

Feature Economy in Dialects of Dutch

Marc van Oostendorp
Meertens Instituut

Linguistics in the Netherlands, 2005

Abstract

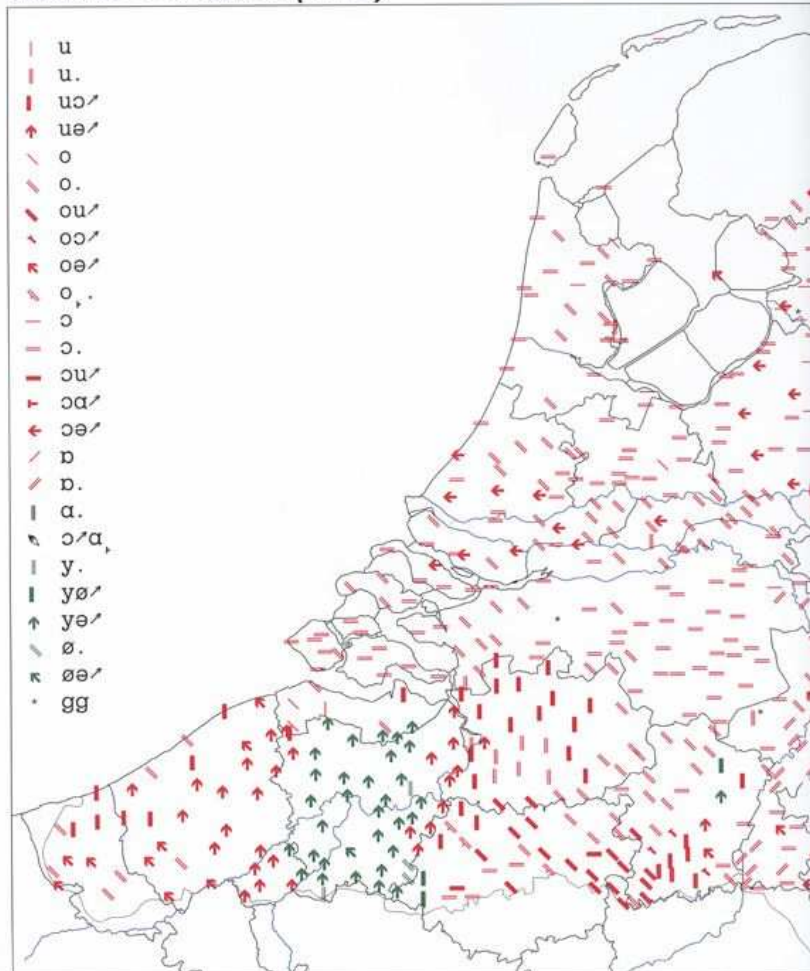
Feature Economy (Clements, 2003) is the hypothesis that “if a feature is used once in a system, it will tend to be used again”. Clements has shown this hypothesis to be a valid generalisation for cross-linguistic databases; in this talk we will examine its value for Dutch dialectology. For this we use the Goeman-Taeldeman-Van Reenen Database, containing a large survey of phonological and morphological variation in the Netherlands. We have derived an inventory of (phonetic) consonants for every dialect, and we show how the feature economy can be tested on a range of features using a small application we have written, DutchInventory. The results seem to confirm the Feature Economy Hypothesis, and by extension, a feature-based view of phonological structure.

1 Structure of FAND

- Based on the Goeman-Taeldeman-Van Reenen Project (GTRP):
- A rather traditional, historically oriented, atlas
- Volumes I (on short vowels) and II (on long vowels) appeared in 2000 and 2002
- Volume III (on consonants) is due in 2005
- FAND is strongly oriented towards history, we do not necessarily know what is e.g. the consonantal system of a modern dialect
- It is hard (if not impossible) to study inventories on the basis of maps of individual segments

- 71 -

kaart nr. 31: boren (1114)



Boren (1114)

Verklaring van de symbolen in de legenda:

u	[u]	ou'	[ou'], [əu]
u.	[u:], [u:], [u:], [θ:], [θ:]]	oa'	[o.,a'], [ə.,a']
uo'	[uo'], [u.o'], [u..o'], [u:ɔ'], [u:ɔ']	oe'	[œ'], [ə.ə'], [ə..ə'], [ə:ə']
ua'	[u,ə'], [u.ə'], [u..ə'], [u.ə'], [u.ə'], [u.,ə'], [u:ə'], [u:ə'], [u:ə']	v	[v:]
o	[o], [o:], [o:], [u:]	v.	[v:], [v:], [o:], [o:']
o.	[o:], [o:], [o:], [o:], [o:], [o:], [o:], [o:], [o:], [u:], [u:]	a	[a:]
ou'	[o,u'], [o:u'], [o:u']	o/a	[o/a:]
oo'	[o:ɔ:], [u:ɔ']	y	[y:]
		ye'	[y:ə']
		ye'	[yə'], [y.ə'], [y.ə'], [y..ə']

2 Using a database

2.1 Principles of GTR

- a database of approximately 1,000,000 words
- from 612 traditional dialects spoken in the Netherlands and Flanders.
- data are collected between 1980-1995
- Database can be accessed at <http://www.meertens.knaw.nl/projecten/mand/>
- Project had as its ideal (in 1978) to answer “research questions for dialectologists in the next few decades” (Goeman & Taeldeman, 1996)
- These were supposed to be questions from ‘formal linguistics and language typology’

2.2 A few example questions from 1978:

1. To what extent are sound systems organized following principles of harmony/economy? (N.B. this is an old research question dating back to the Prague circle)
2. How frequent do we find certain distinctions (e.g. in the case of vowels: [front]-[back] and in the case of front vowels: [\pm round])
3. Are there any recurrent patterns of allophonisation?
4. Do we find a correlation between the functional load of a phoneme and its stability?

2.3 Structure of the database

The following is an example of an entry in GTRP:

(1)	1.	2.	3.	4	5
	E192p	130	6n dr7a2.t	'n draad	a thread
	E192p	131	dr7a2.<d6n	draden	threads
	E192p	132	6n dr7a.>ts2i	'n draadje	a thread-DIM
	E192p	133	6n dr7o5po2l4	'n druppel	a drop
	E192p	134	6n do7y_f	'n duif	a pigeon
	E192p	135	do7y_v6n	duiven	pigeons
	E192p	136	6n do7y_fi	'n duifje	a pigeon-DIM

The three leftmost columns are actually in the database. In column 1 we find the so-called Kloeke-code, a geographical code which represents

a city, town or village in the Netherlands (Kloeker-code E192 represents Utrecht), and in column 2 we find the question number (these thus are questions 130-136). In column 3 there is the actual answer to the question in K-IPA, a pre-SAMPA rendering of IPA in ASCII letters.

Letters correspond to IPA symbols, numbers and other symbols are diacritics (except '6'=ə)

The fourth column gives the actual 'question', i.e. the Dutch word which had to be translated into the dialect, and the fifth column gives an English gloss.

3 Feature economy

3.1 Principles

Clements (2003) revives 'an old research question dating back to the Prague circle' on the basis of macrotypology (i.e. the UPSID database, cf. van de Weijer & Hinskens (2004)):

- (2) a. "A sound S will have a higher than expected frequency in languages that have another sound T bearing one of its features, and vice versa"
- b. If a language L_1 had /p/ and L_2 does not, the chance that L_1 also has /b/ is larger than the chance that L_2 has the same sound

Compare the following two hypothetical consonant inventories:

1.

		[labial]	[coronal]	[velar]
[-son]	[-vc]	p	t	k
	[+vc]	b	d	g
[+son]		m	n	ŋ
2. {t, g, ŋ, ʃ, k}

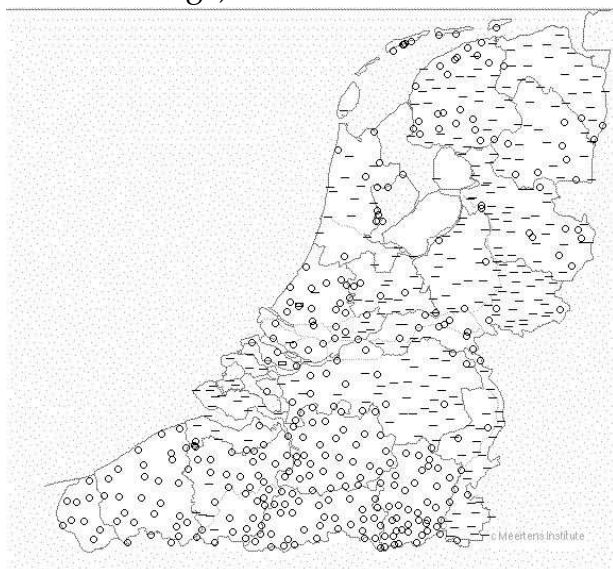
- If phonetics (perception) would be an organizing principle of sound inventories, one might expect (2) to be predominant
- Assuming Feature Economy, we expect (1) to be predominant
- This issue cannot be studied on the basis of one language system alone, it is a typological question by definition

Q: To what extent are the consonant inventories of Dutch dialects organized according to principles of feature economy?

3.2 Method

- We automatically derived phonetic segment inventories for each of the 612 dialects in the GTRP database
- Average size of inventories: 208 distinct segments
- Standard Dutch has approximately 50 ‘phonemes’
- Therefore, the transcriptions are presumably phonetic

Examining the inventories, we discover there is a difference between the Netherlands and Flanders (dots represent dialects with fewer transcriptions than average, horizontal lines dialects with more transcriptions than average):



Probably this is due to a difference in methodology between transcribers. Some of them draw 70 distinctions on average (i.e. a phonological distinction, others over 350.)

3.3 The structure of a segment

- There is no reason to expect that Dutch dialects will differ in the number of e.g. /b/'s
- We will therefore have to concentrate on more fine-grained distinctions if we want to study Feature Economy effects

A simple way of doing this, given the structure of the database, is by looking at diacritics.

- One diacritic is { for dentality (i.e. K-IPA $\tau\{$ corresponds to IPA $\tau\}$)

- This only occurs on coronals {t, d, s, n, l, z}

Given Feature Economy, we hypothesize the following:

- (3) The chance of having $\underset{\bar{n}}{d}$ is bigger in dialects with $\underset{\bar{n}}{t}$, than it is in dialects without $\underset{\bar{n}}{t}$?

3.4 Chance of finding a pair of dentals

Using a set of specifically designed (Python) scripts to search the database of phonetic segment inventories derived from GTRP. Given this, we found the following:

- The chance that we find $\underset{\bar{n}}{d}$ is $(14+11) / 612 = 0.04$.
- The chance that we find $\underset{\bar{n}}{t}$ is $(11+41) / 612 = 0.09$.
- Therefore, the chance that we find $\underset{\bar{n}}{d}+\underset{\bar{n}}{t}$ should be 0.04×0.09
- We thus should find $0.04 \times 0.09 \times 612 = 2$ instances of the combination. But we find 14.

Frequency of dentals:

	$\underset{\bar{n}}{t}$	$\neg\underset{\bar{n}}{t}$	$\chi^2 = 70.43$ $p < 0.001$
$\underset{\bar{n}}{d}$	14	11	
$\neg\underset{\bar{n}}{d}$	41	546	

Obviously, most dialects do not have any dental segment at all. But disregarding these (which are not informative as to our present question), we find that indeed the combination of $\underset{\bar{n}}{d}$ and $\underset{\bar{n}}{t}$ is much more frequent than we expect on the basis of the relative frequencies of each of these sounds individually (but notice that this does not mean that there are more dialects with the combination than there are dialects which have only one of the two dentals).

3.5 Outlook

Because there are such huge differences among transcribers, we are not sure that the effect we found is not due to them.

- A following step should be based on the following definition of Clements:
 - “Feature economy can be quantified in terms of a measure called the economy index.
Given a system using F features to characterize S sounds, its economy index E is given by the expression $E = S/F$ ”

- Are big inventories more economical than big small?
- Initial calculations suggest that they are, although the difference is not big: for large (>203) segment inventories, the economy index is 5.1 on average; for small ones (<204) this is 2.5

Bibliography

- Clements, G.N. (2003). 'Feature Economy in Sound Systems'. *Phonology*, 20, 3: 287–333.
- Goeman, Ton & Johan Taeldeman (1996). 'Fonologie en morfologie van de Nederlandse dialecten. Een nieuwe materiaalverzameling en twee nieuwe atlasprojecten'. *Taal en Tongval*, 48, 1: 38–59.
- van de Weijer, J.M. & F. Hinskens (2004). 'Markedness and Segmental Complexity: A Cross-Linguistic Study'. In: *GLOW Phonology Workshop, Thessaloniki*.