



Between the lines: Newsgroups as an information source for social historians

Sylva Simsova

Mrs Sylva Simsova

DATA HELP, 18 Muswell Avenue, London N10 2EG
tel/fax 0181 883 6351, simsova@simsova.demon.co.uk

Contents

- 1 Introduction
- 2 Use of written material as a source of data
- 3 Mass-Observation methods
- 4 Oral history methods
- 5 Analysis of texts by the content analysis method
- 6 Analysis of texts by the qualitative research method
- 7 Methods chosen for this research
- 8 The sample
- 9 The classification of articles
- 10 UK material
- 11 Contributors
- 12 Contributions
- 13 Content analysis
- 14 Conclusion

Summary

The aim of this pilot study was to show the potential usefulness of Usenet discussions to future social historians and to explore possible methods for analysing the data. A survey of 42 Usenet newsgroups was made in November 1996 to explore the patterns of information formed by the contributors. The 5541 articles written by 2130 persons were analysed to identify UK participants and topics related to the UK; to measure the frequency of contributions and relate it to their authors' interests; to draw a comparison between moderated and non-moderated groups; and to indicate the most commonly discussed topics, using keywords taken from headers. A content analysis was used to check adherence to a point of netiquette and to find out to what extent two selected themes (the forthcoming election and violence in society) occupied the minds of contributors.



1 Introduction

Contributions to Usenet discussion groups will, if we succeed to reliably archive them for posterity, represent a primary source of information to future social historians in the same way as written correspondence and diaries.

This article presents a case study of what selected Usenet groups reveal about the communication behaviour of their participants and about social life in the United Kingdom in November 1996.

2 Use of written material as a source of data

Personal diaries and correspondence, including contributions to the Usenet discussion groups, come under the sociological category of "uncontrolled writings" (Vernon 1939) as distinct from, for instance, replies to questionnaires. Their free text makes it difficult to quantify their contents.

Further, the text does not aim to be objective. As one of the founders of Mass-Observation put it, "they tell us not what society is like, but what it looks like to them". (Madge 1937-38)

Chance plays a part in the availability of material to the researcher, for instance some letters get lost and only some find their way into an archive.

The authors of a collection of documents do not constitute a sample, strictly defined according to a scientific method, they are either self-selected (contributions to the Usenet) or selected by accidents of history (correspondence in an archive).

3 Mass-Observation methods

Mass-Observation was started in response to a letter in the *New Statesman* in 1936 which suggested that "there was considerable scope for an anthropological study of our own civilisation". (Richards 1987)

Its interdisciplinary approach was well suited to the time when "the boundaries of the key disciplines of sociology, anthropology and economic were still sufficiently fluid." (Stanley 1990)

The main method used by Mass-Observation was a "juxtaposition of many observed details of everyday life to build up a composite picture". (Mercer 1989)

The amateur observers used by Mass-Observation were members of a panel of volunteers recruited from respondents to advertisements and from informal contact. It was not claimed that it was quantitatively representative of the British population, but it did cover a wide range of age, occupation and opinion. (Mercer 1989)

Its methods were criticised by mainstream sociology for its quantitative shortcomings and sampling. (Mercer 1989)

4 Oral history methods

While Mass-Observation was interested in people's attitudes to current events, oral history was started a couple of decades later as a tool for studying the past. It arose from an interest in local history. Testimonies of witnesses were recorded to document past events which would otherwise have gone unrecorded.

At the collecting stage oral history is not written down. It is recorded on tape and transcribed afterwards.

Oral history methods aim to record facts rather than their meaning for those experiencing them. It studies the whole text rather than individual words.

5 Analysis of texts by the content analysis method

Content analysis was started at about the same time to identify the often hidden meaning of texts. It originated in media studies. (Berelson 1952, Stone 1966)
The samples for analysis consist of written texts. The actual content is considered less important than the words used to express it.

Researchers establish a set of categories and decide on words which belong to the categories. The words are tagged and their frequency counted. (Silverman 1993)

This quantitative method can extract subtle meanings which would otherwise escape researchers studying merely the texts' subject content.

6 Analysis of texts by the qualitative research method

Where historians study the content of recorded messages, researchers using qualitative methods "seek to identify the meaning of the message as a whole.."(Keele 1995)

One of the methods resembles that of content analysis. It consists of breaking up the text into sections, perhaps even words which represent the participants' categories.

The method differs from that of content analysis in using categories derived from the texts rather than those predefined by the researcher.(Coffey 1996, Silverman 1993)

7 Methods chosen for this research

The methods chosen for this research resemble those of the Mass-Observation and of content analysis.

The framework of the sample was defined by the researcher (newsgroups, number of days), but the contributors were self-selected (those who happened to write to the newsgroup on the particular day).

The Mass-Observation gathering of data was restarted in the 1980s. A panel of new volunteer contributors was recruited through the national and local press. (Sheridan 1991)

It is possible that the method of downloading Usenet contributions suggested in this exploratory study could be used to supplement the data collected by letters, questionnaires and diaries in the traditional way.

8 The sample

About 200 contributions or a week's output from each of the 42 selected groups were downloaded. A full list can be found in Table 4.

The groups were selected from over 2000 available on the Demon server.

From among the "uk.*" groups 16 were considered for inclusion. Of these 7 were rejected before downloading (uk.answers, uk.events, uk.finance, uk.media, uk.org.community, uk.community.social-housing) and 3 after (uk.people.deaf, uk.people.disability, uk.d-i-y). The reasons for rejection were low numbers of contributions, excessive specialisation or duplication of other material.

In the "soc.culture.*" groups 7 were considered for inclusion and 1 rejected after downloading (soc.culture.multicultural). The reason for rejection was that there happened to be no contributions at the time.

In the "soc.history.*" groups 7 were considered for inclusion and 3 rejected before downloading (soc.history.war.misc, soc.history.medieval, soc.history.science). The reason for rejection was that they were considered too specialised.

In the "soc.*" groups 7 were considered and 1 excluded after downloading (so.rights.human). The reason was that it did not deal much with the UK. One group was retained in spite of its strong US bias (soc.veterans).

In the "alt.*" groups 28 were considered for inclusion, 4 rejected before downloading (alt.chrisnet.second-coming.real-soon-now, alt.war, alt.society.deaf, alt.support.disabled.caregivers), and five after downloading (alt.macdonalds, alt.politics.correct, alt.politics.greens, alt.feminism, alt.politics.nationalism). Some of the rejected groups had too few or too many entries or did not refer to the UK situation.

In the "talks.*" groups only 1 was considered for inclusion. The "talks.*" groups often have rather ephemeral contributions. The "talk.euthanasia" was selected because no other group dealt with the subject.

Finally the sample looked as follows:

The first part of the sample consisted of 6 groups from the "uk.*" range, 6 groups from the "soc.culture.*" range, 4 groups from the "soc.history.*" range, and 6 groups from the "soc.*" range dealing with the issues of the day. All of these groups were selected to represent mainstream thinking because they were moderated.

The second part of the sample consisted of 20 groups from the "alt.*" range. Not being moderated they represented alternative thinking.

The number of articles in the groups varied between 3 and 289, the average being 133. The total number was over five and half thousand.

The sizes of the files containing the downloaded newsgroups were between 11k and 1360k, with an average of 389k.

The average length of articles was between 1k and 6k. In one group the average was 44k due to the inclusion of lengthy reports.

9 The classification of articles

The downloaded articles can be classified in two ways: by the level of editorial control and by subject matter.

The "uk.*" and "soc.*" newsgroups were subject to editorial control, the "alt.*" newsgroups were not edited. The unedited ones contained errors and ephemeral material which would not be so useful to historians. On the other hand they revealed spontaneous uninhibited thinking on the part of the contributors.

The subject matter fell roughly into six categories: 8 dealing with culture (2 in "alt.*", 6 in "soc.*"), 9 with environment (8 in "alt.*", 1 in "uk.*"), 5 with feminism (2 in

"alt.*", 3 in "soc.*"), 7 with history (3 in "alt.*", 4 in "soc.*"), 3 with old people (2 in "soc.*", 1 in "talk.*"), 10 with politics (4 in "alt.*", 1 in "soc.*", 5 in "uk.*").

10 UK material

Identification of material dealing with the UK was not easy as the English language newsgroups were used by people from all English speaking countries. Three indicators of UK material were identified:

The first indicator was the newsgroups with "uk.*" in the title. By implication material accepted by them should deal with the UK.

The second indicator was the email addresses of correspondents. Increasingly, however, internet users in the UK are using international hosts and cannot therefore be identified from their addresses.

About 21% of email addresses had a ".uk" ending. The majority (76%) were in the "uk.*" newsgroups, a small number of them (16%) in the "soc.*" newsgroups and hardly any (6%) in the "alt.*" newsgroups.

Political newsgroups had the largest number of ".uk" email addresses (44%), followed by the newsgroups dealing with culture (24%).

The third indicator was the frequency of names of parts of the United Kingdom used as words in the contributions: UK(187), Britain (200), England(234), Scotland(188), Wales(133), Ulster(13). About half of the instances were found in the "soc.*" newsgroups and only 37% in the "uk.*" newsgroups, the reason presumably being that it was taken for granted that discussion of any subject in the "uk.*" newsgroups referred to the UK situation. The names of parts of the United Kingdom were used most frequently in the cultural newsgroups (42%).

11 Contributors

The total number of contributors was 2130, of whom 1225 contributed one article, 787 between 2 and 7 articles (i.e. up to 1 a day), 80 between 8 and 14 articles (i.e. up to 2 a day) and 38 between 15 and 55 articles (i.e. up to 3 or more a day).

The single article contributions were mostly in the "soc.*" newsgroup and the "alt.*" group. The most frequent contributors sent their articles to several different groups. The most frequent combination of subjects was culture and politics.

One could roughly say that the "uk.*" groups were written to at a more or less constant frequency, while the frequency of "alt.*" and "soc.*" decreased with the number of postings and the mixed category increased with the number of postings. There were 21 contributors who wrote on average between 3 and 8 times a day. These were investigated in depth.

They posted most frequently to UPCON (19% of the 714 articles) and APEQA (14%). They also posted to UPMIS, UPELE and ASEAR.

From the address and the organisation name accompanying their articles they could be classified into three types : academic, computer and communication staff, private. The highest number of contributors was in the private category (43%) and among the computer and communication staff (38%). Only 19% were from the academic world.

According to their address 48% were from the UK.

The enthusiasm of some of these correspondents could be indicative of a permanent feature of their life, though some of them wrote only occasionally during temporary leisure. The email addresses of the top 11 correspondents were checked in dejanws in September 1997 to see if they were still writing as ardently as a year before. Bearing in mind that the comparison could be influenced by the dejanws not indexing all the newsgroups, the result for one week was as follows: 3 contributors did not write anything, 3 wrote between 2 and 4 times, 2 between 10 and 13. Only three maintained their high rate of postings between 23 and 28 articles a week.

12 Contributions

The sample contained 5541 articles. One quarter dealt with politics, another quarter with culture, yet another quarter with social problems, such as women or the elderly, one eighth with history and one eighth with the environment. The articles had 1836 headers of which 1502 were used in one group, 126 in two groups, 20 in three groups, 4 in four groups and 1 in five groups. Most headers were used in newsgroups of the same category ("u.*" or "a.*"). Some correspondents posted across the boundaries between categories: 14 were in the "u.*" and "s.*" newsgroups and 14 in the "s.*" and "a.*" newsgroups. Only 1 was in the "u.*" and "a.*" groups. This indicates a commonality of interests between the participants of the various groups.

Tables 1 and 2 show the most frequently used headers for all the groups combined and for the groups with a specific UK interest. The score was calculated as a percentage of all headers in the group or combination of groups.

Table 3 gives the terms used in the headers. They are arranged in rough subject groups and provide an indication of the range of contributors' interests.

A brief listing of the top two or three under each subject group, extracted from the table, suggests that the minds of contributors were filled with forebodings and unhappy memories:

- culture: race, slavery
- environment: barefoot, vegetarians
- feminism: abortion, marriage
- history: holocaust, Halloween, Irish holocaust
- medicine: suicide, aids
- old: retirement, euthanasia
- politics: election, IRA, anarchy
- psychology: violence, responsibility
- recreation: tartan, food, knotwork
- religion: Christian, Fall
- science: climate prediction, toxicity
- society: victim disarmament, gun, arrest

13 Content analysis

Using the content analysis method, the text of the articles was subjected to a word count of selected words to ascertain three points:

First, to what extent did the contributors abide by the rules of netiquette. Two swearwords were selected ("s" and "f") and their occurrence in the sample counted:

About 9% of articles in the "s.*" newsgroup and the same percentage in the "a.*" newsgroups contained one of the two words. Analysed by subject matter the feminist newsgroups had the highest incidence of the words (21% of articles), followed by the old people newsgroups (13%) and politics (10%). The highest incidence was in SMEN (73%), followed by AANGE (54%) and ANGS (41%).

Second, to what extent did the contributors show an interest in the then forthcoming election. The occurrence of the word "election" in the sample was counted.

About 49% of articles in the "u.*" newsgroups used the word "election". In the "a.*" newsgroups it was used by only 12% of the articles and in the "s.*" groups by only 4%. Analysed by subject matter the highest incidence of the word was in politics (35%), followed by environment (22%). In the feminist newsgroups the incidence was very low - 1% only.

Third, a group of words was selected to investigate the theme of violence: gun, kill, murder, bomb. As the discussion on the banning of guns was in progress at the time the incidence of these words was rather high.

In the "u.*" newsgroups the words were used by three quarters (74%) of the articles. In the "a.*" and "s.*" groups they were used by 20% articles in each. Newsgroups dealing with politics had the highest incidence of the violence words (54%), followed by history (29%) and feminism (28%). Two newsgroups had more than one word per article : UPCON (165%), SHLIV (110%). Other newsgroups which scored high on the theme of violence were SMEN (87%), UPMIS(77%), UENV(73%). The newsgroups which scored low on the theme of violence were UPELE (2%), SCWEL (3%), AWSUP (3%), ALBAR (3%). Three newsgroups scored zero: SCCOR, SPOL, SRET.

Due to the lack of equipment, the methods of content analysis could only be used in a simplified version. For this exploratory research the count was done by the search/replace command of the word processor.

14 Conclusion

The purpose of this survey was threefold:

Firstly, to draw attention to a seemingly ephemeral method of communication and to its value in the stream of the collective memory. With fewer people writing personal letters or "letters to the Times", the Usenet may eventually become a rare source of information about what people in our time talked about.

Secondly, to provide some data about the behaviour of contributors on the Usenet. The frequency of postings and cross-postings suggests an information net created spontaneously by the users. There is a need for further case studies to establish the most common patterns of information behaviour.

Thirdly, to show how the method of analysing Usenet contributions relates to similar methods used elsewhere, particularly in Mass-Observation and in content analysis. It is hoped that this pilot study will point the way for further studies of the Usenet on a larger scale and with better technical equipment. (Simsova 1995, Simsova 1996)

References:

- (Berelson 1952)
BERELSON B.R. *Content analysis in communications research*. The Free Press, 1952.
- (Coffey 1996)
COFFEY A. and ATKINSON P. *Making sense of qualitative data...* Sage, 1996, p.5.
- (Keele 1995)
KELLE U. ed. *Computer-aided qualitative data analysis*. Sage, 1995 p.167.
- (Madge 1937-38)
MADGE C. and HARRISON F. *First year's work 1937-38 by Mass Observation*. Lidsey, Drummond, 1938, p.66.
- (Mercer 1989)
MERCER N. *Mass-Observation 1937-40...* Univ. of Manchester Faculty of Economic and Social Studies, 1989, p.3,9,56.
- (Richards 1987)
RICHARDS J. and SHERIDAN D. *Mass-observation at the movies*. Routledge, 1987, p.2
- (Sheridan 1991)
SHERIDAN D. *The Tom Harrison Mass-Observation archive...* Univ. of Sussex Library, 1991, p.3-9.
- (Silverman 1993)
SILVERMAN D. *Interpreting qualitative data...* Sage, 1993, p.10.
- (Simsova 1995)
SIMSOVA S. *Representation of cultures on the Usenet : a survey*. Online Information 95 Proceedings, p 437-454;
- (Simsova 1996)
SIMSOVA S. *Some characteristics of the users of Czech mailing lists dealing with librarianship and computer science*. Focus on Internat. and Comparative Librarianship vol 27,no 3, Dec 1996, p 154-158.
- (Stanley 1990)
STANLEY L. *The archeology of a 1930s Mass-Observation project*. Univ. of Manchester Dept. of Sociology, 1990, p.37.
- (Stone 1966)
STONE P.J. *The General enquirer: a computer approach to content analysis*. Cambridge, Ma.: The MIT Press, 1966
- (Vernon 1939)
VERNON P.E. in BARTLETT, F.C. *The study of society*. 1939, p.199-200.

Table 1 Top headers for all newsgroups combined - 20% and over

newsg	tot	items	
1	223	20.5	45 Not Same Species Marriages???
1	78	20.5	16 vegetarians eating meat????
1	48	20.8	10 Right to Die
1	48	22.9	11 Assisting suicide is malpractice
1	21	23.8	5 Where in time should I start?
1	78	24.4	19 Engineered Soya? (was Bombay duck)
2	269	24.9	67 An Englishman's home is no longer his castle
2	88	25.0	22 UK proposed ban on handguns -reactions?
1	262	25.6	67 Gun Control after Dunblane
1	46	26.1	12 Interested in Reenactment Groups
1	25	32.0	8 Land For Building In Cornwall
1	218	44.0	96 Victim disarmament in the UK
1	73	45.2	33 Skull in Boulder images
1	40	50.0	20 Where did you *barefoot* today?
2	45	57.8	26 After the Fall
1	204	69.6	142 Thoughts on Abortion

Table 2 Top headers for "uk.*" and "soc.*" - 10% and over

newsg	tot	items	
UPELE	207	11.1	23 If it is 1st May.....
SCCEL	204	11.3	23 Celtic Knotwork prior to Christianity
SCCOR	25	12.0	3 Hal an Tow
UENW	78	12.8	10 Foxhunting
SCSCO	220	13.2	29 Why Tartan and Other Questions
UGLOC	15	13.3	2 Spontaneous Marxist demos in Hamburg
SCBRI	206	13.6	28 Violent America (UK Bans handguns)
UPCON	218	13.8	30 UK SHOOTERS SHAFTED BY CHARLATANS!
UPCON	218	15.6	34 UK Pistol Ban for America Awaits Tragedy
UPELE	207	19.8	41 Election Law Reform?
UENW	78	20.5	16 vegetarians eating meat????
UENW	78	24.4	19 Engineered Soya? (was Bombay duck)
UPMIS	262	25.6	67 Gun Control after Dunblane
SCCOR	25	32.0	8 A Question About Land For Building In Cornwall
UPCON	218	44.0	96 VICTIM DISARMAMENT IN THE UK
UPENV	63	84.1	53 An Englishman's home is no longer his castle

Table 3 Terms used in headers - part1

cul	98	fem	839	god	182
cul	49 race	fem	146 abortion	god	39 Christian
cul	28 slavery	fem	100 Marriage	god	28 fall
cul	5 Jews	fem	91 women	god	12 flood
cul	4 skin colour	fem	86 wives	god	10 atheist
cul	3 stereotypes	fem	53 feminism	god	9 God
cul	3 nationalism	fem	52 men	god	8 Mormon
cul	2 language	fem	46 genital mutilatn	god	7 prophecy
cul	2 diversity	fem	43 conception	god	6 survivalist
cul	2 black	fem	29 homeless	god	6 beast
		fem	28 rape	god	5 Trinity
env	73	fem	24 sexism	god	5 satan
env	26 barefoot	fem	15 child	god	5 Old testament
env	17 environment	fem	14 male bashing	god	5 evolution
env	16 vegetarians	fem	10 polygamy	god	5 Bible
env	4 trees	fem	9 status of women	god	4 pope
env	4 greenpeace	fem	9 sex drive	god	4 nativist
env	2 hedgerow	fem	9 lesbian	god	4 good girls
env	2 earth	fem	9 lovers	god	2 psalm
env	2 conservation	fem	8 spanked	god	2 prayer
		fem	8 erotic	god	2 pagan
		fem	5 porn	god	2 Muslim
		fem	5 honoring women	god	2 Jesus
		fem	4 naked	god	2 faith bashing
		fem	4 College	god	2 evil
		fem	4 clitoris	god	2 catholicism
		fem	3 seduction	god	2 blasphemy
		fem	3 male intrigue	god	2 baptism
		fem	3 fertility		
		fem	3 domestic violence		
		fem	2 nags		
		fem	2 gay		
		fem	2 friends		
		fem	2 fashion		
		fem	2 dating		
		fem	2 dad		
		fem	2 circumcision		
		fem	2 boys		

Table 3 Terms used in headers - part 2

his 257	old 19	psy 192
his 27 holocaust	old 7 retirement	psy 39 violent
his 26 Halloween	old 6 euthanasia	psy 31
responsibility		
his 24 Irish holocaust	old 4 veterans	psy 13 logic
his 22 history	old 2 pre-retirement	psy 11 headstone
his 18 druids		psy 10 humanism
his 17 Hansard	pol 379	psy 9 seniors
his 16 masonic	pol 46 election	psy 9 creationism
his 15 RU-486	pol 45 Tories	psy 9 accent
his 12 reenactment	pol 42 IRA	psy 8 Cook
his 12 Nazi	pol 28 anarchy	psy 6 insanity
his 9 civil war	pol 22 liberals	psy 6 angst
his 6 famine	pol 21 natural law	psy 5 control
his 5 ice age	pol 20 proportional repr	psy 4 survival
his 5 historian	pol 20 byelection	psy 4 self abuse
his 5 colonialism	pol 14 party	psy 4 independence
his 5 potato famine	pol 12 European Union	psy 3 tolerance
his 4 song	pol 10 war	psy 3 no touch
his 4 Roswell	pol 10 socialism	psy 3 behaviour
his 4 Gunpowder Plot	pol 10 draft	psy 3 alone
his 3 Salem Witch	pol 8 president	psy 2 tragedy
his 2 Zeus	pol 7 civil rights	psy 2 silence
his 2 mythology	pol 6 representation	psy 2 moral
decision		
his 2 Gulf war	pol 5 vote	psy 2 missing
his 2 enlightenment	pol 5 Labour	psy 2 exploring
his 2 clan	pol 5 equality	psy 2 death
his 2 civilization	pol 4 right-wing	
his 2 civilisation	pol 4 demos	
his 2 braveheart	pol 4 democracy	
his 2 archeology	pol 4 communist	
	pol 3 PC bigots	
med 62	pol 3 EuroCon	
med 14 suicide	pol 2 royalfamily	
med 12 medical	pol 2 policies	
med 9 aids	pol 2 peace	
med 7 hearing aid	pol 2 oppression	
med 4 placenta	pol 2 constitution	
med 3 malaria	pol 2 capitalist Party	
med 3 Health care	pol 2 alternative gov	
med 2 sick	pol 2 affirmative act	
med 2 plagues		
med 2 hunger		
med 2 hormone		
med 2 cigarette		

Table 3 Terms used in headers - part 3

rec 210	sci 174	soc 427
rec 29 tartan	sci 43 climate predict	soc 115 victim
disarmnt		
rec 26 food	sci 41 toxicity	soc 68 gun
rec 25 knotwork	sci 13 phone	soc 38 arrested
rec 19 soya	sci 10 Me-109E	soc 34 pistol
rec 11 rugby	sci 10 chemical agents	soc 24 lifers
rec 10 foxhunting	sci 7 oil production	soc 16 kilt
rec 9 libraries	sci 6 reptiles	soc 14 tax
rec 9 film	sci 6 airships	soc 11 bomb
rec 7 slam	sci 5 satellites	soc 11 press
relations		
rec 6 watchfires	sci 5 fossil	soc 9 assasination
rec 6 olympics	sci 4 patent	soc 8 land
rec 5 bonfires	sci 4 computer	soc 8 building
rec 4 pipe band	sci 3 Toyota	soc 7 underlord
rec 4 lanterns	sci 3 ship	soc 6 marxist
rec 3 sugar	sci 2 TW800	soc 6 legislation
rec 3 slang	sci 2 science	soc 6 knives
rec 3 New Year	sci 2 rockets	soc 5 loan sharks
rec 3 motorist	sci 2 netscape	soc 5 insurance
rec 3 latin	sci 2 kitten	soc 5 finance
rec 3 holidays	sci 2 helicopter	soc 4 third world
rec 3 concert	sci 2 constellations	soc 4 police
killing		
rec 3 aesthetics		soc 4 justice
rec 2 wrestlers		soc 3 names
rec 2 winter		soc 3 horn
rec 2 teacher		soc 3 cost of
living		
rec 2 recipes		soc 3 Communities
rec 2 hurling		soc 2 terrorism
rec 2 hunt		soc 2 opinion poll
rec 2 golf		soc 2 Media bias
rec 2 fishing		soc 2 lawsuit
		soc 2 impeachment
		soc 2 generation X

Table 4 List of newsgroups in the sample

AACT	alt.activism
AANA	alt.anarchism
AANGE	alt.anger
AANGS	alt.angst
ABPRO	alt.bible.prophecy
ACAT	alt.catastrophy
ADISC	alt.discrimination
AEAR	alt.earth_summit
AFIND	alt.feminism.individualism
AHFUT	alt.history.future
AHIF	alt.history.what-if
AHLIV	alt.history.living
ALBAR	alt.lifestyle.barefoot
AOEAR	alt.org.earth-first
APEQA	alt.politics.equality
ASEAR	alt.save.the.earth
ASFUT	alt.society.futures
ASUR	alt.survival
AWSUP	alt.women.supremacy
SCBRI	soc.culture.british
SCCEL	soc.culture.celtic
SCCOR	soc.culture.cornish
SCIRI	soc.culture.irish
SCSCO	soc.culture.scottish
SCWEL	soc.culture.welsh
SFEM	soc.feminism
SHIF	soc.history.what-if
SHIS	soc.history.
SHLIV	soc.history.living
SHMOD	soc.history.moderated
SMEN	soc.men
SPOL	soc.politics
SRET	soc.retirement
SVET	soc.veterans
SWOM	soc.women
TEUT	talk.euthanasia
UENV	uk.environment
UGLOC	uk.gov.local
UPCON	uk.politics.constitution
UPELE	uk.politics.electoral
UPENV	uk.politics.environment
UPMIS	uk.politics.misc