In: C. Paradis et al (eds.) (1998), Papers in Sociolinguistics. Québec: Éditions Nota bene, 137-147.

DANGEROUS AGGREGATIONS
A CASE STUDY OF DUTCH (N) DELETION

Hans Van de Velde & Roeland van Hout
Center for Language Studies, University of Nijmegen

ABSTRACT

This paper focuses on the question which factors rule the variation patterns of (n) deletion in Dutch. A variable rule analysis seems to provide the requested list of strong and clear internal and external factors. It turns out, however, that the speakers need to be included in the analysis as a separate factor, not only to account for quantitative differences between speakers, but especially to account for interaction effects between linguistic factors and individual speakers. This result shows that it is wrong to assume as a standard that community grammars can be grasped by simply aggregating data over speakers. Variation studies must explicitly pay attention to the occurrence of interaction patterns between speakers and linguistic factors. In fact, many variation patterns may exist, just because of the presence of different types of speakers, each having different grammars.

1. INTRODUCTION

In Dutch, syllable final /n/ can be deleted after a shwa. This deletion process is defined in Section 2, where the more important studies on this phenomenon are also discussed. These studies show that (n) deletion is far more variable than is suggested by the phonological and phonetic descriptions of Dutch. We collected new data from a group of 18 speakers of southern standard Dutch. A standard variable rule analysis was applied to the data, resulting in two external and four internal linguistic factors. The outcomes are described in section 3. We felt, however, that this standard treatment was failing, because it masked the real differences between speakers. This masking effect is brought about by the received procedure of aggregating data over speakers, replacing the latter by their social characteristics. In section 4, we will show that a speaker factor needs to be included in the analysis of (n) deletion, not only to account for quantitative differences between speakers but especially because of interaction effects between conditioning linguistic factors and individual speakers. In variable rule analysis, it is common to assume that no such interactions exist. In section 5, we will cluster the 18 speakers in four

In: C. Paradis et al (eds.) (1998), Papers in Sociolinguistics. Québec: Éditions Nota bene, 137-147.

different groups each having different patterns in the linguistic conditioning of (n) deletion. Section 6 discusses the implications of our findings. One conclusion is that variation studies have to pay more attention to the assumptions that underlie standard variable rule analysis. Often, community grammars are assumed to exist, without a shred of evidence. Instead, it is preferable to open up the analysis and the interpretation of the data for the possible co-existence of different types of speakers having different grammars in one and the same speech community. A second conclusion is that the range of statistical analyses applied in variation studies needs to be widened.

2. (N) DELETION

In this study, we focus on the two varieties of standard Dutch: southern standard Dutch as it is spoken in Flanders and northern standard Dutch as it is spoken in the Netherlands (cf. Van de Velde, Van Hout & Gerritsen 1997). In standard Dutch, syllable final /n/ can be dropped after a shwa: /en/ => /e/. Exceptions to this rule are the article 'een' and –according to Booij (1995:140)– verbal stems ending in /en/ (e.g., ik teken 'I draw'). In descriptive and prescriptive studies on standard Dutch pronunciation it is stated that deletion of (n) is more common than realization. In previous studies the phenomenon of (n) deletion was mainly discussed in word final position. We will do the same in our study. Two main types of (n) need to be be distinguished: (1) in monomorphemic words where (n) is part of a free morpheme, which implies that word final /en/ has no morpheme status, e.g., *negen* 'nine', *molen* 'mill', *boven* 'above'; (2) in polymorphemic words where (n) is part of a bound morpheme, which implies that word final /en/ has morpheme status, e.g., plural nouns *boek-en* 'books', plural verbs *zij kijk-en* 'they watch', infinitives kijk-en 'to watch', past participles *ge-kek-en* 'watched', adjectives *goud-en* 'golden'.

		[+n]	%	chi^2
1938	monomorphemic	14/35	40.0	7.015*
	polymorphemic	43/217	19.8	
1982	monomorphemic	9/41	22.0	6.589*
	polymorphemic	19/223	8.5	

Table 1: realization of (n) in Dutch movies of 1938 and 1982, split up by morphological status (N=28, n=516)

According to Koefoed (1979) and Hinskens (1992:336), (n) deletion is a postlexical process which is independent of morphological status. Van Oss & Gussenhoven (1984) mainly agree but suggest that

In: C. Paradis et al (eds.) (1998), Papers in Sociolinguistics. Québec: Éditions Nota bene, 137-147.

there may be individual differences, and that –especially in reading style– some speakers distinguish between monomorphemic and polymorphemic words. In Table 1 our results of a study on (n) deletion in four Dutch movies of 1938 and 1982 are presented. In both periods 14 speakers were selected. We tried to transcribe 20 occurrences of (n) for every speaker. The frequencies in Table 1 show that in both periods (n) is deleted more in polymorphemic than in monomorphemic words, indicating that (n) deletion is not a postlexical process.

In 1958 Ollevier conducted a quantitative study of (n) deletion in southern standard Dutch (see Pauwels 1969). Although this study lacks methodological accuracy, its results can be considered indicative. A majority of speakers shows a very clear tendency: either towards realization of (n) or towards deletion. The amount of (n) deletion appeared to be partly related to the geographical origin of the speaker. Furthermore, extreme stylistic intraspeaker differences were observed. In northern standard Dutch too, large differences between spontaneous speech and reading style are observed (Van Oss & Gussenhoven 1984, Voortman 1994): (n) deletion occurs more frequently in spontaneous speech than in reading style. In reading style the influence of orthography is strong, resulting in more realization of (n).

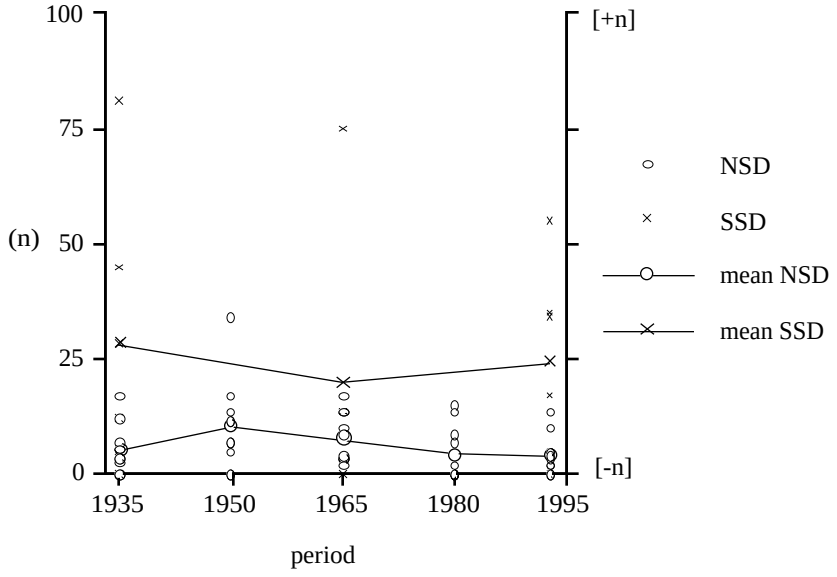

Figure 1: (n) split up by community and period (Van de Velde 1996)

In: C. Paradis et al (eds.) (1998), Papers in Sociolinguistics. Québec: Éditions Nota bene, 137-147.

In Van de Velde (1996) the differences between northern and southern standard Dutch are studied by means of an analysis of radio broadcasts from the period 1935-1993. For an overview of the design of this study we refer to Van de Velde, Gerritsen & Van Hout (1996) and Van de Velde, Van Hout & Gerritsen (1997). Per speaker 30 realizations of (n) were transcribed, equally spread over plural nouns, infinitives and plural verbal forms. The results of this study are presented in Figure 1. The individual scores are plotted with a small symbol. The mean scores split up by community and period are plotted with a larger symbol and connected with a line. Figure 1 shows obvious differences between the Netherlands and Flanders. There is more (n) deletion in northern standard Dutch (NSD) than in southern standard Dutch (SSD). There are no changes over time. An analysis of variance confirms these observations (community: F=13.172. df=1,42, p=.001; period: F=.0078, df=2,42, p=.925; interaction: F=0.426, df=2,42, p=.656). In the Netherlands the amount of realization of (n) is low. In Flanders interspeaker variation is high. This is very striking as southern standard Dutch is an abstract standard which shows little variation for other phonological variables (Van de Velde, Van Hout & Gerritsen 1997:381).

Summarizing the studies on (n) deletion, it has to be concluded that the results are not straightforward. Claims that (n) deletion is a post-lexical rule are contradicted. In a quantitative sense there are large differences between speakers. Even the impact of specific linguistic environments is different. Van Oss & Gussenhoven (1984) concluded that two types of speakers needed to be distinguished on the basis of the right hand environment. Deleters show the highest amount of (n) deletion before a pause, while in the same context inserters display the lowest amount of deletion.

The study of Van de Velde (1996) aimed primarily at studying differences between northern and southern standard Dutch over time. Therefore linguistic factors were kept under control as much as possible, excluding the possibility of studying the role of linguistic factors in greater detail. The only way to get more insight in the factors ruling (n) deletion was to collect additional data from the broadcast corpus. We decided to limit this additional data collection to southern standard Dutch, as this variety obviously shows more variation than northern standard Dutch. The results are presented in the next sections.

3. VARIABLE RULE ANALYSIS

Of all 18 speakers of southern standard Dutch all occurrences of post schwa word final /n/ were taken into account. The number of

In: C. Paradis et al (eds.) (1998), Papers in Sociolinguistics. Québec: Éditions Nota bene, 137-147.

observations per speaker ranged between 70 and 162 (n=2063). Two external factors were distinguished: period (1935, 1965, 1993) and programme type (royal reports vs. sports commentaries). Four linguistic factors were distinguished in analysing the context in which the linguistic variabele (n) occurred: 1. morphological status: monomorphemic vs. polymorphemic; 2. left hand environment: dental (/t/, /d/, /n/) vs. non-dental; 3. right hand environment: vowel (=V) vs. consonant (=C) vs. pause (=P); 4. phonological boundary: ≤ phonological phrase vs. ≥ intonational phrase (cf. Nespor & Vogel 1984).

Factor	chi^2	df	p	p log reg
EXTERNAL				
Period	10.42	2	**	***
Type	160.87	1	***	***
INTERNAL				
Morphological status	9.30	1	**	*
Preceding segment	15.67	1	***	*
Following segment	101.63	2	***	***
Phonological boundary	0.67	1		**

Table 2: Pearson chi^2 analysis separate factors and p-values logistic regression (backstep) all factors.

The results for the separate analyses on these six factors are given in Table 2. The factors were cross-tabulated with the non-deletion / deletion frequency of (n), and a (Pearson) chi-square analysis was applied to determine the strength of the dependency. For these frequency analyses the data of the 18 speakers were collapsed. All factors except phonological boundary are significant. This gives the following interpretation: period: more realization of (n) in 1935 than in 1965 and 1993; programme type: more realization of (n) in royal reports than in sports commentaries; morphological status: more realization of (n) in monomorphemic words; left hand environment: more realization of (n) when preceded by a dental consonant; right hand environment: more deletion of (n) before a consonant: C < (V=P).

The chi-square analyses look at the separate contributions of the factors on (n) deletion. In a variable rule analysis the total set of factors can be put together in one model of analysis, with the consequence that the estimated contribution of factors may change. We applied the technique of logistic regression which is the same as variable rule analysis (Sankoff 1988, Rietveld & Van Hout 1993, Van Hout & Muysken 1994). All six factors were considered to be categorical, the method for selecting factors was the stepwise procedure and the statistical package applied was SPSS. The right hand

In: C. Paradis et al (eds.) (1998), Papers in Sociolinguistics. Québec: Éditions Nota bene, 137-147.

part of Table 2 gives the significance of the factors entered in the analysis (in the column p log reg). It is obvious that this analysis is not completely in line with the one presented at the left hand side. Phonological boundary now appears to influence (n) deletion in a significant way, and the significance levels of three other factors differ between the two analyses. It should also be noted that, as usual, interaction effects between factors were not included in the analysis (in theory there are 32 different interaction effects possible with six factors). Variable rule analysis offers the possibility of including these interactions, but its complicated application is disfavoured (Sankoff 1988). Furthermore, the explained variance of all factors together is not particularly satisfactory (12.9%). We decided to conduct additional analyses, in order to be sure that no speaker effects were overlooked through aggregating or collapsing the data over speakers.

4. THE SPEAKER FACTOR

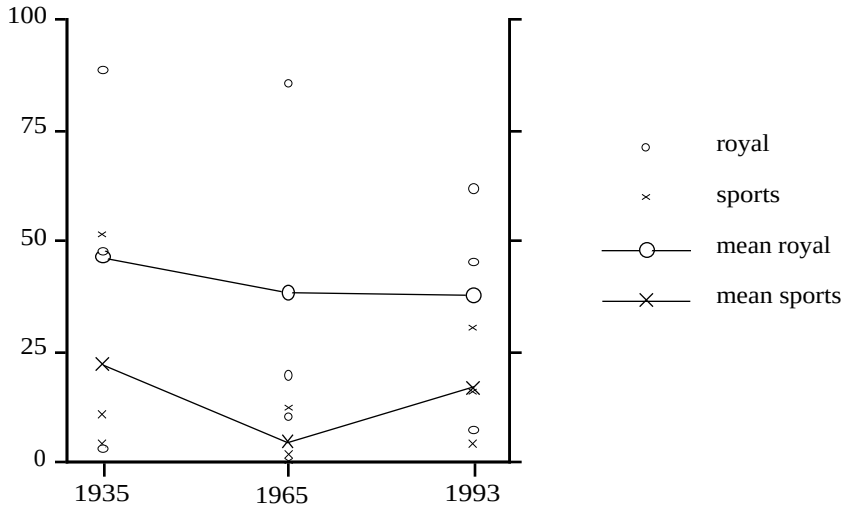

Figure 2: index scores (n) in SSD split up by programme type

In section 2 it was already observed that large differences between speakers may show up. This is confirmed for our own data by the pattern of individual index scores and mean scores split up by programme type which are visualised in Figure 2. There are large interspeaker differences in the amount of (n) deletion in southern standard Dutch. An analysis of variance with the external factors as independent variables shows that neither the effect of period (F=1.424,

In: C. Paradis et al (eds.) (1998), Papers in Sociolinguistics. Québec: Éditions Nota bene, 137-147.

df=2,12, p=.760), nor programme type (F=3.709, df=1,12, p=.078) is significant, which contradicts the result of the variable rule analysis. This contradictory result again points to a degree of interspeaker variation which overshadows other sources of variation. The importance of the interspeaker effect can be pointed out in two ways.

First, we calculated the chi-square for the speaker factor: chi^2=775.06, df=17, p<.001. In comparison with the other effects (see Table 2), the size of the speaker factor is very strong, implying a high degree of association between the speaker factor and the amount of (n) deletion. Second, in Figure 3 the index scores of the 18 Flemish speaker are split up by the right hand environment (vowel, consonant and pause). The lines definitely do not run parallel, indicating a strong interaction effect between the speaker and the linguistic factor of right hand environment.

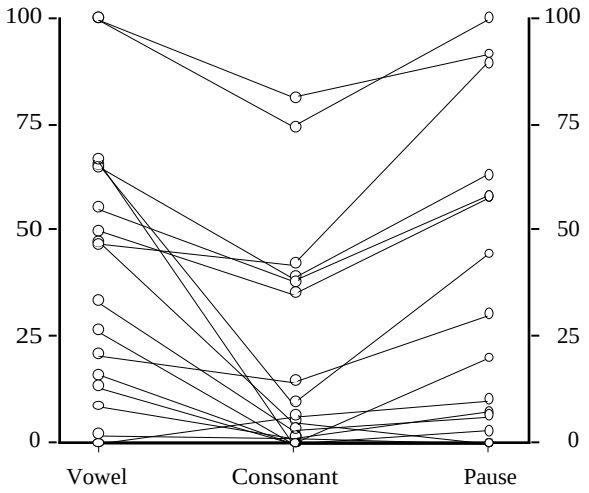

Figure 3: Index scores split up by speaker and right hand environment

5 . A RE-INTERPRETATION OF (N) DELETION

How homogeneous is our group of speakers? Is it only a matter of degree or do we need to distinguish different types of speakers? Different speaker types mean that speakers react differently to linguistic factors. However, morphological status, for instance, cannot be taken as a starting point for a division of speakers. Only 10.2% of the observations of (n) in our corpus are in monomorphemic words. In addition, these occurrences are unequally spread over the speakers. The strongest linguistic factor is obviously the right hand environment (see

In: C. Paradis et al (eds.) (1998), Papers in Sociolinguistics. Québec: Éditions Nota bene, 137-147.

Table 2). So, it seems to be logical to take this factor as a basis for classifying speakers. The two-way distinction by Van Oss & Gussenhoven (1984) for northern standard Dutch is too simple for the southern data. For southern standard Dutch at least four types of speakers can be distinguished in our corpus of 18 speakers, as is illustrated by the classification in Table 3.

	vowel	consonant	pause
non-realisers			
60	.00	.00	.00
61	2.08	1.25	.00
liaisoners			
53	13.04	.00	.00
54	16.00	.00	.00
55	66.67	.00	3.13
59	33.33	3.17	6.52
68	47.50	5.56	.00
deleters			
51	100.00	81.48	91.67
52	55.56	38.30	58.62
56	65.22	39.13	63.16
57	100.00	74.68	100.00
62	26.67	.00	20.00
64	50.00	35.29	58.06
66	8.82	1.28	7.14
67	65.71	9.59	44.44
pausers			
58	21.05	14.77	30.56
63	47.37	42.22	90.00
65	.00	6.35	10.20

Table 3: A typology of speakers of SSD on the basis of right hand environment, index scores indicate the amount of realization of (n)

Non-realisers C = V = P (= zero)
These speakers do not have an underlying /n/. (n) is deleted in all contexts. If [n] shows up, it is an intrusive n, filling the hiatus between shwa and a vowel-initial clitic (Booy 1995:171).

Liaisoners V > (P = C)
These speakers have an underlying /n/, almost exclusively showing up before vowels. Before consonant and pause (n) is not realised. It is a sort of liaison, resulting in the optimal CVCV structure. As (n) deletion is a postlexical rule for these speakers, there cannot be an effect of

In: C. Paradis et al (eds.) (1998), Papers in Sociolinguistics. Québec: Éditions Nota bene, 137-147.

morphological status. This is confirmed by a logistic regression analysis on data from this group of speakers.

Deleters C < (V = P)

These speakers realise (n) in all contexts. They reveal a high amount of realization, except speaker 66, who possibly belongs to another category. A following consonant is the context favouring (n) deletion most. In a previous analysis (Van de Velde & Van Hout 1996) this type of speaker was labelled 'realizer', but it was already suggested they had a deletion rule. A logistic regression reveals an effect of morphological status.

Pausers P > (V = C)

Speaker 63 is a very clear case. (n) is realized most before a pause, and becomes a kind of filled pause. A phonetic/phonological variable has turned into a discourse marker. A logistic regression reveals an effect of morphological status.

6. CONCLUSION

It is quite astonishing that a variable phenomenon like (n) deletion has hardly been studied in Dutch (socio)linguistics. The analysis presented in section 5 has to be considered as preliminary but at the same time it indicates that the variation pattern of this linguistic variable is a very remarkable one and that speakers behave differently, probably according to different grammars. Further research is necessary to get a grip on the factors ruling (n) deletion in Dutch. Therefore, many more data per speaker, covering a wide range of speaking styles are needed.

Because of the speaker differences, it turned out to be wrong to collapse or aggregate data over speakers. The basic assumption for such a procedure is violated by the presence of interaction effects between individual speakers and conditioning linguistic factors. This problem is mentioned explicitly by Sankoff (1988:992) and he refers to Rousseau & Sankoff (1978) for an algorithm to group speakers. Kay (1978) pointed out that there are patterns of language change marked by interactions between speakers and conditioning linguistic factors, whereas there is no straightforward community grammar with only one variable rule shared by all speakers.

Nowadays, this discussion has been completely forgotten in variation studies. Standardly, the real speaker is hidden behind dangerous aggregation levels, the assumed absence of linguistic interaction is never tested, and, commonly, too many factors are made part of the analysis. The standard approach in variation studies should be the other way around: there is interaction. Speaker differences and complex variation patterns marked by interaction effects can be

In: C. Paradis et al (eds.) (1998), Papers in Sociolinguistics. Québec: Éditions Nota bene, 137-147.

studied better when sociolinguists are prepared to use a larger variety of analytic statistical techniques. It is unwise to rely only on variable rule analysis / logistic regression. Using variable rule analysis is not such a self-evident choice as most variationists tend to think and a standard application may even generate results that do not explain the variation patterns at all.

REFERENCES

Booij, G. (1995), *The Phonology of Dutch*, Oxford: Clarendon Press.

Hinskens, F. (1992), *Dialect Levelling in Limburg. Structural and Sociolinguistic Aspects*, Nijmegen: PhD Dissertation.

Kay, P. (1978), "Variable rules, community grammar, and linguistic change", in D. Sankoff, (ed.), *Linguistic variation. Models and methods*, New York: Academic Press, pp. 71-84.

Koefoed, G. (1979), "Paradigmatische invloeden op fonetische processen", *Glot Special*, pp. 51-72.

Nespor, M. & I. Vogel (1984), *Prosodic Phonology*, Dordrecht: Foris.

Pauwels, J. (1969), "Eind-n na toonloze vocaal in Zuid-Nederland", *Taal en Tongval* 21, pp. 216-218.

Rietveld, A. & R. van Hout (1993), *Statistical techniques for the study of language and language behaviour*, Berlin: Mouton de Gruyter.

Rousseau, P. & D. Sankoff (1978), "A solution to the problem of grouping speakers", in D. Sankoff, (ed.), *Linguistic variation. Models and methods*, New York: Academic Press, pp. 97-117.

Sankoff, D. (1988), "Variable rules", in U. Ammon, N. Dittmar & K.J. Mattheier (eds.), *Sociolinguistics. An international handbook of the science of language and society*, Berlin: Walter de Gruyter, pp. 984-997.

Van de Velde, H. & R. van Hout (1996), "Between French liaison and English -t,d deletion: an intermediate (socio)linguistic status for Dutch (n) deletion?" Paper presented at NWAV25, October 17-20, 1996, Las Vegas.

Van de Velde, H., M. Gerritsen, & R. van Hout (1996), "The devoicing of fricatives in standard Dutch. A real-time study based on radio recordings", *Language Variation and Change* 8, pp. 149-175.

Van de Velde, H., R. van Hout & M. Gerritsen (1997), "Watching Dutch change: A real time study of variation and change in standard Dutch pronunciation", *Journal of Sociolinguistics* 1, pp. 361-391.

Van Hout, R. & P. Muysken (1994), "Modeling lexical borrowability", *Language Variation and Change* 6, pp. 39-62.

Van Oss, F. & C. Gussenhoven (1984), "De Nederlandse slot-n in het nieuws", Gramma 14, pp. 261-271.

Voortman, B. (1994), *Regionale variatie in het taalgebruik van notabelen. Een sociolinguïstisch onderzoek in Middelburg, Roermond en Zutphen*, Amsterdam: IFOTT.