © 2008-2080 jost nickel
+?+ (Hexen)polster © 2008-2080 jost nickel
+?+ (Hunde)pere © 2008-2080 jost nickel
+?+ (H□mer)pere © 2008-2080 jost nickel
+?+ (Kemerlein)pl□me © 2008-2080 jost nickel
+?+ (Kuopper)ēiche © 2008-2080 jost nickel
```

Figure 24. Errors from conversions of different character encodings in the MySQL database. Source: the authors.

This kind of error has to be detected during the conversion phase, otherwise they will be perpetuated and may make some data unusable.

3.2.13. Linguistic/Extrinsic/Data Record

Data records are the "compulsory depositaries" of all the errors in the data creation and transformation processes, and this includes ontological, epistemic, user input and data conversion errors. However, even when the creation and conversion are impeccable, uncertainties may arise from the records per se. This is the case, in linguistic regards, of homographs and polysemic words. In the scope of our collection, formed by harvesting multiple sources, differences in details among records also plays its role in uncertainty creation. Figure 25 shows an example of different levels of detail about lemmas originated from TEI-XML entries.

hauptlemma_normalized	hauptlemma_orig	hauptlemma_unique	nebenlemma_normalized	nebenlemma_orig	nebenlemma_unique
		dienstbar			
		dienstbar			
		dienstbar			
		dienstbar			
		dienstlich			
undienstbar	(un)dienstbar				
Dienstbarkeit	Dienstbark-eit				
Dienstbarkeit	Dienstbark-eit				
Dienstbarkeit	Dienstbark-eit		Dienstbarigkeit	Dienstbarigk-eit	
Dienstbarkeit	Dienstbark-eit				
Dienstbarkeit	Dienstbark-eit				
Dienstbarkeit	Dienstbark-eit				
Dienstbarkeit	Dienstbark-eit				
Përgpaudienstbarkeit	(Përg—pau)dienstbark-eit				
		dienstbärig			
		dienstbärig			
		dienstbärig			
		dienstbar			
		dienstbärig			
		dienstbärig			
		dienstbärlich			
		dienstbärlich			
		dienstbärlich			
		dienstbärlich			
		dienstbärlich			

Figure 25. Differences in details about lemmas from the TEI-XML records. Source: the authors.

4. Discussion

Our analysis and exemplification of spatial and temporal aspects of uncertainty in the DBÖ collection has offered an abundance of insights on contributing factors and sources, highlighting also the sheer extent of heterogeneity in this legacy dataset. Although previously-established taxonomies for classifying uncertainties were consulted, it was only with the devising of our own taxonomy that the various dimensions and categories in our data could be fully captured and the necessity of our research purpose confirmed. What has become apparent, however, is the continuous course of remedying, but at the same time introducing new types of uncertainties in the data transformation process, despite the availability and use of guidelines, standards and manual corrections. In particular, as far as spatial dimensions are concerned, the constantly evolving dynamics and possible changes in borders, names of places, regions, territories or other geopolitical changes in the real world may affect and impact not only historical, but also current datasets.

We understand that much of what was illustrated in this paper is common and may happen to all collections formed through a reasonable amount of time, as is the case in heritage and historical documents. With regards to similar studies on spatial or temporal uncertainties in non-standard lexical collections, to our knowledge to date, no such taxonomy exists. As a result, our taxonomy can serve as an example for similar datasets and a starting point for related developments. With regards to similar efforts in the digital humanities field, our taxonomy adds to the pool of existing uncertainty taxonomies (e.g., ProvideDH). It draws its higher classes on wider concepts of uncertainties (i.e., intrinsic, extrinsic), but—unlike other taxonomies—devotes its specific subclasses to the fine-grained categories unique to digitised lexicographic collections (e.g., data conversion, data record, etc.).

A first (and perhaps the easiest) step in remedying the types of uncertainties identified and classified would be to treat them as simple record errors. "Being certain about our uncertainties", however, is not the same as, nor does it even come close, to remedying them. Because of the fact that the collection comprises—as has been shown—many different databases and formats, it would take time, and sometimes a huge amount of manual work to correct each identified instance of the categories shown. Some of them would be more easily identified or corrected, such as spatial ones, if we could use a geographical gold standard for addressing ambiguities [60]. The temporal uncertainties, on the other hand, could be analyzed under certain rules and decisions on how to deal with incomplete information, as is the case of the vague references to centuries. Linguistic uncertainties are harder to fix, as the collection itself contains a lot of indigenous and dialect words, making the comparison and mapping with well-defined instruments, such as the available dictionaries, rather unsatisfactory and incomplete.

5. Conclusions

One piece of learning we have taken from this process, nevertheless, concerns the precautions and best practices that may prevent (or at least diminish) such cases from appearing in future collections and databases that are being constantly built in the scope of cultural and linguistic aspects under our particular digital humanities umbrella. Although many processes of data gathering, inputting and conversion are inherently ad hoc, the possible abstractions and generalizations may serve as a warning to the prospective tasks of maintaining huge textual, imagetic and multimedia collections, as those currently being developed. The majority of computer database collections have been formed in the last few decades, and cases of collections formed throughout long periods—in our case, a whole century—are key to understanding the long-term consequences of each and every decision regarding data maintenance. One could say that "creating uncertainty is one of our main certainties", but keeping it, then, as its lowest and acceptable level is an inextricable goal of data humanists.

The distinction to be made by DH researchers should be when uncertainty is a source of insight, allowing better representation of the complexities of reality in the scope of post normal science [80], and when uncertainty is solely a by-product of bad decisions and/or lack of standards for procedures in dealing with data sources, processes and transformation. We identify our intrinsic dimensions as somewhat related to the former, and extrinsic dimensions definitely related to the latter. Interdisciplinary teams that aggregate different backgrounds are also key in fulfilling tasks and developing the skills expected in this fascinating data-driven landscape that is being unveiled for humanities research.

References

1. Nowotny, H.; Scott, P.B.; Gibbons, M.T. *Re-Thinking Science: Knowledge and the Public in An Age of Uncertainty*; John Wiley & Sons: New York, NY, USA, 2013.
2. Nowotny, H. *The Cunning of Uncertainty*; Polity Press: Cambridge, MA, USA, 2015.
3. Nowotny, H. The radical openness of science and innovation. Why uncertainty is inherent in the openness towards the future. *EMBO Rep.* **2015**, *16*, 1601–1604. [CrossRef] [PubMed]
4. Dow, S.C. Uncertainty about Uncertainty. In *Foundations for New Economic Thinking*; Palgrave Macmillan: London, UK, 2012; pp. 72–82.
5. Downey, H.K.; Slocum, J.W. Uncertainty: Measures, Research, and Sources of Variation. *Acad. Manag. J.* **1975**, *18*, 562–578.
6. Taylor, J. *Introduction to Error Analysis, the Study of Uncertainties in Physical Measurements*, 2nd ed.; University Science Books: New York, NY, USA, 1997.
7. Kuhlthau, C.C. A principle of uncertainty for information seeking. *J. Doc.* **1993**, *49*, 339–355. [CrossRef]
8. Shackle, G.L.S. *Uncertainty in Economics and Other Reflections*; Cambridge University Press: Cambridge, UK, 2010.

9. Weiss, C. Expressing scientific uncertainty. *Law Probab. Risk* **2003**, *2*, 25–46. [CrossRef]

10. Stigler, S.M. *The History of Statistics: The Measurement of Uncertainty before 1900*; Harvard University Press: Cambridge, MA, USA, 1986.

11. Bammer, G.; Smithson, M. (Eds.) *Uncertainty and Risk. Multidisciplinary Perspectives*; Earthscan: London, UK, 2008.

12. Stirling, A. Risk, Uncertainty and Precaution: Some Instrumental Implications from the Social Sciences. In *Negotiating Environmental Change: New Perspectives from the Social Sciences*; Berkhout, F., Leach, M., Scoones, I., Eds.; Edward Elgar Publishing: Cheltenham, UK, 2003; pp. 33–76.

13. Austrian Academy of Sciences. ACDH. Methods and Innovation. Core Unit 4. Available online: https://www.oeaw.ac.at/acdh/about/core-units/core-unit-4 (accessed on 28 June 2019).

14. Open Innovation. ÖAW—Exploration Space. Available online: http://openinnovation.gv.at/portfolio/oeaw-exploration-space/ (accessed on 28 June 2019).

15. Wandl-Vogt, E.; Kieslinger, B.; O'Connor, A.; Theron, R. exploreAT! Perspektiven einer Transformation am Beispiel eines lexikographischen Jahrhundertprojekts. In *DHd2015. Von Daten zu Erkenntnissen. 23. bis 27. Februar 2015, Graz. Book of Abstracts*; Austrian Centre for Digital Humanities: Wien, Austrian.

16. Abgaz, Y.; Dorn, A.; Piringer, B.; Wandl-Vogt, E.; Way, A. Semantic Modelling and Publishing of Traditional Data Collection Questionnaires and Answers. *Information* **2018**, *9*, 297. [CrossRef]

17. Benito, A.; Losada, A.G.; Therón, R.; Dorn, A.; Seltmann, M.; Wandl-Vogt, E. A Spatio-temporal Visual Analysis Tool for Historical Dictionaries. In Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality, Salamanca, Spain, 2–4 October 2016.

18. Benito, A.; Losada, A.G.; Therón, R.; Dorn, A.; Wandl-Vogt, E. Creating Meaningful Narratives in Collections of Historical Lexical Data. *GI Forum* **2018**, *6*, 50–57. [CrossRef]

19. Dorn, A.; Wandl-Vogt, E.; Abgaz, Y.; Benito Santos, A.; Therón, R. Unlocking Cultural Conceptualisation in Indigenous Language Resources: Collaborative Computing Methodologies. In Proceedings of the LREC 2018 Workshop "CCURL2018—Sustainable Knowledge Diversity in the Digital Age", Miyazaki, Japan, 12 May 2018.

20. *Open Innovation Strategy for Austria*; Federal Ministry of Science, Research and Economy (bmwfw) and Federal Ministry for Transport, Innovation and Technology (bmvit): Vienna, Austria, 2015.

21. Wandl-Vogt, E. Wie man ein Jahrhundertprojekt zeitgemäß hält: Datenbankgestützte Dialektlexikografie am Institut für Österreichische Dialekt- und Namenlexika (I DINAMLEX) (mit 10 Abbildungen). In *Bausteine zur Wissenschaftsgeschichte von Dialektologie/Germanistischer Sprachwissenschaft im 19. und 20. Jahrhundert. Beiträge zum 2. Kongress der Internationalen Gesellschaft für Dialektologie des Deutschen*; Ernst, P., Ed.; Praesens Verlag: Wien, Austria, 2008; pp. 93–112.

22. Welcome to the PROVIDEDH CHIST-ERA Project Site. Available online: https://providedh.eu (accessed on 28 June 2019).

23. Theron, R.; Wandl-Vogt, E.; Edmond, J.; Mazurek, C. Progressive Visual Decision Making for Digital Humanities (PROVIDEDH): Conceptual outline and first results. In Proceedings of the European Association for Digital Humanities Conference (EADH 2018), Galway, Ireland, 7–9 December 2018.

24. Fisher, P.; Comber, A.; Wadsworth, R. Approaches to Uncertainty in Spatial Data. In *Qualité de l'Information Géographique (Traité IGAT)*; Devillers, R., Jeansoulin, R., Eds.; Hermes/Lavoisier: Paris, France, 2005; pp. 9–64.

25. Therón, R.; Losada, A.G.; Benito, A.; Santamaría, R. Toward supporting decision-making under uncertainty in digital humanities with progressive visualization. In Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality, Salamanca, Spain, 24–26 October 2018.

26. Benito, A.; Rodriguez, A.; Therón, R. Visual approaches to uncertainty in DH. Presentation at the workshop "Uncertainties in Humanities Research Datasets", Maynooth, Ireland, 6 March 2019. co-hosted by the Humanities Research Institute at Maynooth University, the Trinity College Dublin Centre for Digital Humanities and the consortium members of the PROVIDEDH project.

27. Galbraith, J.K. *The Age of Uncertainty*; Houghton Mifflin: Boston, MA, USA, 1977.

28. Regan, H.M.; Colyvan, M.; Burgman, M.A. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecol. Appl.* **2002**, *12*, 618–628. [CrossRef]

29. Fox, R.C. Medical Uncertainty Revisited. In *Handbook of Social Studies in Health and Medicine*; Albrecht, G.L., Fitzpatrick, R., Scrimshaw, S.C., Eds.; SAGE Publishing: Thousand Oaks, CA, USA, 2000; pp. 409–425.

30. Hoffmann, V.H.; Trautmann, T.; Schneider, M. A taxonomy for regulatory uncertainty—application to the European Emission Trading Scheme. *Environ. Sci. Policy* **2008**, *11*, 712–722. [CrossRef]

31. New World Encyclopedia. "Uncertainty". Available online: http://www.newworldencyclopedia.org/p/index.php?title=Uncertainty&oldid=993112 (accessed on 28 June 2019).

32. Smithson, M. *Ignorance and Uncertainty: Emerging Paradigms*; Springer Science & Business Media: Berlin, Germany, 2012.

33. Bammer, G.; Smithson, M.; Goolabri Group. The Nature of Uncertainty. In *Uncertainty and Risk: Multidisciplinary Perspectives*; Smithson, M., Bammer, G., Eds.; Routledge: Abingdon, UK, 2012; pp. 289–303.

34. Shattuck, L.G.; Lewis Miller, N.; Kemmerer, K.E. Tactical Decision Making Under Conditions of Uncertainty: An Empirical Study. In Proceedings of the Human Factors and Ergonomics Society. Annual Meeting, San Antonio, TX, USA, 19–23 October 2009.

35. Diehl, A.; Yang, B.; Das, R.D.; Chen, S.; Andrienko, G.; Andrienko, N.; Dransch, D.; Keim, D. User-Uncertainty: A Human-Centred Uncertainty Taxonomy for VGI through the Visual Analytics Workflow. In Proceedings of the VGI Geovisual Analytics Workshop, colocated with BDVA 2018, Konstanz, Germany, 19 October 2018.

36. Lovell, B.E. A Taxonomy of Types of Uncertainty. Ph.D. Dissertation, Portland State University, Portland, OR, USA, 1995.

37. Vullings, W.; de Vries, M.; de Borman, L. Dealing with uncertainty in spatial planning. In Proceedings of the 10th AGILE International Conference on Geographic Information Science, Aalborg, Denmark, 13–17 August 2007.

38. Cressie, N.; Wikle, C.K. *Statistics for Spatio-Temporal Data*; John Wiley & Sons: New York, NY, USA, 2015.

39. Aigner, W.; Miksch, S.; Müller, W.; Schumann, H.; Tominski, C. Visualizing time-oriented data—A systematic view. *Comput. Graph.* **2007**, *31*, 401–409. [CrossRef]

40. Kissling, W.D.; Ahumada, J.A.; Bowser, A.; Fernandez, M.; Fernández, N.; Alonso Garcia, E.; Guralnick, R.P.; Isaac, N.J.B.; Kelling, S.; Los, W.; et al. Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biol. Rev.* **2018**, *93*, 600–625. [CrossRef] [PubMed]

41. Binder, F.; Entrup, B.; Schiller, I.; Lobin, H. Uncertain about Uncertainty: Different ways of processing fuzziness in digital humanities data. In Proceedings of the Digital Humanities 2014 Conference Abstracts EPFL—UNIL, Lausanne, Switzerland, 8–12 July 2014.

42. Edmond, J. Managing Uncertainty in the Humanities: Digital and Analogue Approaches. In Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality, Salamanca, Spain, 24–26 October 2018.

43. Track 13. Uncertainty in Digital Humanities. Available online: https://2018.teemconference.eu/uncertainty-digital-humanities (accessed on 28 June 2019).

44. Martin-Rodilla, P.; Gonzalez-Perez, C. Representing Imprecise and Uncertain Knowledge in Digital Humanities: A Theoretica l Framework and ConML Implementation with a Real Case Study. In Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality, Salamanca, Spain, 24–26 October 2018.

45. Senabre Hidalgo, E. Dotmocracy and Planning Poker for Uncertainty Management in Collaborative Research: Two Examples of Co-creation Techniques Derived from Digital Culture. In Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality, Salamanca, Spain, 24–26 October 2018.

46. Wandl-Vogt, E.; Dorn, A.; Piringer, B. The unknThe unknown and the uncertain. A data discovery journey from an analogous data collection to an interactive exploration space. In Proceedings of the First European Association for Digital Humanities Conference (EADH 2018), Galway, Ireland, 7–9 December 2018.

47. Fisher, P.F. Models of uncertainty in spatial data. *Geogr. Inf. Syst.* **1999**, *1*, 191–205.

48. Couclelis, H. The Certainty of Uncertainty: GIS and the Limits of Geographic Knowledge. *Trans. GIS* **2003**, *7*, 165–175. [CrossRef]

49. Fusco, G.; Caglioni, M.; Emsellem, K.; Merad, M.; Moreno, D.; Voiron-Canicio, C. Questions of uncertainty in geography. *Environ. Plan. A Econ. Space* **2017**, *49*, 2261–2280. [CrossRef]

50. Züfle, A.; Trajcevski, G.; Pfoser, D.; Renz, M.; Rice, M.T.; Leslie, T.; Delamater, P.; Emrich, T. Handling Uncertainty in Geo-Spatial Data. In Proceedings of the 2017 IEEE 33rd International Conference on Data Engineering (ICDE), San Diego, CA, USA, 19–22 April 2017.

51.　GIS-Wörterbuch. Available online: https://support.esri.com/en/other-resources/gis-dictionary/term/9ac5d78f-2a00-4c24-81ba-346ad51bf302 (accessed on 28 June 2019).

52.　Wörterbuch der bairischen Mundarten in Österreich (WBÖ). In *Bayerisches Wörterbuch: I*; Verlag der Österreichischen Akademie der Wissenschaften: Wien, Austria, 1970.

53.　*Arbeitsplan und Geschäftsordnung für Das Bayerisch-Österreichische Wörterbuch*; Archive of the Austrian Academy of Sciences: Vienna, Austria, 1912.

54.　Piringer, B.; Wandl-Vogt, E.; Abgaz, Y.; Lejtovicz, K. Exploring and exploiting biographical and prosopographical information as common access layer for heterogeneous data facilitating inclusive, gender- symmetric research. In Proceedings of the Biographical Data in a Digital World, Linz, Austria, 6–7 November 2017.

55.　Scholz, J.; Hrastnig, E.; Wandl-Vogt, E. A Spatio-Temporal Linked Data Representation for Modeling Spatio-Temporal Dialect Data. In Proceedings of the Workshops and Posters at the 13th International Conference on Spatial Information Theory (COSIT), L'Aquila, Italy, 4–8 September 2017.

56.　Schopper, D.; Bowers, J.; Wandl-Vogt, E. dboe@TEI: Remodelling a database of dialects into a rich LOD resource. In Proceedings of the Text Encoding Initiative Conference, Lyon, France, 28–31 October 2015.

57.　Wandl-Vogt, E. Datenbank der Bairischen Mundarten in Österreich @ Electronically Mapped. Projektbeschreibung. 2012. Available online: https://dboema.acdh.oeaw.ac.at/projekt/beschreibung/ (accessed on 28 June 2019).

58.　Wandl-Vogt, E. Von der Karte zum Wörterbuch—Überlegungen zu einer räumlichen Zugriffsstruktur für Dialektwörterbücher dargestellt am Beispiel des Wörterbuchs der bairischen Mundarten in Österreich (WBÖ). In Proceedings of the Atti del XII Congresso Internazionale di Lessicografia, Torino, Italia, 6–9 September 2006.

59.　Wandl-Vogt, E. Mapping Dialects. Die Karte als primäre Zugriffsstruktur für Dialektwörterbücher. *Wiener Schriften Geographie Kartographie* **2006**, *17*, 87–89.

60.　Wandl-Vogt, E.; Kop, C.; Nickel, J.; Scholz, J. Database of Bavarian Dialects (DBÖ) electronically mapped (*dbo@ema*). A System for Archiving, Maintaining and Field Mapping of Heterogeneous Dialect Data for the Compilation of Dialect Lexicons. In Proceedings of the XIII Euralex International Congress, Barcelona, Spain, 15–19 July 2008.

61.　Scholz, J.; Bartelme, N.; Fliedl, G.; Hassler, M.; Mayr, H.C.; Nickel, J.; Vöhringer, J.; Wandl-Vogt, E. Mapping Languages—Erfahrungen aus dem Projekt dbo@ema. In Proceedings of the Angewandte Geoinformatik 2008—Beiträge zum 20. AGIT-Symposium, Salzburg, Austria, 2–4 July 2008.

62.　Bartelme, N.; Scholz, J. Geoinformationstechnologien zur Analyse des Raum- und Zeitbezugs bei Dialektwörtern. In *Fokus Dialekt Analysieren—Dokumentieren—Kommunizieren*; Bergmann, H., Glauninger, M.M., Wandl-Vogt, E., Winterstein, S., Eds.; Georg Olms Verlag: Hildesheim, Germany, 2010; pp. 65–78.

63.　Scholz, J.; Lampoltshammer, T.J.; Bartelme, N.; Wandl-Vogt, E. Spatial-temporal Modeling of Linguistic Regions and Processes with Combined Indeterminate and Crisp Boundaries. In *Progress in Cartography. Lecture Notes in Geoinformation and Cartography*; Gartner, G., Jobst, M., Huang, H., Eds.; Springer: Cham, Switzerland, 2016; pp. 133–151.

64.　Hrastnig, E. A Linked Data approach for Digital Humanities. Master's Thesis, Technische Universität Graz, Graz, Austria, January 2018.

65.　Collection Explorer. Available online: https://exploreat.acdh-dev.oeaw.ac.at/exploreAT-collectionexplorer (accessed on 28 June 2019).

66.　GitHub. Acdh-Oeaw/Exploreat-Collectionexplorer. Available online: https://github.com/acdh-oeaw/exploreAT-collectionexplorer (accessed on 28 June 2019).

67.　Smith, B. Fiat Objects. *Topoi* **2001**, *20*, 131–148. [CrossRef]

68.　Rill, B. Böhmen und Mähren. In *Geschichte im Herzen Mitteleuropas*; Casimir Katz Verlag: Gernsbach, Germany, 2006.

69.　GeoPy. Available online: https://geopy.readthedocs.io/en/stable/# (accessed on 28 June 2019).

70.　Welcome to Nominatim. Available online: https://nominatim.openstreetmap.org/ (accessed on 28 June 2019).

71.　Wikipedia. The Free Encyclopedia. "Date Format by Country". Available online: https://en.wikipedia.org/wiki/Date_format_by_country (accessed on 28 June 2019).

72. Dobrin, L.; Sicoli, M. Why cultural meanings matter in endangered language research. In *Reflections on Language Documentation 20 Years after Himmelmann 1998*; McDonnell, B., Berez-Kroeker, A.L., Holton, G., Eds.; University of Hawaii Press: Honolulu, HI, USA, 2018; pp. 41–54.

73. Belew, A.; Chen, Y.; Campbell, L.; Barlow, R.; Hauk, B.; Heaton, R.; Walla, S. The World's Endangered Languages and their Status. In *Cataloguing the World's Endangered Languages*; Campbell, L., Belew, A., Eds.; Routledge: Abingdon, UK, 2018; pp. 85–249.

74. International Phonetic Alphabet. Available online: http://www.internationalphoneticalphabet.org/ipa-sounds/ipa-chart-with-sounds (accessed on 28 June 2019).

75. Reichel, S. Handbuch zum Zeichensatz SMFTeuthonista. 2003. Available online: https://web.archive.org/web/20040724105759/http://www.sprachatlas.phil.unierlangen.de/materialien/Teuthonista_Handbuch.pdf (accessed on 28 June 2019).

76. Bachmaier, R.; Kramer, U. Symbole der Wiener Teuthonista und der IPA im Vergleich. 2009. Available online: https://web.archive.org/web/20120306040434/http://www.oeaw.ac.at/dinamlex/Teutho_IPA.pdf (accessed on 28 June 2019).

77. UTF-8 Encoding Table and Unicode Characters. Available online: https://www.utf8-chartable.de/unicode-utf8-table.pl (accessed on 28 June 2019).

78. About the Unicode®Standard. Available online: https://unicode.org/standard/standard.html (accessed on 28 June 2019).

79. Character Set Encoding Basics. Available online: https://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=IWS-Chapter03#79e846db (accessed on 28 June 2019).

80. Funtowicz, S.; Ravetz, J. Post-normal science. In *Companion to Environmental Studies*; Castree, N., Hulme, M., Proctor, J.D., Eds.; Routledge: Abingdon, UK, 2018; pp. 443–447.