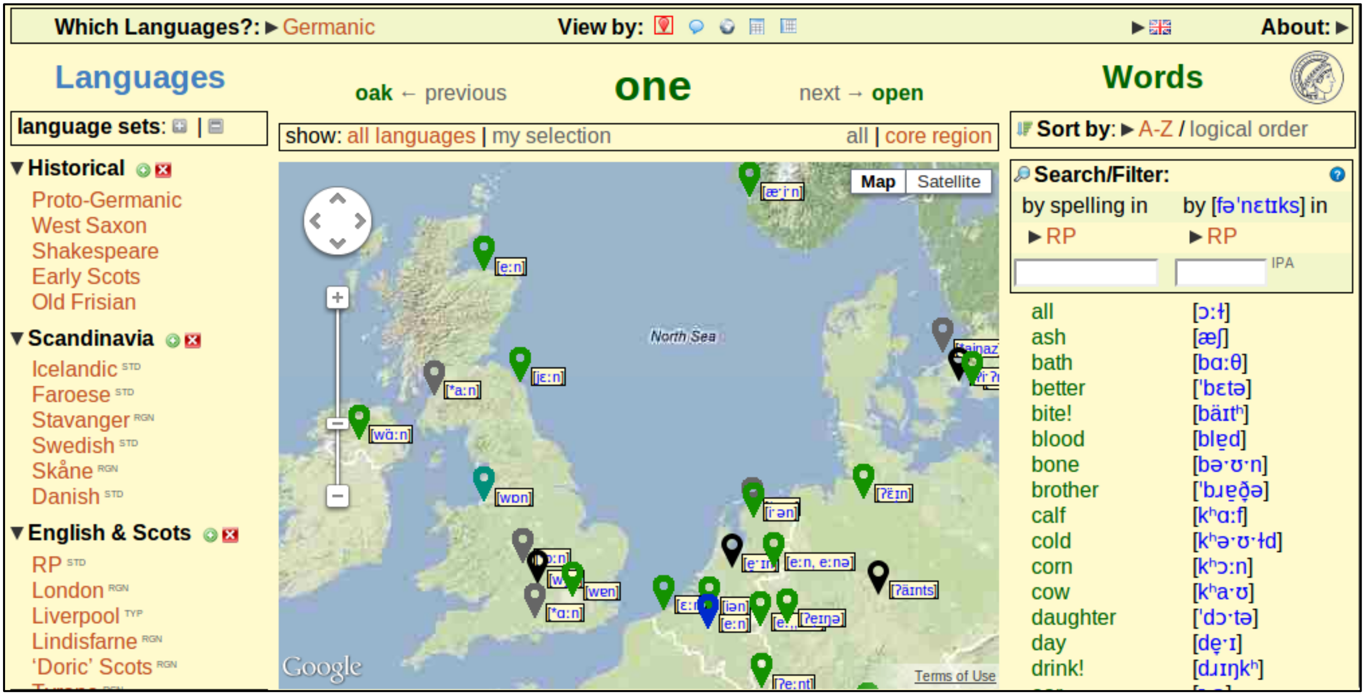
Presenting the ‘Sound Comparisons’ Website



Which Languages?: ▶ Germanic   View by: ▶   About: ▶

Languages   oak ← previous   **one**   next → open   Words

language sets: ⊞ | ⊟

show: all languages | my selection   all | core region

Sort by: ▶ A-Z / logical order

Search/Filter:

by spelling in ▶ RP   by [fə'nɛɪks] in ▶ RP

□   □ IPA

▼ **Historical**
- Proto-Germanic
- West Saxon
- Shakespeare
- Early Scots
- Old Frisian

▼ **Scandinavia**
- Icelandic STD
- Faroese STD
- Stavanger RGN
- Swedish STD
- Skåne RGN
- Danish STD

▼ **English & Scots**
- RP STD
- London RGN
- Liverpool TYP
- Lindisfarne RGN
- ‘Doric’ Scots RGN

all	[ɔːɫ]
ash	[æʃ]
bath	[bɑːθ]
better	['bɛtə]
bite!	[bäɪtʰ]
blood	[blɐ̠d]
bone	[bə·ʊ·n]
brother	['bɹɐ̠ð̞ə]
calf	[kʰɑːf]
cold	[kʰə·ʊ·ɫd]
corn	[kʰɔːn]
cow	[kʰa·ʊ]
daughter	['dɔ·tə]
day	[de̞·ɪ]
drink!	[dɹɪŋkʰ]

Paul Heggarty & Jakob Runge

Dept. of Linguistics

Max Planck Institute for Evolutionary Anthropology, Leipzig

heggarty@eva.mpg.de   jakob_runge@eva.mpg.de

1. BACKGROUND

- Research context:
  - Measuring divergence in phonetics.
  - Between related languages, dialects and accents.

- Major effort of:
  - Data collection: recordings — c. 120 words in c. 350 language varieties.
  - Data analysis: detailed phonetic transcription.

- Aim of *websites*: make those data and analyses available and useful to:
  - Scientific community of linguists, as a training and research resource.
  - Native-speakers of (esp. endangered) language varieties covered, for raising awareness, understanding, prestige, revitalisation (?).

1.1 Research Context

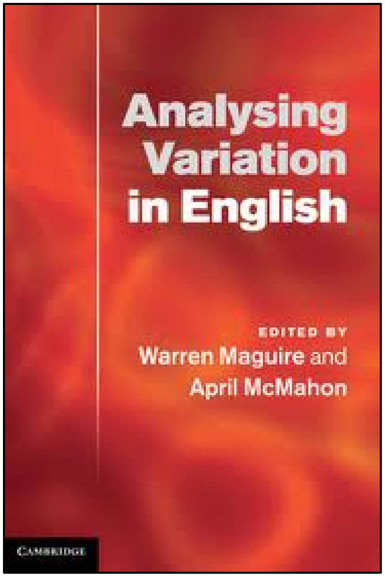
- Input data for a technique for quantifying divergence in phonetics (as precisely as possible).
- Determined data-set: phonetic sample → list of cognates (not meanings).
- Applications: in dialectology, historical linguistics, sociolinguistics.

Maguire, W., & McMahon, A.M.S. eds. 2011. *Analysing Variation in English*. Cambridge: Cambridge University Press.

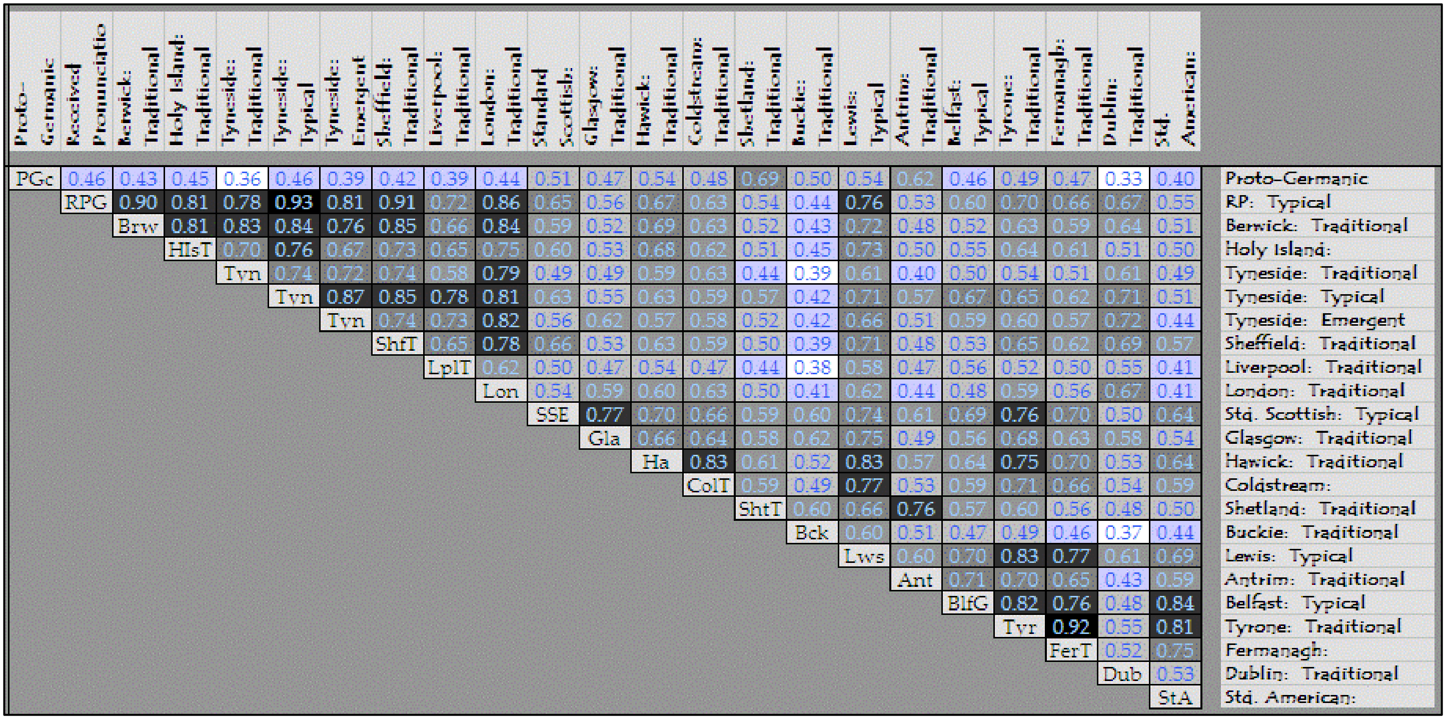
Heggarty, P., Maguire, W., & McMahon, A.M.S. 2010. Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories. *Proceedings of the Royal Society B: Biological Sciences* Cultural and Linguistic Diversity(365): p.3829–3843.

Maguire, W., McMahon, A.M.S., Heggarty, P., & Dediu, D. 2010. The past, present and future of English dialects: quantifying convergence, divergence and dynamic equilibrium. *Language Variation and Change* 22(1): p.69–104.

McMahon, A.M.S., Heggarty, P., McMahon, R., & Maguire, W. 2007. The sound patterns of Englishes: representing phonetic similarity. *English Language and Linguistics* 11(01): p.113.

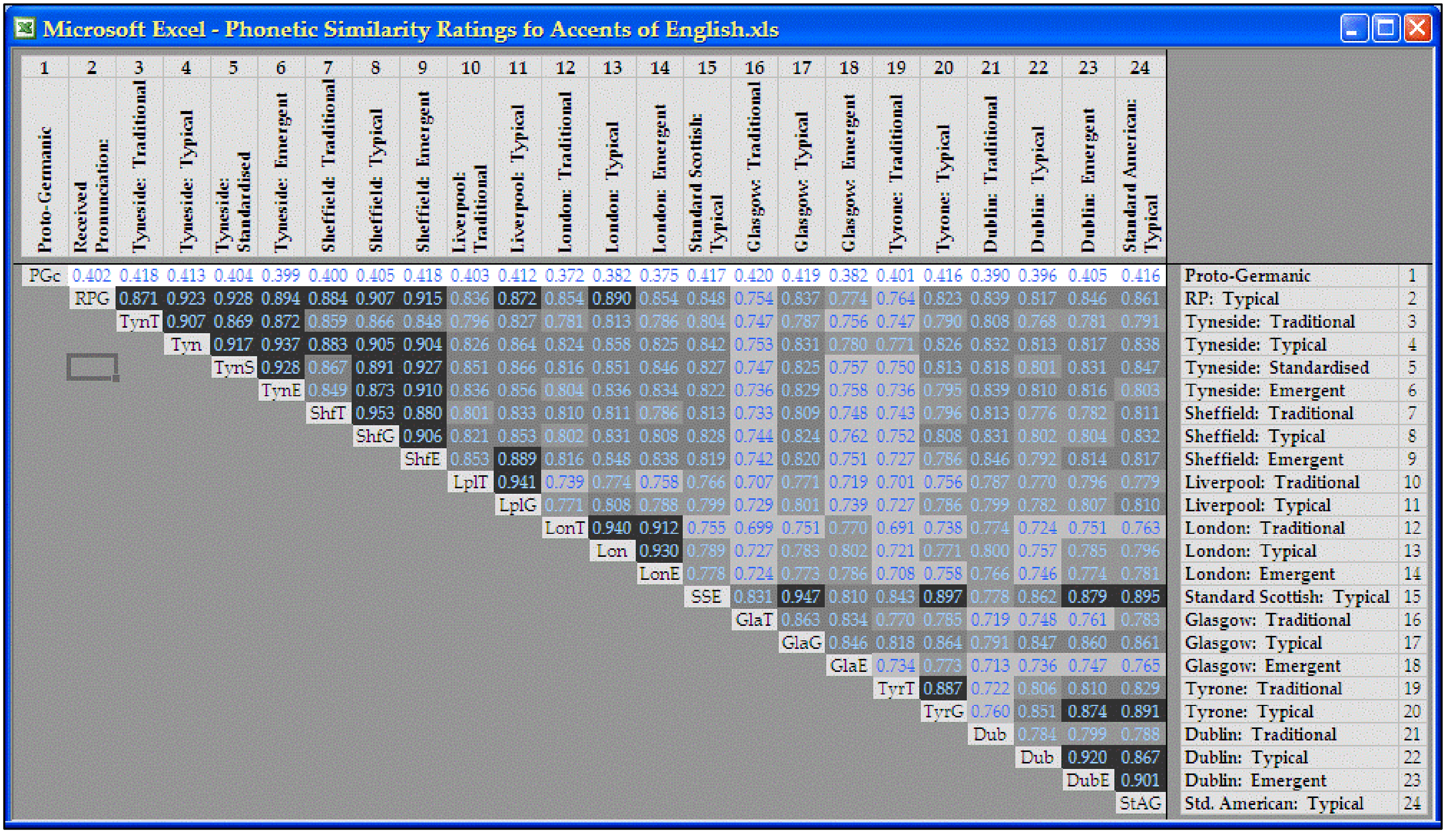


1.2 Divergence Measures: Single Cognate



Proto- Germanic	Received Pronunciatio	Berwick: Traditional	Holy Island: Traditional	Tyneside: Traditional	Tyneside: Typical	Tyneside: Emergent	Sheffield: Traditional	Liverpool: Traditional	London: Traditional	Standard Scottish:	Glasgow: Traditional	Hawick: Traditional	Coldstream: Traditional	Shetland: Traditional	Buckie: Traditional	Lewis: Typical	Antrim: Traditional	Belfast: Typical	Tyrone: Traditional	Fermanagh: Traditional	Dublin: Traditional	Std. American:	
PGc	0.46	0.43	0.45	0.36	0.46	0.39	0.42	0.39	0.44	0.51	0.47	0.54	0.48	0.69	0.50	0.54	0.62	0.46	0.49	0.47	0.33	0.40	Proto-Germanic
	RPG	0.90	0.81	0.78	0.93	0.81	0.91	0.72	0.86	0.65	0.56	0.67	0.63	0.54	0.44	0.76	0.53	0.60	0.70	0.66	0.67	0.55	RP: Typical
		Brw	0.81	0.83	0.84	0.76	0.85	0.66	0.84	0.59	0.52	0.69	0.63	0.52	0.43	0.72	0.48	0.52	0.63	0.59	0.64	0.51	Berwick: Traditional
			HIsT	0.70	0.76	0.67	0.73	0.65	0.75	0.60	0.53	0.68	0.62	0.51	0.45	0.73	0.50	0.55	0.64	0.61	0.51	0.50	Holy Island:
				Tyn	0.74	0.72	0.74	0.58	0.79	0.49	0.49	0.59	0.63	0.44	0.39	0.61	0.40	0.50	0.54	0.51	0.61	0.49	Tyneside: Traditional
					Tyn	0.87	0.85	0.78	0.81	0.63	0.55	0.63	0.59	0.57	0.42	0.71	0.57	0.67	0.65	0.62	0.71	0.51	Tyneside: Typical
						Tyn	0.74	0.73	0.82	0.56	0.62	0.57	0.58	0.52	0.42	0.66	0.51	0.59	0.60	0.57	0.72	0.44	Tyneside: Emergent
							ShfT	0.65	0.78	0.66	0.53	0.63	0.59	0.50	0.39	0.71	0.48	0.53	0.65	0.62	0.69	0.57	Sheffield: Traditional
								LplT	0.62	0.50	0.47	0.54	0.47	0.44	0.38	0.58	0.47	0.56	0.52	0.50	0.55	0.41	Liverpool: Traditional
									Lon	0.54	0.59	0.60	0.63	0.50	0.41	0.62	0.44	0.48	0.59	0.56	0.67	0.41	London: Traditional
										SSE	0.77	0.70	0.66	0.59	0.60	0.74	0.61	0.69	0.76	0.70	0.50	0.64	Std. Scottish: Typical
											Gla	0.66	0.64	0.58	0.62	0.75	0.49	0.56	0.68	0.63	0.58	0.54	Glasgow: Traditional
												Ha	0.83	0.61	0.52	0.83	0.57	0.64	0.75	0.70	0.53	0.64	Hawick: Traditional
													ColT	0.59	0.49	0.77	0.53	0.59	0.71	0.66	0.54	0.59	Coldstream:
														ShtT	0.60	0.66	0.76	0.57	0.60	0.56	0.48	0.50	Shetland: Traditional
															Bck	0.60	0.51	0.47	0.49	0.46	0.37	0.44	Buckie: Traditional
																Lws	0.60	0.70	0.83	0.77	0.61	0.69	Lewis: Typical
																	Ant	0.71	0.70	0.65	0.43	0.59	Antrim: Traditional
																		BlfG	0.82	0.76	0.48	0.84	Belfast: Typical
																			Tyr	0.92	0.55	0.81	Tyrone: Traditional
																				FerT	0.52	0.75	Fermanagh:
																					Dub	0.53	Dublin: Traditional
																						StA	Std. American:

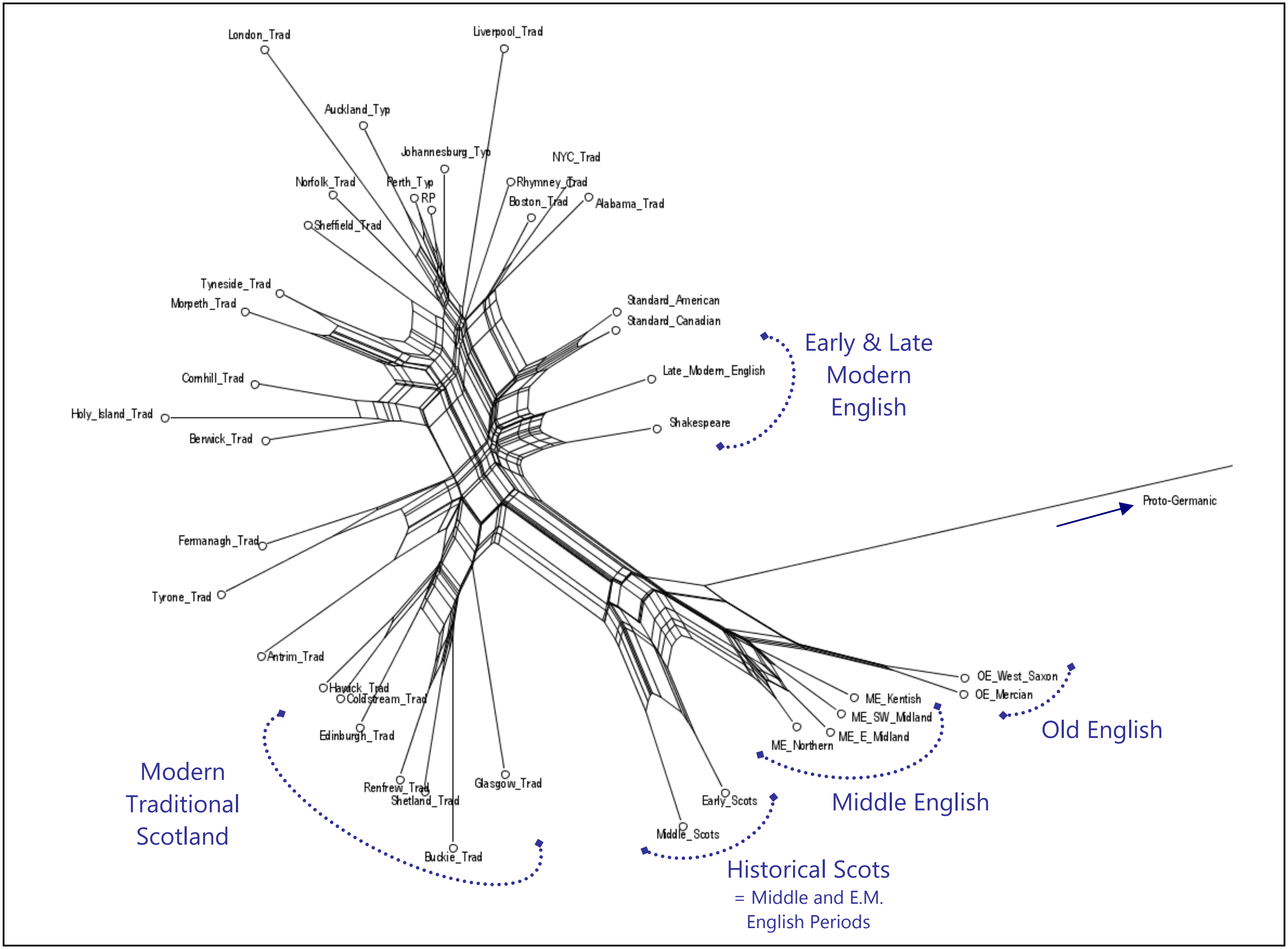
1.3 DIVERGENCE MEASURES: ENTIRE REFERENCE LIST



Microsoft Excel - Phonetic Similarity Ratings fo Accents of English.xls

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		
Proto-Germanic	Received Pronunciation:	Tyneside: Traditional	Tyneside: Typical	Tyneside: Standardised	Tyneside: Emergent	Sheffield: Traditional	Sheffield: Typical	Sheffield: Emergent	Liverpool: Traditional	Liverpool: Typical	London: Traditional	London: Typical	London: Emergent	Standard Scottish: Typical	Glasgow: Traditional	Glasgow: Typical	Glasgow: Emergent	Tyrone: Traditional	Tyrone: Typical	Dublin: Traditional	Dublin: Typical	Dublin: Emergent	Standard American: Typical		
PGc	0.402	0.418	0.413	0.404	0.399	0.400	0.405	0.418	0.403	0.412	0.372	0.382	0.375	0.417	0.420	0.419	0.382	0.401	0.416	0.390	0.396	0.405	0.416	Proto-Germanic	1
	RPG	0.871	0.923	0.928	0.894	0.884	0.907	0.915	0.836	0.872	0.854	0.890	0.854	0.848	0.754	0.837	0.774	0.764	0.823	0.839	0.817	0.846	0.861	RP: Typical	2
		TynT	0.907	0.869	0.872	0.859	0.866	0.848	0.796	0.827	0.781	0.813	0.786	0.804	0.747	0.787	0.756	0.747	0.790	0.808	0.768	0.781	0.791	Tyneside: Traditional	3
			Tyn	0.917	0.937	0.883	0.905	0.904	0.826	0.864	0.824	0.858	0.825	0.842	0.753	0.831	0.780	0.771	0.826	0.832	0.813	0.817	0.838	Tyneside: Typical	4
				TynS	0.928	0.867	0.891	0.927	0.851	0.866	0.816	0.851	0.846	0.827	0.747	0.825	0.757	0.750	0.813	0.818	0.801	0.831	0.847	Tyneside: Standardised	5
					TynE	0.849	0.873	0.910	0.836	0.856	0.804	0.836	0.834	0.822	0.736	0.829	0.758	0.736	0.795	0.839	0.810	0.816	0.803	Tyneside: Emergent	6
						ShfT	0.953	0.880	0.801	0.833	0.810	0.811	0.786	0.813	0.733	0.809	0.748	0.743	0.796	0.813	0.776	0.782	0.811	Sheffield: Traditional	7
							ShfG	0.906	0.821	0.853	0.802	0.831	0.808	0.828	0.744	0.824	0.762	0.752	0.808	0.831	0.802	0.804	0.832	Sheffield: Typical	8
								ShfE	0.853	0.889	0.816	0.848	0.838	0.819	0.742	0.820	0.751	0.727	0.786	0.846	0.792	0.814	0.817	Sheffield: Emergent	9
									LplT	0.941	0.739	0.774	0.758	0.766	0.707	0.771	0.719	0.701	0.756	0.787	0.770	0.796	0.779	Liverpool: Traditional	10
										LplG	0.771	0.808	0.788	0.799	0.729	0.801	0.739	0.727	0.786	0.799	0.782	0.807	0.810	Liverpool: Typical	11
											LonT	0.940	0.912	0.755	0.699	0.751	0.770	0.691	0.738	0.774	0.724	0.751	0.763	London: Traditional	12
												Lon	0.930	0.789	0.727	0.783	0.802	0.721	0.771	0.800	0.757	0.785	0.796	London: Typical	13
													LonE	0.778	0.724	0.773	0.786	0.708	0.758	0.766	0.746	0.774	0.781	London: Emergent	14
														SSE	0.831	0.947	0.810	0.843	0.897	0.778	0.862	0.879	0.895	Standard Scottish: Typical	15
															GlaT	0.863	0.834	0.770	0.785	0.719	0.748	0.761	0.783	Glasgow: Traditional	16
																GlaG	0.846	0.818	0.864	0.791	0.847	0.860	0.861	Glasgow: Typical	17
																	GlaE	0.734	0.773	0.713	0.736	0.747	0.765	Glasgow: Emergent	18
																		TyrT	0.887	0.722	0.806	0.810	0.829	Tyrone: Traditional	19
																			TyrG	0.760	0.851	0.874	0.891	Tyrone: Typical	20
																				Dub	0.784	0.799	0.788	Dublin: Traditional	21
																					Dub	0.920	0.867	Dublin: Typical	22
																						DubE	0.901	Dublin: Emergent	23
																							StAG	Std. American: Typical	24

1.4 Visualisations: Divergence of English Through Space & Time



Early & Late Modern English

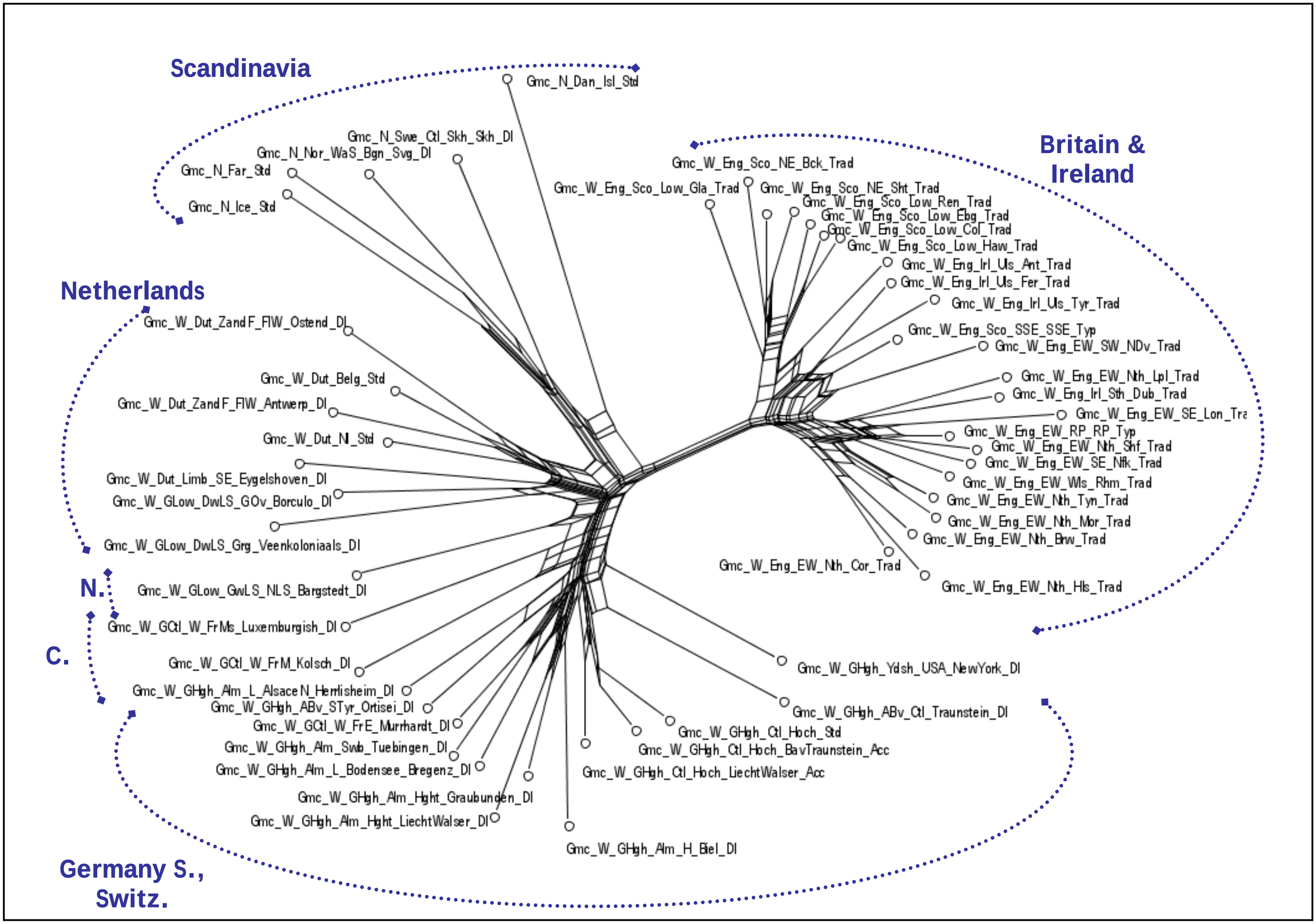
Old English

Middle English

Modern Traditional Scotland

Historical Scots
= Middle and E.M. English Periods

1.5 Phonetic Divergence Between Dialects in Germanic


Scandinavia
Gmc_N_Dan_Isl_Std
Gmc_N_Swe_Ctl_Skh_Skh_Dl
Gmc_N_Nor_WaS_Bgn_Svg_Dl
Gmc_N_Far_Std
Gmc_N_Ice_Std
Britain &
Ireland
Gmc_W_Eng_Sco_NE_Bck_Trad
Gmc_W_Eng_Sco_Low_Gla_Trad
Gmc_W_Eng_Sco_NE_Sht_Trad
Gmc_W_Eng_Sco_Low_Ren_Trad
Gmc_W_Eng_Sco_Low_Ebg_Trad
Gmc_W_Eng_Sco_Low_Col_Trad
Gmc_W_Eng_Sco_Low_Haw_Trad
Gmc_W_Eng_Irl_Uls_Ant_Trad
Gmc_W_Eng_Irl_Uls_Fer_Trad
Gmc_W_Eng_Irl_Uls_Tyr_Trad
Gmc_W_Eng_Sco_SSE_SSE_Typ
Gmc_W_Eng_EW_SW_NDv_Trad
Gmc_W_Eng_EW_Nth_Lpl_Trad
Gmc_W_Eng_Irl_Sth_Dub_Trad
Gmc_W_Eng_EW_SE_Lon_Tra
Gmc_W_Eng_EW_RP_RP_Typ
Gmc_W_Eng_EW_Nth_Shf_Trad
Gmc_W_Eng_EW_SE_Nfk_Trad
Gmc_W_Eng_EW_Wls_Rhm_Trad
Gmc_W_Eng_EW_Nth_Tyn_Trad
Gmc_W_Eng_EW_Nth_Mor_Trad
Gmc_W_Eng_EW_Nth_Brw_Trad
Gmc_W_Eng_EW_Nth_Cor_Trad
Gmc_W_Eng_EW_Nth_Hls_Trad
Netherlands
Gmc_W_Dut_ZandF_FlW_Ostend_Dl
Gmc_W_Dut_Belg_Std
Gmc_W_Dut_ZandF_FlW_Antwerp_Dl
Gmc_W_Dut_Nl_Std
Gmc_W_Dut_Limb_SE_Eygelshoven_Dl
Gmc_W_GLow_DwLS_GOv_Borculo_Dl
Gmc_W_GLow_DwLS_Grg_Veenkoloniaals_Dl
N.
Gmc_W_GLow_GwLS_NLS_Bargstedt_Dl
Gmc_W_GCtl_W_FrMs_Luxemburgish_Dl
C.
Gmc_W_GCtl_W_FrM_Kolsch_Dl
Gmc_W_GHgh_Alm_L_AlsaceN_Herrlisheim_Dl
Gmc_W_GHgh_ABv_STyr_Ortisei_Dl
Gmc_W_GCtl_W_FrE_Murrhardt_Dl
Gmc_W_GHgh_Alm_Swb_Tuebingen_Dl
Gmc_W_GHgh_Alm_L_Bodensee_Bregenz_Dl
Gmc_W_GHgh_Alm_Hght_Graubunden_Dl
Gmc_W_GHgh_Alm_Hght_LiechtWalser_Dl
Gmc_W_GHgh_Alm_H_Biel_Dl
Gmc_W_GHgh_Ctl_Hoch_LiechtWalser_Acc
Gmc_W_GHgh_Ctl_Hoch_BavTraunstein_Acc
Gmc_W_GHgh_Ctl_Hoch_Std
Gmc_W_GHgh_ABv_Ctl_Traunstein_Dl
Gmc_W_GHgh_Ydsh_USA_NewYork_Dl
Germany S.,
Switz.

1.6 PROJECT ORIGINS AND DEVELOPMENT

- 2004-2007 funding: Arts and Humanities Research Council, UK
  - Linguistics, University of Sheffield:
    Quantitative Methods in Comparative Linguistics
  - Linguistics, University of Edinburgh:
    Sound Comparisons: Dialect and Language Comparison and Classification by Phonetic Similarity
- 2006-2009 funding: Leverhulme Trust, UK
  - *McDonald Institute for Archaeological Research*,
    University of Cambridge:
    Languages and Origins in Europe
- 2011-2015 funding: Max-Planck-Gesellschaft
  - Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig.

1.7 Data and Earlier Websites

- Language recordings collected since 2000, continuing whenever possible.
  - *Sounds of the Andean Languages*
    www.quechua.org.uk/sounds
  - *Accents of English from Around the World*
    www.soundcomparisons.com
  - Regional dialects and languages of *Germanic* (+ Romance, Balto-Slavic)
    www.languagesandpeoples.com

- *Any* language family can easily be added, now that system is set up…

1.8 People

- **Website**: originally by Heggarty, but now completely recreated by Jakob Runge (Uni Leipzig).

- **Phonetic transcriptions**:
  - English and Germanic: Warren Maguire (Edinburgh).
  - Andes: Heggarty, Scott Sadowsky (UFRO, Chile).

- **Data collection**:
  - English dialects: Warren Maguire.
  - All others: Paul Heggarty.

- **Initial funding/direction**: April McMahon (Aberystwyth).

- Hundreds of native-speakers!

1.9 Aims of *Website*

- Make use of databases of recordings and phonetic transcriptions.
- Fundamental purpose: compare pronunciations of ‘same’ cognates.
- To serve two user groups together.
- Speakers of language varieties concerned, i.e. general public.
  - Esp. for endangered language varieties.
  - Regional languages/‘dialects’ of main European families.
  - Indigenous languages of Peru, Bolivia, Ecuador: Quechua and Aymara.
  - Part of wider website to support literacy, through understanding and uptake of proposed standard orthography problems.
- Linguistics researchers.
  - Make database valuable and searchable for their ends.

2. Features

- User-friendly, accessible, no specialised linguistic knowledge needed:
  - Sound files.
  - Maps.

- Powerful research tool:
  - IPA transcriptions.
  - 'Linguistically informed' functions.

- User can tailor site to specific interests:
  - Select languages / words / sounds.
  - Select any combination of these.

2.1 Languages Axis

- Map view, includes:
  - Zooming on selected regions.
  - Only showing selected languages of interest.

- Add transcriptions of any known historical varieties of a language.

- Add (hypothesised) phonetics for a family’s proto-language.

  e.g. → Compare all modern reflexes of Proto-Germanic word-initial /t/.

2.2 WORDS AXIS: SEARCH/FILTER FUNCTIONS

- By spelling, i.e. graphemes (or sequences)…
  - In any language variety that does have a standard orthography.

- By sounds, i.e. symbols in IPA…
  - In transcription of any language variety in database.
  - Including IPA diacritics, e.g. vowel length [ː].

- In both, ‘advanced search’ features:
  - Results filtered in real-time as search string is typed.
  - ‘Regular expressions’ to search for contexts, e.g. ʃ$ = word-final [ʃ], etc.

- Add family- or language-specific data to search by:

  e.g. Wells’ (1982: 127-67) “lexical sets” for English dialectology.

  e.g. Use upper case for: C, V, archiphonemes N, R, etc.

2.3 Combined Selections: Words and Languages

- Compare on one screen multiple selected words *and* languages.

  e.g. Numerals 1 to 10 — In all languages.

  e.g. All words for body parts — In all Scandinavian varieties.

  e.g. All words with <r> in English spelling — In all English varieties.

  e.g. All words that contain [ɫ] in RP — In all English varieties.

  e.g. All words that had Proto-Germanic [k] — In all Continental Germanic.

2.4 Website User Language: Multilingual Support

- ‘Outreach’: promote awareness and understanding of regional languages.
  → Make site available in such languages themselves.

- Collaborative: enter translations of site language remotely online.
  (Password protected.)

3. Web Policies

- Free, collaborative (site language translations), open to new families!

- Ensure that website functions in all browsers.
- Sound files available in two formats: .mp3 and .ogg.

- No static webpage at all: all pages generated in real time, ‘on the fly’.
- ‘Links’ and ‘addresses’ are just queries to underlying database.
- Words and languages selected appear in address line,
  so can be typed in to search/filter all pages previously visited.
- Linked data (‘semantic web’).

3.1 Some Technical Data…

- Total size of programme:
  - Only 6 MB of code (+ images + sound files).
  - 6129 lines of PHP.
  - 1038 lines of Javascript.
  - 506 lines of SQL.

- Which technologies?
  - PHP to generate the website on demand.
  - MySQL as database backend for PHP script.
  - Javascript for more powerful features and speed.

- Any technical questions?

  → Ask Jakob Runge.

3.2 LINKS TO OTHER RESOURCES ON LANGUAGES COVERED

- Link to entries on same language varieties in:
  - Wikipedia, Ethnologue, Glottolog/LangDoc, LLMap, Multitree.
- Problems:
  - In different site languages, names of languages to link to are different.
  - Use ISO language codes wherever possible.
  - Solution thanks to Lexvo and Sebastian Nordhoff (MPI-EVA).
- Dialects/accents very sporadically and inconsistently present, no ISO codes.
  - Some proposals available, otherwise need to create *ad hoc* links.

4. Future Plans

4.1 Extend Existing Databases

- Structure now established, no further programming needed.
- Can now extend coverage to:
  - More *site* languages.
  - More *data* languages within the families already covered.
  - More families / regions.

4.2 ADDING NEW FAMILIES

- For each new family, data required:
  - List of languages (by classification?/by region?), lat/long co-ordinates.
  - List of 'pan-family' cognates (or meanings) for that family.
  - Sound recordings.
  - Phonetic transcriptions (in Unicode fonts).

4.3 A NEW WEBSITE FOR THE *INTERCONTINENTAL DICTIONARY SERIES*

- Re-launch the *Intercontinental Dictionary Series*.
  - Begun by Mary Ritchie Key, 1960s.
  - Now managed by Linguistics Dept, MPI-EVA, Leipzig. http://lingweb.eva.mpg.de/ids

- Also essentially comparative, but in lexis: list of meanings, not cognates.
- A much bigger list: 1450 meanings, structured in semantic categories.
- As also used for: World Loanword Database: http://wold.livingsources.org.

- A couple of hundred minority/endangered languages worldwide.

- Transcriptions to be updated to IPA.
- No original sound recordings, but now add recordings where possible.

4.4 Feedback, Co-Operation?

- Any feedback, suggestions on features?
- Interest in using our structure to showcase *your* data?
- Please let us know…

REFERENCES

Heggarty, P., Maguire, W., & McMahon, A.M.S. 2010. Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories J. Steele, P. Jordan, & E. Cochrane (eds). *Proceedings of the Royal Society B: Biological Sciences* Cultural and Linguistic Diversity(365): p.3829–3843.

Maguire, W., & McMahon, A.M.S. eds. 2011. *Analysing Variation in English*. Cambridge: Cambridge University Press.

Maguire, W., McMahon, A.M.S., Heggarty, P., & Dediu, D. 2010. The past, present and future of English dialects: Quantifying convergence, divergence and dynamic equilibrium. *Language Variation and Change* 22(1): p.69–104.

McMahon, A.M.S., Heggarty, P., McMahon, R., & Maguire, W. 2007. The sound patterns of Englishes: representing phonetic similarity. *English Language and Linguistics* 11(01): p.113.

Wells, J.C. 1982. *Accents of English 1: An Introduction*. Cambridge: Cambridge University Press.