The Bergen Corpus of London Teenage Language

They just gonna analyse like length of words and that.

Wicked, innit!
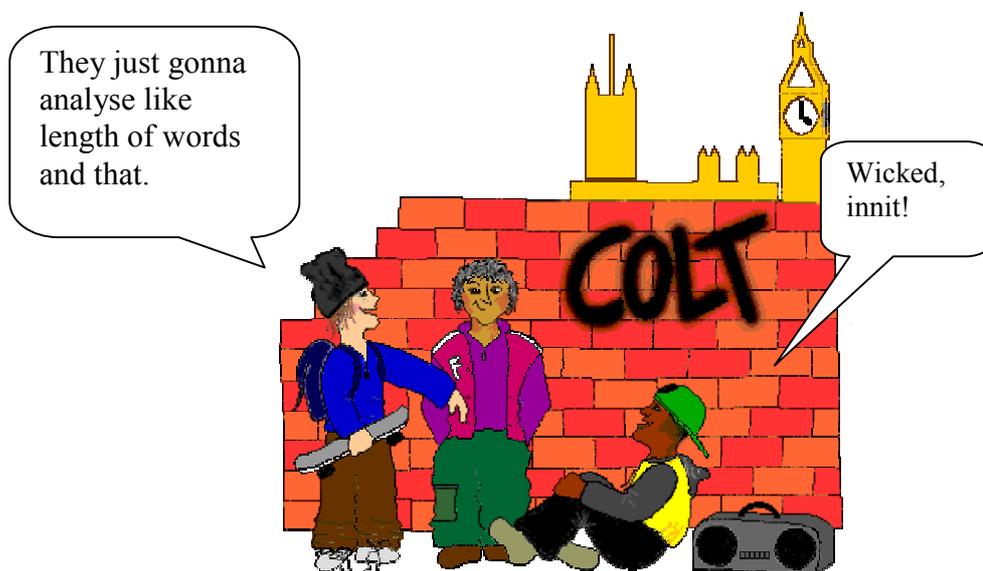
COLT

Introduction

COLT is the first large corpus of British English teenage talk made available for research worldwide. The reason for compiling the corpus was that teenage language so far had been largely unexplored, and we were convinced that the study of spontaneous teenage talk would give new insights into language development and language change. The reason for restricting the corpus collection to London was the assumption that new trends predominate among teenagers in large capitals, from where they can be expected to spread to other areas.

Corpus compilation

The techniques used for collecting COLT were modelled on the principles adopted for the collection of the British National Corpus (BNC). The main recording took place in April/March 1993, and in September 1993, this material was supplemented by recordings collected in a school in the Greater London Metropolitan area. The recordings were made by 31 volunteering 13-17 year old boys and girls ('recruits') with various backgrounds who went to school in five socially different school boroughs.

The entire material of roughly half a million words was orthographically transcribed by trained transcribers, employed by the Longman Group for transcribing the BNC, and a copy of this version of COLT was incorporated in the BNC. Subsequently, the orthographically transcribed material was submitted to Bergen for editing, which involved correcting misinterpreted talk, reducing the number of <unclear> passages, and adding untranscribed talk. This version was then tagged for word classes by a research team at Lancaster university in the same way as the BNC.

Finally, the orthographic version of COLT was digitised by the Bergen team as a preparation for text-sound alignment, which was carried out by SoftSound, St Albans.

The COLT speakers: social background and conversational settings

Information about the speakers' social background and the conversational settings is based on the the logbooks that the recruits were requested to use. For convenience, distinction is made between speaker-specific information (speakers' age, gender, social class, etc) and conversation-specific information (location and setting). For each conversation, the speaker/setting information is given in the text header.

Speaker-specific information: Age and gender

COLT is specifically designed to represent the language of London teenagers. Nevertheless, the speakers that are actually classified with respect to age range from 1 to 59 years old, which is due to the occasional presence of some of the recruits' younger and older family members and to the presence of teachers in some of the conversations. By way of illustration, the speakers have been grouped into six different age groups in Figure 1, preadolescence (0-9), early adolescence (10-13), middle adolescence (14-16), late adolescence (17-19), young adults (20-29) and older adults (30+):
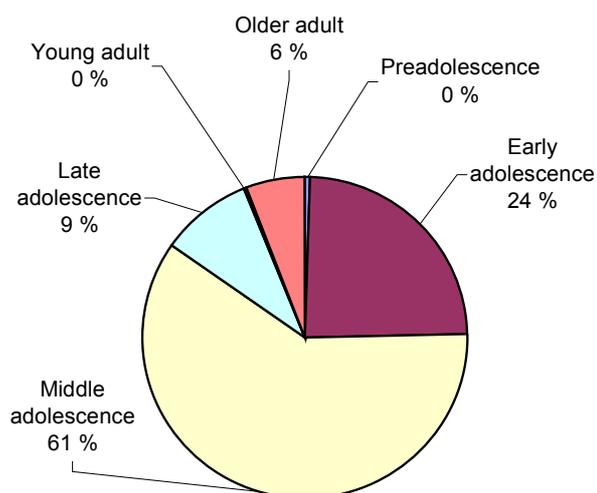
Figure 1: Distribution of COLT text material in the various age groups

Only three of these age groups, early, middle and late adolescence, can be said to represent the 'core' of COLT-informants and the target group of the project. 85 per cent of the corpus material comes from speakers within these age groups. The other age groups are represented to varying degrees. The preadolescent group accounts for a very small amount of text (1,855 words, 0.46 %), and the same goes for the young adult group (1,138 words, 0.28 %). The older adult group mainly comprises the recruits' parents and, to a lesser degree, their teachers. This group contributed about six per cent of the corpus material, which amounts to 23,055 words.

As regards gender, girls and boys contributed roughly the same amount of text: the male speakers about 51.8 per cent (230,616 words) and the female speakers 48.2 per cent (214,215 words).

Speaker-specific information: Social class

The social classification divides the recruits into three (instead of originally five) different social groups, labelled 'high', 'middle' and 'low'. The social class index is based on information that the 31 COLT recruits provided by filling out a personal <u>data sheet</u>. Three pieces of information from the data sheet are used as indicators of social class: residential area, parents' occupation and whether the parents are employed or unemployed. Residential area and parents' occupation constitute social indices in their own right, while the employed/unemployed distinction is used as a slight modification of the occupational index. As this information was provided for no other speakers than the recruits, only the recruits and their families are classified with respect to social class.

The information regarding parents' occupation is treated in accordance with *The Standard Occupational Classification* (Office of Population Censuses and Surveys (OPCS): 1991). Each parent has been classified by the standard categories I-V, except for a single, unclassifiable recruit who did not provide any information regarding parents' occupation. The OPCS classification gives a detailed list of how to categorise each single occupation, and each profession falls into one of the following broad categories, known as 'social classes':

I	Professional etc occupations
II	Managerial and technical occupations
III	Skilled occupations
IV	Partly skilled occupations
V	Unskilled occupations (ibid:12)

Some recruits reported that the parents neither worked nor had a profession. They were given the same occupational score as those belonging to class V. Due to the increase in the number of families with both parents working, as well as a gradual process of levelling of the sex roles, it was natural to include the mother's occupation in the calculation of the social class index, and the two occupations have been weighed equally.

Parents' occupation was considered a more important indicator of social class than area of residence. Information regarding parents' occupation is specific to the single recruit, while area of residence is shared with other recruits. No borough of London is consistent as regards the social class of its inhabitants. We therefore weighted the two indicators differently by multiplying the occupational score by 2 before calculating the Recruit's total score.

As only recruits and their families are classified, only about 50 per cent of the corpus material can be assigned a social group value. The material that has been classified is evenly distributed across the three social groups:
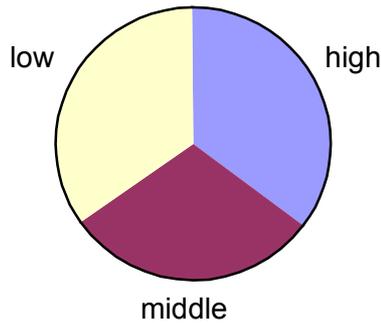
Figure 2: Distribution of COLT text material in the various social groups

Speaker-specific information: Names and anonymity

For the sake of anonymity, last names, addresses and telephone numbers have been erased. We hold that first names ought not to be changed in the COLT transcripts, as they carry social and ethnic information which it is generally difficult to match with fictitious names. The speakers represent a broad range of socioeconomic backgrounds and ethnic groups, which is reflected, to a certain extent, in their first names. Some corpus builders use fictitious names with a similar stress pattern and number of syllables as the real name. We found that, although such a policy is commendable, it could not be successfully implemented on COLT. In the first place, we do not believe it is possible to fulfill all the criteria in a sufficient number of cases, not even with straightforward English names. Considering, then, that there is a large number of ethnic minority speakers in COLT, the task becomes even more demanding.

Conversation-specific information: The London boroughs

The COLT recruits attended five different schools located in five different school boroughs, notably the Inner London boroughs Hackney, Tower Hamlets and Camden, the Outer London borough Barnet, and Hertfordshire, represented by a private boarding school. Hertfordshire is located within the London Metropolitan area, as defined in the *Key statistics for local authorities, Great Britain* (Office of Population Censuses and Surveys:1994). The borough classification incorporates information on borough of residence and school borough of the recruits. The residential boroughs include Barnet, Brent, Camden, Enfield, Hackney, Hertfordshire, Islington, Richmond, Tower Hamlets and Westminster. In most cases borough of residence is identical with school borough, but some recruits attended school in a borough

different from their residential borough. This applies in particular to the Hertfordshire boarding school pupils. Since teenagers are generally assumed to adapt to peer group norms of linguistic behaviour, and are more likely to identify with classmates than their parents, the school borough classification appears to yield a more satisfactory and reliable indicator of group membership. The classification of each recruit with respect to school borough and residential borough is given in the Personal data survey.

Conversation-specific information: Conversational settings

Most of the COLT conversations take place in or near the school or the home of the recruit. Table 1 contains a complete survey of the settings: ,

Table 1: Settings of the COLT conversations

appartment	boarding_school_house	bus
car	Church_Street	classroom
class_room	entrance_to_home_flats	flat
games_hall	girls'_changing_room	gym_room_-_school_playground_-_classroom
home	home/friend's_house	home_of_speaker_11
house	house_of_speaker_3	in_a_park
in_the_house_of_speaker_3	music_room	outside
outside_home	outside_pub	park
Peter's_house	playground	pub
respondent's_room_at_boarding_school	restaurant	school
school_classroom	school_dining_room	school_sixth_form_common_room
school_dinner_hall	school_form_room	school_library
school,_outside	school_playground	
school_study	science_laboratory	shop
street	street_-_bus_-_classroom	study_room_at_boarding_school
walking_to_shop	way_to_school	

Settings in connection with school and home are the most common and account for about 48 per cent and 32 per cent of the material, respectively.

Header information

The speaker-specific and conversation-specific information is listed in the header of each of the COLT text files. The word-class tagged version contains a slightly different set of codes. The following survey illustrates the header information given in the text files:

<REG>	**S**	*region: south*
% reference number // title		
<REF>	**B140402**	*number of text/file*

```
<TIT>      ?                              title
% date // time
<DAT>      ?                              date
<TIM>      ?                              time
% recording // input device
<DEV>      wlk                            walkman
% duration of conversation
<DUR>      6.23                           minutes.seconds in decimal code
<WC>       1324                           word count
% details of conversation // locality - county & town
<CTY>      Greater London                 county
<LOC>      Barnet                         borough where conversation takes place
% setting // activity
<SET>      classroom                      setting
<ACT>      art lesson                     activity
% type of discourse // superfield // subject // individual sub
<TYP>      conversation                   type of discourse
<SUP>      ?                              superfield
<SUB>      ?                              subject
<IND>      ?                              individual subject
% spontaneity factor // audience
<SPN>      3                              3 indicates highest spontaneity
<AUD>      4+                             number of participants and/or people present
<BOB>      ?                              miscellaneous information regarding the conversation
% details of participant // identifier // name // gender
<IDT1>     1                              1 is always the recruit carrying the walkman
<NAM1>     Alex                           name of the recruit
<GEN1>     m                              gender
% age // first language // dialect
<AGE1>     14                             age
<LAN1>     BrE                            language
<DIA1>     London                         dialect
% occupation // education // social group
<OCC1>     student                        occupation
<EDU1>     still studying                 education
<SOC1>     1                              social group
% relationship to respondent // other relationships
<REL1>     respondent                     indicates that this speaker is the recruit (respondent)
<OTH1>     friend2 friend8 pupil2    this speaker is a friend of speakers 2 and 8, and a pupil of speaker 12
<BOB1>     ?                              miscellaneous information regarding the speaker
% details of participant // identifier // name // gender    information about the next speaker
<IDT2>     2
<NAM2>     Marc
<GEN2>     m
% age // first language // dialect
<AGE2>     14
<LAN2>     BrE
<DIA2>     London
% occupation // education // social group
<OCC2>     student
<EDU2>     still studying
<SOC2>     ?
% relationship to respondent // other relationships
<REL2>     friend                         speaker 2 is a friend of the recruit
<OTH2>     friend3 friend8 friend9 pupil12    - and a friend of speakers 3, 8, 9, also a pupil of speaker 12
<BOB2>     ?
% details of participant // identifier // name // gender        etc
<IDT8>     8
<NAM8>     Daniel
<GEN8>     m
```

```
% age // first language // dialect
<AGE8>    13
<LAN8>    BrE
<DIA8>    London
% occupation // education // social group
<OCC8>    student
<EDU8>    still studying
<SOC8>    ?
% relationship to respondent // other relationships
<REL8>    friend
<OTH8>    friend2 pupil12
<BOB8>    ?
% details of participant // identifier // name // gender
<IDT12>   12
<NAM12>   ?
<GEN12>   m
% age // first language // dialect
<AGE12>   35
<LAN12>   BrE
<DIA12>   London
% occupation // education // social group
<OCC12>   teacher
<EDU12>   ?
<SOC12>   ?
% relationship to respondent // other relationships
<REL12>   teacher
<OTH12>   teacher2 teacher8
<BOB12>   ?
```

Mark-up and indexing

Table 3 describes the transcription conventions used in the orthographic corpus and the tagged corpus. Worth noting is the dual use of the full stop (.) and the comma (,), which can indicate either as a sentence boundary or a pause or both.

Table 2: COLT mark-up

ORTHOGRAPHIC	TAGGED	COMMENTS
, . ? !	see CLAWS6 tagset	sentence-like boundaries; also continuing, terminating, questioning, and exclamatory intonation
CAPS	CAPS	sentence beginnings
=	<trunc> text </trunc>	incomplete word
,	see CLAWS6 tagset	brief pause
.	see CLAWS6 tagset	medium pause
...	<pause dur=3>	long pause
... (5)	<pause dur=5>	pause 5 seconds
<nv> laugh </nv>	<vocal desc=laugh>	non-verbal sound
<name>	&name;&NP1;	name
<address>	&address;&NN1;	address
<unclear>	<unclear>	unintelligible speech
<unclear> (5)	<unclear dur=5>	unintelligible speech 5 seconds
[text]	<ptr=P001> text <ptr=P002>	single overlap
[{}]	non-existent	double overlap
<singing> text </>	<shift new=singing> text <shift>	paralinguistic features
(hairdryer on)	<event desc="hairdryer on">	contextual comment

<??> text </>	<uncertain> text </uncertain>	uncertain transcription
(sic)	<sic> text </sic>	awkward pronunciation

Paralinguistic features and non-verbal sounds

A large number of original and innovative paralinguistic features are applied, examples being <mimicking a monkey>, <mimicking a yobbo hooligan>, <mimicking Yorkshire accent> and <teasing>. The markup convention for these is <paralinguistic feature> text </> in the orthographic version, and in the tagged version <shift new=paralinguistic feature> text <shift>. In addition, as a reflection of the humoristic and playful nature of many of the conversations, COLT contains a relatively large number of non-verbal sounds that are, presumably, less prevalent in corpora of adult English. These include <nv>blowing air through lips</nv>, <nv>mimicking gorilla noises</nv>, <nv>mimicking shaving noise</nv>, and so on. These are marked with the tag <nv> in the orthographic corpus, and the tag <vocal desc=> in the tagged corpus.

The prosodic version

A representative selection of COLT conversations, amounting to approximately 150,000 words, has been transcribed prosodically. A survey of the prosodically transcribed conversations that are available is given below. The texts were selected on the basis of three criteria: age, gender and social group. The following prosodic markers were used:

bold type	nucleus
\	fall
/	rise
V	fall-rise
Λ	rise-fall
-	level
#	tone-unit boundary

Prosodically transcribed files

File name
B132401.pro
B132402.pro
B132403.pro
B132404.pro
B132405.pro
B132406.pro
B132407.pro
B132408.pro
B132409.pro
B132501.pro
B132502.pro
B132503.pro
B132504.pro
B132601.pro
B132602.pro
B132603.pro
B132605.pro
B132606.pro
B132607.pro
B132609.pro
B132610.pro
B132611.pro
B132612.pro
B132613.pro
B132614.pro
B132615.pro
B132616.pro
B132617.pro
B132701.pro
B132702.pro
B132703.pro
B132704.pro
B132705.pro
B132706.pro
B132707.pro
B132708.pro
B133101.pro
B133201.pro
B133202.pro
B133203.pro
B133301.pro
B133302.pro

B133401.pro
B133501.pro
B133602.pro
B133701.pro
B133702.pro
B133703.pro
B133704.pro
B133705.pro
B133801.pro
B133901.pro
B133902.pro
B135901.pro
B135902.pro
B135903.pro
B135904.pro
B135905.pro
B135906.pro
B135907.pro
B136101.pro
B136102.pro
B136103.pro
B136104.pro
B136105.pro
B136106.pro
B136107.pro
B136108.pro
B136401.pro
B136402.pro
B136403.pro
B136404.pro
B136405.pro
B136406.pro
B136407.pro
B136408.pro
B136409.pro
B136410.pro
B136411.pro
B136501.pro
B136502.pro
B136503.pro
B137701.pro
B137801.pro
B137802.pro
B137803.pro

B137804.pro
B137901.pro
B137902.pro
B137903.pro
B137904.pro
B138001.pro
B138901.pro
B138902.pro
B138903.pro
B138904.pro
B138905.pro
B138906.pro
B138907.pro
B139001.pro
B139002.pro
B139003.pro
B139201.pro
B139301.pro
B139302.pro
B139303.pro
B139304.pro
B139305.pro
B139306.pro
B139307.pro
B139308.pro
B141101.pro
B141102.pro
B141103.pro
B141104.pro
B141105.pro
B141106.pro
B141107.pro
B141601.pro
B141602.pro
B141603.pro
B141604.pro
B141605.pro
B141606.pro
B141701.pro
B142101.pro
B142102.pro
B142103.pro
B142104.pro
B142105.pro

B142106.pro	B142305.pro	B142606.pro
B142201.pro	B142306.pro	B142607.pro
B142202.pro	B142307.pro	B142608.pro
B142301.pro	B142601.pro	B142701.pro
B142302.pro	B142602.pro	B142702.pro
B142303.pro	B142603.pro	B142703.pro
B142304.pro	B142604.pro	

COLT-based research

The COLT material has already been the source of a large number of studies, including several PhD dissertations. At the English department, University of Bergen, we have concentrated on what immediately struck us as charactertistic as we began studying the teenage conversations: the frequent use of pragmatic particles (eg *cos*, *I think*, *like*, *sort of*), tags (*innit*, *okay*, *right*), taboo words and slang. Other studies were devoted to nonstandard features of grammar (eg auxiliary deletion, negative concord, double negation) and interactional style (eg conflict talk). Nearly all the studies reflect a sociolinguistic approach.

COLT-based publications (as of April 2000)

Andersen, G. 1995. Omission of the primary verbs BE and HAVE in London teenage speech - a sociolinguistic study. Unpublished MA thesis. Department of English, University of Bergen.

Andersen, G. 1997a. *They gave us these yeah, and they like wanna see like how we talk and all that.* The use of *like* and other discourse markers in London teenage speech. In U-B. Kotsinas, A-B. Stenström & A-M. Karlsson (eds). *Ungdomsspråk i Norden.* Stockholm: MINS 43: 82-95.

Andersen, G. 1997b. *They like wanna see like how we talk and all that.* The use of *like* as a discourse marker in London teenage speech. In M. Ljung (ed). *Corpus-based studies in English.* Amsterdam: Rodopi, 37-48.

Andersen, G. 1997c. *I goes you hang it up in your shower, innit? He goes yeah.* The use and development of invariant tags and follow-ups in London teenage speech. Unpublished paper presented at the First Language Variation Workshop, Reading UK.

Andersen, G. 1998a. The pragmatic marker *like* from a relevance theoretical perspective. In A. Jucker & Y. Ziv (eds). *Discourse markers: descriptions and theory.* Amsterdam: John Benjamins, 147-170.

Andersen, G. 1998 b. Pragmatic markers and the semantics/pragmatics distinction. Unpublished paper. University of Bergen.

Andersen, G. Forthcoming a. Are tag questions questions? Evidence from spoken data. Paper presented at the 18th ICAME Conference. Belfast, UK, May 1998.

Andersen, G. Forthcoming b. The role of the pragmatic marker *like* in the identification of explicatures. Paper presented at the 6th International Pragmatics Conference. Reims, France, July 1998.

Andersen, G. 1999. *Pragmatic markers and sociolinguistic variation: a corpus-based study.* Unpublished PhD dissertation, University of Bergen.

Andersen, G. 2000. The role of the pragmatic marker *like* in utterance interpretation. In G. Andersen & T. Fretheim. 2000 (eds).

Andersen, G. & T. Fretheim. 2000. Introduction. In G. Andersen & T. Fretheim. 2000. (eds).

Andersen, G. & K. Hasund. 1996. COLT on TACT: A demonstration of TACT as applied to the Bergen Corpus of Teenage Language. Poster and software demonstration at the ALLC/ACH Conference, University of Bergen. Published online, http://www.ach.org/ACH_Posters/colt.html

Andersen, G. & T. Fretheim (eds). 2000. *Pragmatic markers and propositional attitude.* Amsterdam: John Benjamins.

Andersen, G. & A-B. Stenström. 1996. COLT: A progress report. *ICAME Journal* 20: 133-136.

Berland, U. 1997. Invariant tags: Pragmatic functions of *innit, okay, right* and *yeah* in London teenage conversation. Unpublished MA thesis. Department of English, University of Bergen.

Bynes, A. 1998. A corpus-based study of expletive use among London teenagers. Unpublished MA thesis. Department of English, University of Bergen.

Erhardt, S. 1998. Strategien von Gesprächskontrolle bei jugendlichen männlichen und weiblichen Sprechern. Unpublished MA thesis. Department of English, University of Jena.

Erman, B. 1996. *Just wear the wig innit!* From identifying and proposition-oriented to itensifying and speaker-oriented: grammaticalization in progress. In *Proceedings from the XVIth Scandinavian conference of linguistics.*

Erman, B. 1997. *Guy's just such a dickhead* The context and function of *just* in teenage talk. In U-B Kotsinas, A-B. Stenström & A-M. Karlsson (eds). *Ungdomsspråk i Norden.* Stockholm: MINS 43: 96-110.

Guenther, U. Forthcoming. What's funny to whom? Humour as a conversational strategy in everyday talk. A study based on the BNC and COLT. PhD thesis. University of Freiburg.

Haslerud, V. & A-B. Stenström. 1994. COLT: Mark-up and trends. *Hermes Journal of Linguistics*, 55-70.

Haslerud, V. & A-B. Stenström. 1995. The Bergen Corpus of London Teenager Language (COLT). In G. Leech, G. Myers & J. Thomas (eds). *Spoken English on computer*. London: Longman, 235-242.

Hasund, K. 1996. COLT conflicts: Reflections of gender and class in the oppositional turn sequences of London teenage girls. Unpublished MA thesis. Department of English, University of Bergen.

Hasund K. 1998a. From Woman's Place to Women's Places: Class-determined variation in the verbal disputes of London teenage girls. In A. Despard (ed). *A Woman's Place: Women, Domesticity and Private Life.* Kristiansand: Norwegian Academic Press.

Hasund, K. 1998b. Protecting the innocent: The issue of informants' anonymity in the COLT corpus. In A. Renouf (ed): *Explorations in Corpus Linguistics.* Amsterdam: Rodopi, 13-27.

Hasund, I. K. Forthcoming a. *Pragmatic particles in English and Norwegian: A contrastive study of* like *and* liksom. Doktorgradsprosjekt. Engelsk institutt, Universitetet i Bergen/Høgskolen i Agder.

Hasund, I. K. Forthcoming b. Fisefine fjortisjenter sier "lissom" hele tiden: Tro og viten om småord i talespråket. *Images of otherness. Tradition and identifications,* red. av L. A.Ytrehus. *Kulturtekster* 16. Bergen: Senter for Europeiske kulturstudier.

Hasund, I. K. Forthcoming c. Markers of epistemic stance: a contrastive study of the pragmatic particles *like* in English and *liksom* in Norwegian. *Modalität und mehr*, red. av O. Letnes & H. Vater. *Fokus*. Trier: Wissenschaftlicher Verlag Trier.

Hasund, I. K. Forthcoming d. Engelsk i norsk ungdomsspråk: noen resultater fra en slangundersøkelse. *Språknytt*.

Hasund, K.& A-B. Stenström. 1997. Conflict talk: A comparison of the verbal disputes of adolescent females in two corpora. In M. Ljung (ed). *Corpus-based studies in English.* Amsterdam: Rodopi, 119-133.

Kotsinas, U.-B., Karlsson, A.-M. & A.-B. Stenström. 1997. (eds). *Ungdomsspråk i Norden.* Stockholm: MINS 43.

Monstad, K. 1998. The grammaticalisation of *sort of* and *kind of* in young and old Londoner's speech. Unpublished MA thesis. Department of English, University of Bergen.

Mosaker, H-M. 1998. Qualifying utterances: A study of epistemic phrases in COLT and the BNC. Unpublished MA thesis. Department of English, University of Bergen.

Paradis, C. Forthcoming. *It's well weird.* Degree modifiers of adjectives: the nineties. Paper presented at the 18th ICAME conference. Belfast, UK, May 1998.

Stenström, A-B. 1995. Taboos in teenage talk. In G. Melchers & B. Warren (eds). *Studies in Anglistics*. Stockholm: Almqvist & Wiksell International, 71-80.

Stenström, A-B. 1997a. Tags in Teenage Talk. In U. Fries, V. Müller & P. Schneider (eds). *From Ælfric to the New York Times. Studies in English Corpus Linguistics*. Amsterdam: Rodopi, 139-148.

Stenström, A-B. 1997b. *Can I have a chips please? - Just tell me what one you want*: Nonstandard grammatical features in London teenage talk. In J. Aarts et al (eds). *Studies in English Language and Teaching*. Amsterdam: Rodopi, 141-152.

Stenström, A-B. 1998a. From sentence to discourse: *cos(because)* in teenage talk. In A. Jucker & Y. Ziv (eds). *Discourse markers: descriptions and theory*. Amsterdam: John Benjamins, 127-146.

Stenström, A-B. 1998b The intensifier *well* - a recent development? Poster presented at the 6th International Pragmatics Conference. Reims, France, July 1998.

Stenström, A-B. 1999a. Teenage talk: Chatting about boys and discussing taboo habits. In R. Horowitz (ed). *Developing understanding of the world through talk and text.* The University of San Antonio, Texas.

Stenström, A-B. 1999b. Chatting about boys and other important things. *In Proceedings from the 16th International Congress of Linguistics in Paris 1997.*

Stenström, A-B. 1999c.*He was really gormless - She's a bloody crap.* Girls, boys and intensifiers. *Out of Corpus*. Amsterdam: Rodopi.

Stenström, A-B. 1999 d. Corpora and youth language. In P. Pietilä (ed). *Language Teaching and Linguistic Research. AFinLA Year book 57,* 105-116.

Stenström, A-B. 2000. *It's enough funny man*. Intensifiers in Teenage Talk. To appear in J. Kirk (ed). *Corpora Galore. Analyses and Techniques in Describing English.* Amsterdam: Rodopi.

Stenström, A-B. In press. Teenage talk: Chatting about boys and discussing taboo hatibs. In R. Horowitz (ed*). Developing understanding of the world through talk and text.* The University of San Antonio, Texas.

Stenström, A-B. Forthcoming a. The construction of identity in London teenage talk. To appear in J. Androutsopoulos and G. Georgakopoulou (eds). *Discourse constructions of youth identities*. Amsterdam: Benjamins.

Stenström, A-B. Forthcoming b. From slang to slanguage: a description based on teenage talk. To appear in T. Kis (ed). *Mi a szleng*?. Debrecen: Kossuth Lajos University Press.

Stentröm, A-B. Forthcoming c. Oh! Oh gossip! Gossip! Festschrift (confidential).

Stenström, A-B & Andersen, G. 1996. More trends in teenage talk: A corpus-based investigation of the discourse items *cos* and *innit*. In C. Percy, C. Meyer & I. Lancashire (eds). *Synchronic corpus linguistics*. Amsterdam: Rodopi, 189-203.

Stenström, A-B & L. E. Breivik. 1993. The Bergen Corpus of Teenage Talk. *ICAME Journal* 17, 128.

Stenström, A-B & V. Haslerud. 1994. Transcribing COLT: Mark-up and trends. *Hermes. Journal of linguistics* 13, 55-70.

Stenström, A-B. Forthcoming d. with G. Andersen & K. Hasund. *Trends in Teenage Talk: Corpus compilation., analysis and findings.* To appear in the series Studies in Corpus Linguistics. Amsterdam: John Benjamins.

Straume, A. 1998. The variable (t) in London adolescent speech. A sociolinguistic study of a phonological variable. Unpublished MA thesis, Department of English, University of Bergen.

Tandberg, A. 1996. *Innit* from a grammatical and pragmatic point of view. Unpublished MA thesis, Department of English, University of Bergen.

Aas, H. J. 1998. Conversational storytelling in COLT: A study of three stories from the perspective of Systemic Functional Linguistics. Unpublished MA thesis. Department of English, University of Bergen.

Personal data sheet

Personal data

All information will be treated confidentially. Recruits, their family and conversation partners are guaranteed full anonymity.

Recruit number___18_____

A)   Area of residence (London borough) Brent_____

Postcode_NW10 4AJ_____

B)   Have you ever lived in any other part of England? If so,

where?_____and for how long?_____

C)   Mother's occupation__Teacher_____

Is she currently employed? Y/N Yes_____

D)   Father's occupation__Graphic designer_____

Is he currently employed? Y/N___Yes_

Please remember to take this form with you when you are giving back the personal stereo etc. to Ms Haslerud

Thanks very much for your co-operation!

Personal data survey

Recruit number	Recruit name	Text files	Sex	Age	School borough	Residential borough	Postcode	Also lived in	Mother's occupation	Currently employed?	Father's occupation	Currently employed?	Social group
1	Peter	B132401-B132504	M	14	Hackney	Hackney	E97HT		Teacher	Yes	Doctor(GP)	Yes	2
2	Josie	B132601-B132913	F	14	Hackney	Hackney	6AP		Housewife	No	Plumber	Yes	3
3	Robert	B133101-B133602	M	15	Hackney	Hackney	N16 OEU		Receptionist	Yes	Carpenter	Yes	3
4	Cassie	B133701-B134202	F	15	Hackney	Hackney	N16		Yoga teacher/gardener	Yes	Gardener	Yes	3
5	Anthony	B134401-B134602	M	15	Hackney	Hackney	OJR		College student	Yes	British telecom	Yes	3
6	Grace	B134801-B135207	F	14	Hackney	Stoke Newington	OSX			No		No	3
7	Richard	B135301-B135306	M	13	Tower Hamlets	Tower Hamlets	E3 5LG			No		Yes	
8	Michael	B135601-B135603	M	13	Tower Hamlets	Bow	G3 4NS	Essex (2 years)	Credit control	No	Lorry driver	Yes	3
9	Craig	B135701-B135809	M	13	Tower Hamlets	Tower Hamlets	E2		Advertiser	No	Pub owner	No	3
10	Anthony	B135901-B135907	M	13	Tower Hamlets	Bow	45E		Admin officer	Yes	Student	No	2
11	Barry	B136101-B136108	M	14	Tower Hamlets	Bow	2RS		Waitress	No	Market trader	Yes	3
12	Mark	B136301-B136304	M	13	Tower Hamlets	Bow	2EN		Social worker	Yes	Taylor	Yes	2
13	Carla	B136401-B136503	F	13	Camden	Westminster	W9 2NU			No	Science engineer	Yes	1
14	Sarah	B136601-B136903	F	13	Camden	Brent	NW6 6SB	Devon (1year)	Housewife	No	Computer work	Yes	2
15	Leon	B137101-B137202	M	13	Camden	Camden	NW1 9BJ		Teacher	Yes			1
17	Selum	B137701-B138301	M	13	Camden	Camden	NW0 2BA	Rowleyway (2 yrs)	Secretary	Yes	Tayler	Yes	2
18	Matthew	B138501-B138604	M	13	Camden	Brent	NW10 4AJ		Teacher	Yes	Graphic designer	Yes	2
19	Jonathan	B138901-B139003	M	17	Camden	Camden	NW3 3SV				Trade finance	Yes	1
21	Skonev	B139201-B139308	M	12	Camden	Islington	N19		Housewife	No	History lecturer	Yes	1
22	Terry	B139401-B139614	M	14	Barnet	Barnet	EN4 8PN		Housewife	No	Dress cutter	Yes	2
23	Robin	B139701-B139809	M	14	Barnet	Southgate	N14 5RY		Student (course)	No	Financial consultant	Yes	1
24	Eddie	B140201-B140303	M	15	Barnet	Barnet	N14 5RG		Teacher	Yes	Retired teacher	No	1
25	Alex	B140401-B140607	M	14	Barnet	Barnet	N14		Managing director	Yes	Engineer	Yes	1
26	Caroline	B140701-B140904	F	14	Barnet	Hackney	N4 2LX	Southgate (4 yrs)	Housewife	No	Computer programmer	Yes	2
27	Cassie	B141001-B141003	F	15	Barnet	Enfield	2 RG		Clerk	Yes	Civil servant	Yes	2
28	Danny	B141101-B141501	M	13	Hertfordshire	Islington	3JS	Enfield(6 years)	Pubowner	Yes	Pubowner	Yes	2
29	Jock	B141601-B142005	M	16	Hertfordshire	Richmond		Scotland(14 yrs)	Housewife	No	Loss control coordinator	Yes	1
30	Alistair	B142101-B142202	M	15	Hertfordshire	Essex	COg 4NX		Caterer	Yes	Insurance broker	Yes	1
31	Kath	B142301-B142307	F	17	Hertfordshire	Hertfordshire	EN5 4RL		Farmer	Yes	Farmer	Yes	1
33	Catriona	B142601-B142802	F	16	Hertfordshire	North London	N6 5BY		Artist	Yes	Artist	Yes	1

COLT tagset (CLAWS 6)

APPGE possessive pronoun, pre-nominal (*e.g. my, your, our*)

AT article (*e.g. the, no*)

AT1 singular article (*e.g. a, an, every*)

BCL before-clause marker (*e.g. in order (that), in order (to)*)

CC coordinating conjunction (*e.g. and, or*)

CCB adversative coordinating conjunction (*but*)

CS subordinating conjunction (*e.g. if, because, unless, so, for*)

CSA as (*as conjunction*)

CSN than (*as conjunction*)

CST that (*as conjunction*)

CSW whether (*as conjunction*)

DA after-determiner or post-determiner capable of pronominal function (*e.g. such, former, same*)

DA1 singular after-determiner (*e.g. little, much*)

DA2 plural after-determiner (*e.g. few, several, many*)

DAR comparative after-determiner (*e.g. more, less, fewer*)

DAT superlative after-determiner (*e.g. most, least, fewest*)

DB before determiner or pre-determiner capable of pronominal function (*all, half*)

DB2 plural before-determiner (*both*)

DD determiner (*capable of pronominal function*) (*e.g. any, some*)

DD1 singular determiner (*e.g. this, that, another*)

DD2 plural determiner (*e.g. these, those*)

DDQ wh-determiner (*which, what*)

DDQGE wh-determiner, genitive (*whose*)

DDQV wh-ever determiner (*whichever, whatever*)

EX existential there

FO formula

FU unclassified word

FW foreign word

GE germanic genitive marker (*' or 's*)

IF for (*as preposition*)

II general preposition

IO of (*as preposition*)

IW with, without (*as prepositions*)

JJ general adjective

JJR general comparative adjective (*e.g. older, better, stronger*)

JJT general superlative adjective (*e.g. oldest, best, strongest*)

JK catenative adjective (*able in be able to, willing in be willing to*)

MC cardinal number, neutral for number (*two, three*)

MCGE genitive cardinal number, neutral for number (*two's, 100's*)

MCMC hyphenated number (*40-50, 1770-1827*)

MC1 singular cardinal number (*one*)

MC2 plural cardinal number (*e.g. sixes, sevens*)

MD ordinal number (*e.g. first, second, next, last*)

MF	fraction, neutral for number (*e.g. quarters, two-thirds*)
ND1	singular noun of direction (*e.g. north, southeast*)
NN	common noun, neutral for number (*e.g. sheep, cod, headquarters*)
NN1	singular common noun (*e.g. book, girl*)
NN2	plural common noun (*e.g. books, girls*)
NNA	following noun of title (*e.g. M.A.*)
NNB	preceding noun of title (*e.g. Mr., Prof.*)
NNJ	organisation noun, neutral for number (*e.g. council, department*)
NNJ2	organisation noun, plural (*e.g. goverments, committees*)
NNL1	singular locative noun (*e.g. island, street*)
NNL2	plural locative noun (*e.g. islands, streets*)
NNO	numeral noun, neutral for number (*e.g. dozen, hundred*)
NNO2	numeral noun, plural (*e.g. hundreds, thousands*)
NNT1	temporal noun, singular (*e.g. day, week, year*)
NNT2	temporal noun, plural (*e.g. days, weeks, years*)
NNU	unit of measurement, neutral for number (*e.g. in, cc*)
NNU1	singular unit of measurement (*e.g. inch, centimetre*)
NNU2	plural unit of measurement (*e.g. ins., feet*)
NP	proper noun, neutral for number (*e.g. IBM, Andes*)
NP1	singular proper noun (*e.g. London, Jane, Frederick*)
NP2	plural proper noun (*e.g. Browns, Reagans, Koreas*)
NPD1	singular weekday noun (*e.g. Sunday*)
NPD2	plural weekday noun (*e.g. Sundays*)
NPM1	singular month noun (*e.g. October*)
NPM2	plural month noun (*e.g. Octobers*)
NPO	vague words (e.g. *whatshisname*)
PN	indefinite pronoun, neutral for number (*none*)
PN1	indefinite pronoun, singular (*e.g. anyone, everything, nobody, one*)
PN2	indefinite pronoun, plural (*ones, others*)
PNQO	objective wh-pronoun (*whom*)
PNQS	subjective wh-pronoun (*who*)
PNQV	wh-ever pronoun (*whoever*)
PNX1	reflexive indefinite pronoun (*oneself*)
PPGE	nominal possessive personal pronoun (*e.g. mine, yours*)
PPH1	3rd person sing. neuter personal pronoun (*it*)
PPHO1	3rd person sing. objective personal pronoun (*him, her*)
PPHO2	3rd person plural objective personal pronoun (*them*)
PPHS1	3rd person sing. subjective personal pronoun (*he, she*)
PPHS2	3rd person plural subjective personal pronoun (*they*)
PPIO1	1st person sing. objective personal pronoun (*me*)
PPIO2	1st person plural objective personal pronoun (*us*)
PPIS1	1st person sing. subjective personal pronoun (*I*)
PPIS2	1st person plural subjective personal pronoun (*we*)
PPX1	singular reflexive personal pronoun (*e.g. yourself, itself*)
PPX2	plural reflexive personal pronoun (*e.g. yourselves, themselves*)
PPY	2nd person personal pronoun (*you*)
PT	taboo words (*e.g. bollocks, bonkers*)
RA	adverb, after nominal head (*e.g. else, galore*)
REX	adverb introducing appositional constructions (*namely, e.g.*)
RG	degree adverb (*very, so, too*)

RGQ	wh-degree adverb (*how*)
RGQV	wh-ever degree adverb (*however*)
RGR	comparative degree adverb (*more, less*)
RGT	superlative degree adverb (*most, least*)
RL	locative adverb (*e.g. alongside, forward*)
RP	prep. adverb, particle (*e.g. about, in*)
RPK	prep. adv., catenative (*about in be about to*)
RR	general adverb
RRQ	wh- general adverb (*where, when, why, how*)
RRQV	wh-ever general adverb (*wherever, whenever*)
RRR	comparative general adverb (*e.g. better, longer*)
RRT	superlative general adverb (*e.g. best, longest*)
RT	quasi-nominal adverb of time (*e.g. best, longest*)
TO	infinitive marker (*to*)
UH	interjection (*e.g. oh, yes, um*)
VB0	be base form (*finite i.e. imperative, subjunctive*)
VBDR	were
VBDZ	was
VBG	being
VBI	be infinitive (*To be or not... It will be...*)
VBM	am
VBN	been
VBR	are
VBZ	is
VD0	do base form (*finite*)
VDD	did
VDG	doing
VDI	do infinitive (*I may do... To do...*)
VDN	done
VDZ	does
VH0	have base form (*finite*)
VHD	had (*past tense*)
VHG	having
VHI	have infinitive
VHN	had (*past participle*)
VHZ	has
VM	modal auxiliary (*can, will, would, etc.*)
VMK	modal catenative (*ought, used*)
VV0	base form of lexical verb (*e.g. give, work*)
VVD	past tense of lexical verb (*e.g. gave, worked*)
VVG	-ing participle of lexical verb (*e.g. giving, working*)
VVGK	-ing participle catenative (*going in be going to*)
VVI	infinitive (*e.g. to give... It will work...*)
VVN	past participle of lexical verb (*e.g. given, worked*)
VVNK	past participle catenative (*e.g. bound in be bound to*)
VVZ	-s form of lexical verb (*e.g. gives, works*)
XX	not, n't
ZZ1	singular letter of the alphabet (*e.g. A, b*)
ZZ2	plural letter of the alphabet (*e.g. A's, b's*)

Punctuation tags

YEX	punctuation tag - exclamation mark (*!*)
YQUO	punctuation tag - quotes (*''*)
YBL	punctuation tag - left bracket (*(*)
YBR	punctuation tag - right bracket (*)*)
YCOM	punctuation tag - comma (,)
YDSH	punctuation tag - dash (-)
YSTP	punctuation tag - full-stop (.)
YLIP	punctuation tag - ellipsis (...)
YCOL	punctuation tag - colon (*:*)
YSCOL	punctuation tag - semicolon(*;*)
YQUE	punctuation tag - question mark (*?*)

Nonverbal sounds in COLT

<nv>barking</nv>
<nv>belch</nv> Appendix 1 Non-verbal sounds in
<nv>blowing air</nv>
<nv>blowing air through lips</nv>
<nv>blowing nose</nv>
<nv>blurgh</nv>
<nv>bomb sounds</nv>
<nv>boomph</nv>
<nv>breathe</nv>
<nv>breathing heavily</nv>
<nv>burp</nv>
<nv>chewing</nv>
<nv>choke</nv>
<nv>clap</nv>
<nv>clapping</nv>
<nv>clears throat</nv>
<nv>clicking fingers</nv>
<nv>clicks tongue</nv>
<nv>cough</nv>
<nv>coughing</nv>
<nv>crying</nv>
<nv>deep breath</nv>
<nv>drinking</nv>
<nv>drinking sound</nv>
<nv>drowning noises</nv>
<nv>eating crips</nv>
<nv>gasp</nv>
<nv>giggle</nv>
<nv>hissing sound</nv>
<nv>howl</nv>
<nv>humming</nv>

<nv>humming+singing</nv>
<nv>imitates cat licking</nv>
<nv>imitating computer beep</nv>
<nv>imitating gun noise</nv>
<nv>imitating gun fire</nv>
<nv>imitates vomiting</nv>
<nv>kiss</nv>
<nv>kissing microphone</nv>
<nv>laugh</nv>
<nv>makes car sounds</nv>
<nv>makes creaking door noise</nv>
<nv>makes drunken sounds and a pretend belch</nv>
<nv>makes incantation sounds</nv>
<nv>makes running noises</nv>
<nv>making odd noises</nv>
<nv>making sucking noises</nv>
<nv>making vomiting noises</nv>
<nv>meow</nv>
<nv>mimicking</nv>
<nv>mimicking banjo</nv>
<nv>mimicking blowing handkerchief</nv>
<nv>mimicking bringing up phlegm</nv>
<nv>mimicking buzzing sound</nv>
<nv>mimicking car sounds</nv>
<nv>mimicking chicken sound</nv>
<nv>mimicking clearing throat</nv>
<nv>mimicking cry</nv>
<nv>mimicking dog barks</nv>
<nv>mimicking engine revving</nv>
<nv>mimicking explosion</nv>
<nv>mimicking gorilla noises</nv>
<nv>mimicking kissing sound</nv>
<nv>mimicking licking sound</nv>
<nv>mimicking microphone noises</nv>
<nv>mimicking person speaking with mouth full</nv>
<nv>mimicking Scandinavian accent from the Muppet Show</nv>
<nv>mimicking sound effect</nv>
<nv>mimicking sound of biting</nv>
<nv>mimicking shaving noise</nv>
<nv>mimicking squeaking</nv>
<nv>mimicking squeaking noise</nv>
<nv>mimicking vomiting</nv>
<nv>moan</nv>
<nv>moaning</nv>
<nv>noise for paws</nv>
<nv>noise in microphone</nv>
<nv>panting</nv>
<nv>purring noises</nv>
<nv>raspberry</nv>
<nv>running sound</nv>

<nv>scream</nv>
<nv>sigh</nv>
<nv>singing</nv>
<nv>singing+humming</nv>
<nv>sharp intake of breath</nv>
<nv>shout</nv>
<nv>slurps</nv>
<nv>slurping noises</nv>
<nv>shouting</nv>
<nv>sneeze</nv>
<nv>sniff</nv>
<nv>sound effect</nv>
<nv>sound of drinking with a straw</nv>
<nv>spitting cat sound</nv>
<nv>spitting sound</nv>
<nv>squawking sound</nv>
<nv>squeals</nv>
<nv>sucking noise</nv>
<nv>sucking then purring noises</nv>
<nv>sulking</nv>
<nv>vomiting noises </nv>
<nv>whine</nv>
<nv>whining</nv>
<nv>whinge</nv>
<nv>whinging</nv>
<nv>whistling</nv>
<nv>yawn</nv>
yelping sound