

The comprehension of grammatical and natural gender: a cross-linguistic experiment

ANTON BATLINER

Abstract

This study examines the semantic interpretation of man and its anaphoric pronoun he and of their equivalents in German, Danish, and Swedish within a cross-linguistic experimental design. These languages differ systematically in respect of the features that define the grammatical and natural gender of these items. The hypothesis that the items are more or less pseudogeneric to the same extent that their features are more or less sex-marking could be confirmed in two experiments where subjects rated the acceptability of sentences containing these items. In conclusion, the pros and cons of such an 'experimental linguistic' approach are discussed.

1. Introduction

It has been shown by several authors that in (American) English, the supposedly generic terms *man* and *he* are in fact often pseudogeneric, i.e. they are interpreted as indicating a male rather than a female referent.¹ But is it possible to apply these — and similar — results generally to other languages? As in many other linguistic disciplines it seems that such hidden generalizations extending from 'homo americanus' to 'homo erectus' are also made in the field of 'language and sex' (cf. Smith 1982: 85). That may be partly due to neglect or ignorance of studies written in other languages, but often such studies simply do not exist, as in this case: to our knowledge, the use of, for example, equivalents of generic *man* in other languages has never been dealt with — at least not on an empirical basis. If generalizations are made, they are made implicitly (as mentioned above) or derived on an intuitive basis, as in Pusch (1979: 89ff.). Yet intuition as the only investigative method has serious drawbacks, as for example Dubois and Crouch (1975) have shown. (Moreover, Pusch's persuasive examples for a pseudogeneric use of male terms like *der*

Mensch in German are drawn from translations of English originals.)

On the other hand, the German *Mensch* is indeed a good candidate for a pseudogeneric use; Suttner (1889:79) has questioned its generic use, as has Gabelentz (1969: 235f.): 'Das grammatische Geschlecht bringt es nun weiter mit sich, dass wir Deutschen nie eine Frauensperson als einen Menschen und nicht so leicht einen Mann als eine Person bezeichnen.' Jespersen (1968: 231) concludes,

It is often desirable, and even necessary, in speaking of living beings to have words which say nothing about sex and are equally applicable to male and female beings. Such a word is German *mensch*, Dan. and Norw. *menneske*, Sw. *människa*, though it is curious that grammatically *mensch* is masculine (whence Germans in some connexions hesitate to use it about a woman), *människa* is feminine, and *menneske* neuter. In English *man* has from the oldest times been used for both sexes, but as it may also be used specifically of the male sex, ambiguity and confusion sometimes result (...). The generic singular *man* sometimes means both sexes (...) and sometimes only one [i.e. the male sex. A.B.] (...). This is decidedly a defect in the English language (...).

It can be seen from this quotation that in this respect there are interesting and systematic differences between some closely related languages, i.e. English, German, Danish, and Swedish (Norwegian will be excluded from further analysis: it is similar to Danish, and therefore Danish can stand for both languages; moreover, the Norwegian diglossia would complicate matters too much). This constellation of data will be considered as the starting point for a cross-linguistic experiment. The possibly unique systematic difference between the languages in question is a necessary prerequisite — otherwise no strictly parallel experimental design could be considered; and their close relationship is necessary too — otherwise strictly parallel translations would be not only theoretically but also practically impossible.² From now on, 'MAN' will be used for the English generic *man* and its equivalents in the other languages. In all the languages considered, the generic pronoun is usually male, with some marginal exceptions, cf. for German Duden (1966: 628ff.) and Werner (1975: 51), for Danish Hansen (1967: 72ff.), and for Swedish Thorell (1973: 28f.). In Swedish, *han eller hon*, an equivalent to *he/she*, is getting more common nowadays, cf. Blume (1982); the same is true for Danish *han eller hun*. In this paper, we will restrict our investigation to the use of the anaphoric personal pronoun that refers to MAN, henceforth called 'HE'. In English, the sex of the referent of MAN can be unspecified or male, but not female; the anaphoric pronoun *he* is male. In the other languages, MAN is unspecified and can refer to both men and women; but in German, the gender of *der Mensch* is masculine, and its anaphoric

pronoun *er* is masculine as well. In Danish, the gender of *mennesket* (MAN plus definite article) is neuter, and its anaphoric pronoun *det* is neuter as well. In Swedish, the gender of *människan* (MAN plus definite article) is neither masculine nor feminine, but utrum, yet its anaphoric pronoun *hon* is female. These data are classified in Table 1; we have enclosed in boxes the parts of the matrix where the relevant features differentiate. In Table 1 and further on, 'm' and 'f' are used both for 'masculine/feminine' and for 'male/female'. We can see, that for Swedish, Jespersen mixed up the gender of *människan* and the gender of its anaphoric pronoun *hon*; both must be kept apart.

Later we will see that there are linguistic reasons for specifying the features in Table 1 as more or less sex-marking. What we want to investigate is whether the relationship between grammatical gender and natural gender is arbitrary or whether grammatical gender is used to indicate natural gender. So for the sake of the argument, we will for the moment NOT differentiate between grammatical and natural gender in our hypotheses. It is for this reason that it doesn't really matter now whether we use 'male/female' or 'masculine/feminine'. However, we take it for granted that a relationship between natural and grammatical gender has been proven experimentally for English generic *man* and *he*, in other words that these items are at least sometimes semantically interpreted.

The alternate main hypothesis we want to test is that AS FOR THE GENERIC USE OF MAN OR OF MAN PLUS HE, ANY KIND OF GENDER IS INTERPRETED SEMANTICALLY. Testing this main hypothesis, we can find an answer to the following three questions:

Table 1. *Feature matrix: grammatical and natural gender of MAN and HE*

	(a) Possible sex of referent		(b) Anaphoric pronoun			(c) Gender of MAN		
	m	f	m	n	f	m	n	f
Common features	0 +	0 -						
English	<div style="border: 1px solid black; padding: 5px; display: inline-block;"> - + - + - + </div>	<div style="border: 1px solid black; padding: 5px; display: inline-block;"> - + - + - + </div>	+ - -			0 0 0		
German			<div style="border: 1px solid black; padding: 5px; display: inline-block;"> + - - - + - - - + </div>			<div style="border: 1px solid black; padding: 5px; display: inline-block;"> + - - - + - 0 - 0 </div>		
Danish								
Swedish								

(a) Is a generalization extended from the English generic *man* to the other languages permissible, cf. the different features in Table 1(a)?

(b) Is (in this special case) the anaphoric personal pronoun in the other three languages interpreted semantically, as it is in English, cf. the different features in Table 1(b)?

(c) Has the gender of MAN — in this case manifested by the gender of the definite article — any influence on the semantic interpretation, cf. the different features in Table 1(c)?

We can reformulate the main hypothesis as specific alternate hypotheses for these three questions:

(a) Sentences containing MAN denoting a male referent are more acceptable in English than in the other languages, and, vice versa, sentences containing MAN with a female referent are less acceptable in English than in the other languages. Henceforth this hypothesis is labeled 'H_{MAN}'.

(b) Sentences containing MAN plus anaphoric HE denoting a male referent are more acceptable in German, less acceptable in Danish, and least acceptable in Swedish. For female referents, the scale is inverted. Henceforth this hypothesis is labeled 'H_{HE}'.

(c) With regard to the sentences containing MAN without HE, not only is there a difference between English and the other languages, but these sentences are also more acceptable in German if they denote a male referent than they are in Danish and Swedish; vice versa, they are less acceptable if they denote a female referent. From now on this hypothesis is labeled 'H_{MAN*}'.

The last two hypotheses are not independent of each other. We will come back to this point in the discussion below.

Our main hypothesis is roughly the same as MacKay and Fulkerson's (1979:661) '*pronominal dominance hypothesis*, whereby the lexical meaning of a pronoun determines the interpretation of its antecedent', and the null hypothesis is similar to their '*pronominal surrogate hypothesis*, whereby the nature of an antecedent completely determines the interpretation of a pronoun' (1.c). As we also want to deal with MAN and moreover with a grammatical category that does not exist in English, namely the grammatical gender existing in the other languages, the hypothesis has to be put a bit differently: it has to be assumed that not only all kinds of gender function as indicators of grammatical congruency or are simply arbitrary but also that they are interpreted as carriers of semantic information.

The experimental philosophy which this approach is based on is simple: what we want to test on an intersubjective, nonintuitive basis is the acceptability in its widest sense, including grammaticality, of sentences. If the experimental manipulations — i.e. in this case the systematic differ-

ences between the four languages — should result in varying acceptability of the test items, then the latter is attributed to the former — all other things being equal, of course. To guarantee this condition, we have to make two strictly speaking contrafactual assumptions:

(i) Translations can be 'equivalent'.

(ii) The samples from the different languages are taken from the 'same' populations.

For (i), no exact, objective criteria exist. It is commonly assumed that translations are never entirely equivalent, but as these languages are closely related, and as the cultural background is roughly the same — in Whorf's (1956: 138) terminology, they are all 'Standard Average European Languages' — the degree of deviation can be minimized by using translations carefully scrutinized by several competent speakers of the respective languages, and by considering not the value of the rating of one single sentence, but the difference between the values of two or more sentences, cf. the analysis of the data obtained below. To guarantee (ii), the subjects are all students of linguistics, i.e. they are taken from corresponding populations, therefore education, age, etc., are roughly the same, and we have tried to control the possible intervening variable of conservatism. For the details, see below.

2. Methods

2.1. Sentence rating

The first part of the experiment was based on 45 sentences; 21 were filler items — more or less acceptable — and not analyzed further; the remaining 24 consisted of four groups of six sentences each:

(i) Sentences with an inappropriate *but*, as in *George is stupid, but ugly*, henceforth called 'L-S' ('logic sentences').

(ii) Sentences with an inappropriate *hier-jetzt-ich-origo* (cf. Bühler 1934: 102ff.); as in *Yesterday, I'll travel to Rome*, henceforth called 'O-S' ('origo sentences').

(iii) Sentences with inappropriate role distribution, as in *The lieutenant reproved the general*, henceforth called 'C-S' ('conventionalism sentences').

(iv) Sentences containing MAN, henceforth called 'M-S': the actual test items. These sentences are given in Table 2, in order of language.

From an originally larger number of sentences formulated in German, those were selected for which equivalent translations in the other languages could be found. All of these were checked by at least three

Table 2. *Test items containing 'IAN*

Language	Item no.	Sentence
English	M1	In many countries, man is legally obliged to provide for wife and children.
	M2	Exposure to radiation can lead to impotence in man.
	M3	Man often alters some of his habits, when he becomes a father.
	M4	Nowadays, man enters into a matrimonial union with a man less than in the past.
	M5	During menstruation, man tends to be in a changeable mood.
	M6	Man should not smoke, while he is pregnant.
German	M1	In vielen Staaten ist der Mensch gesetzlich dazu verpflichtet, für Frau und Kinder zu sorgen.
	M2	Strahlenschäden können beim Menschen zu Impotenz führen.
	M3	Der Mensch ändert oft einen Teil seiner Gewohnheiten, wenn er Vater wird.
	M4	Heutzutage bindet sich der Mensch weniger als früher durch Heirat an einen Mann.
	M5	Während der Menstruation neigt der Mensch zu Stimmungsschwankungen.
	M6	Der Mensch sollte nicht rauchen, solange er schwanger ist.
Danish	M1	I mange lande er mennesket lovmæssigt forpligtet til at forsørge kone og børn.
	M2	Strålebeskadigelser kan føre til impotens hos mennesket.
	M3	Mennesket ændrer ofte en del af sine vaner, når det blir far.
	M4	Nu om dage binder mennesket sig mindre end tidligere ved ægteskab til en mand.
	M5	Under menstruationen er mennesket tilbøjelig til at svinge stemningsmæssigt.
	M6	Mennesket burde ikke ryge, så længe det er gravid.
Swedish	M1	I många länder är människan enligt lag förpliktad att sörja för hustru och barn.
	M2	Hos människan kan strålningsskador leda till impotens.
	M3	Människan ändrar ofta en del av sina vanor, när hon blir far.
	M4	Idag ingår människan mindre ofta än förr äktenskap med en man.
	M5	Under menstruationen tenderar människan att vara lynnig.
	M6	En människa borde inte röka medan hon är gravid.

competent native speakers. Also, pilot tests followed by informal interviews were conducted in order to check the suitability of the test items. As can be seen from Table 2, the M-Ss are arranged in pairs of sentences, each containing a semantic component [male] or [female]: M1 versus M4: *wife* versus *man*, M2 versus M5: *impotency* versus *menstruation*, M3 versus M6: *father* versus *pregnant*. The last pair additionally contained the pronoun HE. All of them are relevant for H_{MAN}, the first two pairs are

relevant for H_{MAN^*} , and the last pair is relevant for H_{HE} . Therefore, the first two will be analyzed together and the last one separately. The sentences are all to be interpreted generically in so far as MAN in English is used without any article, in the other languages with the definite article. Only the Swedish M6 is exceptional, as *människan* in this sentence would indicate an individual female referent.³ The number of the test items was kept small in order to avoid undue length of the experiment and to minimize any awareness of the experimental task. This was also making a virtue out of necessity, as it is quite difficult to find suitable sentences that can be translated equally well into the other languages. Of course it could be possible that the acceptability of the paralleled M-Ss in the various languages is rated slightly differently for other reasons than those under discussion. But we will assume that these differences are either negligible or randomly distributed in the languages in question. Furthermore we do not choose the rating of single paralleled M-Ss as the basis of our comparison, but the differences of rating between corresponding sentences within one language. Thus we can eliminate any possible influence of overall tendencies in the respective samples to score differently. Of course all the M-Ss are, or at least could be, more or less unacceptable insofar as there is a semantic conflict between GENERIC MAN and the NONGENERIC contexts restricted to either male or female referents. Yet what we are interested in is not the question whether these sentences are acceptable or unacceptable but the various degrees of acceptability or unacceptability which are attributed to them. In other words what we want to find out is the value ' d_1 ' with results from subtracting the rating of M3 from the rating of M6 and the value ' d_2 ' resulting from subtracting the ratings of M1 plus M2 from the ratings of M4 plus M5.

The order of sentences was randomized but balanced: there were three pages with 15 sentences each, and two sentences out of each group (M-S, C-S, O-S, and L-S) on every page. The test was anonymous. Detailed written instructions in the respective languages were given in order to avoid questions by the subjects that might have been answered differently by the various supervisors. The central passage of the instructions for this part of the experiment reads as follows:

This section comprises 45 English sentences. They have been selected in such a way that some of them should sound quite impossible, others quite normal, and others may be rated between these two extremes. Unlike sentences such as 'He went in the morning for working', the deviant sentences are often NOT grammatically incorrect. Usually, they are sentences that sound peculiar because certain elements do not fit together, for example because they contradict our experience, as for instance in the sentence: 'The girl met the elephant in order to go with it to the cinema.' One doesn't go to the cinema with an elephant, does one?

You are now requested to rate each sentence on the scale given next to the sentences, circling the appropriate mark. The scale reads:

This sentence sounds to me

Totally
normal

++

Normal

+

More or less
normal

0

Strange

-

Totally
impossible

--

For later analysis, each rating was replaced by a numerical value: 0 for 'Totally normal', 1 for 'Normal', 2 for 'More or less normal', 3 for 'Strange', and 4 for 'Totally unacceptable'.⁴

2.2. *The conservatism scale*

To control a possibly intervening variable of conservatism, a second test was conducted after the sentence rating test. As mentioned above, we had to assume that the subjects belong to the 'same' population. Schmidt et al. (1973) have shown that there is a positive correlation between misogyny — which could be considered as just another term for sexism — and conservatism. If then sexism, or more generally conservatism, can influence the rating of the sentences, it could be the case that it is not the different linguistic acceptability of the M-Ss that is decisive, but rather the fact that more conservative subjects rate the sentences differently from less conservative.⁵ Even more important is that the Scandinavian subjects might generally be less conservative than the English and German subjects — a consideration which is on the face of it not too wide of the mark. For this reason we have to take into account that the differences predicted by our hypotheses could be at least in part a result of this intervening variable; and that would jeopardize our findings. With respect to the samples under consideration we can assume that the conservatism tests available are 'culture-free' insofar as they are all designed for Western societies. However, we cannot take it for granted that they are also 'language-free' i.e. that they are strictly comparable if composed in different languages. There are considerable differences between the tests in various languages with regard to length, items, wording, etc. As such a 'language-free' test could not be found, and as it would have exceeded the scope of this paper to design an appropriate test, we chose the following approach: as a starting point we decided on a German adaptation (Schneider and Minkmar 1972) of the conservatism scale created by Wilson and Patterson (1968). This adaptation had been validated for a corresponding sample (German students). The scale in this case does not consist of the usual, sometimes rather lengthy, statements, but of simple

lexical items or noun phrases. This format is an improvement '(...) in so far as it reduces the influence of cognitive processes, task conflict, grammatical confusion and social desirability' (Wilson and Patterson 1968: 265); what was equally important for our task, it made translation easier and less equivocal.⁶

From the 40 items used by Schneider and Minkmar (1972), 23 were selected that could be translated equally well into the other languages. Example (1) lists the items in the English wording. Items used in the Wilson and Patterson (1968) scale are asterisked, and items with the wording slightly modified are asterisked twice.

- (1) *Chastity, *Patriotism, **Bible, Virgin brides, **Self-discipline, *Inborne conscience, Banning of extreme political parties, **Sunday observance, **Divine providence, Obedience toward parents, Strict principles; **Nudism, Artificial insemination, *Suicide, Nude photographs, Homosexuality, **Striptease, **Underage car drivers, *Working mothers, Adultery, *Socialism, **Legal abortion, *Divorce.

The items were ordered randomly within the restriction of an alternating pattern of affirmatively (the first eleven items in Example [1]) and negatively scored items. The central part of the instructions for this section reads as follows:

Here, we have collected some concepts with which everyone is, no doubt, already familiar. We would like to know how comparable groups in various countries react to them. In order to make it easier for you to record your attitude, we have, for each concept, given two grades of approval, one noncommittal, and two of disapproval; you know the symbols from the first section. Please circle the symbol that corresponds with your attitude:

- + + if your attitude to the given concept is strongly positive
- + if your attitude is on the whole positive
- 0 if you are unable to decide, i.e. your attitude is basically neutral
- if your attitude to the given concept is on the whole negative
- if it is strongly negative

The numerical values that were given to the ratings are the same as in the first section, i.e. 0 for + +, etc.

Now the crucial point is, we can rank the four languages on an ordinal scale 'LG' putting English fourth (in which a male interpretation of MAN is most likely), German third, Danish second, and Swedish first (in which a male interpretation of MAN is least likely). We can compute a correlation between LG and the conservatism scale KX: $r_{LG,KX}$ on the one hand, and a correlation between KX and the differences d_1 and d_2 on the

Table 3. *Number, sex, and age of subjects*

	N	f	m	-20	21-23	24-26	27-29	30-32	33-35
English	43	34	9	39		3	1		
German	24	15	9	8	13	2	1		
Danish	12	6	6	4	4		1	2	1
Swedish	36	20	16	6	21	9			

other hand: $r_{d_1, KX}$ and $r_{d_2, KX}$. If these correlations are high enough and if there exists a positive correlation between LG and the differences d_1 and d_2 — as predicted by our hypotheses — then we can rule out the possible influence of an intervening variable ‘conservatism’ by computing the partial correlations $r_{d_1, LG - KX}$ and $r_{d_2, LG - KX}$.

2.3. *Subjects and procedure*

Our subjects were students of linguistics at the universities of Edinburgh, Munich, Copenhagen, and Gothenburg. The experiments were conducted in connection with introductory linguistics courses attended by the subjects and were supervised by the teachers. No time limit was given; approximately 15–25 minutes were needed to complete the test. Non-native speakers and subjects older than 35 were excluded from the analysis, as were incomplete questionnaires. The number, sex, and age of the subjects can be seen from Table 3.

3. *Results and discussion*

The statistics were computed with SPSS at the Leibnitz Rechenzentrum in Munich. Because of the nature of our variables, we decided on non-parametric tests and correlations. The 0.05 level of significance was chosen in advance; however, for every difference the p-value is given as well. To save space we will restrict the documentation and discussion to the results relevant for the alternate hypotheses. In Table 4, the mean \bar{x} and the standard deviation s are given; the figures for M1 to M6 serve only to show the values and range and are not discussed in detail; for reasons already mentioned we take into consideration only the overall mean of the 24 test items XX, the conservatism scale KX, the differences $d_1 = M6 - M3$, $d_2 = (M4 + M5)/2 - (M1 + M2)/2$, and $d_1 - d_2$. (We will explain below why we also want to discuss this last difference.)

Table 4. Mean values and standard deviations

	English		German		Danish		Swedish	
	\bar{x}	s	\bar{x}	s	\bar{x}	s	\bar{x}	s
M1	1.04	1.21	2.00	1.41	2.33	1.55	2.55	0.96
M2	1.07	1.16	1.29	1.42	2.08	1.67	1.80	1.06
M3	1.30	1.03	1.79	1.31	2.83	1.46	3.30	0.85
M4	3.46	0.70	3.00	0.93	2.66	1.55	3.38	0.54
M5	3.60	0.49	1.83	1.40	2.33	1.49	2.44	0.80
M6	3.69	0.51	2.00	1.38	1.91	1.37	1.41	1.33
XX	2.50	0.40	2.38	0.52	2.43	0.54	2.67	0.35
KX	2.19	0.53	2.03	0.42	1.81	0.31	2.20	0.38
d ₁	2.39	1.11	0.21	1.21	-0.92	1.24	-1.89	1.52
d ₂	2.47	0.96	0.77	0.85	0.29	0.78	0.73	0.84
d ₁ -d ₂	-0.08	1.10	-0.56	1.74	-1.21	1.74	-2.62	1.60

To start with we would like to show that two possibly intervening variables — conservatism and sex of subject — did not have any decisive influence.

3.1. *The conservatism scale*⁷

In two-tailed Mann-Whitney tests two of the differences between the languages were significant for KX: English \gg Danish ($p < .008$), and Swedish \gg Danish ($p < .003$). (From now on, ' \gg ' is to be read as 'is significantly greater than'.) Only the first one is in line with the differences we predicted. The only significant correlations between KX and XX or any of the subgroups L-S, O-S, C-S, and M-S in any of the samples or in a sample 'ALL' (all the samples taken together) are the following: in German, $r_{KX,C-S} = .44$ ($p < .016$), and $r_{KX,L-S} = .38$ ($p < .032$); in ALL, $r_{KX,C-S} = .20$ ($p < .016$). However all these correlations are low and therefore of little interest. The only significant correlation between KX and d₁ and d₂ occurred in Danish $r_{KX,d_2} = .59$ ($p < .02$). On the other hand in ALL the correlations between the language scale LG and d₁ and d₂ respectively are not only significant but also sufficiently high: $r_{d_1,LG} = .83$ ($p < .001$) and $r_{d_2,LG} = .59$ ($p < .001$); these correlations are consistent with our hypotheses. As there is no substantial correlation between KX and d₁ or d₂ (with one single exception in Danish) we do not have to compute the partial correlations $r_{d_1,LG-KX}$ and $r_{d_2,LG-KX}$. The negative correlation $r_{d_2,KX}$ in Danish could indicate an influence of conservatism on the experimental task. Other correlations with KX are either very low or nonexistent. We can therefore assume that as regards the sentence rating, the conservatism of the subjects had no or very little influence. Only one

correlation indicated that the subject's awareness of the experimental task in connection with different degrees of conservatism might have influenced his or her rating. Unfortunately this significant negative correlation between KX and a d -value occurs in Danish where it is not easy to interpret, as in Danish MAN and HE are both neuter and therefore obviously can't be pseudogeneric. How can we explain that less conservative Danish subjects had to be attributed a higher d_2 value than more conservative ones? Maybe we could assume, in favor of our design, that in the other languages the different genders overrode any possible effect of conservatism? Even if we cannot totally exclude conservatism as an intervening variable, we can nevertheless conclude that conservatism does not unduly favor our alternate hypotheses. (We will come back to this problem in the discussion below.)

3.2. *Sex of subjects*

No significant differences between the sexes are to be found for d_1 and d_2 ; for KX only in ALL, $f \gg m$ ($p < .010$); for XX, only in German, $f \gg m$ ($p < .003$). There appears to be a slight general tendency for women to score higher than men. This is in accordance with sociolinguistic research, which shows that women often score themselves as more conservative language users than men, and it is also in accordance with Wilson and Patterson (1968: 266), who state that 'The sex difference is consistent, but very small, and may be disregarded for most practical purposes. The correlation with age is more significant, and where comparisons across age groups are contemplated, it would be necessary to use age-standardized scores.' In our experiment, age was controlled by keeping it relatively constant throughout the samples. Martyna (1978: 136) and MacKay and Fulkerson (1979: 671) reported an effect of subject sex that might be explained — as they suggest — by a different pattern of imagery in response to the experimental task. However, this interpretation presupposes a certain kind of awareness of the experimental task. We didn't observe any significant effect of subject sex on d_1 and d_2 , so we might consider this as a proof of our assumption that the subjects indeed rated nothing but the acceptability of the test sentences.


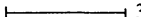












3.3. *Sentence rating*

For a start we have to stress that we cannot interpret the 'raw data' d_1 , d_2 , and $d_1 - d_2$ WITHIN languages, our design allowing us only to discuss their

differences BETWEEN languages.⁸ This explains the organization of Table 5, where the first three columns show the hypotheses and the predicted differences. In the fourth column, the p-values obtained by a one-tailed Mann-Whitney test are given. We will disregard for the moment the fifth column and the underlining in the third column. The sixth column shows graphically and numerically the differences of the mean d-values between the languages. We can see that all the differences are in the predicted direction and that all but one are significant; therefore we can accept H_{MAN} and H_{HE} in every respect, where H_{MAN^*} can be only partly accepted; for d_2 , German \gg Danish, but not German \gg Swedish.

H_{MAN} can be confidently accepted; it is consistent with all the other experimental results that were achieved for the pseudogeneric use of English *man*. As for the acceptance of H_{HE} , a caveat has to be made: M3 and M6 also contain MAN; the difference in the degree of acceptability of MAN that is parallel to that of HE could therefore be an additive factor in the case of German, not in the case of Danish and Swedish. As MAN in the latter two languages is either neuter or utrum its gender cannot contribute to a semantic interpretation. By computing the 'difference of

Table 5. Predicted and obtained differences

Hypothesis	d	Predicted difference	Sex-marking features			Difference of mean d-values between the languages
			p <	obl.	fac.	
(a) H_{MAN}	d_1	<u>Engl.</u> \gg Germ.	.000	2	0	 2.18
		<u>Engl.</u> \gg Dan.	.000	2	0	 3.31
		<u>Engl.</u> \gg <u>Swed.</u>	.000	3	0	 4.28
	d_2	<u>Engl.</u> \gg Germ.	.000	1	0	 1.70
		<u>Engl.</u> \gg Dan.	.000	1	0	 2.18
		<u>Engl.</u> \gg <u>Swed.</u>	.000	1	0	 1.74
(b) H_{HE}	d_1	<u>Germ.</u> \gg Dan.	.007	0	2	 1.13
		<u>Germ.</u> \gg <u>Swed.</u>	.000	1	2	 2.10
		Dan. \gg <u>Swed.</u>	.036	1	0	 .97
(c) H_{MAN^*}	d_2	<u>Germ.</u> \gg Dan	.047	0	1	 .48
		<u>Germ.</u> \gg <u>Swed.</u>	.481	0	1	 .04
(d)	$d_1 - d_2$	<u>Germ.</u> \gg Dan.	.151	0	1	 .65
		<u>Germ.</u> \gg <u>Swed.</u>	.000	1	1	 2.06
		Dan. \gg <u>Swed.</u>	.005	1	0	 1.41

the differences' $d_1 - d_2$, cf. Table 4 and Table 5(d) — note again that only this comparison BETWEEN languages is permissible — we can try to eliminate the influence of MAN and to arrive at a 'corrected' value. We can see in Table 5(d) that these differences are still in the predicted direction and that they are not inconsiderable. Two of these differences are significant; however, as d_1 and d_2 are values for sentences with different semantic content, this cannot be taken as a positive proof, but only as a kind of indicator. (We will come back below to some other factors that might be relevant.) Note by the way that with respect to d_2 , a 'female' interpretation is favored for the Swedish *hon*, this being in contrast to 'male' interpretations for the German *er* and the English *he*.

We can give no clear and simple explanation why H_{MAN^*} is only partly acceptable. If we assumed that — contrary to our expectations — grammatical gender in this case were fully arbitrary, then the reason for Danish subjects having a lower d_2 than German and Swedish ones might be the negative correlation $r_{d_2, KX} = .59$ (cf. the discussion of the conservative scale above). That could explain the one instance where H_{MAN^*} is acceptable.⁹ On the other hand the results of our second experiment show that the gender of German *der Mensch* in fact is not interpreted arbitrarily, cf. below.

As a result we could say that although the lexemes in question (MAN and HE) can be defined by semantic (*male/female*) as well as by syntactic (*masculine/feminine*) features, the subjects tended to interpret the items on a semantic rather than syntactic basis.

3.4. A fresh look at the features

Until now we have disregarded the fact that the GENERAL semantic status of these features differs. Table 6 below is similar to Table 1; however, now we indicate different 'degrees of sex-marking' by different boxes.

(i) The features enclosed by solid lines indicate in (nearly) all cases the sex of possible referents, i.e. they are OBLIGATORILY sex-marking. English *man* can never be interpreted as [+ female, – male]; the pronouns *he/she* in English and *han/hon* in Swedish are sex-marking with some minor exceptions: *she*, for example, can refer to 'sexless' referents like ships or cars, and Swedish *stjärna* 'star', for example, is referred to by *hon*. (For a more comprehensive account, cf. Palmer 1971: 37f., 189f; and Blume 1982, from whom we took over the distinction between obligatorily and facultatively sex-marking features.) One might expect that *hon* as the anaphoric pronoun of MAN is interpreted generically — however, this is not the case, as we have already seen. Still, the 'semantic load' of *hon*, so

Table 6. *MAN* and *HE*: different degrees of sex-marking

	(a) Possible sex of referent		(b) Anaphoric pronoun			(c) Gender of MAN		
	m	f	m	n	f	m	n	f
Common features	0	0						
	+	-						
English			+	-	-	0	0	0
German	-	+	+	-	-	+	-	-
Danish	-	+	-	+	-	-	+	-
Swedish	-	+	-	-	+	0	-	0

to speak, is intrinsically greater than that of Danish neuter *det* or of the Germ. masculine *er*, as we will see now.

(ii) The broken lines indicate FACULTATIVELY sex-marking features: the German article *der* and the pronoun *er* denote in most cases sexless referents with morphologically marked gender (e.g. *der Baum* 'the tree' — *er*; *der Tisch* 'the table' — *er*). Congruence between this semantic feature and the gender of the article and the pronoun normally occurs only when the referent is male or female. (For some interesting exceptions, cf. below.) From a functional point of view, therefore, the features defining these items hold a somewhat intermediate position between grammatical and natural gender. (We do not make any difference here between the degrees of sex-markedness of *der* and *er*, although it seems likely that the article is less sex-marked than the pronoun, cf. below the comment on examples (2), (3), and (4).

(iii) The features in Table 6 which are not enclosed by any lines do not mark sex at all, i.e. they are neuter or not specified with respect to semantic differences.

In the third and fifth columns of Table 5 the different semantic status discussed in (i), (ii), and (iii) is indicated through different ways of underlining (obligatory, facultative, unmarked), e.g. two facultatively sex-marking features (in der and er), and one obligatorily sex-marking feature (in Swedish *hon*) contribute to constitute the difference d_1 : German > Swedish.

As our hypotheses predicted only differences on an ordinal level, we didn't have to take into account that obligatorily sex-marking features might have a greater influence than the other two kinds. (We could therefore disregard a possible influence of the masculine possessive pronouns *seine* in the German M3 and *his* in the English M3 as well.)

Nevertheless a close look at the status of the features as it is described above reveals two interesting facts:

(i) If obligatorily sex-marking features are involved, the obtained differences are always significant and fairly large; if only facultatively sex-marking features are involved, the differences are smaller and not always significant. We can therefore state the 'truism' that semantic features are in fact to be interpreted semantically. But obviously this is no self-evident tautology — otherwise so much effort would not have been spent to prove this proposition for English generic *man* and *he*.

(ii) There also seems to be an additive effect: if we divide the kinds of differences shown in Table 6(a), (b), and (c) into two groups — one with obligatorily, the other with facultatively sex-marking features as relevant parameters — then we see that the more features involved, the greater the difference is. If we compute the mean of the differences of mean d -values between the languages with one obligatorily sex-marking feature, we get $\bar{x} = 1.65$; for two features, we get $\bar{x} = 2.74$; the mean of the only difference involving three features is 4.28. The same holds for the second group: if we have one relevant facultatively sex-marking feature, $\bar{x} = .26$; for two features the difference is 1.13. (Note that we disregarded *seine* in the German M3 and *his* in the English M3, as it is difficult to calculate their influence on the value of d_1 . In any case the rough ratio here wouldn't change.)

4. The second experiment

Compared with other features we didn't get equally clearcut results for the differences where the German facultatively sex-marking features were involved, cf. especially Table 5(c). For this reason we designed another WITHIN-language experiment with two German samples. The samples (students of introductory courses in linguistics), the design, and the instructions were basically comparable with the first experiment, with the following modifications: presumably because the samples were too homogeneous we didn't get any manifest interpretable results in the conservative test, and consequently we omitted it. The six sentences containing *der Mensch* were replaced by three dummy sentences and by the three sentence pairs shown in Table 7: M7(a) is identical with M4, and M9(b) is identical with M3; M8 could be translated as, 'Man has to restrain himself as long as he lives as $\left\{ \begin{smallmatrix} \text{daughter} \\ \text{son} \end{smallmatrix} \right\}$ with his parents.' One sample was given the (a) version denoting a female referent, the other sample the (b) version denoting a male referent. As alternate hypotheses, we assumed that the

Table 7. *Test items containing MAN, second experiment*

Item	Sentence
M7 $\left\{ \begin{smallmatrix} (a) \\ (b) \end{smallmatrix} \right\}$	Heutzutage bindet sich der Mensch weniger als früher durch Heirat an eine(n) $\left\{ \begin{smallmatrix} \text{Mann} \\ \text{Frau} \end{smallmatrix} \right\}$
M8 $\left\{ \begin{smallmatrix} (a) \\ (b) \end{smallmatrix} \right\}$	Der Mensch muß sich einschränken, solange er als $\left\{ \begin{smallmatrix} \text{Tochter} \\ \text{Sohn} \end{smallmatrix} \right\}$ im Elternhaus wohnt.
M9 $\left\{ \begin{smallmatrix} (a) \\ (b) \end{smallmatrix} \right\}$	Der Mensch ändert oft einen Teil seiner Gewohnheiten, wenn er $\left\{ \begin{smallmatrix} \text{Mutter} \\ \text{Vater} \end{smallmatrix} \right\}$ wird.

sentences in the (a) version would be rated significantly less acceptable than the same sentences in the (b) version.

For the same reasons as mentioned above with regard to the first experiment, the number of test items was kept small. In M7 there is one lexical item with a facultatively sex-marking feature (*der*), in M8 there are two (*der*, *er*), and in M9 there are three (*der*, *er*, and *seiner*). Because the respective semantic content of the sentences is different, the sentences are not strictly comparable as concerns effects of additivity. The distribution of age within the samples was as follows: -20 years: 5 (sample a)/13 (sample b), 21-23: 10/13, 24-26: 5/2, 27-29: 0/1.

The results are shown in Table 8. One-tailed Mann-Whitney tests were

Table 8. *Predicted and obtained differences, second experiment*

Subjects and their number	Sen- tence	Hypothesis: $a \gg b$				p <	Difference of mean values between the samples
		(a) \bar{x}	s	(b) \bar{x}	s		
(i) females and males $N_a = 20$ $N_b = 29$	M7	3.25	.55	2.24	1.35	.004	————— 1.01
	M8	3.25	.44	2.45	1.12	.002	————— .80
	M9	2.85	.67	1.65	1.23	.000	————— 1.20
(ii) females $N_a = 14$ $N_b = 21$	M7	3.36	.63	2.38	1.36	.017	————— .98
	M8	3.29	.47	2.67	.97	.016	————— .62
	M9	2.79	.80	1.81	1.17	.006	————— .98
(iii) males $N_a = 6$ $N_b = 8$	M7	3.00	.00	1.87	1.36	.026	————— 1.13
	M8	3.17	.41	1.87	1.36	.017	————— 1.30
	M9	3.00	.00	1.25	1.39	.004	————— 1.75

conducted with the female and male subjects taken together (i), with the female subjects only (ii), and with the male subjects only (iii), cf. the first column in Table 8. The mean \bar{x} and the standard deviation s for the (a) version is given in the third column, for the (b) version in the fourth column. In the fifth column the probability obtained is listed; the sixth column shows the differences of the mean values graphically and numerically. As for (i), we can see that all the differences are in the predicted direction and significant at the chosen level. We can therefore conclude that in German the facultatively sex-marking features of MAN and HE are in fact interpreted semantically, i.e. as indicators of the sex of the referent.

Part (ii) and (iii) show that if we look at the sexes separately, the differences are still significant in all cases; however, the mean values are smaller for females than for males; this could possibly be traced back to the role of self-imagery. In any case none of the differences between the sexes are significant. The p-values for a two-tailed Mann-Whitney test, male versus female subjects, for sample (a) are for M7: $< .14$, for M8: $< .58$, for M9: $< .54$, and for sample (b) for M7: $< .38$, for M8: $< .15$, and for M9: $< .24$. In (i) and (ii), no effect of additivity can be observed; in (iii), there is an effect; however, it is fairly small. This could be explained in two ways: either additivity is not relevant or, what seems more likely, the different semantic content of the sentences is an intervening variable that we didn't control.¹⁰

5. General discussion

5.1. A new aspect

We have shown experimentally that gender in the cases considered is not interpreted arbitrarily. However, one doesn't have to leave the realm of linguistics proper (that is, nonexperimental linguistics) to show that this is sometimes the case. A quotation from Lyons (1968: 286f.) shall serve as a starting point:

[If in French 'le nouveau professeur' refers to a woman, then] Neither *le nouveau professeur est beau* (which necessarily refers to a man) nor **Le nouveau professeur est belle* (which is ungrammatical) is possible. The 'conflict' is resolved with a sentence like *Elle est belle, le nouveau professeur* ('She is beautiful, the new teacher.').

Lyons goes on to remark that this and further examples suggest that grammatical cohesion within the noun phrase is stronger than it is

between subject and predicate. Here again it is worthwhile to look at German, because in this language three different genders exist that are always morphologically marked, unless neutralization takes place. In the other languages under consideration, it is always possible to use a relative pronoun which is not specified with respect to gender, cf. English *that*, French *qui*, and Danish/Swedish *som*. In German, there are only gender-specific relative pronouns as *der*, *die*, *das*. This particular point can be illustrated by the examples below. (As there are no English equivalents for those particular constructions, we tried to give approximate, however awkward, translations.)

(2) *gender:neuter/sex:female*

Schau dir dieses *Mädchen* an, $\left\{ \begin{array}{l} \left\{ \begin{array}{l} \text{das} \\ *die \end{array} \right\} \text{ so gut Tennis spielt.} \\ \text{wie gut } \left\{ \begin{array}{l} es \\ sie \end{array} \right\} \text{ Tennis spielt.} \end{array} \right\}$

'Look at this *girl*, $\left\{ \begin{array}{l} \text{who plays tennis so well.} \\ \text{how well she plays tennis.} \end{array} \right\}$,

(3) *gender:feminine/sex:male*

Hast du diese *Memme* gesehen, $\left\{ \begin{array}{l} \left\{ \begin{array}{l} die \\ *der \end{array} \right\} \text{ einfach davonläuft?} \\ \text{wie } \left\{ \begin{array}{l} sie \\ er \end{array} \right\} \text{ einfach davonläuft?} \end{array} \right\}$

'Have you seen this *coward*, $\left\{ \begin{array}{l} \text{who just runs away?} \\ \text{the way he just runs away?} \end{array} \right\}$,

(4) *gender:masculine/sex:female*

Hast du diesen *Drachen* gesehen,

$\left\{ \begin{array}{l} \left\{ \begin{array}{l} der \\ *die \end{array} \right\} \text{ ihm keine Ruhe läßt?} \\ \text{wie } \left\{ \begin{array}{l} er \\ sie \end{array} \right\} \text{ ihm keine Ruhe läßt?} \end{array} \right\}$

'Have you seen this *shrew*, $\left\{ \begin{array}{l} \text{who doesn't leave him alone?} \\ \text{they way she doesn't leave him} \\ \text{alone?} \end{array} \right\}$,

In these and some (though not many) other instances, the gender of the noun does not agree with the sex of the referent. The relative pronoun and the article nowadays MUST show congruence of gender, otherwise the sentence would be ungrammatical; the anaphoric personal pronoun may show agreement with gender OR sex. In any case, the respective sentences

will have slightly different connotations and they also may have different degrees of acceptability. For these cases, some grammars try to account for these sentences by claiming a tendency toward grammatical congruency, cf. Dal (1952: 167f.):

In der heutigen Sprache verwendet man bei neutralen Bezeichnungen für weibliche Personen meistens die Femininform des persönlichen und des possessiven Pronomens, während das Relativpronomen grammatische Kongruenz zeigen muß (...). Jedoch dringt formale Kongruenz immer mehr durch.

Also cf. Duden (1966: 630): 'Die Berücksichtigung des natürlichen Geschlechtes war früher üblicher, ist heute jedoch seltener (...)'. Obviously, some time ago, semantic congruency was more common, cf. Paul (1920: 265), who quotes Goethe and Wieland: '(...) vgl. Fälle wie *eines Frauenzimmers, die sich am artigsten gegen mich erwiesen hat* (Goe.); *die hässlichste meiner Kammermädchen* (Wieland); (...)'. Grimm and Grimm (1878: 88) observed a tendency contrary to the one stated by Dal (1952) and Duden (1966): 'FRÄULEIN, *n. und f., das genus in adj. und pron. schwankend, doch so dasz allmählich f. überwiegt.*' We can only guess why this earlier tendency and the nonarbitrary approach toward gender (cf. the quotations by Gabelentz and Jespersen in the first section) changed to the contrary: it could well be that this is another case of the structuralist *a priori* belief in the *arbitraire du signe* that is reflected in the modern grammars here. (This explanation might suggest itself if one looks, for example, at Royen 1929, who collected several prestructuralist theories about the origin and the classification of gender — all of them decidedly nonarbitrary.) However, if Gabelentz was right, and *Mensch* could hardly ever refer to females, then there also must have been some change in language use. It is obvious that at least our German subjects interpreted the gender of *der Mensch* more arbitrarily — though still in a pseudogeneric way — than Gabelentz would have expected.

Let us now go back to examples (2), (3), and (4). It could be possible that, as a result of the different connotations connected with these sentences, the events referred to are interpreted differently. Considering this possibility we may arrive at the pivotal point of this matter. Hörmann (1978: 460–496) in his chapter on 'Ebenen und Vektoren des Verstehens' summarizes the state of research on sentence comprehension, arguing that sentences are not comprehended or stored in their original form, but rather after they have undergone some transformations, that is after they have been interpreted semantically, after some generalizations have been made and some images evoked. This concept is basically also applicable to the question whether grammatical gender is arbitrary or pseudogeneric, and whether it is always the one or the other. The images evoked may

vary considerably, sometimes being quite abstract (even if the referent is human), sometimes rather concrete, always depending on specific contexts and situations: *der Mensch* (and its equivalents in the other languages) may sometimes be comprehended as an abstract, generic term for human beings, thus functioning as a universal quantifier, but sometimes it may be comprehended as a pseudogeneric term; in our test items, the context certainly favors a pseudogeneric interpretation. *Der Ehepartner* ('the spouse') may be interpreted relationally and therefore more abstractly and generically than, let's say, *der Boxer* ('the boxer'), in which case a clear and concrete — male — image is evoked, even if all these expressions were meant as generic terms. (In the last two instances, the male term can be generic as well; however, there exist female counterparts: *die Ehepartnerin*, *die Boxerin*.) For this reason we should not only look at some features defined by linguists, but also take into consideration the process of comprehension itself. In the introduction above we emphasized that the features in Table 1 are to be interpreted literally. It may be that the ontological (semantic) status of these features (which has not been discussed) is invariable in isolated noun phrases or in citation forms; however, this status may be very much altered and in the end constituted by the context and the process of comprehension. (For an argumentation along similar lines, cf. MacKay 1980.)

5.2. *Limitations*

The first limitation to be mentioned results from our relatively small set of variables. In spite of the precautions we have taken, the results may thus be influenced by some idiosyncrasies of the items which were not controlled. For this reason we must not generalize beyond the scope of our variables. Of course we can take the results as plausible starting points for working hypotheses, for example, 'French *homme* behaves like English *man* in Table 1(a), and thus we can assume that it is more pseudogeneric than, let us say, German *Mensch*.' With a greater set of variables one might try to unravel the influence of the particular variables (e.g. the gender of MAN and HE) and thus to achieve better control of additivity.

With a greater number of subjects and a more balanced relation between males and females, a possible effect of subject sex may show up more clearly. Similarly we cannot generalize beyond the population from which we selected our samples — students of linguistics. (Strictly speaking, we cannot even generalize beyond our samples, as they were not really randomly selected. However, this is a common dilemma.) Maybe

we can make assumptions about a population of students in general, but it is very likely that age and education are relevant factors, and thus we can only make guesses about their influence on our experimental task.

Another moot point is whether the conservatism test could control any possibly intervening variables and whether our finding is really true, namely that there is no positive correlation between conservatism and normatism (the tendency to rate more sentences as less acceptable). The latter may be correct only for our relatively homogeneous samples but not for heterogeneous ones which vary considerably, for example, with respect to age and education.

Another very important limitation was dealt with above: the role of the context and the images evoked have not been taken into account.

5.3. *Concluding remarks*

In so far as the limitations just mentioned are concerned, most of the objections against 'the great "he/she" battle' raised by Dubois and Crouch (1978: 8) hold for our study as well. Still, to our knowledge we were the first to be concerned with languages other than English both on an experimental basis and with a cross-linguistic aspect. On the other hand, we had to keep constant as many variables as possible, which means that within the range of our task the limitations with regard to age, education, etc., should not be considered as serious drawbacks but rather as necessary prerequisites.

From our results we must not draw the conclusion that 'all instances of MAN and HE are pseudogeneric'; nevertheless we can conclude that 'at least some instances are', and that this holds not only for English but also for German and Swedish. These instances of pseudogeneric use can be traced back to LINGUISTIC factors, i.e. the different degree of sex-marking of the features involved. Thus far our results are consistent with a weak version of linguistic relativity along the lines of, for example, Carroll (1963: 12).

Once more we want to stress that our design was to serve a special cross-linguistic task, and that we were only concerned with different degrees of semanticity, not with additivity or other factors. The former turned out to be quite stable — it featured in all but one of the differences. As for additivity, we are now at least in a better position to formulate some hypotheses that can be tested.

To sum up our approach we can say that we took as a starting point the common assumption (which is discussed and criticized in the studies referred to in note 1) that the status of features attributed to MAN and

HE (cf. Tables 1 and 6) is EITHER purely syntactic (expressing congruence) OR semantic (denoting the sex of the possible referent), and only rarely something in between. We questioned this dichotomy and were able to prove on an experimental basis that the syntactic features of MAN and HE are interpreted semantically much more often than some linguists would have us believe. However, there remains reason to believe that this 'experimental linguistic' approach must be supplemented by a different one that is able to deal not only with the OUTCOME of comprehension, but with the PROCESS of comprehension itself.¹¹

Received 21 August 1984

Institut für Nordische Philologie
University of Munich
Geschwister-Scholl Platz 1
8000 Munich 22
West Germany

Notes

1. Cf. for example the conceptual papers by Bate (1975), Beardsley (1975), Bodine (1975), and Stanley (1977), and the experimental papers by Schneider and Hacker (1973), Martyna (1978), Moulton et al. (1978), and MacKay and Fulkerson (1979). As for more references, cf. Silveira (1980), who sums up the arguments and the different experimental designs.
2. Representative examples of the pseudogeneric use of *man* and its equivalents in the other languages could possibly be found in frequency dictionaries. Such dictionaries exist for German, cf. Rosengren (1972, 1977); for Danish, cf. Mægaard and Ruus (1978); and for Swedish, cf. Allén (1970ff.). For each language there are several hundred entries for *Mensch*, *menneske*, and *människa*, respectively. In the respective concordances, a context of up to 120 characters is given, i.e. approximately 4–10 words before and after the item. We could only find one single instance of clearcut pseudogeneric use. Obviously, pseudogeneric use of the items in question is not as frequent in these languages as it is in English; to discover further instances, the contexts were too small.
3. In this context, however, *en människa* is generic as well, possibly having slightly different connotations. If we look at the ratings of Swedish M3 and M6 in Table 4, we can see that M6 contributes in roughly the same way to the value of d_1 as does M3. (For Swedish M4 and M5, it does not matter whether a nongeneric interpretation is also possible which might unduly favor H_{MAN^*} : the respective results, cf. in Table 5 d_2 : German \gg Swedish, are not significant anyway.)
4. Pilot studies have shown that the 'specific' nature of the deviant sentences had to be pointed out. In the instructions, the sentences were therefore characterized rather vaguely as 'often not grammatically incorrect, but contradicting our experience'.
5. To our knowledge, studies that relate an attitude of 'conservatism' to the rating of linguistic phenomena are rare; Ross (1979), for example, assumes such a correlation; however, he does not apply any inferential statistics. Brown and Gilman's (1960)

classic study has shown that there can be a comparable correlation between conservatism and the choice of the formal or the informal form of address.

6. We do not want to hide our doubts about the general value of such attitude tests; however, they did prove to be useful for controlling the variables in question, cf. Schmidt et al. (1975: 50f.), thus serving our purpose exactly.
7. In Table 9, the reliability coefficients (Guttman split-half) obtained for the items are given for the separate samples and for a sample 'ALL' (all the samples taken together). In comparison,

Table 9. *Obtained and predicted reliability coefficients in the conservatism test*

	N	English	German	Danish	Swedish	ALL
Obtained	23	.85	.75	.75	.82	.80
Predicted	40		.84			.88
Predicted	50	.92				.90

Wilson and Patterson (1968) obtained a coefficient $r = .94$ for 50 English items, and Schneider and Minkmar (1972) a coefficient $r = .86$ for 40 German items. The lower coefficients of KX can be explained as a consequence of the lower number of test items: the respective coefficients for KX predicted by the Spearman-Brown Profecy Formula for 40 or 50 items are also given in the table for English, German, and ALL. We had to put up with the lesser degree of reliability, because otherwise our samples could not have been strictly comparable: that is, tested in a language-free way. Although the comparability of individuals may not exactly meet the standards, the samples are comparable — and this, after all, is the purpose of KX.

8. One example may explain this point: because there should be an additional effect of HE for d_2 , in German d_2 should be larger than d_1 ; however, it is the other way round. Presumably the content of M6 is so evidently true — smoking is dangerous, after all — that its general acceptability partly overrides the incongruency of *er* and *schwanger*. If we assume that this holds for the other languages as well, we can ignore this effect, comparing the d-values only BETWEEN languages.
9. The German and Swedish values for d_2 are approximately equal. The difference between Danish and Swedish cannot be taken into account in the inferential statistics, because it is not in the predicted direction.
10. As an informal test of reliability we can compare the ratings of the identical sentences M7(a) and German M4 in the first experiment: 3.25 versus 3.00; and of M9(b) and German M3: 1.65 versus 1.79. These values are indeed quite close together.
11. Without the help of many people this paper could not have been written. Among others I am especially grateful for their assistance to Gillian Brown, Alan Kemp, Inger Berggren, Ib Poulsen, Thomas Becker, Beatrice Primus, Hartmut Günther, and Klaus Netter.

References

- Allén, S. (1970ff.). *Nusvensk Frekvensordbok*, 1–3. Stockholm: Almqvist and Wiksell.
 Bate, B. A. (1975). Generic man, invisible woman: language, thought, and social change. *The University of Michigan Papers in Women's Studies* 2, 83–95.

- Beardsley, E. L. (1975). Referential genderization. *The Philosophical Forum (Women and Philosophy)* 5 (1-2) (Fall-Winter 1973-1974), 285-293.
- Blume, H. (1982). Schwedisch *han/hon* und seine Entsprechungen im Deutschen. *Skandinavistik* 12, 137-151.
- Bodine, A. (1975). Androcentrism in prescriptive grammar: singular 'they', sex-indefinite 'he', and 'he or she'. *Language in Society* 4, 129-146.
- Brown, R., and Gilman, A. (1960). The pronouns of power and solidarity. In T. A. Sebeok (ed.), *Style in Language* 253-276. Cambridge, Mass.: MIT Press.
- Bühler, K. (1934). *Sprachtheorie*. Stuttgart: Fischer. (Reprinted 1965.)
- Carroll, J. B. (1963). Linguistic relativity, contrastive linguistics, and language learning. *International Review of Applied Linguistics in Language Teaching* 1, 1-20.
- Dal, I. (1952). *Kurze deutsche Syntax*. Tübingen: Niemeyer.
- Dubois, B. L., and Crouch, I. (1975). The question of tag questions in women's speech: they don't really use more of them, do they?↓ *Language in Society* 4, 289-294.
- , and Crouch, I. (1978). Introduction to special issue: American Minority Women in Sociolinguistic Perspective. *International Journal of the Sociology of Language* 17, 5-15.
- Duden (1966). *Duden. Grammatik der deutschen Gegenwartssprache*, 2nd ed. Mannheim: Bibliographisches Institut.
- Gabelentz, G. von der (1969). *Die Sprachwissenschaft*. Tübingen: Narr. (First edition Leipzig 1891.)
- Grimm, J., and Grimm, W. (1878). *Deutsches Wörterbuch. Vierten Bandes erste Abtheilung. Erste Hälfte*. Leipzig: Hirzel.
- Hansen, A. (1967). *Moderne Dansk. II*. Copenhagen: Grafisk Forlag.
- Hörmann, H. (1978). *Meinen und Verstehen*. Frankfurt am Main: Suhrkamp.
- Jespersen, O. (1968). *The Philosophy of Grammar*. London: Allen and Unwin. (First edition 1924.)
- Lyons, J. (1968). *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- MacKay, D. G. (1980). Language, thought and social attitudes. In H. Giles, W. P. Robinson, and P. M. Smith (eds.), *Language: Social Psychological Perspectives*, 89-96. Oxford: Pergamon.
- , and Fulkerson, D. C. (1979). On the comprehension and production of pronouns. *Journal of Verbal Learning and Verbal Behavior* 18, 661-673.
- Mægaard, B., and Ruus, H. (1978). DANwORD. Hyppighedsundersøgelser i moderne dansk: Baggrund og materiale. *Danske Studier* 73, 42-70.
- Martyna, W. (1978). What does 'he' mean? Use of the generic masculine. *Journal of Communication* 28, 131-138.
- Moulton, J., Robinson, G. M., and Elias, C. (1978). Sex bias in language use. 'Neutral' pronouns that aren't. *American Psychologist* 33, 1032-1036.
- Palmer, F. (1971). *Grammar*. Harmondsworth: Penguin.
- Paul, H. (1920). *Prinzipien der Sprachgeschichte*. Halle: Neimeyer. (Fifth edition.)
- Pusch, L. F. (1979). Der Mensch ist ein Gewohnheitstier, doch weiter kommt man ohne ihr. — Eine Antwort auf Kalverkämpers Kritik an Trömel-Plötz' Artikel über 'Linguistik und Frauensprache'. *Linguistische Berichte* 63, 84-102.
- Rosengren, I. (1972, 1977). *Ein Frequenzwörterbuch der deutschen Zeitungssprache*. 1,2. Lund: CWK Gleerup.
- Ross, J. R. (1979). Where's English? In C. Fillmore, D. Kempler, and W. S.-Y. Wang (eds.), *Individual Differences in Language Ability and Language Behavior*, 127-163. New York: Academic Press.

- Royen, G., O.F.M. (1929). *Die nominalen Klassifikations-Systeme in den Sprachen der Erde*. Mödling bei Wien: Administration des 'Anthropos'.
- Schmidt, H. D., Brunner, E. J., and Schmidt-Mummendey, A. (1975). *Soziale Einstellungen*. Munich: Juventa.
- , Schmerl, C., Krameyer, A., Wagner, A., Steinbach, D., and Schmidt-Mummendey, A. (1973). *Frauenfeindlichkeit. Sozialpsychologische Aspekte der Misogynie*. Munich: Juventa.
- Schneider, J. W., and Hacker, S. L. (1973). Sex role imagery and use of the generic 'man' in introductory texts: a case in the sociology of sociology. *The American Sociologist* 8, 12–18.
- Schneider, J. and Minkmar, H. (1972). Deutsche Neukonstruktion einer Konservatismus-skala. *Diagnostica* 18, 37–48.
- Silveira, J. (1980). Generic masculine words and thinking. *Women's Studies International Quarterly* 3, 165–178.
- Smith, P. M. (1982). Review of *Women and Men Speaking: Frameworks for Analysis*, by Cheri Kramarae. *Journal of Language and Social Psychology* 1, 83–86.
- Stanley, J. P. (1977). Gender-marking in American English: usage and reference. In A. P. Nilsen, H. Bosmajian, H. L. Gershuny, and J. P. Stanley (eds.). *Sexism and Language*, 43–74. Urbana: National Council of Teachers of English.
- Suttner, B. von (1889). *Das Maschinenzeitalter. Zukunftsvorlesungen über unsere Zeit. Von Jemand*. Zürich: Schabelitz.
- Thorell, O. (1973). *Svensk grammatik*. Stockholm: Esselte Studium AB.
- Werner, O. (1975). Zum Genus im Deutschen. *Deutsche Sprache* 3, 35–58.
- Whorf, B. J. (1956). *Language, Thought, and Reality. Selected Writings Edited and with an Introduction by John B. Carroll*. Cambridge, Mass.: Technology Press of MIT.
- Wilson, G. D., and Patterson, J. R. (1968). A new measure of conservatism. *British Journal of Social and Clinical Psychology* 7, 264–269.