

## **Actualization of a Counter-Change: Contractions on the Internet**

Susan C. Herring  
*University of Texas at Arlington*  
herring@uta.edu

[The] cultural and functional determinants [...] of change derive from the general conditions of language and are, whenever a given language is adequately documented, by and large open to investigation.  
(Cosieriu 1952: 83; quoted in Andersen 1989)

### **1. Introduction**

Traditional studies of language change face severe methodological limitations: Older stages of a language tend to be sporadically and discontinuously documented, and those texts that survive through the ages inevitably lack socio-cultural context. Recently, however, a new source of data has become available that has the potential to provide a more complete picture of how, when, and why languages change: computer-mediated communication via the Internet. The Internet constitutes a self-contained context for authentic social interaction, and preserves a typed record of past interaction, including archives that date from the early days of computer networking. This makes it possible, at least within the domain of the Internet itself, to observe a language continuously over time in its social contexts of use, and thereby to investigate “the cultural and functional determinants of change”, in Cosieriu’s (1952) words, in new and more comprehensive ways.

Among the questions that the Internet can help illuminate is one at the very core of historical linguistic inquiry: What causes a linguistic innovation to arise (termed the ‘actuation problem’ by Weinreich, Labov, and Herzog 1968) and spread (termed ‘actualization’ by Andersen 1973)? One influential model proposes that the source of an innovation determines its subsequent spread; specifically, that a language-internal innovation, motivated by naturalness considerations, will follow a different actualization path from a language-external, socially motivated innovation (Romaine, cited in Andersen 1997). A general principle claimed to determine this spread is similarity between the source of the innovation—as unmarked (internal) or marked (external)—and the linguistic contexts (unmarked or marked) in which the innovation first appears and from which it spreads gradually to contexts with the opposite markedness value (Andersen 1997; Timberlake, this workshop). Andersen (1997) terms this the Principle of Markedness Assimilation.

This paper asks how the principle of markedness assimilation fares in accounting for language change on the Internet, given the availability of continuous or near-continuous language use data that have accumulated through online communication over the past 25 years. The principle of markedness assimilation is based on empirically-testable claims, three of which are of particular interest here. First, given sufficient textual evidence to establish that a linguistic change has taken place, the principle of markedness assimilation claims that patterns of spread are evident in the linguistic data, from marked to unmarked contexts or vice versa. Second, it claims that there is a relationship between the trajectory of spread and the source of the innovation (internal or external). Third, it claims that the actualization of a change, once underway, is linear and unidirectional (Andersen, 1997). Conversely, I argue in this paper that in the case of contraction in English on the Internet, the linguistic data do not pattern clearly according to markedness values, and thus the markedness assimilation principle sheds little light on why or how change in their use takes place; rather, the explanation for change resides in text-external social facts. Moreover, change in contraction use over a roughly continuous 23-year period on the Internet is not unidirectional, but rather appears to reverse its directionality mid-way.

These observations raise questions about the nature of Internet language data and call into question the usefulness of the markedness principle as a broad explanation for language change. They point instead to the explanatory importance of social factors, in keeping with a growing body of linguistic research from a socio-historical perspective. I conclude by suggesting that the popularity of “structure-internal” explanations for linguistic change reflects a need to compensate for contextually-impoverished historical data, and that by providing continuous, socially-contextualized data, the Internet has the potential to bring about a paradigm shift in historical linguistics, one that places greater emphasis on social factors in accounting for language variation and change.

## **2. Contractions in Internet English**

Evidence for these claims comes from the use of contracted two-word expressions in written English in discussion forums on the Internet. The contractions in question comprise three basic structural types: pronoun + auxiliary (e.g. I am --> I'm), auxiliary + 'not' (e.g. do not --> don't), and auxiliary + auxiliary (e.g. must have --> must've).<sup>1</sup>

I examined all instances of contraction as defined above in periodic samples taken from two e-mail-based Internet discussion groups: MsgGroup, which operated between 1975 and 1986, and the Linguist List (Linguist-L), which began in December 1990 and continues to the present time.<sup>2</sup> The groups are similar in user demographics (highly educated, middle-class, English-speaking professionals, predominantly male) and purpose (serious discussion of issues relating to the groups' respective professional foci). The groups differ in the subject matter of discussion—computer message system design as compared to linguistics—and in the time period during which they arose and were active: MsgGroup was one of the first e-mail discussion groups on the ARPANET (the

predecessor of the Internet), at a time when only computer science professionals had access to e-mail, while Linguist-L came into being as part of a later surge in interest in online discussion groups that took place when Internet access started to become widely available to scholars in academic institutions.<sup>3</sup> Complete archives of all messages posted to each group are publicly available on the Internet.

Contraction, a feature of spoken language traditionally considered to be motivated by ease of production, has since the 17th century been infiltrating the written language as part of a general drift towards the colloquialization of written English (Biber & Finegan 1989).<sup>4</sup> This drift continues up to the present time: Mair and Hundt (1997) found that the frequency of contraction has continued to increase in written English during the past 30 years. There is therefore good reason to expect to find evidence of this trend in e-mail communication, all the more so in that e-mail is widely considered to be more “oral” and informal in style than other forms of writing (Cho forthcoming; Hale 1996; Maynor 1994). Furthermore, some researchers maintain that the computer medium itself encourages structural reduction—abbreviation, contraction, and ellipsis—as a means to conserve keystrokes and increase the speed of typing (Cho forthcoming; Ferrara et al. 1991; Murray 1990). According to this view, one might expect the Internet context not simply to reflect, but actually to accelerate, the increase in contraction use over time.

However, such is not the case—at least not on the Linguist List. The rates of contraction over time in the two discussion groups are charted in Figure 1. Simple linear regression (dotted) lines show the overall trends in each group.

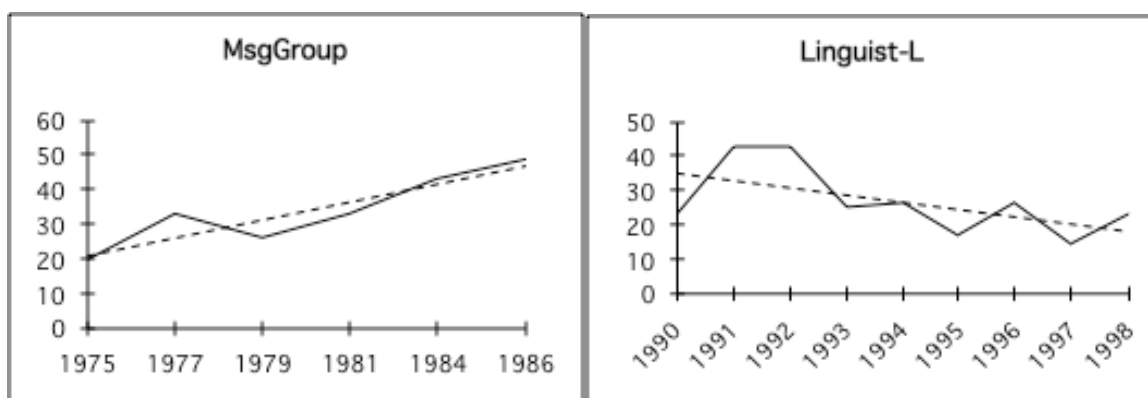


Figure 1. Rate of contraction in two discussion groups over time

Contraction rate (calculated as the percentage of possible contraction sequences that are actually contracted) increases steadily between 1975 and 1986 in messages posted to MsgGroup, as expected. However, the direction of change is the opposite between 1990 and 1998 in Linguist List messages; that is, contraction use decreases. Indeed, the overall rate of contraction in 1997-1998 is almost the same as in 1975, such that if one only had access to data at these two points in time, one might infer (mistakenly) that little change in frequency of contraction had taken place during the intervening 23 years.

This overall result is initially puzzling. The pattern in the older group poses no problem: It can be interpreted as a natural, internally-motivated, and hence ‘unmarked’ change. Andersen credits Dressler (1980) with the observation that “lenition and other obscuration innovations are favored by unmarked environments” (1997, p. 14) such as casual styles. Given Herring’s (1998) finding that language use on MsgGroup became increasingly informal over the 11 years of the group’s existence, the rise in contraction in MsgGroup appears to support the Principle of Markedness Assimilation. But what of the apparent reversal of this trend in the Linguist List?

In what follows, I attempt to account for this counter-trend first in terms of structural patternings in the linguistic data—which are hypothesized, following Andersen (1997), to show markedness agreement—and subsequently in terms of social and situational factors associated with the Linguist List and the larger Internet cultural context. It is found that although individual forms vary in the frequency with which they contract, attempts to group the forms into morphological or lexical categories reveal few clear patterns and thus little evidence of spread across marked or unmarked contexts. A more plausible explanation for the counter-trend involves social factors such as message purpose (an increase in announcements over interactive messages), gender of users (women decline more sharply in their use of contractions than men), and group-internal cultural dynamics that led to an increase in level of overall formality in Linguist List discourse.

This proposed explanation—that contraction loss (a marked change) has an external (marked) source—appears to bring the results of the analysis back into alignment with the markedness assimilation principle, especially if the increase in contraction on MsgGroup is considered to have a system-internal (e.g. natural) source. However, further evidence is presented that suggests that change in contraction had an external trigger in MsgGroup, as well; thus social factors appear to account better than markedness overall for the actuation and actualization of change in contraction use on the Internet.

## **2.1 Structural evidence**

Previous researchers have ascribed markedness values to the members of paired structural contrasts on the levels of phonology, morphology, syntax, style, medium, and genre (Andersen 1997; Timberlake 1977). In this section, I explore whether there are structural patterns in the Linguist List data that provide an explanation for the decrease in contraction in terms of its association with marked and unmarked contexts.

### **2.1.1 Overall distribution**

Since contractions are made up of lexical items that encode morphological information (e.g. pronouns have person and number; auxiliaries encode morphological tense and sometimes person and number), I coded each lexical item found in the contractions in the Linguist data for these distinctions, as well as for the position it can occupy in a contraction (first member, second member, or either). I then compared this

information with the rate at which each lexical item is contracted in the corpus (number of actual contractions divided by the total tokens of the item). Table 1 ranks the results for the Linguist-L corpus as a whole.

Table 1. Overall rate of contraction by lexical item

Item	Total tokens	# contracted	% contracted	Position in contr.	Word class	Person	Number	Tense
am	312	113	36.2%	2nd	aux	1	sg	pres
do	319	89	27.9%	1st	aux	1, 2, (3)	sg, pl	pres
not	1224	269	22.0%	2nd	neg	--	--	--
does	240	42	17.5%	1st	aux	3	sg	pres
I	1763	241	13.7%	1st	pron	1	sg	--
would	530	69	13.0%	either	aux	1, 2, 3	sg, pl	past
did	78	9	11.5%	1st	aux	1, 2, 3	sg, pl	past
she	96	11	11.5%	1st	pron	3	sg	--
have	880	97	11.0%	either	aux	1, 2, (3)	sg, pl	pres
can	639	55	8.6%	1st	aux	1, 2, 3	sg, pl	pres
we	711	56	7.9%	1st	pron	1	pl	--
it	1210	87	7.2%	1st	pron	3	sg	--
here	161	10	6.2%	1st	adv	--	--	--
had	147	9	6.1%	either	aux	1, 2, 3	sg, pl	past
will	1217	69	5.7%	either	aux	1, 2, 3	sg, pl	pres
you	856	46	5.4%	1st	pron	2	sg, pl	--
is	2954	153	5.2%	either	aux	3	sg	pres
could	170	8	4.7%	1st	aux	1, 2, 3	sg, pl	past
were	141	6	4.3%	1st	aux	2, (3)	(sg), pl	past
there	557	20	3.6%	1st	adv	--	--	--
they	349	11	3.2%	1st	pron	3	pl	--
are	1496	48	3.2%	either	aux	2, 3	(sg), pl	pres
what	415	13	3.1%	1st	wh-pron	--	--	--
he	194	4	2.1%	1st	pron	3	sg	--
has	471	7	1.5%	1st	aux	3	sg	pres
that	1888	19	1.0%	1st	pron	3	sg	--
was	395	4	1.0%	1st	aux	3	sg	past
need	117	1	.9%	1st	aux	1, 2, 3	(sg), pl	pres
should	486	2	.4%	1st	aux	1, 2, 3	sg, pl	past
must	230	1	.4%	1st	aux	1, 2, 3	sg, pl	pres
who	353	1	.3%	1st	wh-pron	--	--	--
N=31	20,599	1570	7.6%	--	--	--	--	--

As Table 1 shows, individual lexical items vary widely in the rate at which they participate in contraction; for example, ‘am’ contracts 36% of the time as compared with ‘that,’ which contracts in only 1% of its occurrences. The first question to ask, therefore, is whether there is a discernible pattern as regards which items favor or disfavor contraction. According to the principle of markedness assimilation, contraction is an unmarked behavior that should be found most often in unmarked contexts. The relevant contexts for the present study, along with the markedness values given in Andersen (1997), are summarized in Table 2 below:

Table 2. Some structural markedness values (from Andersen 1997)

	<i>Unmarked</i>	<i>Marked</i>
Nominal reference	lexical NP	pronoun
Tense	present	past
Number	singular	plural
Person	1st person	2nd person

In addition, the contrast ‘affirmative-negative’ is relevant in English contraction formation. Although Andersen (1997) gives no markedness values for this contrast, he discusses examples of cultural oppositions in which positive values are unmarked and negative values are marked; I infer from this similar markedness values for the grammatical opposition affirmative-negative.

The contraction distributions in Table 1 are inconsistent with the predictions of the markedness assimilation principle with respect to nominal reference. All NPs found in contractions in the Linguist-L data are pronouns; no contracted lexical NPs (as in, e.g., “the book’s now available”) were found in the data.<sup>5</sup> However, one might question the notion that pronouns are marked relative to full NPs in all contexts. In narrative discourse in English, as in many other languages, pronouns are the unmarked (most frequent, most continuous) form of reference; full NPs are marked in that they signal new or contrasting referents (Givón 1983). Pronouns also tend to receive less stress than lexical NPs in continuous spoken discourse, a fact likely related to the propensity of pronouns to undergo reduction through contraction.

A second inconsistency is evident in the fact that negation—in the form of the word ‘not’—strongly favors contraction (22%), moreso than the average of the affirmative items combined. A possible pragmatic explanation for this asymmetry has been advanced by Yaeger-Dror (1997), who proposes that negating is a face-threatening act that speakers seek to downplay through the use of reduced, contracted forms. However, negated forms in English are marked relative to affirmative forms. Thus it is not immediately obvious how Yaeger-Dror’s explanation fits, if at all, with the principle of markedness assimilation.

A closer examination of the columns in Table 1 reveals that second members of contractions contract at a higher rate than first members, although as this pattern is based on only two lexical items, one of which is ‘not’, it is difficult to interpret. Items that can be either first or second members of contractions (such as ‘have’) are distributed evenly down the ranking order. Regarding word class, auxiliaries are slightly more likely than pronouns to contract (7.2% vs. 6.7%) overall, followed by the adverbs ‘here’ and ‘there’ (4.2%) and the two wh- pronouns (1.8%). However, as these distinctions have not been posited to enter into binary markedness oppositions, the differences between them neither support nor refute the principle of markedness assimilation.

Partial support for the principle can be found in the categories of person and number. Leaving aside invariant and partially invariant forms that encode more than one person, first person favors contraction (14.7%) over second person (4.0%). However, third person is not accounted for in the binary system in Table 2, nor does it fit neatly in traditional hierarchies of pronoun markedness which rank third person as least marked (Greenberg 1966, cited in Smith, this workshop) or as most marked (i.e. in face-to-face discourse). In the Linguist-L data, pronouns and auxiliaries that encode third person contract on average at a rate intermediate between first and second person (4.6%<sup>6</sup>), although third person forms are distributed widely over the range. The evidence for number is less ambiguous: Exclusively singular forms contract at a higher rate (7.5%) than plural forms (4.5%), in support of the principle of markedness assimilation.

Tense shows no clear pattern. Although there are more present than past tense auxiliaries in the data, reflecting a general tendency for computer-mediated communication on the Internet to focus on the here-and-now (Collot & Belmore, 1996), present tense forms do not contract more often than past tense forms. On the contrary, if items that contract fewer than five times are excluded, the opposite tendency is found (7.9% present vs. 9.5% past).

Taken together, this structural evidence constitutes mixed support, at best, for the principle of markedness assimilation. Some predictions are partially supported by the overall distribution of contraction across individual lexical items, but other predictions are unsupported or contradicted.

More problematic, the markedness principle has little to say about structural patterns that *do* appear in the data. Pronouns are not a homogeneous category; there is a ranking of pronouns with respect to contraction: I > she > we > it > you > they > he > that, covering a range between 14% and 1%. This order is only roughly and rather unsatisfactorily accounted for by principles of person and number markedness (‘I’ and ‘we’ are both first person but do not rank together; third person forms are spread across the scale; ‘she’ and ‘he’ are structurally similar but contract at different rates, etc.). Auxiliaries span an even wider range, from 22% to .4%, following the order: do > will > can > have > be > need/should/must. It is not clear how markedness could be invoked to explain this order of preference for contraction. Moreover, there is considerable variation within each

auxiliary type (compare the rate of contraction of ‘am’ with that of ‘are’, for example) that cannot be explained in relation to tense. Nor is it evident what structural features the most frequently contracted forms—‘am’, ‘do’, and ‘not’—share. After all the morphological evidence has been considered, one is tempted to conclude that the best predictor of contraction rate is lexical: Some items simply contract more frequently than others, independent of their morphological category membership.

### 2.1.2 Distribution over time

Thus far we have considered structural patterns of contraction in the corpus as a whole, without regard for change over time. However, the markedness assimilation principle also makes diachronic predictions. In the words of Timberlake (this workshop), “change proceeds through a hierarchy of contexts whose ranking seems to be determined by something in the system of language: change develops first in those contexts that are natural or unmarked for the incipient change” (p. 6). In order to determine if individual words increase or decrease in their preference for contraction at different rates in the online discussion groups, I charted the rate of contraction for each item by year, considering separately the contraction formative categories ‘1st position’ and ‘2nd position’ in order to avoid interactions between commonly co-occurring formatives (e.g. ‘I’ and ‘am’; ‘do’ and ‘not’) that could bias the results. Logistic regression lines for the 1st position formatives are shown in Figure 2 and for the 2nd position formatives in Figure 3.<sup>7</sup>

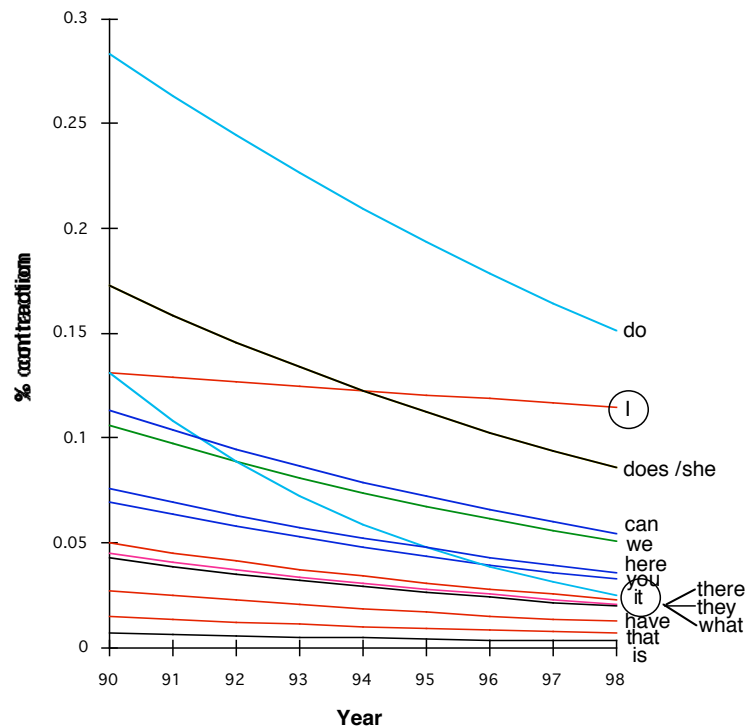


Figure 2. Logistic regression lines for first members of contractions over time



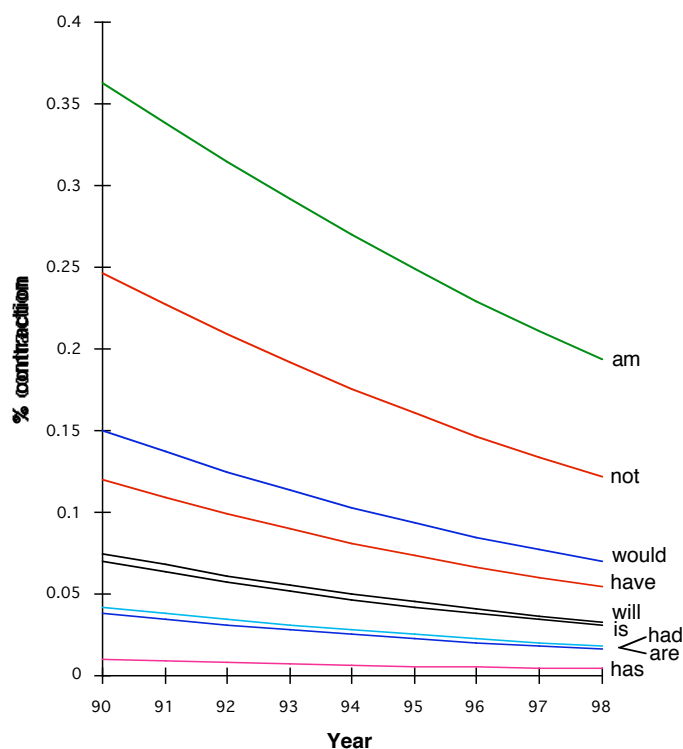


Figure 3. Logistic regression lines for second members of contractions over time

The principle of markedness assimilation predicts that a change will spread at different rates in different linguistic contexts; that is, that it will appear earlier in some contexts than in others. The same prediction can be applied to an innovation that involves a reversal or loss of a linguistic feature: The feature should be lost earlier in some contexts than in others. In addition, the first contexts in which an innovation applies should be harmonious in their markedness with the innovation itself. If we consider the loss of contraction to be a marked innovation (in that it goes against the natural trend towards structural reduction), then we would expect it to be lost first in marked contexts, as defined in the previous section.

Contrary to this prediction, however, Figures 2 and 3 show that for the majority of the formative items, the frequency of contraction decreases uniformly. This is evident from the preponderance of parallel, non-intersecting lines in the figures.<sup>8</sup> The only exceptions appear in Figure 2: the pronouns ‘I’ and ‘it’. ‘I’ shows a more level rate of contraction than the other forms, barely decreasing in its tendency to contract between 1990 and 1998, whereas ‘it’ loses contraction more rapidly than the other forms.

Two forms out of 31 is not overwhelming evidence in support of differential rates of loss of contraction, but it is worth considering whether these exceptions can be explained in terms of markedness. ‘I’ is morphologically unmarked by virtue of being first person and singular, hence the retention of contraction (an unmarked feature) in the

environment of ‘I’ is consistent with the markedness assimilation principle. Since ‘I’ is also the pronoun that participates most frequently in contraction in the corpus, it seems plausible that markedness considerations would favor the retention of contraction with ‘I’. The picture is less clear for ‘it’, however. ‘It’ is singular—an unmarked feature—and third person—more marked than either first or second person in face-to-face discourse, but less marked according to Greenberg’s (1966) pronoun hierarchy (see discussion in Smith, this workshop). There are many other forms in Table 1 that could be considered “more marked” than ‘it’—why, then, should ‘it’ lose contraction first?

Markedness alone, at least on a purely grammatical level, does not explain this pattern. Moreover, the overall diachronic result, according to which contraction declines at a uniform rate over time regardless of morphological context, runs counter to a basic prediction of the markedness assimilation principle. In order to understand why contraction use decreases overall on the Linguist List, and why it does so differently in contractions involving ‘I’ and ‘it’, it is necessary to look to the online social context in which the use of contractions is embedded.

## **2.2 Social factors**

Evidence for social influence on linguistic change comes from a variety of sources, both internal and external to the textual data. In this section, I first consider text-internal evidence, i.e. social factors that can be observed directly from the Linguist List messages. I then bring in evidence regarding the evolution of the group’s social dynamics and the larger Internet culture, based on ethnographic analysis of Linguist-L and the Internet carried out over the past nine years.

### **2.2.1 Message purpose**

According to Biber (1988), contraction in English is a feature of interactional, as opposed to informational, registers.<sup>9</sup> Although e-mail messages posted to an academic Internet discussion group might be considered a single register, not all messages within the genre fulfill the same communicative purpose. On the Linguist List, some messages initiate interaction by requesting information, some respond to other messages, and still others provide unsolicited information in the form of announcements of conferences, jobs, etc. Therefore, one might expect rate of contraction to differ according to message purpose; specifically, more interactive messages—queries and responses—should have a higher rate of contraction than informative messages such as announcements.

This expectation is born out by the results of a Varbrul analysis<sup>10</sup> of the entire corpus, as summarized in Table 3. Overall, messages that ask a question or respond to another message favor the use of contraction (.591) relative to announcements, which strongly disfavor contraction (.161).

Table 3. Varbrul results for contraction by message purpose

Factor Group	Factors	Varbrul weight
Message purpose	query/response	.591
	announcement	.161

p = .000

An example of an interactive response is given in (1) below, in which all possible contractible sequences are contracted (in boldface). Note also the presence in the message of interactive, colloquial expressions such as “But to tell you the truth”.

- 1) Date: Sun, 1 Dec 91 13:08:12 -0500  
 From: macaulay@j.cc.purdue.edu (Monica Macaulay)  
 Subject: NOT

Re Larry Horn’s query: to my knowledge that “Not” originated from “Wayne’s World,” a skit on Saturday Night Live. **It’s** supposed to be a public access program hosted by these two high school boys. **It’s** pretty hilarious, actually. **But to tell you the truth, I’ve** never been sure if they originated that use of “Not” or if it was part of teen slang that they then picked up and popularized.

[1991: 35]

In contrast, contraction is infrequent in announcements. Writers either use full, uncontracted expressions (example 2), or they avoid contexts in which contraction could occur altogether, e.g. through the use of passive constructions which reduce the incidence of personal pronouns (example 3).

- 2) Date: Thu, 3 Dec 1998 12:19:15 +0000  
 From: Chris Makemson <cmakemson@COMPUSERVE.COM>  
 Subject: ALPHABETS OF INDIGENOUS LANGUAGES OF EUROPE

Colleagues,

CEN Project Team 11, in conjunction with the CEN/ISSS Workshop Alpha, is developing a catalogue of the alphabets of the indigenous languages of Europe. (...)

**I would** like to start a 2-month review of this document. The document contains a large number of language alphabets and **I do not** expect people to review the whole set. Rather, I ask you to review the alphabets for languages for which **you would** consider yourself to be an expert. In this way I hope that a large proportion of the material will be reviewed and at the same time you **will not** find it to be an impossible task. (...)

The repertoires are based on facts, namely, dictionaries, grammars, and other materials. If **you are** not happy with the content of an alphabetic repertoire, it is necessary that documentation of the requested changes be sent to the editor. (...) To ensure clarity, the documentation should be sent in hardcopy photocopies to the editor at [editor's name and address].

[1998: 65]

- 3) Date: Sun, 4 Dec 94 02:38:35 CST  
From: susan@utafll.uta.edu (Susan Herring)  
Subject: Final call for abstracts: Computer-mediated discourse analysis

### FINAL CALL FOR ABSTRACTS

GURT Pre-session on  
"Computer-Mediated Discourse Analysis"  
March 8, 1995  
Georgetown University

This one-day symposium will focus on linguistic approaches to computer-mediated discourse (e-mail, computer conferencing, chat, MUDs/MOOs, etc). Abstracts are invited for 20-minute presentations on any application of linguistic methodology to computer-mediated data. Analyses that are pragmatic, sociolinguistic/interactional, textual, or focus on issues of methodology or genre are especially welcome.

Interested persons should submit a 350-500 word abstract briefly describing the proposed presentation by \*December 10, 1994\* to:

[organizer's name and address]

E-mailed submissions are preferred.

[1994:44]

In the announcement in example 3, the expression 'abstracts are invited', for instance, might conceivably have been expressed in a more informal style as 'we're inviting abstracts' or 'we'd like to invite abstracts'.

The fact that announcements disfavor contraction potentially explains the decrease in use of contraction on the Linguist List over time, if it turns out that there are more announcements in the later than in the earlier portion of the sample. This is indeed the case: In the first several years of the Linguist List, questions and responses outnumbered announcements, but by 1995, announcements began to be more frequent. However, statistical analysis shows that contraction decreases in questions/responses and announcements at an equal rate over time.<sup>11</sup> That is, the same degree of decrease is evident regardless of message purpose. Thus, while the increase in announcements is interesting

for what it suggests about a change in preferred activity type on Linguist, it does not in and of itself explain the decrease in contraction.

### 2.2.2 Gender

Another possible explanatory factor is participant gender. Do females and males use contraction differently, and if so, does this affect the pattern of decrease for the group as a whole?

Contradictory claims have been advanced about gender and language use. According to one popular view, women tend to be more interactive in their conversational style than men (e.g. Tannen 1990), while according to another view (which has received much support within variationist sociolinguistics; see e.g. Labov 1994, Trudgill 1974), women use more “standard” linguistic features than men. Thus depending on whether one considers contraction in writing to be primarily a marker of interactive, casual discourse, or a “less standard” feature compared to non-contraction, one could predict that women would contract either more or less often than men.

It turns out that women make significantly less use of contraction than do men in the Linguist corpus overall. The Varbrul results for this comparison are summarized in Table 4.

Table 4. Varbrul results for contraction by gender

Factor Group	Factors	Varbrul weight
Gender	male	.529
	female	.388

p = .000

This overall result, however, masks an interesting diachronic pattern, shown in Figure 4.

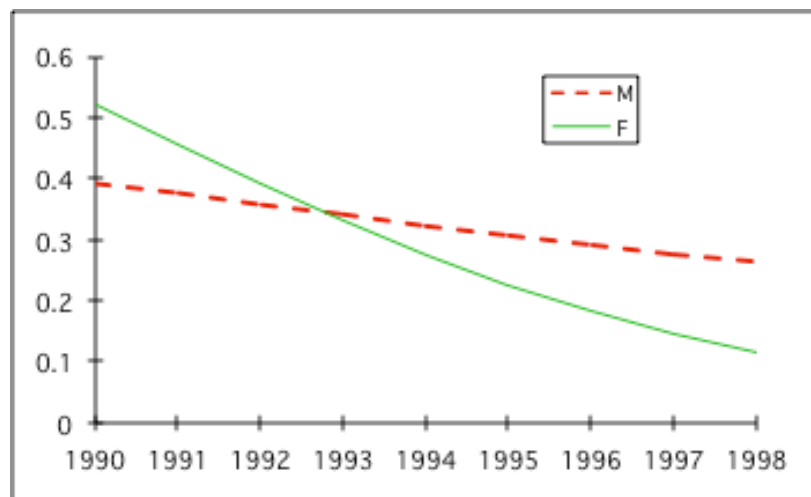


Figure 4. Rate of contraction on the Linguist List over time, by gender

As the logistic regression lines in Figure 4 show, in the early years of the group, women contract more often than men, but their frequency of contraction drops off sharply, such that by around 1993, they contract less frequently than men. This creates a cross-over pattern, one that shows women leading the decrease in contraction in a sharp decline, while for men the decline is more gradual. From the perspective of variationist sociolinguistics, this suggests that the change in question is a “standardizing” one associated with overt prestige, i.e. with the approval of dominant social institutions, since women tend to lead in such changes (Labov 1994; Trudgill 1974).

This result will be argued to play a role in the “counter-change” in use of contractions on the Linguist List. However, it does not suffice to explain the overall decrease in contraction, since the ratio of female to male participants in the group remains roughly stable over time, and men also show a (more gradual) decrease.<sup>12</sup>

### 2.2.3 Group-internal dynamics

Thus far two social factors have been identified that favor the loss of contraction on Linguist-L: information-focused messages and female gender. (We return below to the question of whether these observations can be reconciled with a theory of markedness.) But what causes the increased propensity towards information-focused messages, and what causes women to reverse their contraction preference? Are these two behaviors related? When the disparate facts about contraction that have been presented thus far are considered from the perspective of the social context of the Linguist List, a more coherent picture emerges. According to this picture, the key to understanding the decrease in contraction on the Linguist List resides in viewing contraction use as a marker of informality.

When the Linguist List was started late in 1990, the list owners and moderators, a husband and wife team of linguists, initially adopted a relatively informal tone in their communication to the group. This can be seen in one of their first messages describing the group’s purpose.

#### 4) [Moderators’ Comment:

LINGUIST was begun in order to serve as a forum for the dissemination of information which is relevant to the academic discipline of linguistics. (...) **It is** certainly an electronic bulletin board, but **it is** also a relatively open mailing list (...). The postings may be queries, notices of forthcoming conferences or upcoming job vacancies, comment on issues of theoretical interest, or requests for copies of papers. The postings are moderated, but only to ensure that the content remains linguistic, and that the laws of libel remain unfringed.

Can--should?--LINGUIST do more than this? We **don’t** know. What do you think?]

[Dec. 1990]

In this message, not only do the moderators define the mailing list as “relatively open”, they reinforce this idea by adopting a casual, interactive style at the end, by asking direct questions of the group members, and by using a contraction (“We don’t know”).

The group’s subscribers followed the moderators’ lead. The sample message in (1) above is typical of the relatively informal style of discourse that came to characterize group exchanges by 1991. During that time, contraction use jumped to a level that exceeded that of MsgGroup at its peak. However, informality had other, undesirable consequences— notably the tendency for some participants to become disinhibited and make negative evaluations of other participants’ views. The field of linguistics, then as now, was ideologically divided between so-called “functional” and “generative” approaches to the study of grammar. Only a few months after the group was started, a discussion took place between advocates of these opposing views that evolved into an extended, acrimonious debate—what, in Internet parlance, would later come to be known as a “flame war” (Herring 1992). Over the next few years, other public displays of contentiousness followed.

At the same time, membership in the Linguist List was growing rapidly, and high message volume was becoming a problem for some subscribers. Responses to queries were targeted as a source of the problem—a single query could generate many responses, some of which were redundant. Responses were also a prime site for conflict, i.e. when someone took issue with something someone else said. The moderators therefore instigated a policy that responses to queries should be e-mailed privately to the person being responded to, ostensibly to control message volume. It may or may not have been the moderators’ conscious intention to reduce the likelihood of conflict in the public forum thereby. What is incontrovertible is that their messages to the group began to take on a more formal tone, imposing restrictions rather than inviting openness in Linguist-L discourse. As part of this formal tone, the moderators’ use of contraction decreased.

5)

#### REMINDER

[**We’d** like to remind readers that the responses to queries are usually best posted to the individual asking the question. That individual is then strongly encouraged to post a summary to the list. This policy was instituted to help control the huge volume of mail on LINGUIST; so **we would** appreciate your cooperating with it whenever it seems appropriate.]

[1994]

6)

Please **do not** use abbreviations or acronyms for your conference unless you explain them in your text. Many people outside your area of specialization **will not** recognize them. Also, if **you are** posting a second call for the same event, please keep the message short. Thank you for your cooperation.

[1997]

I propose that the movement towards greater social control necessitated by the growth and propensity for conflict on the Linguist List was the impetus for the decrease in contraction use, via the association of ‘informality’ with ‘disorder’, and thence, by analogy, of ‘order’ with ‘formality’. This association was modeled for the group members by the moderators, in their increasingly formal statements of policy over time. Moderator policies are directly responsible for the decrease in the ratio of queries/responses to announcements; with responses taken off-line, interactive discourse decreased and announcements increased proportionately. Indirectly, the change in moderator policies encouraged a decrease in informal language use, including in use of contractions. This perspective suggests a “change from above” (Labov 1994), that is, a change instigated by a source holding greater social (in this case, institutional) power.

Women’s sharper decrease in contraction use can be understood in this context. Questionnaire studies (e.g. Herring 1992; Herring, Johnson & DiBenedetto 1992) found that female participants in electronic discussion groups at the time tended to respond with greater aversion than men to online conflict. It is conceivable that women on Linguist-L perceived “excessive” informality as leading to contentiousness and thus retreated to a safer, more formal style of posting—including preferring information-focused messages to interactive exchanges. This would account for women’s earlier, and sharper, decrease in contraction use. In so doing, they also would have been complying with the change in posting policy instigated by the moderators, thereby becoming leaders in an institutionally-sanctioned, standardizing change. Middle-class women in modern western societies tend to be judged more harshly than middle-class men for deviance from standard norms, and to use more standard language accordingly.<sup>13</sup> In the male-dominated context of the Linguist List (at the time of this study, 64% of male subscribers contributed on average more than 80% of the words), female subscribers may have felt it necessary to adopt a somewhat formal style in order to be taken seriously as academic professionals.

In contrast, men do not follow the trend toward formality as strongly, and even show signs of resisting it, at least as regards one contraction type. Recall that the pronoun ‘I’ was the only contraction formative item that did not significantly decrease its frequency of contraction over time; this was earlier considered evidence in support of the markedness assimilation principle. However, closer inspection of the contraction ‘I’m’ reveals that women decrease in their frequency of use of this contraction, exactly as they do for the other contractions—that is, there is no difference between the pattern for ‘I’m’ and the patterns for the other contractions, as far as female participants are concerned. Men, however, show an *increase* in frequency of contraction of ‘I’m’ in the later years of the sample, not only in interactive messages (unmarked contexts), but in announcements (marked contexts) as well. Because there are more male than female messages in the sample, the male pattern predominates and causes the slope of change for ‘I’m’ to flatten out. In fact, however, the male pattern is at odds with the female pattern and with the overall diachronic trend. This observation argues against an interpretation solely in terms



of markedness, since no markedness principle, to my knowledge, predicts that men and women will obey markedness constraints differently.

#### 2.2.4 The Internet cultural context

Another possible social explanation for the use of contractions on the Internet makes reference to the larger Internet cultural context. It proposes that the decrease in contraction on the Linguist List, and the general increase in formality of style that accompanied it, reflects a more-or-less conscious resistance on the part of an academic elite to the “informalization” of language on the Internet brought about by the popularization of the Internet, and especially, the opening up of Internet access to ordinary people in the 1990s through the rise of Internet Service Providers (ISPs). As recently as the late 1980s, Internet access in the United States was still largely restricted to the scientific and intellectual elite in universities, corporations, and government institutions. ISPs ushered in a flood of new users, most connecting from their homes, many young (including adolescents), and most accessing the Internet primarily for recreational purposes—interacting in chat rooms, surfing the World Wide Web, etc. These users have engaged in varied and informal linguistic practices, including non-standard orthography and typography (smiley faces; abbreviations such as IMHO, brb, lol; simplification of ‘are you’ to ‘r u’, etc.), which soon began attracting comment both online and in the popular media. While many reactions were bemused, others—including by scholars—were concerned about what they saw as the likely negative effects of Internet writing on the correctness and complexity of the English language (e.g. Baron 1984).

A precedent for resistance of this sort can be found in the history of the English language. Following the invention of the printing press and the subsequent increased availability of printed material, the 17th century saw the “popular” social classes in England gain access to literacy for the first time. Written English subsequently began to incorporate features of popular speech, a trend that continues to the present time (Biber & Finegan 1989). This triggered a backlash in the 18th century, when a number of writers of the educated class began using a hyper-literary style, avoiding “speech-like” features such as contraction. Thus although the frequency of contraction in written English has increased overall starting from the 17th century, this trend shows a temporary reversal during the 18th century. Similarly, the counter-trend in contraction use on the Linguist List may represent a (temporary) backlash against the latest wave of written language popularization facilitated by the Internet.

### **3. Discussion: The Principle of Markedness Assimilation Revisited**

In the previous sections, I argued that morphological markedness considerations reveal no coherent pattern associated with the use and spread of non-contraction on the Linguist List, whereas social considerations reveal patterns that can be brought together in

a coherent account, given sufficient knowledge of the Linguist List and Internet contexts. However, Andersen's (1997) markedness assimilation principle does not restrict itself to grammatical criteria in identifying patterns of actualization, but rather includes contexts of genre and style as well—contexts that could be considered social. Could the observations in section 2.2 be made to fit an account in terms of markedness?

Such an account would presumably be based on the classification of increase in contraction as an unmarked (natural, simplifying) change, and decrease in contraction as a marked (unnatural, complexifying) change. It would predict that increase in contraction has a natural or internal source—ease of production, in the sense of reduction of keystrokes—while decrease in contraction has an external, presumably social source. The spread of the respective changes should then proceed according to the markedness assimilation principle, that is, contraction should spread first in unmarked contexts and be lost first in marked contexts.

Regarding spread, contraction on the Linguist List is favored by interactive messages, and disfavored by purely informational messages. Since interactive messages tend to be more casual (unmarked) in style, and informational messages more formal (marked), this is consistent with the markedness assimilation principle. However, loss of contraction is not actualized first in informational messages, contrary to the diachronic predictions of the principle; rather contraction is lost at an equal rate in interactional and informational messages. Thus message purpose provides mixed evidence for a markedness analysis.

Contraction is further favored by the male gender of the writer and disfavored by female gender. Masculine gender is sometimes argued to be unmarked relative to feminine gender, although this classification is better suited to reference to masculine and feminine linguistic objects than to the behavior of male and female persons, which is highly complex and can be understood from various perspectives (see discussion in section 2.2.2). Granted, however, that we assign men the value of (unmarked) and women the value of (marked); the Linguist-L evidence then appears consistent with the markedness assimilation principle, including diachronically, in that women “lose” contraction sooner and at a faster rate than men do. However, there remains the problem of men's increase in contraction of ‘I'm’, which is marked relative to the overall decrease in contraction on Linguist-L. More problematic yet, women start out contracting much more than men, which—if female gender is marked, and contraction is unmarked—violates the principle of markedness assimilation. In all, markedness seems a less than satisfactory means by which to characterize the gender patterns found in the data.

What, then, of the source of the change? Our analysis has proposed an external, rather than a “natural”, impetus for the decrease in contraction on Linguist-L. The change appears to have been triggered by social problems that arose early in the life of the list—conflict, excessive message volume—and official moderator policy in response to them. At a broader level, this change may have been fueled by an urge on the part of the academic

elite on Linguist-L to distance themselves from popular uses of the Internet and the linguistic informality that ran rampant there. A marked change introduced from outside the language system (a marked source) is consistent with the markedness assimilation principle. Moreover, as noted earlier, it is tempting to posit that the source for the increase in contraction in the earlier discussion list, MsgGroup, was a natural impetus towards reduction and simplification, abetted by the desire of computer users to avoid unnecessary keystrokes, again in support of the predictions of the markedness assimilation principle.

However, there are two problems with this argument. First, if contraction use were favored by computer-mediated environments, we would expect to find a high incidence of contraction in CMC overall. In fact, contraction is relatively infrequent in CMC, even in real-time chat, which in other respects resembles spoken conversational interaction (Ko 1996). The reason for this, I propose, is that typed contraction saves relatively few keystrokes (consider 'I'm' vs. 'I am', 'don't' vs. 'do not', 'you are' vs. 'you're', etc. In each case, only a touch of the space bar is saved by contracting). Thus the functional motivation for contraction is not as strong in typing as it is in speech.

Second, there were other, more compelling, *social* reasons for participants in MsgGroup to contract. Principal among these is that the cultural definition of the Internet changed dramatically between the inception of MsgGroup in 1975, when a handful of computer scientists were developing the Arpanet for the U.S. government, and 1986, by which time the Arpanet had become the Internet and was available for general use, e.g. on many college campuses. Whereas in the early 1970s a computer science professional reported feeling guilty for once "using government property" (i.e. sending an e-mail message to a colleague) to recover an electric shaver he had left at a conference (Hafner & Lyon 1996), by 1986 students were using the Internet to send e-mail for a variety of academic and social purposes. Correspondingly, the overall level of formality of e-mail messages shifted during this time; the Internet as a whole became less formal. Similar to the later popularization of the Internet discussed above, this broad cultural development would presumably have had an effect on anyone participating regularly on the Internet at the time, as was the case for many MsgGroup members. Thus, the most probable explanation for increasing informality in MsgGroup is social, rather than functional pressures imposed by the computer medium itself (Herring 1998). In other words, social factors are needed to explain the changes in contraction use in both groups, not just in the group that manifests the "marked" directionality of change.

The markedness assimilation principle has nothing to say about these larger social and cultural influences. This, in combination with the inability of the principle to account persuasively for the actualization patterns observed in the data, leads this researcher to conclude that the principle of markedness assimilation is not supported by change in contraction use on the Internet.

#### 4. Conclusion

This study set out to assess the explanatory power of an established model of language variation and change—the markedness assimilation principle—as applied to English contraction data from two Internet discussion groups over a 23-year period. The results suggest that change in contraction use need not proceed uniformly or linearly, but may be characterized by countervailing dynamic trends that emerge only when the data permit continuous or near-continuous observation, and that the genesis and spread of such change may only be fully explained with reference to the social context.

At the same time, these results call into question the nature of Internet language data and their suitability for diachronic analysis. Internet data differ from typical data used to adduce evidence for language change in at least two respects that could potentially render a comparison—and hence an evaluation of the markedness assimilation principle—problematic. First, the Internet has not been in existence very long; only a 23-year period was considered in this study. Twenty-three years may not be a sufficiently long time for system-level changes to take place, and markedness assimilation may only apply to changes affecting linguistic systems, rather than including trends that might (or might not) prove to be precursors of system-level change.

Second, Internet language is written, and as such, could be considered a derivative, rather than a primary, modality of language (although many linguists consider it a hybrid between traditional speaking and writing, due to its interactive nature and “oral” features; Cho *et al.* 1991, Ko 1996, Maynard 1994). Moreover, it is a subtype of writing (typing) that occurs only in specific technological environments. Thus it is possible that diachronic patterns in Internet language may not extend to Modern English as a whole and therefore may not qualify as instances of language-wide change.

In response to these potential objections, the following arguments can be advanced. It has been claimed that the pace of change—including as regards language—has accelerated with the implementation of the Internet (e.g. Baron 1984). While 23 years may seem a short time, in this study it manifested two different, statistically significant trends (a trend and a counter-trend). Clearly *some* changes took place in contraction use on the Internet during that time period; these deserve explanation. Moreover, with the spread in popularity of e-mail, chat rooms, instant messaging, newsgroups, and the like, Internet writing is increasingly becoming a core modality of communication for many English speakers, young people as well as scholars. Relatedly, evidence is starting to mount that lexical innovations (such as ‘ping’ and the term ‘on-line’ itself) and word formatives (such as ‘e-’ and ‘cyber-’) are jumping from the Internet into off-line speech and writing; this suggests that it is only a matter of time before a wider range of linguistic structures and usages that originated online impact off-line language.

For those who accept these arguments, the Internet can constitute a valuable source of data for the study of language change. Internet data are much richer than historical texts,

offering (in some cases) complete, continuous records within a particular domain, for which the surrounding social context is still available in the memory of living persons. This abundance may require different models—especially, more socially-sensitive ones—than those that can be devised on the basis of the partial, intermittent, decontextualized textual evidence that tends to remain from older languages. Thus while trends in contraction on the Internet may not be the kind of data for which the markedness assimilation principle accounts best, the fault may not lie with the data. It may also be that markedness assimilation is not the best theory to account for the dynamics of near-continuous, emergent change of the sort that takes place on the Internet. If our goal as historical linguists is to construct theories that account for all types of language change, this challenge cannot be ignored.

## Notes

- <sup>1</sup> The last contraction type is rare in the corpus; consequently, most of the findings in this paper primarily describe the first two types.
- <sup>2</sup> Sampling was done as follows. For MsgGroup, six 50-message samples were extracted from the archive at roughly two-year intervals (see Herring 1998 for further details and discussion of this corpus). For the Linguist List, all messages from the month of December were extracted for each year (or an equivalent subset of the December messages for years with very heavy posting activity), for a total of eight samples.
- <sup>3</sup> The two groups also differ in size: MsgGroup boasted just over 100 members at its peak—a large group for its time—as compared to the Linguist List, which started with 160 and grew to over 8,000 subscribers.
- <sup>4</sup> This generalization applies to the structural types of contraction considered in this paper, i.e. the modern types. Older forms of contraction no longer productive in Modern English are attested as far back as Beowulf; see Fulker (1990).
- <sup>5</sup> One instance of a contracted proper noun—*John's*—occurred in an example sentence being discussed as a linguistic datum.
- <sup>6</sup> This calculation excludes items in Table 1 of which fewer than five are contracted.
- <sup>7</sup> Formatives that could occur in either position were classified according to the actual positions in which they occurred in the data, as either 1st or 2nd position.
- <sup>8</sup> The slopes of all the parallel regression lines are statistically similar. Only the slopes of the intersecting lines—those for 'I' and 'it'—differ significantly from the others ( $p < .05$ ).
- <sup>9</sup> Contraction has a strongly positive loading on Biber's Dimension 1 (interactive vs. informational).
- <sup>10</sup> For a description of this statistical method of variable rule analysis, see Sankoff (1988).
- <sup>11</sup> Logistic regression analysis;  $p < .01$ .
- <sup>12</sup> The percentage of female participants remains stable at around 20%.
- <sup>13</sup> For a critique of this asymmetry, see Spender (1980).

## References

- Andersen, H. 1973. "Abductive and deductive change." *Language* 49, 765-793.
- Andersen, H. 1989. "Understanding linguistic innovations." In *Language Change: Contributions to the Study of Its Causes*, ed. by L. E. Breivik and E. H. Jahr, pp. 5-27. Mouton de Gruyter.
- Andersen, H. 1997. "Linguistic change and the theory of markedness." Paper presented at the XIIIth International Conference on Historical Linguistics, Düsseldorf, Germany.
- Baron, N. 1984. "Computer mediated communication as a force in language change." *Visible Language* XVIII 2, 118-141.
- Biber, D. 1988. *Variation Across Speech and Writing*. Cambridge University Press.
- Biber, D. & E. Finegan. 1989. "Drift and the evolution of English style: A history of three genres." *Language* 65, 487-517.
- Cho, T. Forthcoming. "Linguistic features of electronic mail." In *Computer-Mediated Conversation*, ed. by S. C. Herring.
- Collot, M. & N. Belmore. 1996. "Electronic Language: A new variety of English." In *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*, ed. by S. C. Herring, pp. 1-28. Amsterdam: John Benjamins.
- Ferrara, K., H. Brunner & G. Whittemore. 1991. "Interactive written discourse as an emergent register." *Written Communication* 8(1), 8-34.
- Fulk, R. D. 1990. "Contraction as a criterion for dating Old English verse." *Journal of English and Germanic Philology* 89(1), 1-16.
- Givón, T. (ed.) 1983. *Topic Continuity is Discourse: A Quantitative Cross-Language Study*. Amsterdam: John Benjamins.
- Hafner, K. & M. Lyon 1996. *Where Wizards Stay Up Late: The Origins of the Internet*. New York: Simon & Schuster.
- Hale, C. 1996. *Wired Style: Principles of English Usage in the Digital Age*. San Francisco: HardWired.
- Herring, S. 1992. "Gender and participation in computer-mediated linguistic discourse." Washington, D.C.: ERIC Clearinghouse on Languages and Linguistics. Document no. ED345552.
- Herring, S. 1998. "Le style du courrier électronique: Variabilité et changement." *Terminogramme* 84-85, 9-16.
- Herring, S., D. Johnson, & T. DiBenedetto. 1995. "'This discussion is going too far!' Male resistance to female participation on the Internet." In M. Bucholtz & K. Hall (eds.), *Gender Articulated: Language and the Socially Constructed Self*. New York: Routledge.

- Ko, K-K. 1996. "Structural characteristics of computer-mediated language: A comparative analysis of InterChange discourse." *Electronic Journal of Communication* 6(3), special issue on Computer-Mediated Discourse Analysis, ed. by S. C. Herring.
- Labov, W. 1994. *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.
- Mair, C. & M. Hundt. 1997. "The corpus-based approach to language change in progress." Paper presented at the XIIIth International Conference on Historical Linguistics, Düsseldorf, Germany.
- Maynor, N. 1994. "The language of electronic mail: Written speech?" In M. Montgomery & G. Little (eds.), *Centennial Usage Studies*. Publications of the American Dialect Society Series. University of Alabama Press.
- Murray, D. 1990. "CmC." *English Today* 23, 42-46.
- 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, Vol. 2*, pp. 984-997. Berlin: Walter de Gruyter.
- Spender, D. 1980. *Man Made Language*. London: Routledge & Kegan Paul.
- Tannen, D. 1990. *You Just Don't Understand*. New York: Ballantine.
- Timberlake, A. 1977. "Reanalysis and actualization in syntactic change." In Charles N. Li (ed.), *Mechanisms of Syntactic Change*, pp. 141-177. Austin & London: University of Texas Press.
- Trudgill, P. 1974. *Sociolinguistics*. Harmondsworth, UK: Penguin Books.
- Weinreich, U., W. Labov & M. Herzog. 1968. "Empirical foundations for a theory of language change." In *Directions for Historical Linguistics: A Symposium*, ed. by W. P. Lehman & Y. Malkiel, pp. 95-188. Austin: University of Texas Press.
- Yaeger-Dror, M. 1997. "Contraction of negatives as evidence of variance in register-specific interactive rules." *Language Variation and Change* 9(1), 1-36.