Enrichment of Legacy Language Data: Linking Lexical Concepts in Data Collection Questionnaires on the example of exploreAT!

Amelie Dorn, Barbara Piringer, Yalemisew Abgaz, José Luis Preza Díaz, Eveline Wandl-Vogt

Budapest, Hungary
27.09.2019
DH_Budapest_2019 Conference, 25.–27.09.2019

CC BY NC SA

Presentation outline

- Background
- exploreAT! – Motivation and aims
- Description of questionnaires
- Technical approach
- Workflow
- Challenges
- Conclusion

Background

1911	Establishment of the „Commission for the creation of a Bavarian-Austrian dictionary in Austria“ (WBÖ) at the Austrian Academy of Sciences
1913–1935	Compilation & mailing of 120 questionnaires to the collectors to build up a lexical collection
1993–2011	Digitisation of the collection within the framework of the database project DBÖ
2007–2010	Project „Database of the Bavarian Dialects in Austria electronically mapped“ (dbo@ema)
2015–2020	Project exploreAT!

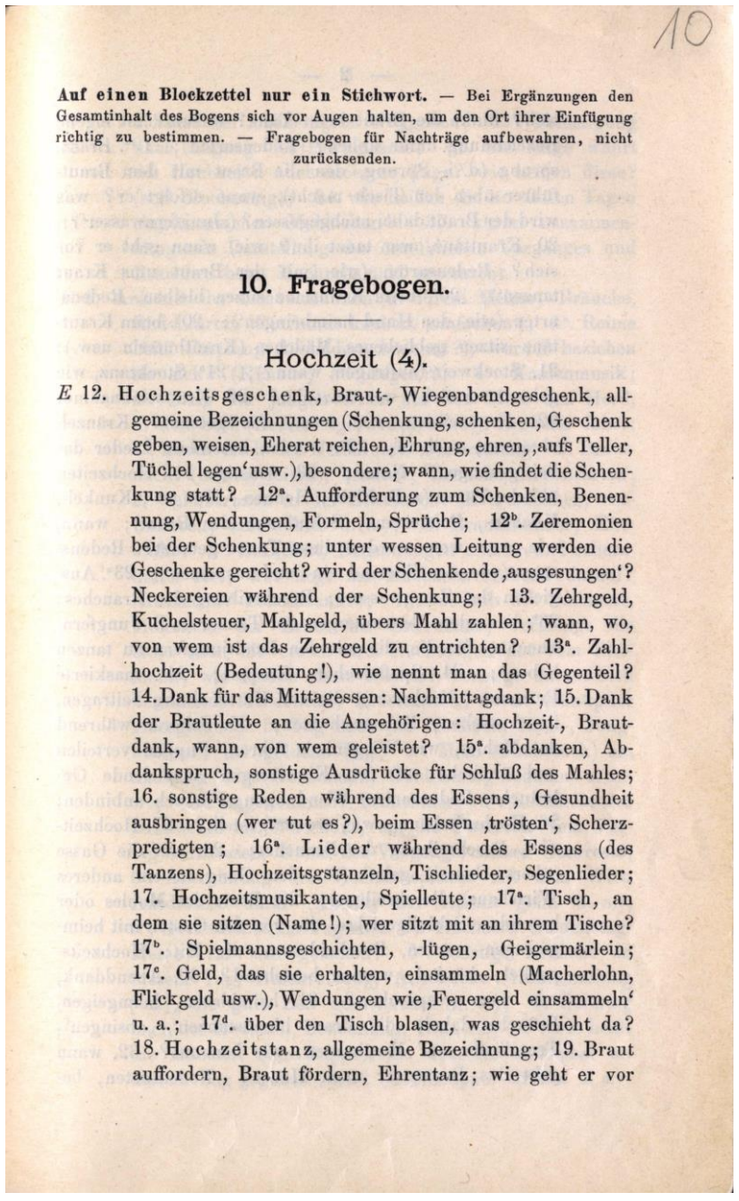
exploreAT!: Motivation and aims

- Exploit this non-standard German language resource from the perspective of semantic technologies, visual analysis tools, citizen science and **cultural lexicography**
- Reveal cultural knowledge contained in the DBÖ and make this information accessible
- Opening up the questionnaires by adding of lexical concepts and their linking to other resources
- Enable and foster to understand rural life in the early 20th century

Questionnaires (1)

- 109 „Systematic questionnaires“
- Dedicated to a **special theme** (focusing on rural life) – but a topic can also span several questionnaires, if necessary
- Most of the questionnaires are **divided in sections** with an emphasis on a specific issue
- Sections contain **numbered questions** that sometimes are partitioned in sub-questions
- Structure of the questionnaires **changed** over time

Questionnaires (2)



10

Auf einen Blockzettel nur ein Stichwort. — Bei Ergänzungen den Gesamtinhalt des Bogens sich vor Augen halten, um den Ort ihrer Einfügung richtig zu bestimmen. — Fragebogen für Nachträge aufbewahren, nicht zurücksenden.

10. Fragebogen.

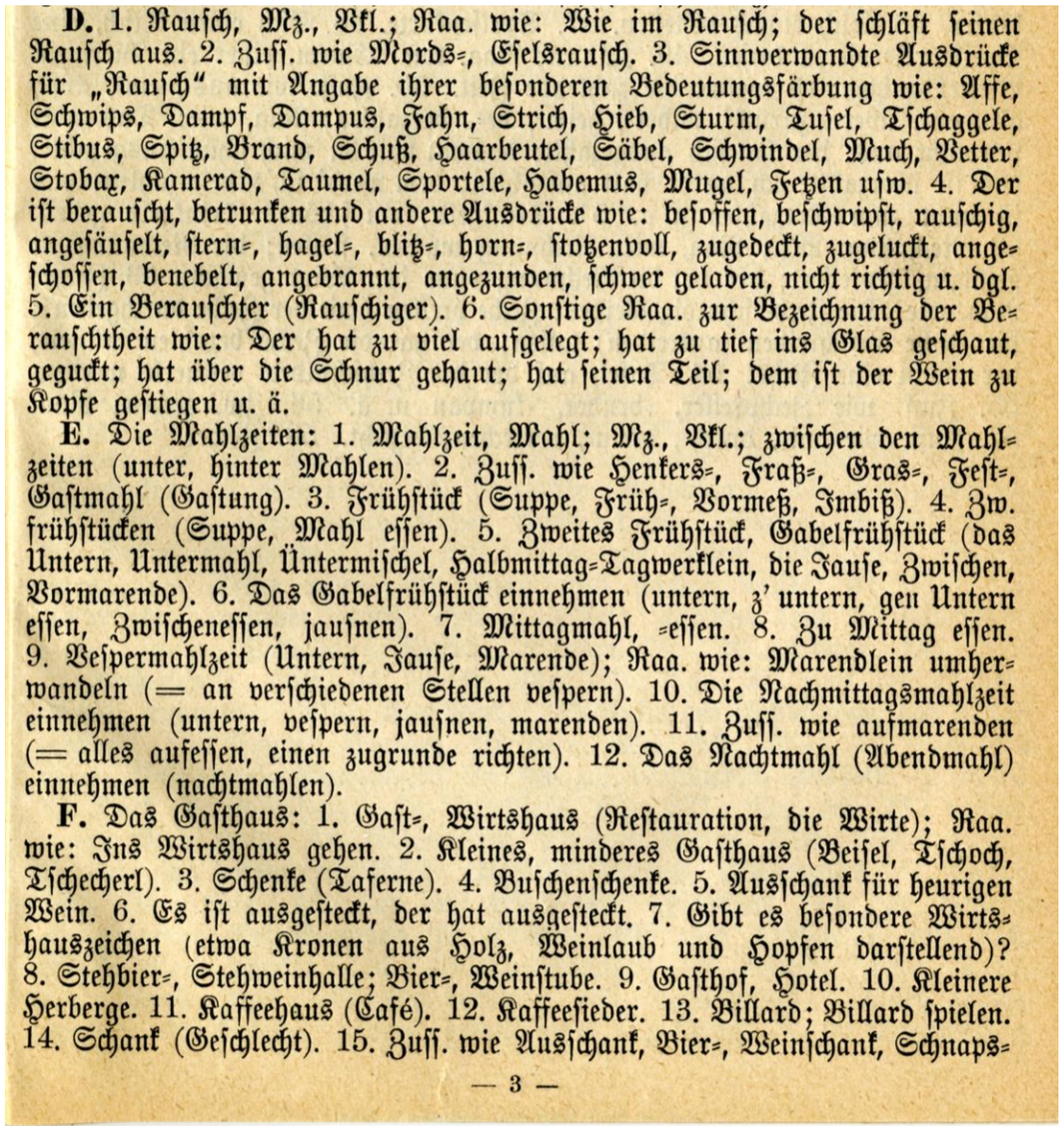
Hochzeit (4).

E 12. Hochzeitsgeschenk, Braut-, Wiegenbandgeschenk, allgemeine Bezeichnungen (Schenkung, schenken, Geschenk geben, weisen, Eherat reichen, Ehrung, ehren, ‚aufs Teller, Tüchel legen' usw.), besondere; wann, wie findet die Schenkung statt? 12ª. Aufforderung zum Schenken, Benennung, Wendungen, Formeln, Sprüche; 12ᵇ. Zeremonien bei der Schenkung; unter wessen Leitung werden die Geschenke gereicht? wird der Schenkende ‚ausgesungen'? Neckereien während der Schenkung; 13. Zehrgeld, Kuchelsteuer, Mahlgeld, übers Mahl zahlen; wann, wo, von wem ist das Zehrgeld zu entrichten? 13ª. Zahlhochzeit (Bedeutung!), wie nennt man das Gegenteil? 14. Dank für das Mittagessen: Nachmittagdank; 15. Dank der Brautleute an die Angehörigen: Hochzeit-, Brautdank, wann, von wem geleistet? 15ª. abdanken, Abdankspruch, sonstige Ausdrücke für Schluß des Mahles; 16. sonstige Reden während des Essens, Gesundheit ausbringen (wer tut es?), beim Essen ‚trösten', Scherzpredigten; 16ª. Lieder während des Essens (des Tanzens), Hochzeitsgstanzeln, Tischlieder, Segenlieder; 17. Hochzeitsmusikanten, Spielleute; 17ª. Tisch, an dem sie sitzen (Name!); wer sitzt mit an ihrem Tische? 17ᵇ. Spielmannsgeschichten, -lügen, Geigermärlein; 17ᶜ. Geld, das sie erhalten, einsammeln (Macherlohn, Flickgeld usw.), Wendungen wie ‚Feuergeld einsammeln' u. a.; 17ᵈ. über den Tisch blasen, was geschieht da? 18. Hochzeitstanz, allgemeine Bezeichnung; 19. Braut auffordern, Braut fördern, Ehrentanz; wie geht er vor

- Questions are rather elaborated
- Question items often contains several questions
- A question item often has different kinds of questions
- Further subdivision of the section with guiding themes

Fig. 1. Systematic questionnaire Nr. 10 (Wedding 4); edited in 1913.

Questionnaires (3)



D. 1. Rausch, Mz., Vtl.; Raa. wie: Wie im Rausch; der schläft seinen Rausch aus. 2. Zuss. wie Mords-, Eselsrausch. 3. Sinnverwandte Ausdrücke für „Rausch" mit Angabe ihrer besonderen Bedeutungsfärbung wie: Affe, Schwips, Dampf, Dampus, Fahn, Strich, Hieb, Sturm, Tusel, Tschaggele, Stibus, Spitz, Brand, Schuß, Haarbeutel, Säbel, Schwindel, Much, Vetter, Stobar, Kamerad, Taumel, Sportele, Habemus, Mugel, Fetzen usw. 4. Der ist berauscht, betrunken und andere Ausdrücke wie: besoffen, beschwipst, rauschig, angesäuselt, stern-, hagel-, blitz-, horn-, stotzenvoll, zugedeckt, zugeluckt, angeschossen, benebelt, angebrannt, angezunden, schwer geladen, nicht richtig u. dgl. 5. Ein Berauschter (Rauschiger). 6. Sonstige Raa. zur Bezeichnung der Berauschtheit wie: Der hat zu viel aufgelegt; hat zu tief ins Glas geschaut, geguckt; hat über die Schnur gehaut; hat seinen Teil; dem ist der Wein zu Kopfe gestiegen u. ä.

E. Die Mahlzeiten: 1. Mahlzeit, Mahl; Mz., Vkl.; zwischen den Mahlzeiten (unter, hinter Mahlen). 2. Zuss. wie Henkers-, Fraß-, Gras-, Fest-, Gastmahl (Gastung). 3. Frühstück (Suppe, Früh-, Vormeß, Imbiß). 4. Zw. frühstücken (Suppe, Mahl essen). 5. Zweites Frühstück, Gabelfrühstück (das Untern, Untermahl, Üntermischel, Halbmittag-Tagwerklein, die Jause, Zwischen, Vormarende). 6. Das Gabelfrühstück einnehmen (untern, z' untern, gen Untern essen, Zwischeneſſen, jausnen). 7. Mittagmahl, -essen. 8. Zu Mittag essen. 9. Vespermahlzeit (Untern, Jause, Marende); Raa. wie: Marendlein umherwandeln (= an verschiedenen Stellen vespern). 10. Die Nachmittagsmahlzeit einnehmen (untern, vespern, jausnen, marenden). 11. Zuss. wie aufmarenden (= alles aufessen, einen zugrunde richten). 12. Das Nachtmahl (Abendmahl) einnehmen (nachtmahlen).

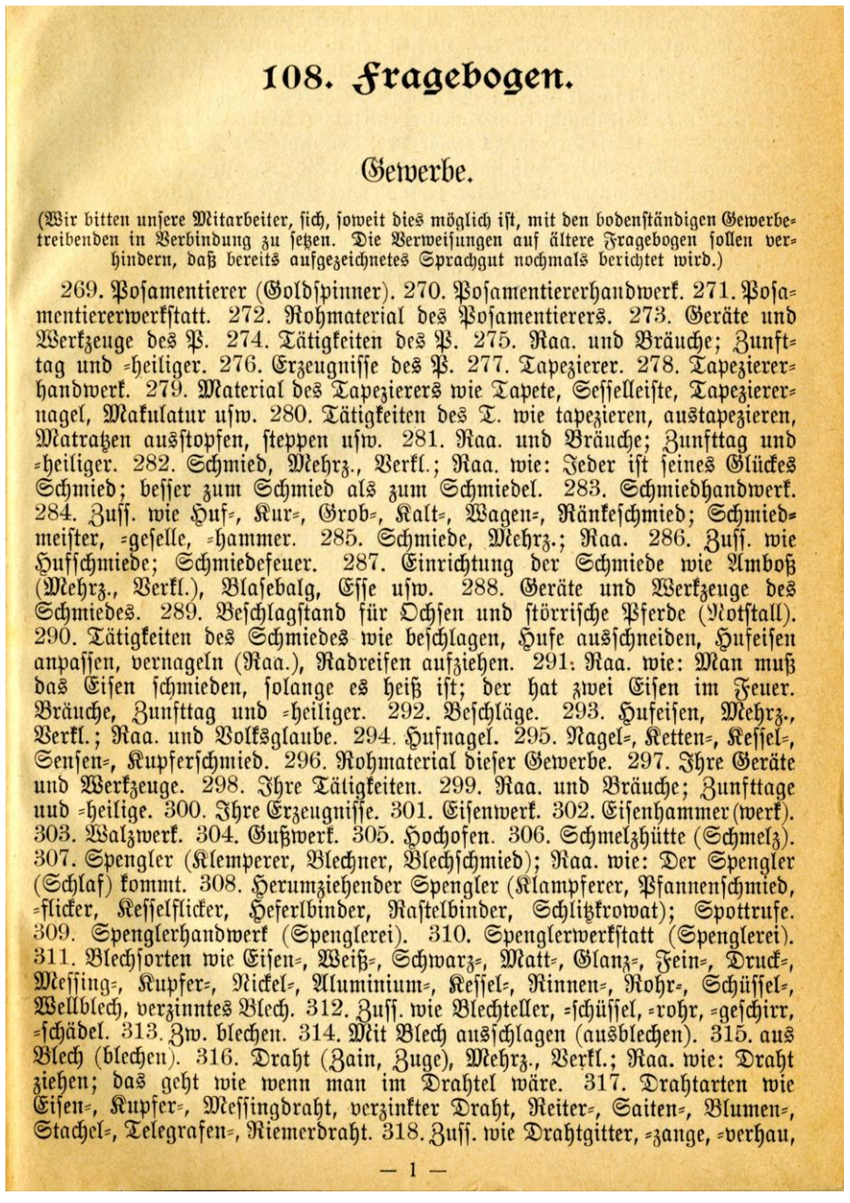
F. Das Gasthaus: 1. Gast-, Wirtshaus (Restauration, die Wirte); Raa. wie: Ins Wirtshaus gehen. 2. Kleines, minderes Gasthaus (Beisel, Tschoch, Tschecherl). 3. Schenke (Taferne). 4. Buschenschenke. 5. Ausschank für heurigen Wein. 6. Es ist ausgesteckt, der hat ausgesteckt. 7. Gibt es besondere Wirtshauszeichen (etwa Kronen aus Holz, Weinlaub und Hopfen darstellend)? 8. Stehbier-, Stehweinhalle; Bier-, Weinstube. 9. Gasthof, Hotel. 10. Kleinere Herberge. 11. Kaffeehaus (Café). 12. Kaffeesieder. 13. Billard; Billard spielen. 14. Schank (Geschlecht). 15. Zuss. wie Ausschank, Bier-, Weinschank, Schnaps-

— 3 —

- Questions are more like key terms
- Abbreviations are common
- Suggestions, leading questions

Fig. 2. Detail from Systematic questionnaire Nr. 97 (Eat and Drink); edited in 1928.

Questionnaires (4)



108. Fragebogen.

Gewerbe.

(Wir bitten unsere Mitarbeiter, sich, soweit dies möglich ist, mit den bodenständigen Gewerbetreibenden in Verbindung zu setzen. Die Verweisungen auf ältere Fragebogen sollen verhindern, daß bereits aufgezeichnetes Sprachgut nochmals berichtet wird.)

269. Posamentierer (Goldspinner). 270. Posamentiererhandwerk. 271. Posamentiererwerkstatt. 272. Rohmaterial des Posamentierers. 273. Geräte und Werkzeuge des P. 274. Tätigkeiten des P. 275. Raa. und Bräuche; Zunfttag und =heiliger. 276. Erzeugnisse des P. 277. Tapezierer. 278. Tapeziererhandwerk. 279. Material des Tapezierers wie Tapete, Sesselleiste, Tapeziernagel, Makulatur usw. 280. Tätigkeiten des T. wie tapezieren, austapezieren, Matratzen ausstopfen, steppen usw. 281. Raa. und Bräuche; Zunfttag und =heiliger. 282. Schmied, Mehrz., Verkl.; Raa. wie: Jeder ist seines Glückes Schmied; besser zum Schmied als zum Schmiedel. 283. Schmiedhandwerk. 284. Zuss. wie Huf=, Kur=, Grob=, Kalt=, Wagen=, Ränkeschmied; Schmiedmeister, =geselle, =hammer. 285. Schmiede, Mehrz.; Raa. 286. Zuss. wie Hufschmiede; Schmiedefeuer. 287. Einrichtung der Schmiede wie Amboß (Mehrz., Verkl.), Blasebalg, Esse usw. 288. Geräte und Werkzeuge des Schmiedes. 289. Beschlagstand für Ochsen und störrische Pferde (Notstall). 290. Tätigkeiten des Schmiedes wie beschlagen, Hufe ausschneiden, Hufeisen anpassen, vernageln (Raa.), Radreifen aufziehen. 291. Raa. wie: Man muß das Eisen schmieden, solange es heiß ist; der hat zwei Eisen im Feuer. Bräuche, Zunfttag und =heiliger. 292. Beschläge. 293. Hufeisen, Mehrz., Verkl.; Raa. und Volksglaube. 294. Hufnagel. 295. Nagel=, Ketten=, Kessel=, Sensen=, Kupferschmied. 296. Rohmaterial dieser Gewerbe. 297. Ihre Geräte und Werkzeuge. 298. Ihre Tätigkeiten. 299. Raa. und Bräuche; Zunfttage und =heilige. 300. Ihre Erzeugnisse. 301. Eisenwerk. 302. Eisenhammer(werk). 303. Walzwerk. 304. Gußwerk. 305. Hochofen. 306. Schmelzhütte (Schmelz). 307. Spengler (Klemperer, Blechner, Blechschmied); Raa. wie: Der Spengler (Schlaf) kommt. 308. Herumziehender Spengler (Klampferer, Pfannenschmied, =flicker, Kesselflicker, Heferlbinder, Rastelbinder, Schlitzkrowat); Spottrufe. 309. Spenglerhandwerk (Spenglerei). 310. Spenglerwerkstatt (Spenglerei). 311. Blechsorten wie Eisen=, Weiß=, Schwarz=, Matt=, Glanz=, Fein=, Druck=, Messing=, Kupfer=, Nickel=, Aluminium=, Kessel=, Rinnen=, Rohr=, Schüssel=, Wellblech, verzinntes Blech. 312. Zuss. wie Blechteller, =schüssel, =rohr, =geschirr, =schädel. 313. Zw. blechen. 314. Mit Blech ausschlagen (ausblechen). 315. aus Blech (blechen). 316. Draht (Zain, Zuge), Mehrz., Verkl.; Raa. wie: Draht ziehen; das geht wie wenn man im Drahtel wäre. 317. Drahtarten wie Eisen=, Kupfer=, Messingdraht, verzinkter Draht, Reiter=, Saiten=, Blumen=, Stachel=, Telegrafen=, Riemerdraht. 318. Zuss. wie Drahtgitter, =zange, =verhau,

— 1 —

- No numbering of title
- Only few words or short phrases per question item
- No subdivisions
- Consecutive numbers

Fig. 3. Systematic questionnaire Nr. 108 (Trade); edited ca. 1932.

Technical approach

- **DBpedia**: Crowd-sourced community effort
- Extracts structured content from the information created in various Wikimedia projects for the Semantic Web
- This information resembles an open knowledge graph, in a machine-readable form
- Standards: Resource Description Framework (RDF) and SPARQL as query language
- **DBpedia Spotlight**: Tool for automatically annotating mentions of DBpedia resources in text, useful for Named Entity Recognition
- it links unstructured information sources to the Linked Open Data cloud through DBpedia

Workflow (1)

Baking Bread	http://dbpedia.org/resource/Baking	0,999999988
Baking Bread	http://dbpedia.org/resource/Bread	0,9999982133
Consanguinity and Ages	http://dbpedia.org/resource/Consanguinity	0,9999998715
Consanguinity and Ages	http://dbpedia.org/resource/Ages_of_Man	0,4632036979
Movement	http://dbpedia.org/resource/International_Red_Cross_and_Red_Crescent_Movement	0,4962919378
Religion and Church	http://dbpedia.org/resource/Religion	0,9909724509
Religion and Church	http://dbpedia.org/resource/Catholic_Church	0,8741306615
Movement	http://dbpedia.org/resource/International_Red_Cross_and_Red_Crescent_Movement	0,4962919378
Tailoring	http://dbpedia.org/resource/Bespoke_tailoring	0,9999938676
Cutlery, scales, weights and measures	http://dbpedia.org/resource/Cutlery	0,9999999246
Cutlery, scales, weights and measures	http://dbpedia.org/resource/Weighing_scale	0,9558299177
Cutlery, scales, weights and measures	http://dbpedia.org/resource/Units_of_measurement	0,6317335139
Cutlery, scales, weights and measures	http://dbpedia.org/resource/Measure_(mathematics)	0,7479263735
Traffic	http://dbpedia.org/resource/Traffic	0,9813941006
Mythological matters	http://dbpedia.org/resource/Mythology	0,999999343
Nose, Ear	http://dbpedia.org/resource/Human_nose	0,5038536109
Nose, Ear	http://dbpedia.org/resource/Ear	0,9999950228
Animals	http://dbpedia.org/resource/Animal	0,918021904
Body parts	http://dbpedia.org/resource/Human_body	0,8271304661
Body parts	http://dbpedia.org/resource/List_of_auto_parts	0,6127437825
Apiculture	http://dbpedia.org/resource/Beekeeping	0,9999998947
Clothing	http://dbpedia.org/resource/Clothing	0,9992486079
Arm , Hand	http://dbpedia.org/resource/ARM_architecture	0,4900467368

Workflow (2)

Consanguinity and Ages	http://dbpedia.org/resource/Ages_of_Man	0.4632036979		
Movement	http://dbpedia.org/resource/International_Red_Cross_a	0.4962919378		
Religion and Church	http://dbpedia.org/resource/Religion	0.9909724509		
Religion and Church	http://dbpedia.org/resource/Catholic_Church	0.8741306615		
Movement	http://dbpedia.org/resource/International_Red_Cross_a	0.4962919378		
Tailoring	http://dbpedia.org/resource/Bespoke_tailoring	0.9999938676		
Cutlery, scales, weights and measures	http://dbpedia.org/resource/Cutlery	0.9999999246		s. Comments
Cutlery, scales, weights and measures	http://dbpedia.org/resource/Weighing_scale	0.9558299177		
Cutlery, scales, weights and measures	http://dbpedia.org/resource/Units_of_measurement	0.6317335139		
Cutlery, scales, weights and measures	http://dbpedia.org/resource/Measure_(mathematics)	0.7479263735		
Traffic	http://dbpedia.org/resource/Traffic	0.9813941006		
Mythological matters	http://dbpedia.org/resource/Mythology	0.999999343		s. Comment
Nose, Ear	http://dbpedia.org/resource/Human_nose	0.5038536109		
Nose, Ear	http://dbpedia.org/resource/Ear	0.9999950228		
Animals	http://dbpedia.org/resource/Animal	0.918021904		
Body parts	http://dbpedia.org/resource/Human_body	0.8271304661		
Body parts	http://dbpedia.org/resource/List_of_auto_parts	0.6127437825		
Apiculture	http://dbpedia.org/resource/Beekeeping	0.9999998947		
Clothing	http://dbpedia.org/resource/Clothing	0.9992486079		
Arm , Hand	http://dbpedia.org/resource/ARM_architecture	0.4900467368		
Arm , Hand	http://dbpedia.org/resource/Hand	0.9848029733		
Human Hair and Beard	http://dbpedia.org/resource/Human	0.9991456628		
Human Hair and Beard	http://dbpedia.org/resource/Hair	0.9998387464		
Human Hair and Beard	http://dbpedia.org/resource/Beard	0.9995603698		

Workflow (3)

Colour	Farben	http://dbpedia.org/resource/Color	Color	Farbe
Companionship, friendship and love	Geselligkeit, Freundschaft und Liebe	http://dbpedia.org/resource/Social_relation	Social relation	Soziale Beziehung
Companionship, friendship and love	Geselligkeit, Freundschaft und Liebe	http://dbpedia.org/resource/Friendship	Friendship	Freundschaft
Companionship, friendship and love	Geselligkeit, Freundschaft und Liebe	http://dbpedia.org/resource/Love	Love	Liebe
Consanguinity and Ages	Blutsverwandtschaft, Altersstufen	http://dbpedia.org/resource/Consanguinity	Consanguinity	Blutsverwandtschaft
Consanguinity and Ages	Blutsverwandtschaft, Altersstufen	http://dbpedia.org/resource/Category:Human_life_stages	---	---
Consanguinity and Ages	Blutsverwandtschaft, Altersstufen	http://dbpedia.org/resource/Childhood	Childhood	Kindheit
Consanguinity and Ages	Blutsverwandtschaft, Altersstufen	http://dbpedia.org/resource/Youth	Youth	Jugend
Consanguinity and Ages	Blutsverwandtschaft, Altersstufen	http://dbpedia.org/resource/Adult	Adult	Erwachsener
Consanguinity and Ages	Blutsverwandtschaft, Altersstufen	http://dbpedia.org/resource/Old_age	Old age	Alter
Coopery	Maurerhandwerk. Binderei	http://dbpedia.org/resource/Cooper_(profession)	Cooper (profession)	Küfer
Cutlery, scales, weights and measures	Eßzeug, Waage, Gewichte und Maße	http://dbpedia.org/resource/Tableware	Tableware	Essgeschirr
Cutlery, scales, weights and measures	Eßzeug, Waage, Gewichte und Maße	http://dbpedia.org/resource/Weighing_scale	Weighing scale	Waage
Cutlery, scales, weights and measures	Eßzeug, Waage, Gewichte und Maße	http://dbpedia.org/resource/Units_of_measurement	Units of measurement	Maßeinheit
Death	Tod	http://dbpedia.org/resource/Death	Death	Tod
Dog and Cat	Hund und Katze	http://dbpedia.org/resource/Dog	Dog	Haushund
Dog and Cat	Hund und Katze	http://dbpedia.org/resource/Cat	Cat	Hauskatze
Eye	Auge (Kopf 5)	http://dbpedia.org/resource/Eye	Eye	Auge
Earth's surface	1. Erdoberfläche. 2. Straße und Weg	http://dbpedia.org/resource/Terrain	Terrain	Relief (Geologie)
Easter week	Die Osterwoche (1)	http://dbpedia.org/resource/Holy_Week	Holy Week	Karwoche
Easter week	Die Osterwoche (2)	http://dbpedia.org/resource/Holy_Week	Holy Week	Karwoche
Easter week	Die Osterwoche (2)	http://dbpedia.org/resource/Easter	Easter	Ostern
Eat and Drink	Essen und Trinken	http://dbpedia.org/resource/Eating	Eating	---
Eat and Drink	Essen und Trinken	http://dbpedia.org/resource/Drinking	Drinking	Trinken

Challenges

- Lack of concepts („Human life stages“)
- Concepts only in English („Human nose“)
- Different focus or aspect of a concept („Baking“; „Water“)
- Different definitions of English/German concepts („Schuster“ vs. „Cordwainer“/„Cobbler“; „Cutaneous conditions“)
- Title is a more of a description of a concept („Period between Midsummer and New Year“)

Conclusion

- Finer divisions can help solving some challenges
- Knowledge of the detailed content of the questionnaires is crucial
- Comparison and verification of the assigned concept can be necessary

Thank you for your attention!

Barbara.Piringer@oeaw.ac.at
Team exploration space @ ACDH-OeAW
Twitter: #explorations4u

Enrichment of Legacy Language Data: Linking Lexical Concepts in Data Collection Questionnaires on the example of exploreAT!

Amelie Dorn, Barbara Piringer, Yalemisew Abgaz, José Luis Preza Díaz, Eveline Wandl-Vogt

Budapest, Hungary
27.09.2019
DH_Budapest_2019 Conference, 25.–27.09.2019

CC BY NC SA