Ollscoil na hÉireann
National University of Ireland

AHRC CECD AHRC CENTRE FOR THE EVOLUTION OF CULTURAL DIVERSITY

Phonetic comparison, varieties, and networks: Swadesh's influence lives on here too.

Jennifer Sullivan and April McMahon,

University of Edinburgh

Outline of presentation

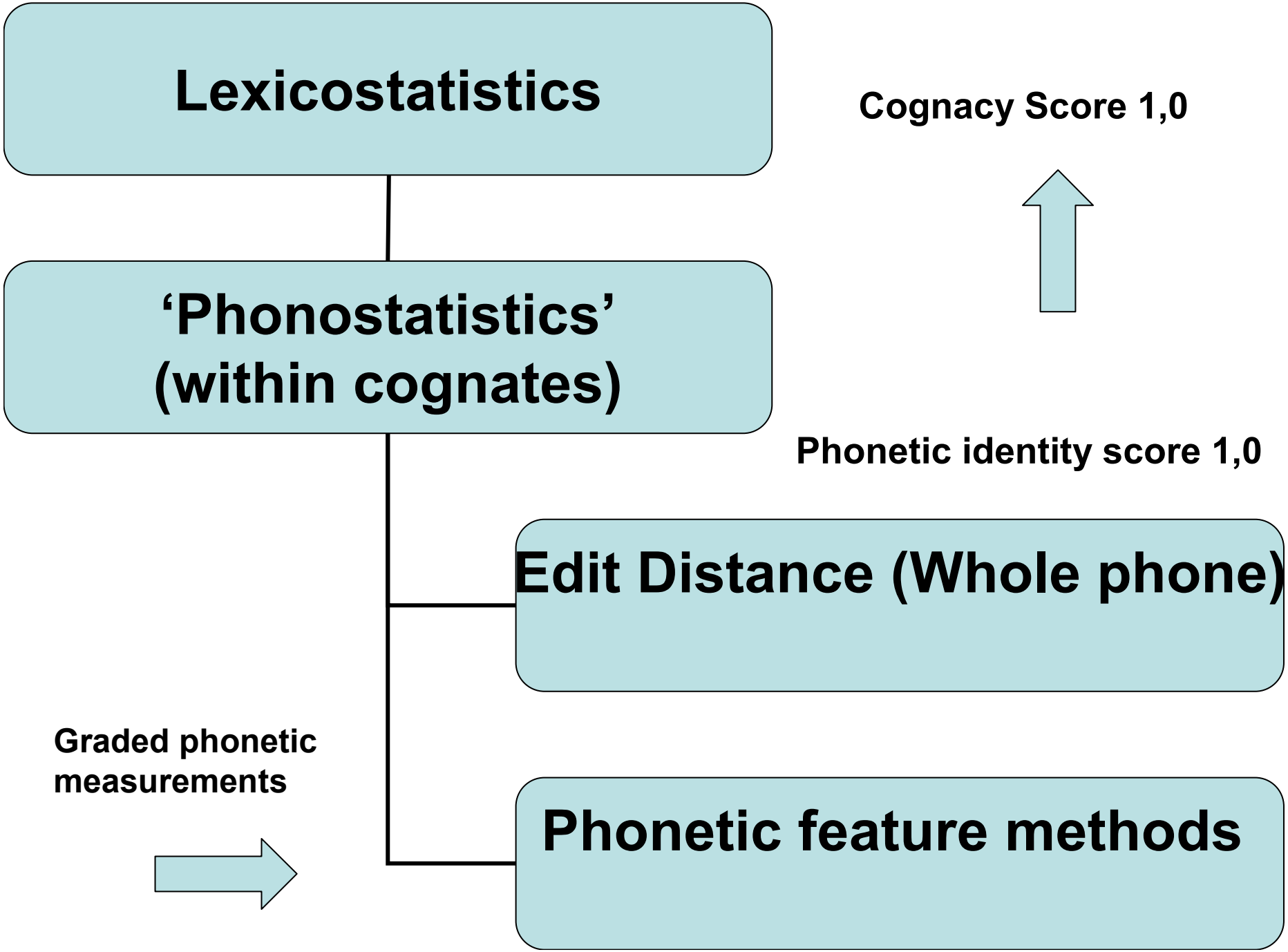
1) The perhaps unexpected relevance of Swadesh here
2) Small-scale comparison of methods measuring phonetic similarity among English/Germanic varieties
3) Implications of results for how we measure phonetic similarity in a synchronic context
4) Begin to tackle question of Chance Phonetic Similarity

Swadesh's Legacy

- **Lexicon**: Ubiquitous 100/200 word lists of basic vocabulary
- Measurement of ***Language*** Distance **(Lexicostatistics)**
- Estimation of dates of ***Language*** splits (Glottochronology)
- **Phonetics**: Papers on **English varieties** and other languages
- Lexicostatistics and Glottochronology equally applied by Swadesh to **Varieties**
- Threshold scores from these techniques for separating Languages from **Varieties** (Swadesh 1950, 1972)

Swadesh's Insights

- Swadesh did **not quantify** phonetic similarity in the manner of Lexicostatistics but interested in English variety **vowel variability** (1947) and explores **isogloss tradition** (1972: 16).
- **"Mesh principle"** (1972: 285-92) argues against ignoring dialect gradation and always assuming clear treelike splits.
- Broached the issue of **chance** in assessing whether languages were related or not.

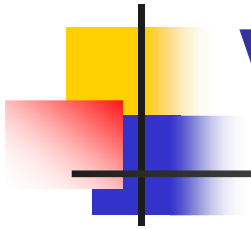

Lexicostatistics
Cognacy Score 1,0
'Phonostatistics'
(within cognates)
Phonetic identity score 1,0
Edit Distance (Whole phone)
Graded phonetic
measurements
Phonetic feature methods

Swadesh 100 list	**Swadesh 200 list**	**Gmc Cognates only**
cold	five	brother
eye	four	daughter
foot	ice	eight
heart	mother	holy
horn	right	home
long	three	nine
mouth		north
one		over
two		six
white		seven
		storm
		swear
		ten
		word

30 word subset

(McMahon et al 2005-07)

Phonetic comparison in Varieties



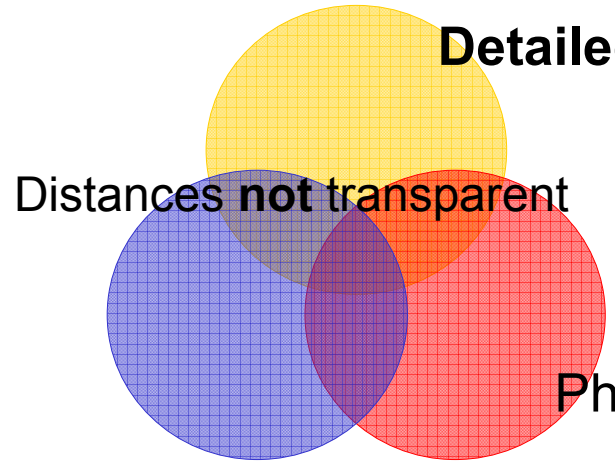
- 2 **Languages**: English, German (Hochdeutsch)
- 4 **Varieties** of English: Std American, RP, Std Scottish, Buckie

Questions

Convergence problem e.g. Kessler (2007), Heeringa (2004)
Do feature methods behave differently or is important information lost?

How much phonetic detail should there be?

Detailed e.g. Heggarty (2000)



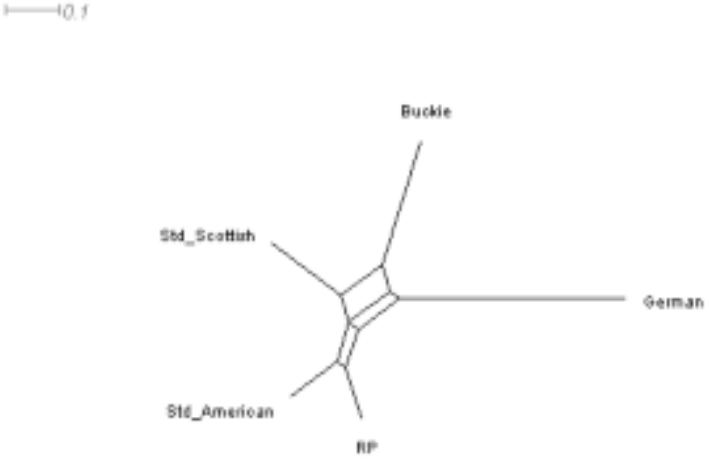
Sparse e.g. Kessler & Lehtonen (2006)

Chance issue unexplored outside historical context

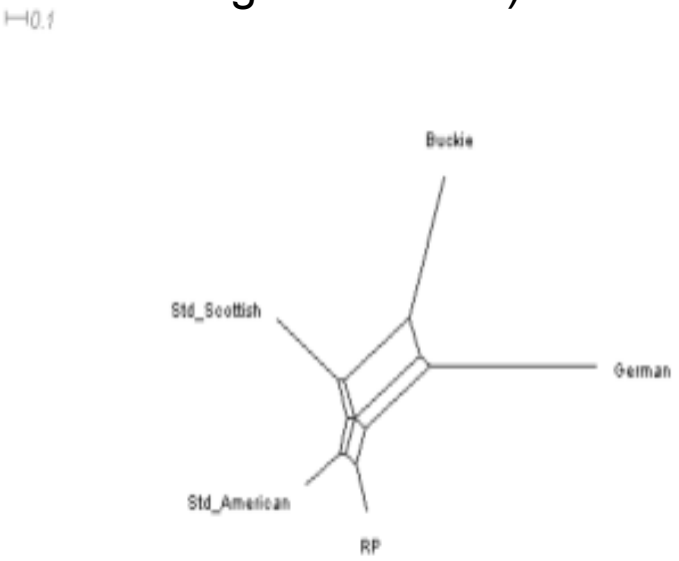
Results-Networks

- Large **convergence** between **Whole Phone** and **Phonetic Feature** methods (especially when aggregate scores used)

Edit Distance (Whole Phone)



Phonetic feature method (Almeida & Braun (1986) original method)



Splitstree-NeighborNet (Huson & Bryant 2006)

Std American vs RP: Similarity/Distance Chasm

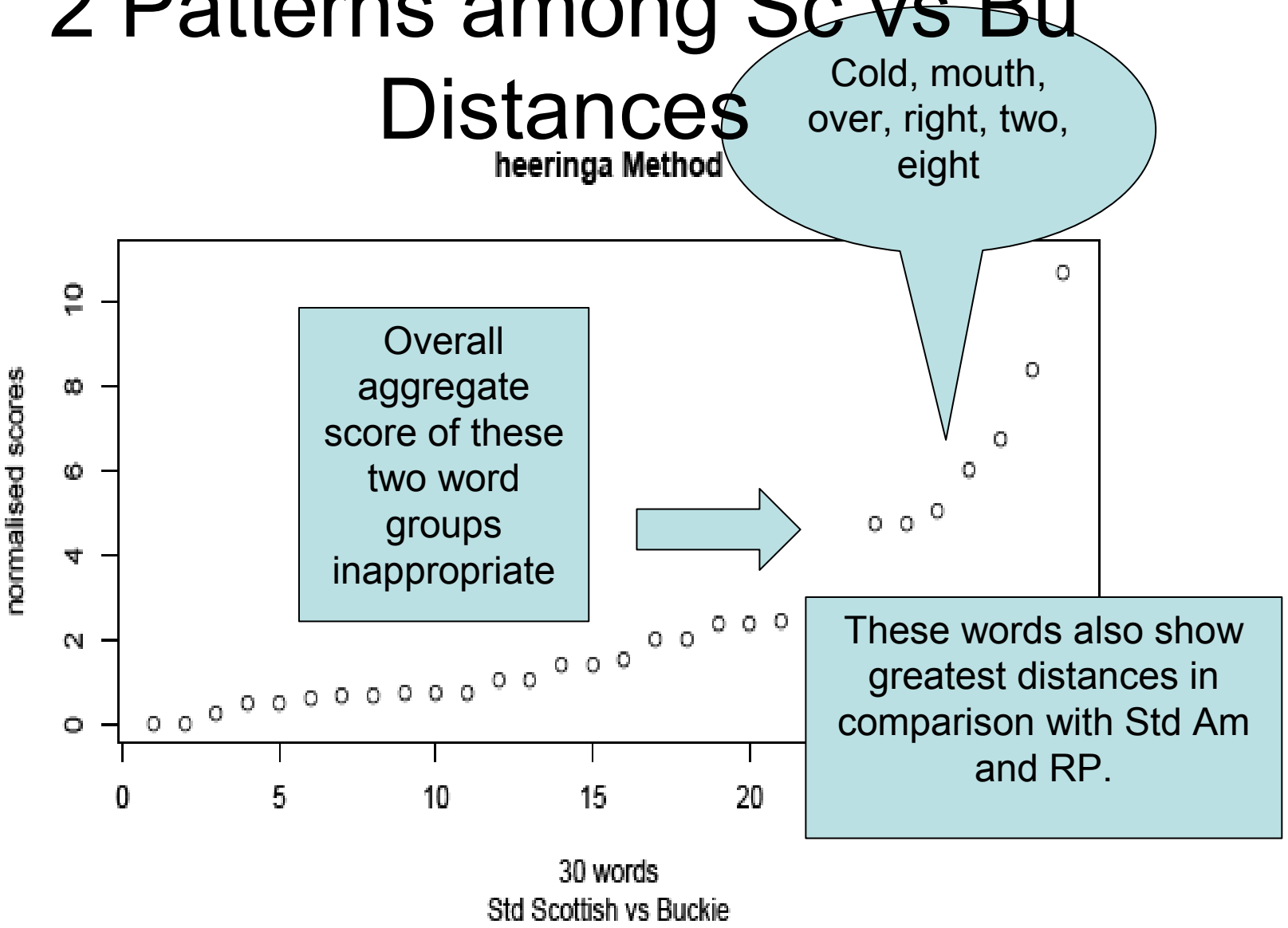
Similarity

- Vowel distances extremely slight overall.
- Always the most similar pair of varieties
- BUT
- Std Dev scores always higher than the mean-aggregate mean score inappropriate.
- Why?

Distance

- **Rhoticity** divide in English varieties (commented on by Swadesh)
- Two-Sample t-test, t -2.599 $p<0.02$
- Heavy weighting of rhoticity-affects impact of subtle phonetic differences e.g. slight vowel differences.

2 Patterns among Sc vs Bu Distances

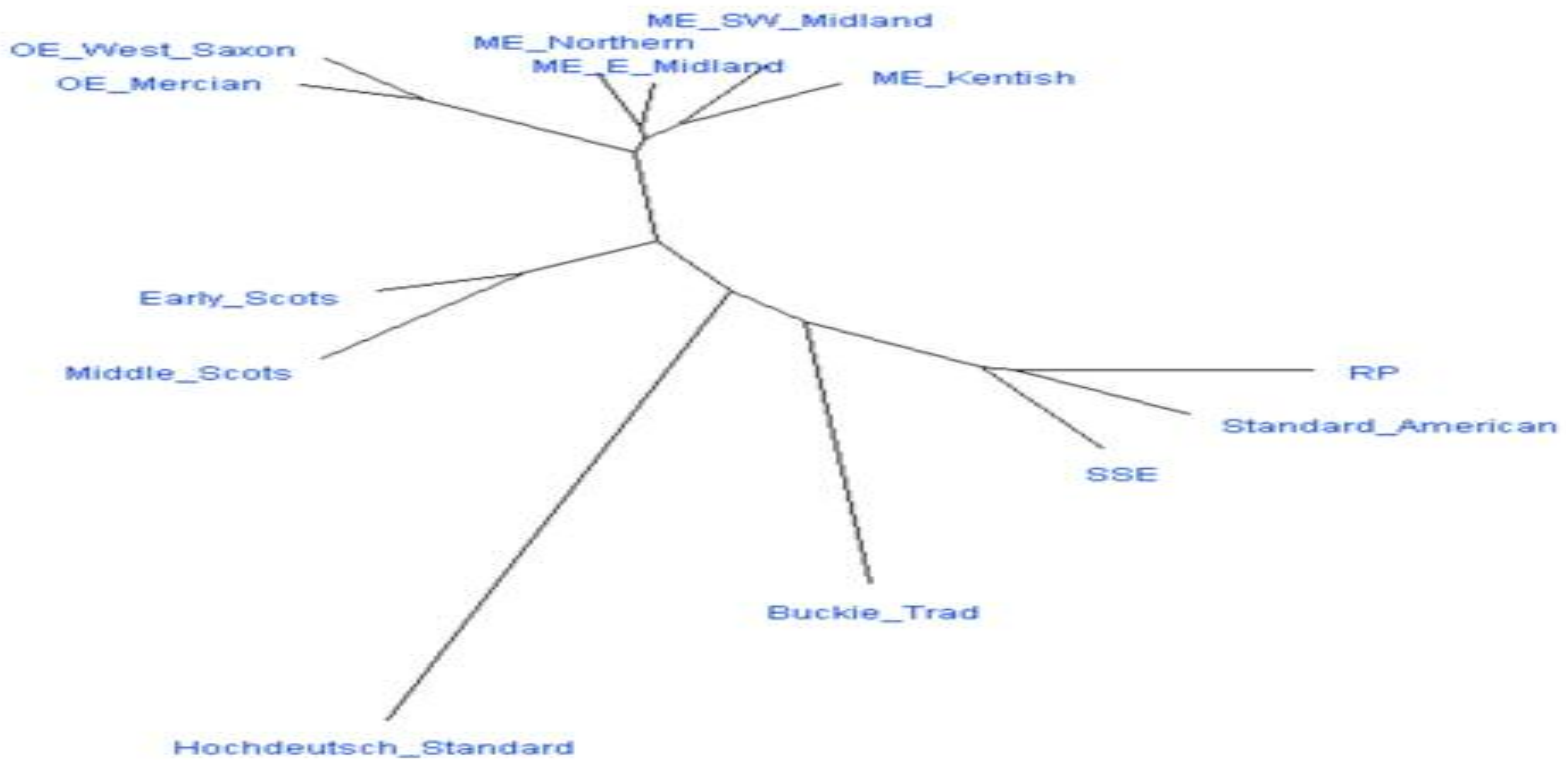


Cold, mouth, over, right, two, eight

Overall aggregate score of these two word groups inappropriate

These words also show greatest distances in comparison with Std Am and RP.

Separate study: Links with Historical Varieties

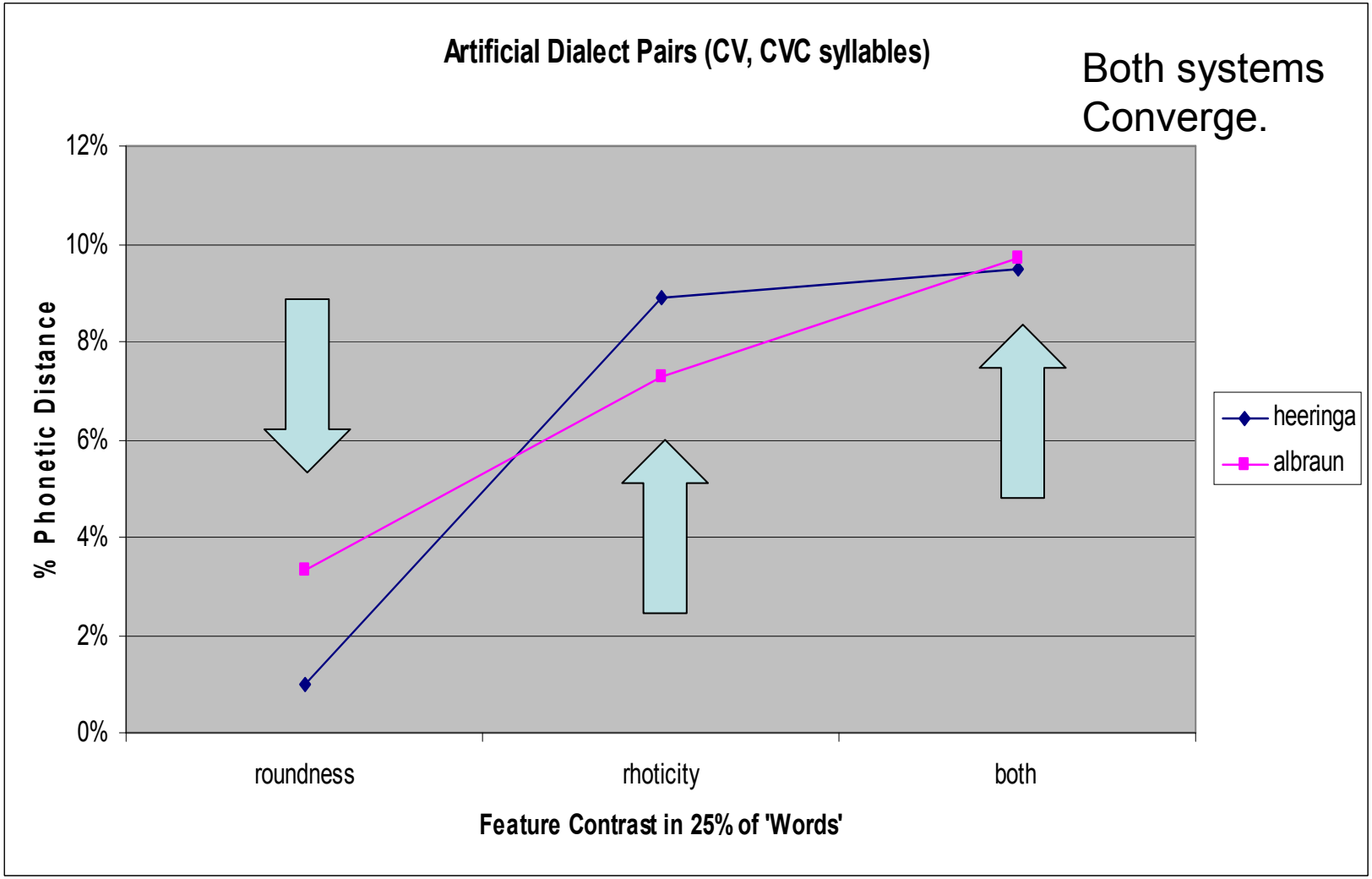


Acknowledgements: April McMahon, Warren Maguire and Paul Heggarty

Differences between systems cancelled out

Original Amsterdam Braun system
Weights roundness
Higher.

Heeringa system
Weights rhoticity
Higher.



Interim Summary

- Convergence of Different methods:

  -Subtle **phonetic feature** differences do not make much impact when alongside heavily weighted elements (e.g. rhoticity).

  -Differences between systems can be cancelled out when features are combined.
- Data may not be *phonetically* unified enough for simple aggregation-Analogy with Borrowed vs Non-borrowed words in the lexicon.

Chance Phonetic Similarity

Approach 1: Permutation testing (Monte Carlo)

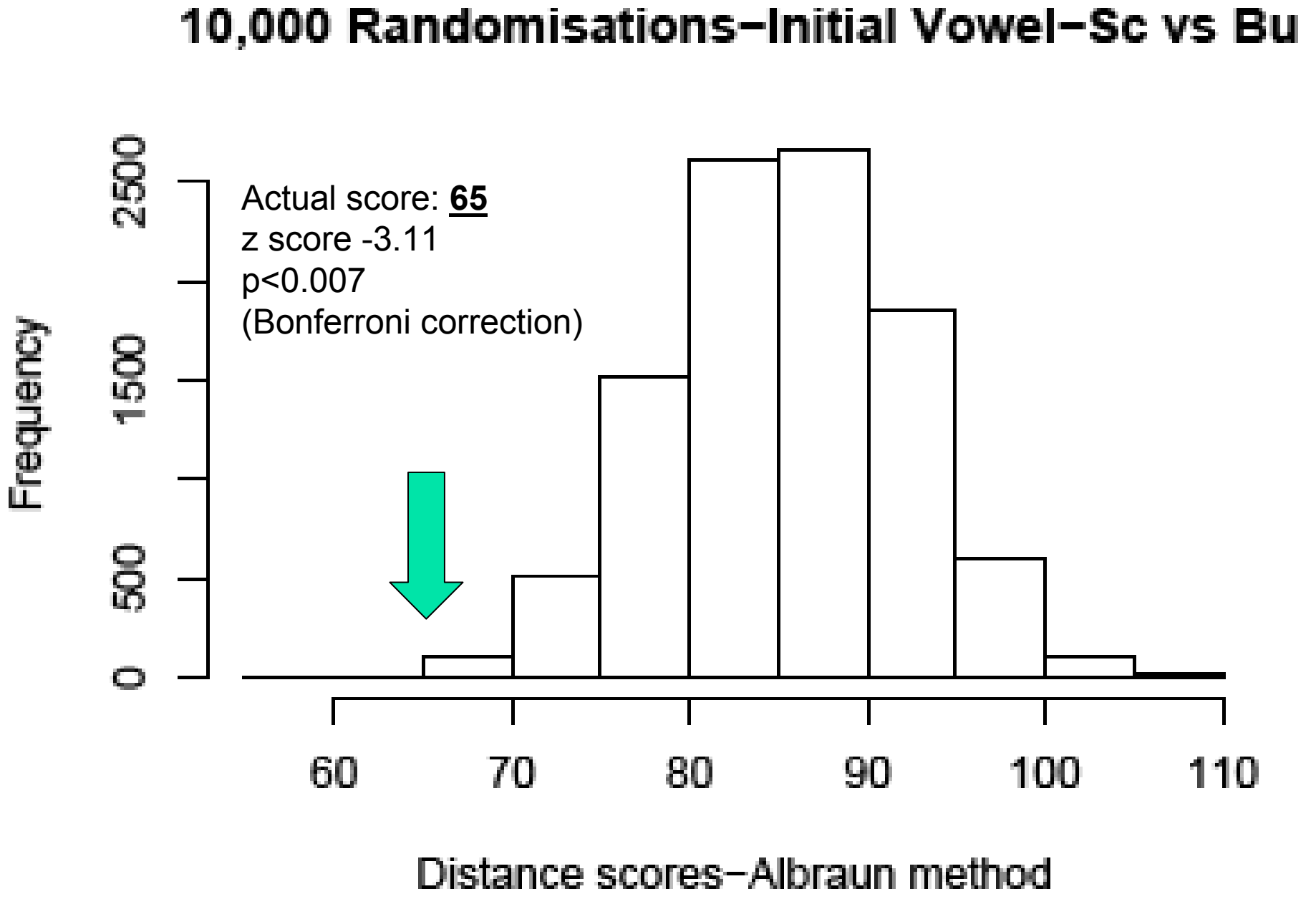
Influenced by Oswalt (1970), Swadesh (1956, 1972) and Baxter & Manaster Ramer (2000).

Previous Studies

- Initial ***consonant***-Historically stable
- Counting consonant '**Matches**'
- Testing **putative language relationships**

Present Work

- Initial ***vowel***-suitable for varieties
- **Sums of distances**
- **Known relationships** but unknown levels of phonetic similarity when cognates are not paired

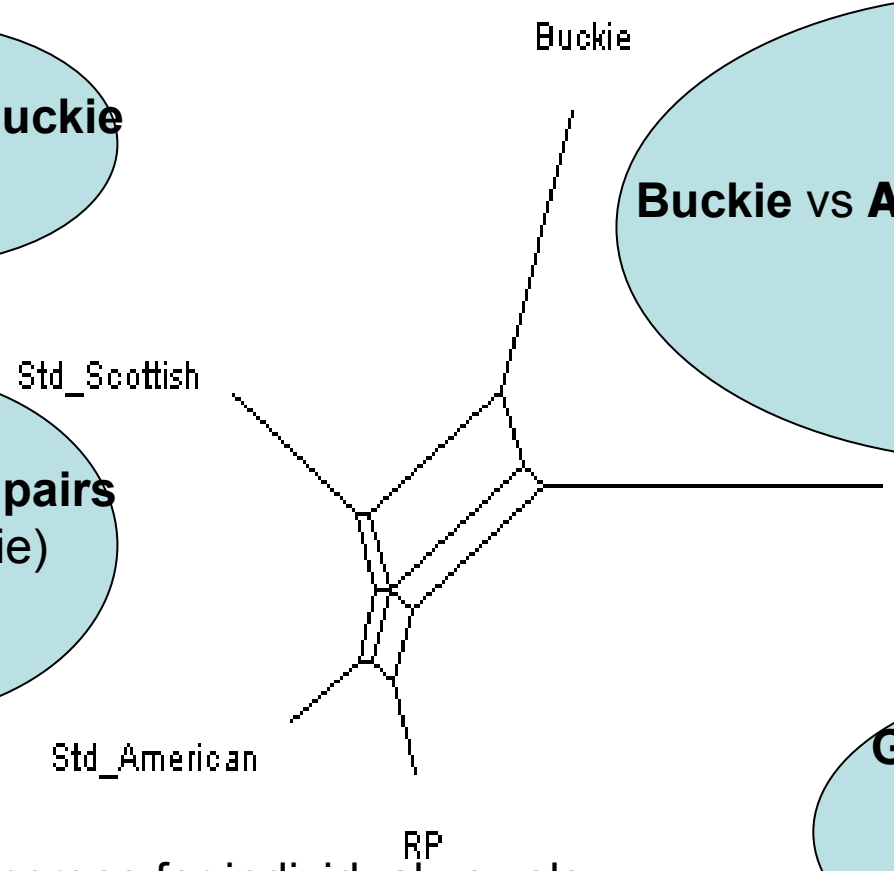

10,000 Randomisations−Initial Vowel−Sc vs Bu
Actual score: 65
z score -3.11
p<0.007
(Bonferroni correction)
Frequency
0
500
1500
2500
60
70
80
90
100
110
Distance scores−Albraun method

H0.1

Bonferroni correction applied in all cases.

Scottish vs **Buckie**
p<0.007

Buckie vs **Am/RP/German** p=0.1
(n.s.)



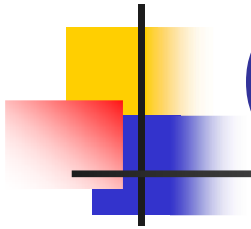
English variety pairs
(except Buckie)
p<0.001

German and **English Pairs** (except Buckie)
n.s.

Similar picture emerges for individual vowels and dipthongs as a unit

BUT Problems with this method… (especially in the context of varieties)

Alternative approaches (under exploration)



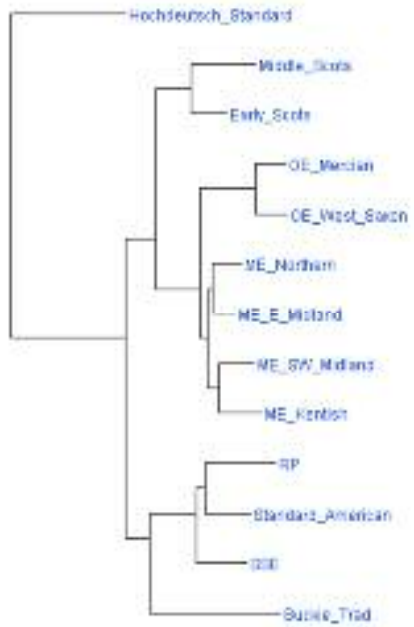
- Is the difference between varieties greater than a baseline of vowel variability modelled on **Drift**?
- Is it surprising that two varieties should share particular vowels given their **frequency** and occurrence **typologically**?
- Are *between*-variety vowel differences greater than known levels of **acoustic variability** *within* a single variety?

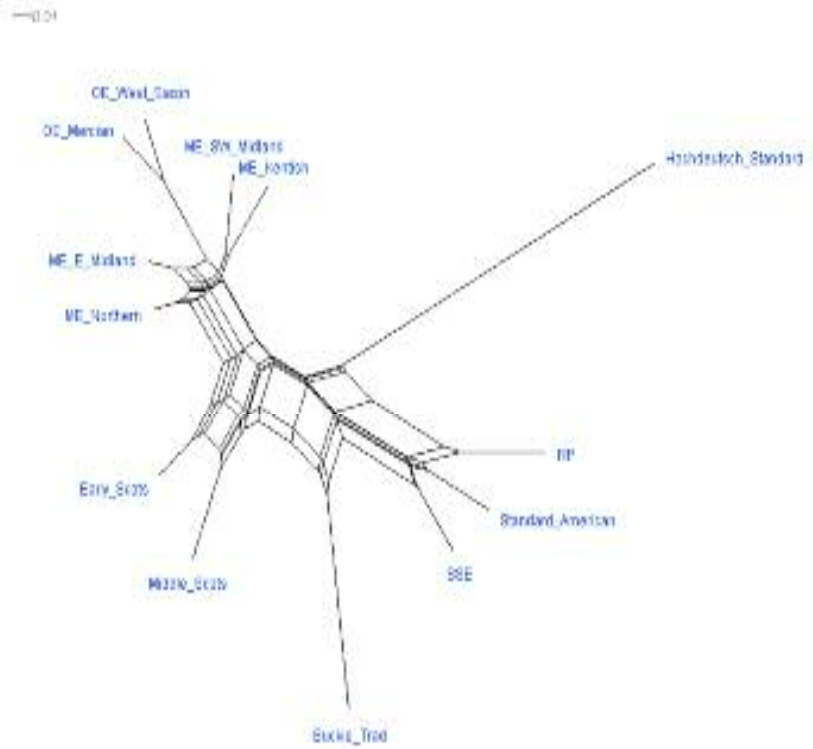
Conclusion

- Methods and ideas of Swadesh very relevant to contemporary work on Synchronic Phonetic Comparison
- 'Phonostatistics'-some current ways of measuring do not maximise subtlety of feature methods.
- Single overall score of phonetic similarity may be inappropriate
- Assessing chance needs to be approached from many angles.

References

- Almeida, Almerindo & Angelika Braun. 1986. 'Richtig' und 'falsch' in phonetischer Transkription: Vorschläge zum Vergleich von Transkriptionen mit Beispielen aus deutschen Dialekten. *Zeitschrift für Dialektologie und Linguistik* LIII-2. 158-72.
- Baxter, William H. & Alexis Manaster Ramer. 2000. Beyond lumping and splitting: Probabilistic issues in historical linguistics. In Renfrew, McMahon & Trask (eds). 2000a, 167-188.
- Forster, Peter & Colin Renfrew (eds.). 2006. *Phylogenetic methods and the prehistory of languages*. Cambridge: McDonald Institute for Archaeological Research.
- Heeringa, Wilbert. 2004. *Measuring dialect pronunciation difference using Levenshtein distance*. Groningen: University of Groningen Doctoral Dissertation.
- Heggarty, Paul. 2000b. Quantifying change over time in phonetics. In Renfrew, McMahon & Trask (eds.). 2000b, 531-562.
- Huson, Daniel & David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Molecular biology and Evolution. 23. 2. 254-67.
- Kessler, Brett. 2007. Word similarity metrics and multilateral comparison. In Nerbonne, Ellison & Kondrak (eds.). 2007a, 6-14.
- Kessler, Brett & Annukka Lehtonen. 2006. Multilateral comparison and significance testing of the Indo-Uralic question. In Forster & Renfrew (eds). 33-42.
- McMahon, April, Warren Maguire & Paul Heggarty. 2005-07. Sound comparions: Dialect and language comparison and classification by phonetic similarity. http://www.soundcomparisons.com/ (Jan 2009)
- Nerbonne, John, T. Mark Ellison & Grzegorz Kondrak (eds.). 2007a. *Proceedings of the Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*. Prague.
- Oswalt
- Renfrew, Colin, April McMahon & Larry Trask (eds.). 2000a. *Time depth in historical linguistics. Vol. 1*. Cambridge: The McDonald Institute for Archaeological Research
- Renfrew, Colin, April McMahon & Larry Trask (eds.). 2000b. *Time depth in historical linguistics. Vol. 2*. Cambridge: The McDonald Institute for Archaeological Research.
- Swadesh, Morris. 1947. On the Analysis of English syllabics. Language 23. 137-50.
- Swadesh, Morris. 1950. Salish internal relationships.International Journal of American Linguistics. 21 121-37.
- Swadesh, Morris. 1956. Problems of long-range comparison in Penutian. Language 32.1. 17-41.
- Swadesh, Morris (ed. Joel Sherzer). 1972. The origin and diversification of language. London. Routledge & Kegan Paul.


Hochdeutsch_Standard
Middle_Scots
Early_Scots
OE_Mercian
OE_West_Saxon
ME_Northern
ME_E_Midland
ME_SW_Midland
ME_Kentish
RP
Standard_American
Dublin_Trad


OE_West_Saxon
OE_Mercian
ME_SW_Midland
ME_Kentish
Hochdeutsch_Standard
ME_E_Midland
ME_Northern
RP
Early_Scots
Standard_American
Middle_Scots
SSE
Buckie_Trad