

WORD CLASSIFICATION, WORD FREQUENCY, AND VOWEL REDUCTION

by F.J. Koopmans-van Beinum and J.H. Harder*

INTRODUCTION

Putting phonological rules to the test of real speech may be a precarious enterprise. Too often we have to conclude that the results of experimental phonetic research do not confirm or explain phonological intuitions. However the availability of a large quantity of phonetic research data presents a challenge and a chance to eradicate part of the mutual suspicion of phoneticians and phonologist with respect to their work.

The phonetic work on Vowel Contrast Reduction (Koopmans-van Beinum, 1980) offers the possibility of such testing because of the existence of a great amount of perception data.

Several authors on phonology (cf. Booij, 1981) mention the fact that words with a low frequency of occurrence will not produce vowel reduction as soon as words with a high frequency of occurrence, and, besides, that vowel reduction is related to grammatical word class as well. To make the notions low and high frequency manageable and to make testing of these rules possible, the original speech material of Koopmans-van Beinum (1980), existing of unstressed words and syllables in free conversation of four speakers (two male and two female) was categorized in grammatical word classes in the present study. For this purpose words were grouped in the closed category of function words at one hand and the open category of content words at the other (cf. Bolinger,

*Part of the research has been done by J.H. Harder in order to obtain her doctoral assignment.

1975; Clark and Clark, 1977; Bloom and Lahey, 1978). The function words, a class of unproductive parts of speech, show a substantially higher frequency of occurrence than the class of productive parts of speech, to which the content words belong.

Thus we can try to test the hypothesis that, in Dutch, function words will show more vowel contrast reduction than content words will do. Zwicky (1972) and Dressler (1974) among others put forward this phenomenon for other languages as well.

METHODS

In the perception tests of the original study on vowel contrast reduction the task of the listeners has been, among other things, to identify vowel sounds segmented from free conversation as one of the twelve Dutch vowels. Here it turned out that only 33% of the vowels could be identified correctly, that is to say in accordance with the vowel intended in the word. In the present study we established for each of the vowel sounds from which word it had been segmented and to which word class this word belonged. Besides we knew of each vowel sound the percentages of correct or incorrect identifications. Since false identifications as such don't say anything about the degree of contrast reduction, this category was divided in [œ]- responses at the one hand (in which the vowel [œ] and the schwa have been put on a par, because acoustically they are nearly identical in Dutch) all other false identifications were grouped together at the other hand.

In this way the first group can be considered as representing complete vowel contrast reduction whereas the second group represents a slight degree of reduction. From an acoustic viewpoint the second group of responses is situated between the intended vowel and [œ] (cf. Koopmans-van Beinum, 1980). So in the present study the responses are divided into three categories: (1) not reduced, (2) slightly reduced, and (3) strongly reduced. In these three categories the scores are classified as follows:

- (1) given response = intended vowel
- (2) given response \neq intended vowel or [œ]
- (3) given response = [œ]

Words containing the vowel [ø] as in the Dutch word beuk, and words containing the vowel [œ] as the Dutch word buk, although being part of the original perception data, are left out of the present study since they might distort the results because of the amount of [œ]- responses to be expected.

Thus the total number of words involved in this study amounts to 400 (viz. 4 speakers x 10 vowels x 10 words per vowel). The number of listeners in the original vowel identification test has been one hundred; together this resulted in 400 x 100 = 40.000 scores. Since the acoustic measurements (Koopmans-van Beinum, 1980) proved that differences in duration between the so-called long vowels and short vowels in Dutch are minimal when appearing in unstressed position in free conversation, we reckoned [ɔ]-, [ɑ]-, and [ɪ]- responses among the correct identifications as well, where respectively [o], [a], and [e] were intended.

RESULTS

Table I renders the distribution of the responses in the three response categories for content words and for function words. The hypothesis of a relatively equal distribution among the categories can be tested by means of a χ^2 -test for two or more test samples

Table I. Absolute numbers and percentages of given identifications distributed over the three response categories for content words and for function words, averaged over 100 listeners.

grammatical word class	correct		incorrect# [œ]		[œ]		total	
content words	117.34	42.21%	102.91	37.02%	57.75	20.77%	278.00	100%
function words	40.36	33.08%	44.74	36.67%	36.90	30.25%	122.00	100%
total	157.70	39.43%	147.65	36.91%	94.65	23.66%	400.00	100%

where the numbers obtained are compared with the numbers expected on account of the hypothesis postulated above.

Application of this test results in a $\chi^2 = 5.01$, from which we may conclude that the present differences in the distribution are significant at a level of 10% ($p < 0.10$ one side probability level for $\chi^2 = 4.605$ with two degrees of freedom).

So there proves to be a tendency that the distribution of the responses in the three categories are not proportionally equal for content words and for function words. It has to be mentioned that the discrepancies are mainly in the category correct and in the category [œ]. Fig. 1 displays in histograms the data of table I.

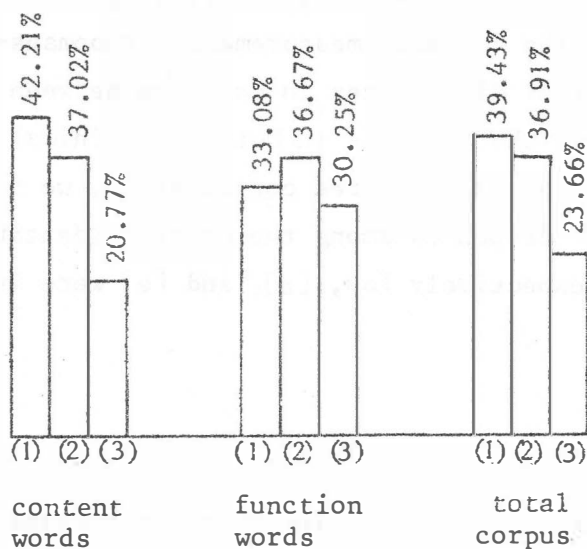


Fig. 1 Percentages of given identifications distributed over the three response categories for content words, for function words, and for the total corpus.

- (1) : correct
- (2) : incorrect, ≠ [œ]
- (3) : [œ]

The results of the original perception test, distributed over the three response categories mentioned above ((1) correct, (2) incorrect but not [ce], and (3) [ce]) display a distribution in percentages of respectively 39.43%, 36.91%, and 23.66%.

We can now divide the grammatical word classes of content words and function words into grammatical subgroups (cf. Bloom and Lahey, 1978) and trace how far the distribution over the response categories for each grammatical subgroup corresponds to the overall distribution, and which subgroups display the greatest deviations.

Table II gives the distribution of the responses for each grammatical subgroup, tested against the overall distribution of respectively 39.43%, 36.91%, and 23.66%. Some grammatical subgroups turn out to deviate significantly from the overall distribution.

However, a χ^2 -test for two or more test samples, applied to the complete distribution in all subgroups together as given in table II, does not yield significant results. For this distribution

$\chi^2 = 15.77$ whereas we need $\chi^2 = 21.06$ for a one side probability level of $p < 0.10$ with 14 degrees of freedom.

Table II. Mean numbers and percentages of given identifications distributed over the response categories, for each grammatical subgroup tested against the overall distribution.
* , ** , *** : deviating significantly at respectively the 10%, 5%, and 2.5% level (df = 2).

	grammatical subgroup	correct		incorrect#[ce]		[ce]		total		χ^2
content words	substantives	39.93	41.16%	32.58	33.59%	24.49	25.25%	97.00	100%	0.49
	adjectives	15.95	39.88%	18.29	45.72%	5.76	14.40%	40.00	100%	5.73*
	main verbs	25.47	53.06%	13.35	27.81%	9.18	19.13%	48.00	100%	7.82***
	adverbs	35.99	38.70%	38.69	41.60%	18.32	19.70%	93.00	100%	1.27
function words	pronouns	12.52	36.82%	9.98	29.35%	11.50	33.82%	34.00	100%	6.08**
	copulas + auxiliaries	11.34	34.36%	9.73	29.48%	11.93	36.15%	33.00	100%	8.74***
	conjunctions	8.92	33.04%	12.03	44.56%	6.05	22.41%	27.00	100%	2.70
	prepositions	7.58	27.07%	13.00	46.43%	7.42	26.50%	28.00	100%	6.67**
	total	157.70	39.43%	147.65	36.91%	94.65	23.66%	400.00	100%	

Fig. 2 displays the data of table II in histograms; in fig. 2a the distribution within the grammatical subgroups of the content words is given, in fig. 2b the distribution within the grammatical subgroups of the function words.

Fig. 2a

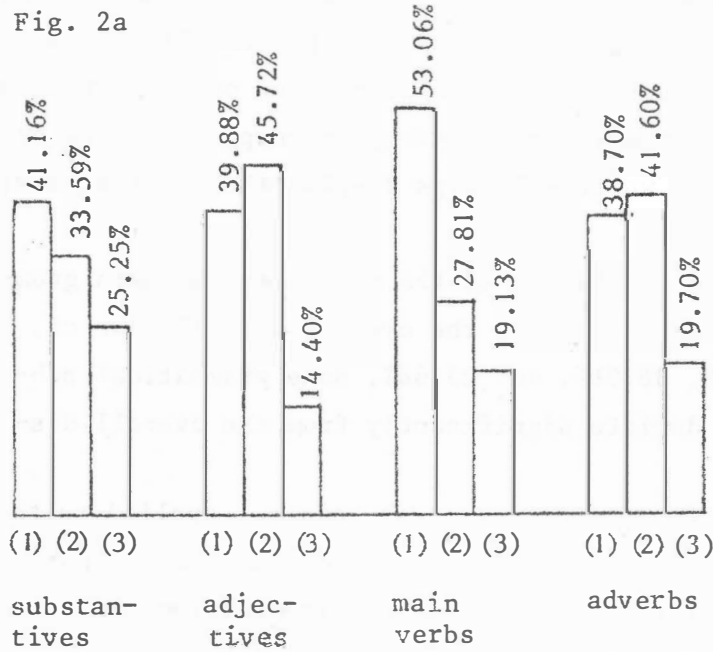


Fig. 2a

Percentages of given identifications for grammatical subgroups within the class of content words.

- (1) : correct
- (2) : incorrect, ≠ [œ]
- (3) : [œ]

Fig. 2b

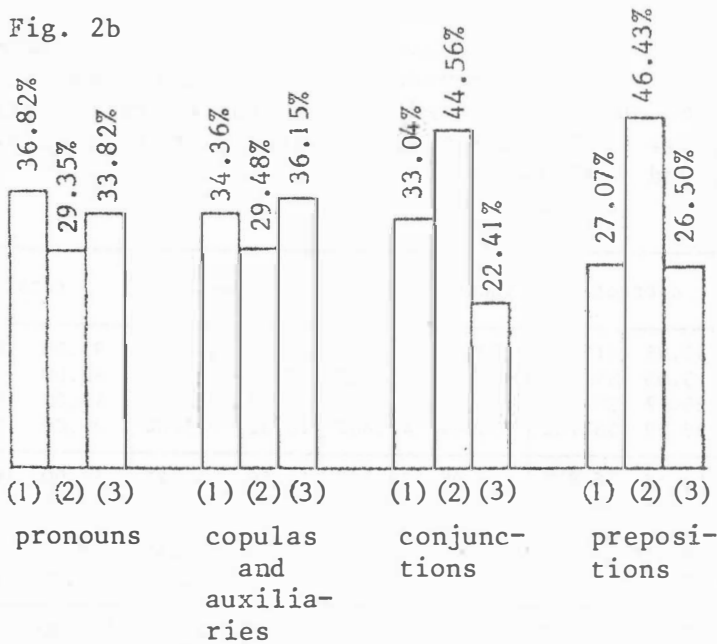


Fig. 2b

Percentages of given identifications for grammatical subgroups within the class of function words.

- (1) : correct
- (2) : incorrect, ≠ [œ]
- (3) : [œ]

CONCLUSIONS

Further inspection of table II results in the following conclusions.

Concerning content words:

- The substantives deviate from the overall distribution only slightly.
- Concerning the adjectives we may conclude that the percentage of correct identifications is indeed in accordance with what was expected, but that both other categories strongly deviate from the expected percentages: slight reduction here frequently appears at the cost of strong reduction, which is in accordance with the prior hypothesis.
- For the group of main verbs the percentage of correct identification (i.e. no reduction) is strikingly high, while the categories of slight and of strong reduction both display a smaller percentages than could be expected on account of the overall distribution.
- The adverbs deviate slightly, mainly since here as well as in the case with the adjectives the percentage of slight reduction is higher than would be expected, and here too at the cost of the category of strong reduction.

Taking this together we may conclude for the four grammatical subgroups in the class of content words, that we are dealing with a tendency into the direction of little or no reduction when comparing the percentages of the given responses to the percentages expected on account of the overall distribution.

In the grammatical subgroups belonging to the class of function words we can see a totally different picture:

- For the pronouns, and for the copulas and auxiliaries the category of strong reduction considerably exceeds expectation at the cost of the category of correct identifications (i.e. no reduction) and the category of slight reduction.
- For the conjunctions and the prepositions the category of strong reduction practically corresponds to expectation, whereas the

category of slight reduction highly dominates at the cost of the category of no reduction.

Taking this together we may conclude for the four grammatical subgroups in the class of function words that in all cases we can establish a shift into the direction of strong reduction, when comparing the percentages of given responses to the percentages expected on account of the overall distribution.

Considering the tendencies mentioned above we may ask whether it is correct to conclude that groups of words having a low frequency of occurrence do not display contrast reduction in their vowels as readily as groups of words of high frequency, as Booij (1981) supposes?

In spite of the fact that not in all cases the results were significant and that often only tendencies were present, we may in effect answer the question in the affirmative. For the speech material to which the present test was applied, consisted solely of vowel sounds segmented from syllables without the main stress. This means that multisyllabic words (especially appearing in the class of content words) in the present study will display more reduction than in a study including the stressed syllables as well. Add to this the fact that the pronouns, a subgroup of strong reduction, can be considered to be underrepresented in this study. Indeed this is the case, when the number is compared with the numbers in the frequency lists of De Jong (1979) for spoken Dutch (in the present study 8.5% as against 25.6% in De Jong). This underrepresentation is the consequence of pronouns often being pronounced enclitically, a form of complete vowel reduction, and because of the absence of any vowel sound not included in the original study on vowel contrast reduction.

These facts therefore may be considered to reinforce the conclusions mentioned above. Related to grammatical word groups it may be claimed justifiably that actual speech material confirms the phonological rule which states that low frequency word groups do

not display vowel reduction as easily as high frequency word groups.

However, the question as to how far this means that within word groups as well the frequency of occurrence among other things defines the degree of vowel reduction, as Fidelholtz (1975) states, is still open to further research.

REFERENCES

- Bloom, L. and Lahey, M. (1978). Language Development and Language Disorders. New York etc., Wiley.
- Bolinger, D.L. (1975) . Aspects of Language. New York, Harcourt, Brace and World.
- Booij, G.E. (1981). Generatieve fonologie van het Nederlands. Utrecht/Antwerpen, Het Spectrum.
- Clark, H. and Clark, E.V. (1977). Psychology and Language. New York, Harcourt Brace.
- Dressler, W. (1974). Methodisches zu Allegro-regeln, in: W. Dressler and F. Mares (eds.), Phonologica 1972. München, Fink Verlag, 219-234.
- Fidelholtz, J. (1975). Word Frequency and Vowel Reduction, in: R.E. Grossman, L.J. San and T.J. Vance (eds.), Papers from the 11th Regional Meeting of the Chicago Linguistic Society. Chicago, Chicago Ling. Society, 200-213.
- Jong, E.D. de (red.) (1979). Spreektaal; woordfrequenties in gesproken Nederlands. Amsterdam, Bohn, Scheltema, and Holkema.
- Koopmans-van Beinum, F.J. (1980). Vowel Contrast Reduction. An acoustic and perceptual study of Dutch vowels in various speech conditions. Amsterdam, Academische Pers B.V., Diss.
- Zwicky, A.M. (1972). On casual speech, in: P.M. Peranteau, J.N. Levi, and G.C. Phares (eds.), Papers from the 8th Regional Meeting of the Chicago Linguistic Society. Chicago, Chicago Ling. Society, 607-615.