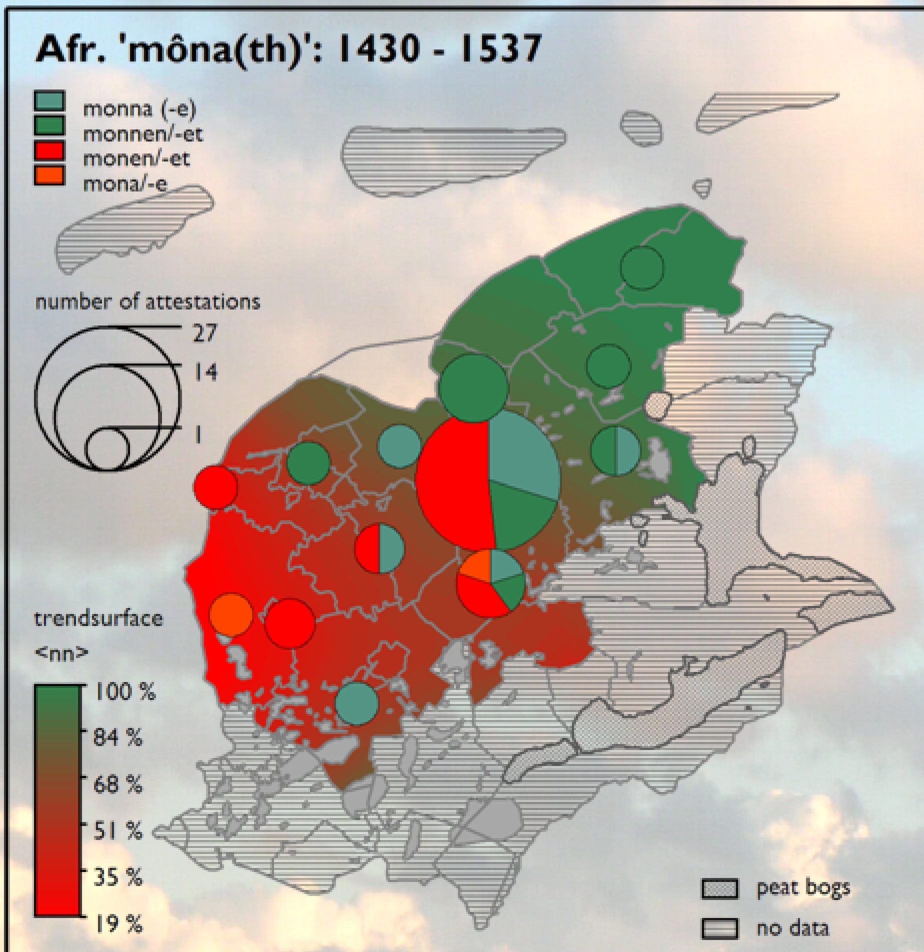
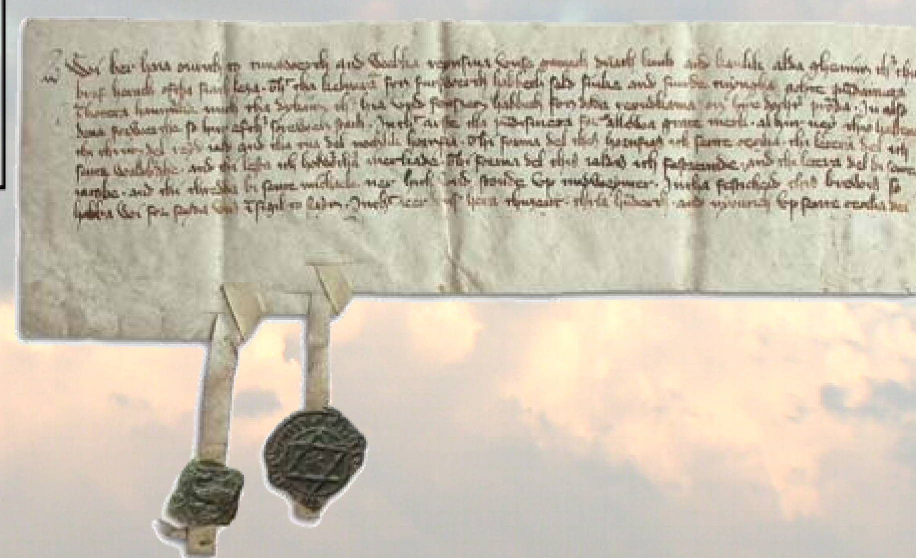


Arjen P. Versloot

Using late-mediaeval sources for linguistic reconstructions



Model I:

One-third of the authors uses variant A and two-thirds variant B. This implies that the variance per author is minimal, for instance, completely predictable after the first observation in a text, but for the corpus, maximal: Charters from three different authors are needed to establish the 1:2 ratio, but per author, one form is sufficient. In this model it is preferable to count per text or even per author;

Model II:

Every author uses both alternatives in the ratio 1:2 in a random alternation in his writings, (for example, within one charter). If this is applied consistently, the individual variance equals the population variance. With a sufficient number of tokens to track the variation with a significant estimation, it is possible to deduce the variation among the entire population. If the inter-speaker variance is zero, it does not matter whether there is a long text from one author or two shorter ones from two authors, assuming that both are representative for that period and region. In this model, token count is the preferred method.

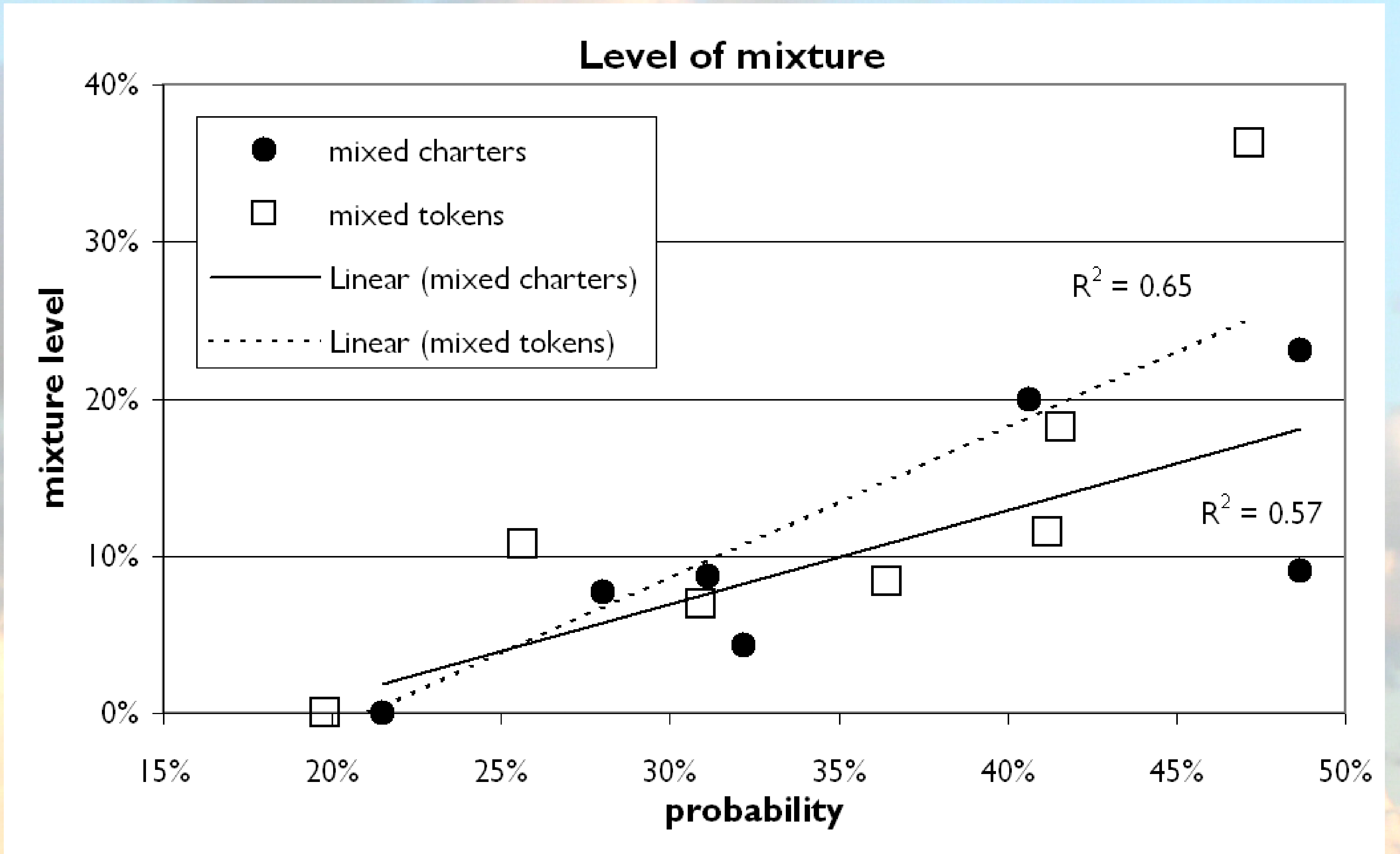
The mixture of forms in one charter is only possible if two conditions are fulfilled:

- The two variants must be co-existing at that time;
- A charter must contain at least two tokens for the relevant form to be able to show different variants in one charter.

seka vs. *seken* 1460-1500	*-a*	*-en*	%-*a*
All tokens in this period	21	30	41%
All charters in this period with an attestation to...	18	19	49%
	both *-a* and *-en*	only *-a* or *-en*	% mixed
Tokens in charters with at least two examples	3	23	12%
Charters with...	1	10	9%

The following cases have been studied:

• *seke* ‘case’	singular -e or ø
• *seka* ‘cases’	plural -a or -en
• *bitalad* ‘paid’	ending with <a> or <e> / <i>
• *wesa* ‘to be’	root vowel <a> or <e>
• *kapad* ‘bought’	unstressed vowel <e> or <i>
• *degum* ‘days’	dative plural ending <Vm> or <en>
• *habbath* ‘have’	ending with <a> or <e> / <i>


Level of mixture
40%
30%
20%
10%
0%
mixture level
mixed charters
mixed tokens
Linear (mixed charters)
Linear (mixed tokens)
$R^2 = 0.65$
$R^2 = 0.57$
15%
20%
25%
30%
35%
40%
45%
50%
probability

Graph 1.9: Level of Mixture, token count.

words	tokens in mixed charters	tokens in charters > 1 token	% mixed tokens	tokens var. 1	tokens var. 2	%overall mixture	variant 1 vs. variant 2	time frame
bitalad	0	14	0%	20	81	20%	<a> - <e/i>	-1481
habbath	3	28	11%	10	29	26%	<a> - <e/i>	-1435
kapad	12	174	7%	137	306	31%	<i> - <e>	all
wesa	6	72	8%	39	68	36%	<a> - <e>	-1471
degum	4	22	18%	42	49	46%	<um> - <en>	all
seke	29	80	36%	58	65	47%	<e> - ø	1430-1480
seka	3	26	12%	21	30	41%	<a> - <en>	1460-1500
	57	416	r =	78,8%				

average /e/ =	12%	r2 =	62,2%
df =	5	p 1-T=	1,7%
n =	7	p 2-T=	3,5%

<ael> & <aell>	<ael>+V	<aell>+V	% <aell>
bitalad/-ath	11	9	45%
bitalia(ne)	9	2	18%
bitalinge	70	6	8%
rest of the charters	224	3	1%

Table 2.11: Skewed distribution of the sequence <aell>.

Table 2.9: Skewed distribution of the sequence <aell>.

	order	% <aell>	log(% aell)
bitala-	**0**	**45%**	**-0,35**
bitalia-	**1**	**18%**	**-0,74**
bitalinge	**2**	**8%**	**-1,10**
rest	**3**	**1%**	**-1,88**

r2 =	p 1-T =	p 2-T =	df =	n =	
0,90	2,5%	5,0%	2	4	%
0,96	1,0%	1,9%	2	4	log(%)