The SLX Corpus of Classic Sociolinguistic Interviews

SOCIOLINGUISTIC VARIABLES
CODING SPECIFICATIONS

http://www.ldc.upenn.edu/Projects/DASL/SLX

Sociolinguistic Variable Survey

Introduction

After the time-aligned transcripts had been created and checked for quality, researchers reviewed the data to identify sociolinguistic variables of interest. The interview subjects do not belong to a coherent speech community in space or time, except the speech community consisting of all English speakers. Thus certain phonological and grammatical variables in this survey were selected because they exist in all English dialects studied to date. They include among others the realization of -ing, t/d deletion and negative concord. Other variables in the survey are associated with certain of the dialects under consideration: habitual 'be' in African American English, for example, or the frication of syllable initial stops in Liverpool speech. Since some of these dialect features are categorically absent from non-speakers of the dialect, not all speakers were coded for all variables; nor were speakers coded exhaustively for any variable. In all, over 90 phonological, phonetic and prosodic and over 60 grammatical and lexical variables have been identified in the corpus. This document defines each of the categories and variables used in the survey.

It is important to note in general that the survey is experimental, non-systematic and principally descriptive. It is not intended to give an exhaustive account of the variation that exists within the corpus. Instead, its purpose is to provide snapshot of intra- and inter-speaker variation in the corpus, and to highlight a broad range of sociolinguistic variables that are attested in the data. It further gives us an idea of how the DASL Project could eventually allow sociolinguists to examine a general variable like t/d deletion across many speech communities.

Within the SLX Corpus, the Sociolinguistic Variable Survey can be explored using the DASLTrans tool. The columns shown in DASLTrans are drawn directly from the original survey for each speaker, and include:

Time_Start	Start time of the utterance containing the variable
Time_End	End time of the utterance containing the variable
Speaker	The name of the speaker
File	The file name the data is from
Tkn#	Token number. This is a unique code that shows the speaker's initial(s), the file number of the speaker that the data is from and then a number that corresponds to the order in which the token was added. So, the 10th token of Adolphus from his file 01 has a tkn# of A1-10.
Category	This is the name for the category that the variable belongs to. (There are fewer categories than variables. See the Sociolinguistic Variable Survey Guide in the corpus documentation for the grouping of variables into categories.)
Variable	This is the name of the variable
Realized as	This is the realization of the variable in IPA transcription, or standard orthography for syntactic variables. The symbol "Ø" represents the absence of a form when null is on variant of a variable (i.e., copula deletion).
Example	This is the utterance surrounding the variable – the context.
Style	This it the style code for the variable in the context of the interview
Notes	This includes comments about any aspect of the variable, as well as definitions for some of the British lexical items.

Phonological/Phonetic Categories and Variables

C O N S O N A N T S

CONS – miscellaneous consonant category
(CCR) - Consonant cluster reduction (includes *t/d* deletion)
(FCD) - deletion of final consonant, especially nasals with subsequent nasalization of preceding vowel (*man* = [mæ̃]; *house* = [haw])
(FCDV) - final consonant devoicing (*boys* = [bojz̥])
(FLAP) - flapping of *t/d* (or other Cs if applicable)
(N-FLAP) - nasal flap; flapping of -*nt*- clusters as [n] or [r̃]
(ING) - *ing* vs. *in'* – [ɪŋ] or [ɪn] or [ɪŋk]
(NG-G) - deletion of /g/ after /ŋ/, *finger* [fiŋɚ]
(ING-REL) - strong release of velar nasal, esp. in –*ing* ending (talk[ɪŋᵏ])
(n-DEL) - deletion of nasal in -*nt*- clusters
(y-DEL) - deletion of /y/, especially after /n/
(hw) - voiceless labiovelar fricative *wh* = [ʍ]
(str) - palatalization and/or lip-rounding of /str/- onsets

FRIC – Fricatives
(dh) - vd interdental fricative /ð/ → [d] or [d͡ð] {stop}; [v] or [ð] {fric}
(th) - voiceless interdental fricative /θ/ → [t] or [t͡θ] {stop}; [f] or [θ] {fric}
(FRIC-STOP) - stopping of fricatives
(s-VOC) - overapplication of voicing of word-final /s/ *(us = [ʊz])*
(J-FRIC) - a fricative realization of affricates
(h-DROP) - deletion of /h/, including replacement with glottal stop. Very unusual in Scotland, except in weak form [ɪm] for *him*
(tr-FRIC) - frication of *tr*- clusters /tr/ → [č]
(PL VCS-FR) - retention of voiceless fricative in plural nouns: *wifes, knifes*

STOP – Stops
(t-GLOT) - /t/ realized as glottal stop *kitten* = [kɪʔn̩] or maybe even [kɪʔ.ʔn̩]
(d-GLOT) - Final /d/ glottalization
(GLOT) - replacement of consonants (except /t/ and /d/) with glottal stop
(ASP) - voiceless stop aspiration/unaspiration, usually aspirated as onset to stressed syllable
(IN-ASP) - lack of aspiration on initial /p, t, k/
(STOP-AFF) - affrication of stops
(STOP-FRIC) - frication of stops

LIQ – Liquids
(l-VOC) - /l/-vocalization

(l-DEL) - /l/-deletion
(r-VOC) - /r/-vocalization
(r-DEL) - /r/-deletion
(r-TRILL) - realization of /r/ as trill
(r-TAP) - realization of /r/ as [ɾ]
(LINKING-r) - in r-less dialects, realization of underlying word-final /r/ before a following vowel. *(I got done ove[r] in town)*
(INTRUSIVE-r) - in r-less dialects, overapplication of the linking-r rule, to word-final environments without underlying /r/. *(I don't know if you[r] 'eard about it)*
(r) - approximant realization of /r/ → [ɹ]
(R-LOSS) - loss of rhoticity. Scottish English is a rhotic dialect, but influence from England is putting pressure on /r/. Loss of pre-consonantal /r/ in eg. *hard, farm, form* is an indicator.

Vowels

VOWEL – general vowel category
(2STRLAX) - laxes or reduces tense secondary stressed vowels, e.g. *happy* [ˈhæˌpij] → [ˈhæˌpə] or [ˈhæˌpɪ] {happY}
(SCHWA) - realization of schwa, excluding (-er) constructions (*comma, about*) {commA}
(L-NAS) - nasalization of vowel preceding /l/
(MID-M) - monophthongal realization of mid vowels /ey/ and /ow/ as [eː] and [oː]
(HI-M) - monophthongal realization of high vowels /iy/ and /uw/ as [iː] and [uː]
(SVT) - short vowel tensing/raising and ingliding (southern drawl), particularly front vowels {These are diphthongs of sorts in that the first part of the vowel is the nucleus and longer than the second vowel or inglide}
(SVL) - Laxing of short vowels, particularly front vowels (*dead* = [dɛ̞d])

VR – vowel + /r/ sequences
(-er) - how the *–er* ending is realized (*worker, letter*, etc) [ɚ] or [ɛr] or [er] {lettER}
(ER) - realization of stressed /ɚ/ as [ɜː] or [ë̞ː] (BritEng) {NURSE}
(iyr) - realization of /iyr/ sequence as [ɪə] or [ʌə] in BritEng {NEAR}
(eyr) - realization of /eyr/ sequence as [eə] or [ɜə] in BritEng; or [ɛr] or [ær] in Southern American Eng {SQUARE}
(eyrV) - intervocalic /ey/ merged with /ɛ/ and /æ/ before /r/. (*Mary-merry-marry* merger)
(ur) - realization of /ʊr/ as [ʊə] or [ɔː] in British English {CURE}
(owr) - realization of /owr/ (*four, sport, hoarse*) {FORCE}
(ohr) - realization of /ɔr/ sequence as [or] or [ɔr] ([ɔə] in BritEng); is /ɔr/ merged with /or/? {NORTH}
(ahr) - realization of /ahr/ sequence; raises and/or backs to [ʌr] or [ɔr]; is /ar/ merged with /ɔr/?; [aə], [ɑə] or [æə] realization in BritEng; [ar] or [ær] in Northern Cities Shift {START}

VL – vowel + /l/ sequences

(FV-l) - merger of front tense and lax vowels before /l/ (*sell = sale*)

(ohl) - realization of /ɔ/ before /l/ as [ɔl], [ɒl], [ɑl] or [ɒ�l]

(al) - realization of /ɑ/ before /l/ as [ɑl], [ɒl], [al] or [ɒ�l]

(el) - realization of /ɛ/ before /l/ as [ɛl], [ɜl], [æl] or [al]

(il) - realization of /ɪ/ before /l/ as [ɪl], [il] or [əl]

(uwl) - realization of /uw/ before /l/

(ul) - realization of /ʊ/ before /l/

VFR – front vowels

(iy) - realization of /iy/ as [iː], [ɪi], or [əɪ]{FLEECE}

(i) - realization of /ɪ/ → [i], [ɪ] or [ɹ], or [ɨ] (or [ɛ] Northern Cities Shift) {KIT}

(ey) - realization of /ey/ diphthong → [eː], [ɛ͡ɪ], [ʌ͡ɪ] or [a͡ɪ]{FACE}

(e) - realization of /ɛ/ as [e], [ɛ] or lowered/centralized to [æ] or [ʌ] in Northern Cities Shift {DRESS}

(IN-EN) - pin/pen merger

(ae) - realization of /æ/ as [a], [æ], [iə] or [ɑː], includes broad-a words in BritEng {TRAP}, {BATH}

(aey) - diphthongization of /æ/ with a front upglide → [æɪ] in Southern American Eng

(ae-NAS) - tensing of short-a before nasals (*man* = [mæ̃n]) {Tense short-a represented as [æ] with diacritic indicating advanced tongue root}

(ae-TNS) - tensing of short-a before voiceless fricatives, voiced stops; Mid-Atlantic phonemic split of /æ/ into tense and lax phonemes

VBK – back vowels

(uw) - realization of /uw/ as [uw], [ɪw], [yw] etc., fronted/rounded/diphghongized {GOOSE}

(juw) - realization of /ju/ class [njɪ͡u] or is yod deleted?

(u) - realization of /ʊ/ as [u] or [ʊ] or [ʏ] or [ɪ]{FOOT}

(ow) - realization of /ow/ as [ow], [oː], [ʌw], [øw] or [ew]{GOAT}

(uh) - realization of /ʌ/ as [ʌ] or [ɔ] in Northern Cities Shift; or [ə] or [ʊ] in BritEng {STRUT}

(ONE) – use of [ɒ] in *one*-words and *once*

(a) - realization of /ɑ/ as [ɑ] or [ɑ̈] (*cot/caught* merger); [a] or [æ] in Northern Cities Shift; or [ɒ] or [ɔ] BritEng {LOT}

(oh) - realization of /ɔ/ as [ɔ], [ɒ�], [o], [ʊə] or [ʌ], [a], includes words that may have short-o before nasals and voiceless fricatives which may be part of /ɑ/ or /ɔ/ phoneme; also *cot/caught* merger {THOUGHT}

(ohw) - /ɔ/ has back upglide (e.g., [ɔwn] for *on*)

(a=oh) - *cot/caught* merger

(a~oh) - realization of /ɔ/ or /ɑ/ phoneme as [ɔ], [ɒ�], [o], [ʊə], [ʌ], [a], [ɑ], [ɑ̈] or [ɒ]; includes words that may have short-o before nasals and voiceless fricatives which may be part of either

/ɑ/ or /ɔ/ phoneme; historically from short-o (spelled with '*o*'); includes short-o before /g/, voiceless fricatives, nasals, and the words *on* and *off* {CLOTH}

DIPH – diphthongs

(ay) - realized as [aɪ], [ɑɪ], (or [a̠ɪ] in Northern Cities Shift); [əɪ] or [aɪ] in *tied/tight/tide* in ScotEng {PRICE}

(ay-C) - /ay/ centralization, usually before voiceless consonants

(ay-M) - /ay/ monophthongization in various environments (before voiced/voiceless segments, in free or checked position)

(aw) - realization of /aw/ as [æʊ], [aʊ] or [ɑʊ] {MOUTH}

(aw-C) - /aw/ centralization, usually before voiceless consonants

(aw-M) - /aw/ monophthongization in various environments (before voiced/voiceless segments, in free or checked position)

(oy) - realization of /oy/ as [ɔɪ], [oɪ] or [ʊɪ] {CHOICE}

(oy-M) - monophthongization of /oy/

S Y L L A B I C / P R O S O D I C / M O R P H O - P H O N O / M I S C E L L A N E O U S

PHON

(AUX-RED) - reduction of aux elements (*he is* → *he's*; *he was* → *he's*)

(-STR-DEL) - unstressed syllable deletion ('*fore = before*; '*til = until*; *gov'ment*)

(META) - metathisis of 2 segments (*ask* = [æks])

(STRSHFT) - stress shift to first syllable (*pólice, cóncrete*, etc)

(PROSODY) - Merseyside non-standard English falling intonation at end of intonational phrase (possibly raising or staying at medium or high pitch)

N O T E S

For vocalic variables, Wells lexical sets appear after the description in {curly brackets}.

Variables that involve an actual phoneme are named with a lower case letter to represent the phoneme involved. Variables with capital letters represent a more abstract relationship than just one phoneme. For instance, (l-VOC) is vocalization of /l/, while (J-FRIC) is fricative realization of affricates (not just /d͡ʒ/).

Grammatical Categories and Variables

ADJ – miscellaneous adjective category
(ADJ-ADV) - Lack of distinction between adjectives and their adverbial counterparts in -ly: *She sings nice.*
(ADJ-COMPSUP) - Non-standard comparative and superlative constructions: *She's more rougher; He's the most roughest.*
(ADJ-THEM) - Use of non-standard them for those: *We didn't have them books.*
(ADJ-WO) - Placing e.g. adjective after the noun, where Standard English would place it before: *I had guineas some.*

AGR – miscellaneous non-standard agreement structures
 (AGR-N-VB) - Singular verb with plural noun or plural verb with singular noun: *There's gangs everywhere; Blazers was in the region.*
(PLURAL) - Singular endings on plural nouns: *16 mile, ten pound; I don't eat much squirrel/call much cow*

DET – miscellaneous non-standard determiner usage; includes use of determiner where usually absent in StEng: *I'm afraid of the water; To get away from the wife.*
(DET-AGR) - Use of *a* where standard English has *an*: *Planted me a acre.*
(DET-DEL) - Deletion of definite or indefinite articles: *I was Ø mechanic; We would go to Ø swimming hole.*

GRAM - Misc grammatical construction that do not fit any of the following categories or variables

---

LEX – regional lexical i.e. non-standard US or British English
(LEX-BRIT) - pan-British lexical item i.e. differs from US English
(VOQ) - verb of quotation, e.g. *say, tell, go, like, be like* etc.

NEG – miscellaneous negative constructions
(NEG-AINT) - Use of non-standard *ain't* in negative constructions: *I ain't ready yet.*
(NEG-CON) - Negative concord. *We didn't have nothing like that.*
(NEG-INV) - Negative inversion: *Didn't nobody laugh.*
(NEG-NEVER) - Extension of StEng *not on any occasion* to *not on this one occasion*, or as emphatic: *He never even said thanks.*
(NEG-NOT) - *not* favored over *n't* constructions: *There's not such a thing as a square go.* (cf. *There isn't such a thing…*)

PRAG – pragmatic features
(ACH) - *ach* or *och* interjection (Scottish)

(LIKE) - use of *like* as a discourse marker
(MAN) - use of *man* as a discourse marker
(NO) - use of *no* as a tag question
(YKNOW) - use of *you know* as a discourse marker
(YSEE) - use of *you see* as a discourse marker

PREP – miscellaneous non-standard uses of prepositions
(PREP-DEL) - Deletion of prepositions: *Close Ø sixty years old.*

PRO – miscellaneous non-standard pronoun usage
(2PPL) - Existence of a second person plural pronoun: *youse, y'all, you together*
(POS-LEV) - Leveling to possessive pronouns to mine paradigm: *That's ourn, that's yourn, that's theirn.*
(PRO-ME) - Use of *me* rather than *my*. *Where's me bike?*
(PRO-REL) - Relative pronouns: *I've got an uncle that's a policeman.*
(PRO-US) - Use of *us* as a singular, i.e. to mean *me*: *Give us a kiss; He tapped us on the shoulder.*
(REFL-IO) - The objective case of personal pronouns (*me, you, her, hi*m) used reflexively as the indirect object of a verb: *I bought me a car*.
Includes (esp. Southern US) uses of indirect object pronouns where reflexives would be impossible: *I'm gonna write me a letter to the President.*
(REFL-LEV) - Leveling of reflexive pronouns to e.g. *hisself, theirselves*
(REFL-NON) - Reflexive pronouns used non-reflexively: *How's yourself today?; Joe and myself*

QUANT – miscellaneous non-standard quantifier usage
(Q-BUT) - Use of but to mean *only* or *no more than*: *And I didn't use but two hundred pound.*
(Q-MUCH) - Use of bare *much* to describe quantity: *I don't eat much ice cream.*

VERB – miscellaneous non-standard verb usage
(3PS-DEL) - Leveling of present tense forms to exclude 3PS ending: *She want it; He don't like it; He come in.*
(3PS-LEV) - Leveling of present tense forms to third person singular ending: *I likes it; We goes home.*
(BARE-PP) - Leveling of past participle of irregular verbs to simple past: *I seen, I been.*
(COP-DEL) - Copula deletion: *They all about the same.*
(DID-EMPH) - Emphatic use of did: *I never did take much castor oil.*
(DONE-PERF) - Perfective done: *I done forgot my hat.*
(DO-PAST) - Use of *done* as past tense form of main verb *do*; auxiliary *do* as *did*: *I don't know what they done; did you work last night?*
(DOUBLE-MODAL) - Combinations of *might* and *could* or *must* and *could,* etc: *I might could do it; He wouldn't could've worked.*
(-ED-REG) - Leveling of irregular past tense forms to bring them into line with regular verbs: *I drawed, I have drawed.*
(GET-PASS) - Use of *got* or *get* to form a passive: *He got beaten up.*

(GOT) - Deletion (or reduction) of auxiliary *have* in possessive *have got* constructions: *It all depends how many children you got.* Other *got* phenomena.
(HAB-BE) - Habitual be: *She always be singing.*
(INF-DEL) - Deletion of infinitival morpheme *to*: *It didn't take much Ø fertilize.*
(NAR-HISTPRES) - Use of present tense to relate narrative events in the past: *We were in the room when all of a sudden Tom walks in.*
(PAST) - Unusual or irregular past and present perfect forms: *I writ, I have writ*
(PAST-LEV) - Leveling of simple past tense form to present perfect: *I have went, I have saw*
(PRES-LEV) - Leveling of present tense form to past tense and present perfect: *I give, I give (past), I have give*
(SUBJUNCT) - Use of indicative tense markings in the subjunctive mood: *If I was raisin' a whole lot.*
(THERE+0) - No tensed verb with there: *There too many people.*
(WAS) - Leveling of simple past tense be to 1st person: *You was right; We was hungry; They was at home.*
(WERE) - Leveling of simple past tense be to 2nd and 3rd person: *I were there all day.*

WO – miscellaneous non-standard word orders
(IF-DEL) - Conditional sentence minus *if*: *One of her kids get hurt, she calls for Kay.*
(WO-CONDINV) - Inversion in conditional clauses: *And if ever you go.*
(WO-DEMFOC) - Focusing of dem pronoun: *That was me, I was left.*
(WO-ITCLEFT) - *It*-clefting. Dummy *it* with *be* followed by NP and rel clause: *That's us that's paying for her.*
(WO-LEFTDIS) - Left dislocation of initial NP, referred to later by anaphor: *My mother, she made us go to Protestant school.*
(WO-PARTXS) - Parataxis, i.e. juxtaposition of clauses without using conjunctions. Differentiate between parataxis as result of fast speech, and parataxis as a grammatical feature: *The reason for that, it didn't take them as much fertilize.*
(WO-RIGHTDIS) - Right dislocation of obj or subj NP: *They take it to heart, the football.*
(WO-WHCLEFT) - *Wh*-clefting. Split clause with one half beginning with *wh*-word: *What I been raised by was common sense.*