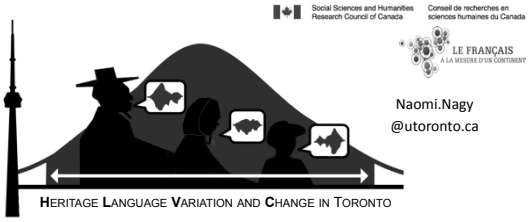


Naomi.Nagy
@utoronto.ca

HTTP://PROJECTS.CHASS.UTORONTO.CA/NGN/HLVC

Documenting Heritage Languages: Methodological Aspects

This presentation has been enabled thanks to the project entitled "*Le français à la mesure d'un continent : un patrimoine en partage*" *under the directorship of France Martineau and supported through a Major Collaborative Research Initiative grant from the Social Sciences and Humanities Research Council of Canada*. The research was supported by SSHRC SRG 410-2009-2330 to Nagy.

Long-range questions

LINGUISTIC:

- Are cross-linguistic generalizations possible about the types of features, structures, rules or constraints that are borrowed earlier and more often ?
- If so, what ?

Nagy 2009

Long-range questions

SOCIOLINGUISTIC:

- How are social factors relevant ?
- Do the same (types of) speakers lead changes in both/all their languages ? (i.e., are the same factors relevant in majority and minority languages?)
- Or do speakers choose to use one language or the other for this social "work" ?

Nagy 2009

Problem

Studies of contact-induced language variation vary widely

in terms of methods & contexts,

inhibiting generalizable findings

Solution

Consistent methods and context,

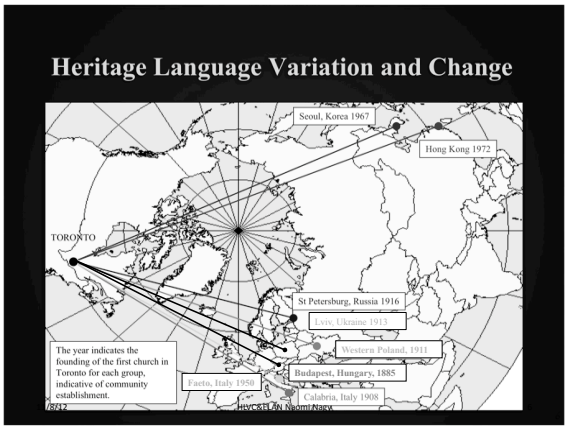
while varying pairs of languages in contact

Conditions necessary to establish the existence of contact-induced change

Paraphrased from Thomason (2001:93-94):

1. Situate the proposed change with respect to its host linguistic system
2. Identify a presumed source of the change
3. Locate structural features shared by the source and recipient languages

   Need to find several structural features that have been influenced! (Nichols 2008:361)
4. Prove that the proposed interference features were not present in the pre-contact variety
5. Prove that the proposed interference features were present in the source variety prior to contact
6. Rule out (or situate) internal motivations
7. Sociolinguists: Replace "features" with "stochastic patterns of variables"



Italian
Chinese
Cantonese
Punjabi
Portuguese
Spanish
Tagalog
Urdu
Tamil
Polish


The language quilt
TORONTO STAR
30 Dec. 2007

11/8/12 HLVC&ELAN Naomi Nagy 7

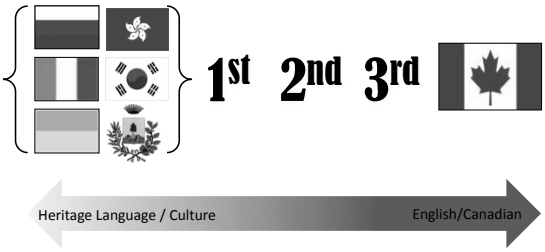
Contrasting demographics

Language	MT speakers (2006 Census)	Ethnic Origin (2006 Census)	Est. in TO	Came from
Italian	194,000	466,000	1908	Calabria
Cantonese	170,000	537,000	1951	Hong Kong
(Polish	80,095	207,495	1911	Eastern Poland)
Russian	66,000	58,505	1916	St. Petersburg, Moscow
Korean	49,000	55,000	1967	Seoul
Ukrainian	27,000	122,000	1913	Lviv
(Hungarian	20,190	53,210	1880	Budapest)
Faetar	<100?	<500?	1950	Faeto, Celle di St. Vito (Apulia Italy)

www40.statcan.ca/l01/cst01/demo12c-eng.htm

11/8/12 HLVC&ELAN Naomi Nagy 8

Expected outcome



11/8/12 HLVC&ELAN Naomi Nagy 9

Participant criteria

(Self-defined) fluent speaker of…
Cantonese
Faetar
Korean
Italian
Russian
Ukrainian

11/8/12 HLVC&ELAN Naomi Nagy 10

Generation

Speaker of…	Generation
Ukrainian	**1st:** born in/near Lviv; moved to Toronto after age 18; in Toronto 20+ years
	2nd: born in Toronto (or came from homeland before age 6); parents qualify as 1st generation
	3rd: born in Toronto; parents qualify as 2nd generation
Italian	**1st:** born in Calabria…
Russian	**1st:** born in Moscow or St. Petersburg…

11/8/12 HLVC&ELAN Naomi Nagy 11

Age group

Languages	Generation	Age
Ukrainian	**1st**: born in homeland; moved to Toronto after age 18; in Toronto 20+ years	60+
		39-59
	2nd: born in Toronto (or came from homeland < age 6); parents qualify as 1st generation	60+
		40-59
		21-39
		<21
	3rd: born in Toronto; parents qualify as 2nd generation	60+
		40-59
		21-39
		<21

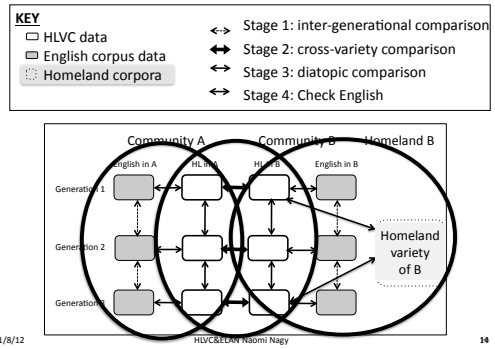
11/8/12 HLVC&ELAN Naomi Nagy 12

Sex

Languages	Generation	Age	Sex
Ukrainian	1st: born in homeland; moved to Toronto after age 18	60+	2 females
			2 males
		39-59	2 females
			2 males
Italian	" "		
Russian	" "		
Korean	" "		
Cantonese	" "		
Faetar	" "		

= 240 speakers

For every variable, 3 kinds of comparisons



Data collection methods

1. Sociolinguistic interview (~1 hour)
2. Ethnic Orientation Questionnaire
3. Picture Description Task

All conversations guided and recorded by native speakers in the heritage language

Ethnic Orientation Questionnaire

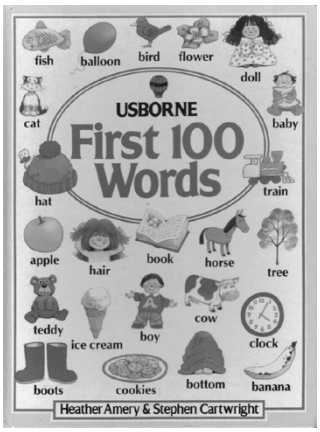
A. Ethnic identity
1. Do you think of yourself as Italian, Canadian or Italian-Canadian?
2. Are most of your friends Italian?
3. Are people in your neighbourhood Italian?...

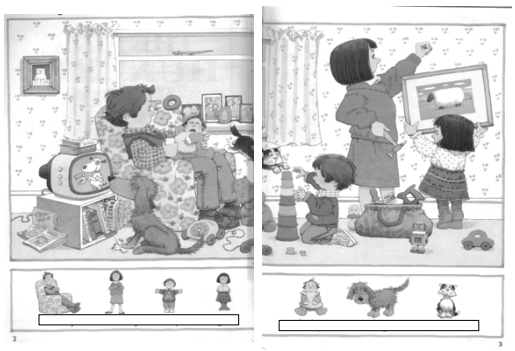
B. Language use
1. Do you speak Italian? How well? How often?
2. Where did you learn Italian? At home? In school?
3. Do you prefer to speak Italian or English?
4. Do you prefer to read and write in Italian or English? …

C. Family language choice
1. What language does your family speak when you get together?
2. What language do your parents prefer to speak?

D. Cultural heritage
E. Media preference
F. Discrimination experience

Adapted from Keefe & Padilla 1987, Hoffman & Walker 2010





Comparative Variationist Analysis

(cf. Labov 1972, Tagliamonte 2006, Walker 2010)

1. Compare rates of variant use across groups
2. Compare constraint effects across groups

Analysis by undergraduate and graduate students and a team of collaborating colleagues:

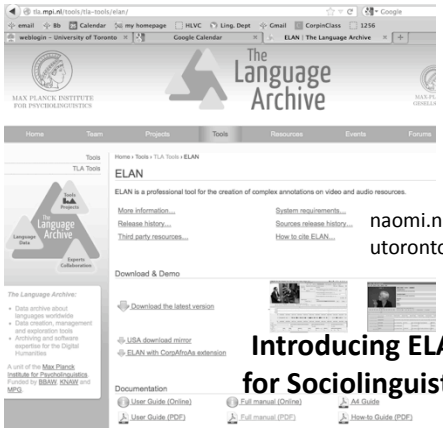
- Yoonjung Kang
- Alexei Kochetov
- James Walker

Cross-variety comparison

(Factor weights in 3 separate regression analyses)

Heritage: Hollett (2011:60), Homeland: Pustovalova (2011:22), English (Nagy et al. (2010)

Factor groups	Homeland	Generation 1	Generation 2	English
Subject continuity	same: 60 switch: 47	same: 63 switch: 42	same: 62 switch: 42	V same (but univ.)
Person & Number	sg. pl. 3: 43 77 2: 62 51 1: 38 48	3: 69 2: 67 1: 35	3: 62 2: 87 1: 40	*not sig.*
Clause type	conjoined: 65 subord.: 47 main: 41	conjoined: 72 main: 49 subord.: 42	conjoined: 71 main: 50 subord.: 35	conjoined > main > *(subord. 0%)*
Negation	neg.: 51 affirm: 50	not sig.	neg.: 67 affirm.: 47	*not sig.*
Gender (grm.)	neuter: 85 none: 52 masc.: 45 fem.: 43	neuter: 82 none: 51 fem.: 58 masc.: 42	neuter: 84 none: 48 fem.: 50 masc.: 55	*not sig.*
Age	older > younger	older > younger	older > younger	*(not examined or expected)*
Sex	male > female	n.s.	male > female	

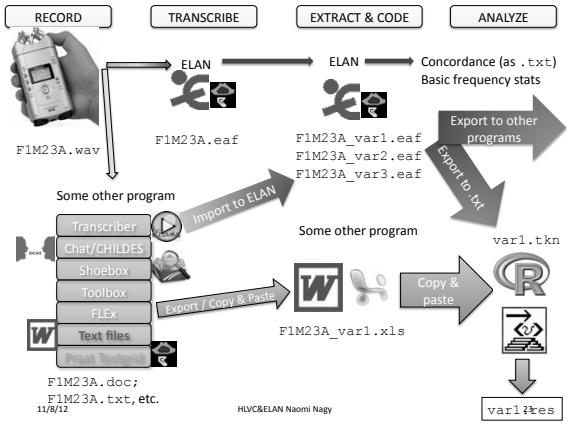


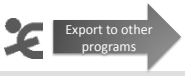
naomi.nagy@utoronto.ca

Introducing ELAN for Sociolinguistics

Helpful links

- Download ELAN & get the manual:
  - http://tla.mpi.nl/tools/tla-tools/elan/
- More info and instructions I've created:
  - http://projects.chass.utoronto.ca/ngn/HLVC **> Software**
- Coding & analysis assignment with step-by-step instructions:
  - http://individual.utoronto.ca/ngn/LIN/courses/LIN351/LIN351_project.htm
- Download Goldvarb:
  - http://individual.utoronto.ca/tagliamonte/goldvarb.htm

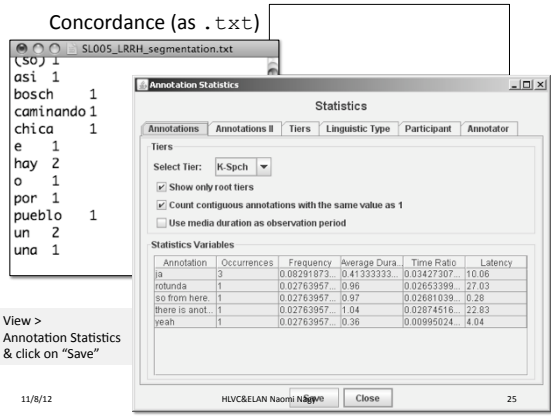




4.2.22. Exporting a document to Shoebox
4.2.23. Exporting a document to Toolbox (UTF-8)
4.2.24. Exporting a document as a tab-delimited text file
4.2.25. Exporting Tiger XML
4.2.26. Exporting CHAT files
4.2.27. Exporting traditional transcript files
4.2.28. Exporting a Praat TextGrid file
4.2.29. Exporting an alphabetical list of words
4.2.30. Exporting a part of a clip
4.2.31. Exporting a SMIL clip
4.2.32. Exporting to QuickTime Text
4.2.33. Exporting to Subtitle Text
4.2.34. Exporting ELAN's document view
4.2.35. Exporting to interlinear text
4.2.36. Exporting to HTML
4.2.37. Exporting to a Filmstrip Image
4.2.38. Exporting Multiple Files
4.2.39. Opening a wave file in Praat
4.2.40. Exporting a selection to a wave file with Praat

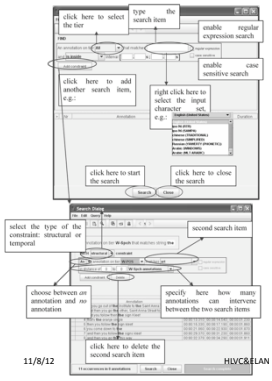
(straight from the ELAN manual)

Concordance (as `.txt`)



View > Annotation Statistics & click on "Save"

Searching

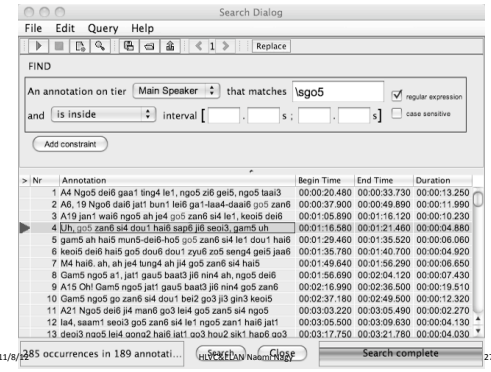


CTRL + F is very powerful

6.1.1. Advanced searching: an example

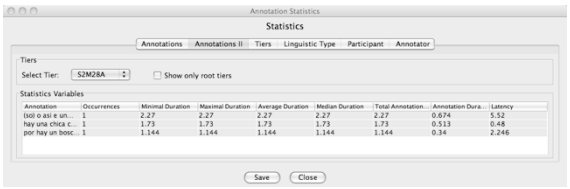
Suppose we are investigating turn taking and we want to find all switches from speaker W to speaker K that don't overlap, with gaps of at most 2 seconds. In order to find this, we fill in the search form as follows:…

Search Results



Basic statistics about
words, durations, pauses, speakers, transcribers…

can be saved as `.txt`



Getting going in ELAN

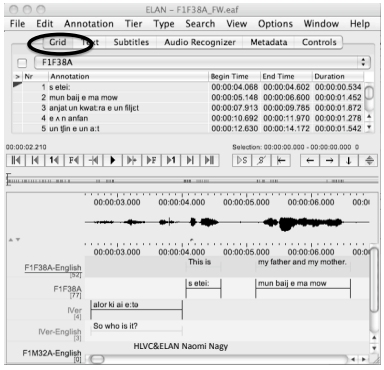
only the 1st time you run ELAN on your computer

1. Start ELAN.
2. "Open" your file from the "File" menu (or choose "New" from the "File" menu).
3. Select the .wav file you want to transcribe if prompted.
4. Choose "AUTO BACKUP > 1 minute" from the File menu.
5. Edit>Preferences>Edit Preferences > Editing > ☑ Enter key commits change in the inline box. (This lets you save an annotation just by hitting Enter before you leave the annotation box.)
6. Immediately save your file.
7. In the Edit menu, choose "Linked files," (unless you already specified the sound file to use). Select the `.wav` file for this speaker.
8. You should now see a soundwave in the center of the `.eaf` window.
   a) If you don't, Command-Click (⌘+Mouse click) on the place where the soundwave should be (it will be showing as a flat horizontal line) and then choose a big number to magnify/Vertical Zoom in by.
   b) If it is still too quiet to hear quickly, use the "Amplify" function to edit the `.wav` in Audacity.
9. Test volume. (menu > System Preferences > Sound > Output or the ◀ icon on the upper right of your monitor.)
10. In the Controls tab, adjust the speed (slider in the bottom right corner).
11. Switch to the Grid or Text tab and select the relevant tier from the pull-down menu.
12. You are ready to (transcribe or) extract and code tokens.

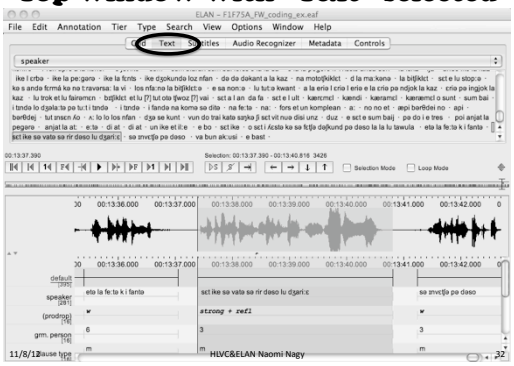
Navigating around in ELAN

- Moving around
  - Annotations are like "bookmarks" to help see where you are in the transcription and for doing text searches.
  - Clicking on an annotation in "Grid" or "Text" will take you to that part of the wave.
  - You may find the Shortcut Keys (section 7.2. of the ELAN manual) helpful.

Time-aligned Transcription

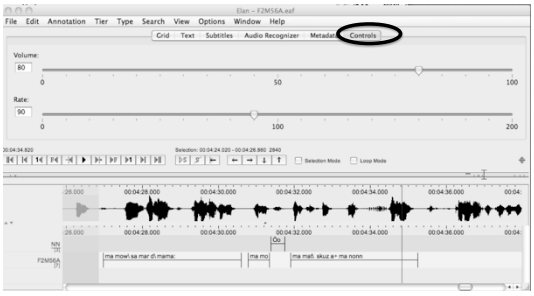


Top window with "Text" selected



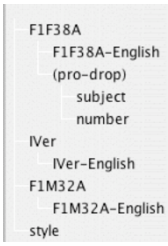
Top window with "Controls" selected

(instead of "Grid")

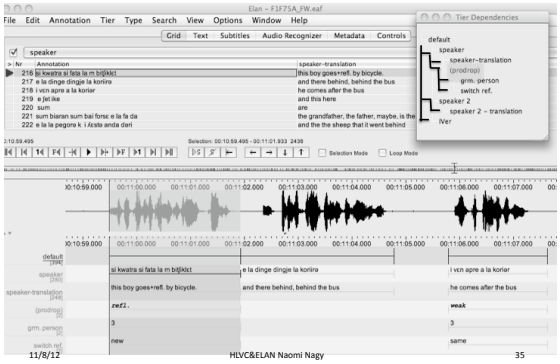


Coding & Extracting

- Linguistic and stylistic factors can be coded directly in ELAN, each on their own tier.
- Exportable data file for analysis
- **Advantages:**
  - **See all the context you need, and hear it, as you code each factor.**
  - **From ELAN, create a .txt file for multivariate analysis using R or Goldvarb or …**
  - **Can (repeatedly) revise codes in ELAN and quickly recreate the data file.**

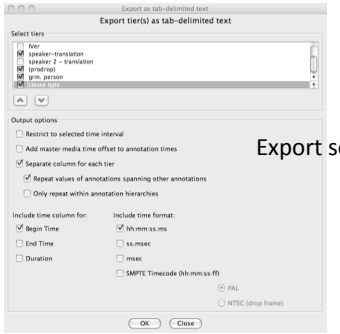


One tier for each variable



Export the data for Goldvarb analysis

File menu
> Export As…



Export settings

(Almost) Ready for 'varbing

- File > Export as… > Tab delimited Text. Make sure the filename specifies the speaker and ends in ".`txt`"
- Select all the tiers that have relevant labels or transcriptions in them.
- Select: √ Separate column for each tier
- Save as a `.txt` file.
- Open the `.txt` file in Excel (use Import, skip directly to "Finish.")

(Unsorted) coding file

Begin Time	F1F75A	F1F75A - translation	(prodrop)	grm. person	clause type
…					
10:59.5	si kwatra si fata la m bitʃiklɛt	this boy goes +refl. by bicycle.	reflex.	3	main
11:01.9	e la dinge dingje la koriirə	and there behind, behind the bus			
11:05.8	i vɛn apre a la koriər	he comes after the bus	weak	3	m
11:08.7	e ʃet ike	and this here			
11:09.8	sum	are			
11:11.4	sum biaran sum bai forsɛ e la fa da	the grandfather, the father, maybe, is the (?)			
11:14.2	e la la pegorə k i ʌɛstə anda dəri	and the the sheep that it went behind	w	3	relativ e clause (r)

dependent variable; independent variable #1; independent variable #2

(Sorted) Coding File

Begin Time - hhmm:ss.ms	F1F75A	F1F75A -translation	(prodrop)	grm. person	clause type
13:18.4	ahjat la at!	there is the cat	0	expl	m
13:30.9	ɔ sɛt i ʌɛstə kə sə fɛtʃə daʃkund ə la la la lu tawula	expl. this it is that refl. makes hide under the the table	both	3	m
11:38.8	kə s andə fɛrmá kə nə tːravərsaː la vi	that has stopped+refl. so-that they can cross the street	refl	4	rel. clause
10:59.5	si kwatra si fata la m bitʃiklɛt	this boy goes+refl. by bicycle.	reflex.	3	main
11:36.0	sɛt ə lu stopːə	this is the stop-sign	strong	3	m
13:37.4	sɛt ike sə vatə sə rir dəso lu ʌɛriɛ	this here refl. (?) under the (?)	strong + refl	3	m
11:14.2	e la la pegorə k and the the sheep that i ʌɛstə and dəri		w	3	r
12:35.7	i tində lo ʌɛplaltə pa i tində			3	m
12:37.8	i tində	he holds	w	3	m

If using Goldvarb, search & replace to create 1-character codes

Goldvarbify it in Excel

formula to create token strings

(pro-drop)	grm. person	clause type	="("&D4&E4&F4&" "&B4&" "&B3
0	e	m	(0em anj at la atː F1F75A
b	3	m	(b3m o sɛt i ʌɛstə kə sə fɛtʃə … F1F75
r	4	r	(r4r kə s andə fɛrmá kə nə tːravərsaː la vi F1F75
r	3	m	(r3m si kwatra si fata la m bitʃiklɛt F1F75
s	3	m	(s3m sɛt e lu stopːə F1F75

dependent variable; independent variable #1; independent variable #2; copy 1 column to a Goldvarb tkn file; useful extra information in your .tkn file

Using an Excel formula to create GV-ready tokens

1. Type this formula in a blank cell in the row of your first token.

C	D	E	F	G	H	I
Newscaster	**word**	**(ING)**	**P.O.S.**	**# sylls.**	**social co**	**="("&E2&F2&G2&H2&" "&D2&" "&C2**
tell us about the terrifying experienc	terrifying	g	a	4	M3M	(ga4M3M terrifying tell us about the terrifying
Telly, how are you doing today?	doing	n	v	2	M3M	(nv2M3M doing Telly, how are you doing today
and how's your son doing?	doing	g	v	2	M3M	(gv2M3M doing and how's your son doing?
(Thanks) so much for talking to us	talking	g	v	2	M3M	(gv2M3M talking (Thanks) so much for talking

2. Use Fill Down (from the Edit menu) to copy it to all the other tokens' rows

3. Make sure that the formula is referring to the correct rows and cells in your spreadsheet.

4. Copy *only* that column to a new .Tkn file in Goldvarb.

Other good things to learn to use

- Vertical zoom & horizontal zoom in the .wav window (Control + click)
- Navigate with "Grid" and "Text" (choose relevant tier from pull-down menu)
- Control speed and volume of playback in "Controls"
- "Loop mode" for playback
- List of "shortcuts" from the View menu (key combos)
- Change order of tiers
- Delete annotation (select it, Option+D)
- Change size of annotation (select it, then Option+Drag edge with mouse)
- Templates to set up tiers for many files

Create spiffy WWW examples

CuPED

Customizable Presentation of ELAN Documents

Users' Manual
Draft · February 2010

For CuPED version 0.3.14
http://sweet.artsrn.ualberta.ca/cdcox/cuped/

CuPED turns an .eaf into .html

Scrolls and highlights each line of text as it plays

More about ELAN

http://tla.mpi.nl/tools/tla-tools/elan/

- You can add an unlimited number of annotations to audio and/or video streams.
  - An annotation can be a sentence, word or gloss, a comment, translation or a description of any feature observed in the media.
  - Annotations can be created on multiple layers, called tiers.
  - Tiers can be hierarchically interconnected.
- An annotation can either be time-aligned to the media or it can refer to other existing annotations.
- The textual content of annotations is always in Unicode and the transcription is stored in an XML format.
- ELAN provides several different views of the annotations, each view is connected and synchronized to the media.
- ELAN delegates media playback to an existing media framework, like Windows Media Player, QuickTime or JMF (Java Media Framework). As a result a wide variety of audio and video formats is supported and high performance media playback can be achieved.
- ELAN is written in the Java programming language and the sources are available for non-commercial use. It runs on Windows, Mac OS X and Linux.

References

Farley, C. & D. Lister. 2007. *Greater Toronto's language quilt*. Toronto Star. Dec. 30, 2007.

Hoffman, M. & J. Walker. 2010. Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *LVC* 22:37-67.

Hollett, Meghan. 2010. Heritage Russian in Toronto: Generational change and **subject omission**. CVC IV, Memorial University of Newfoundland.

Keefe, S. & A. Padilla. 1987. *Chicano Ethnicity*. Albuquerque, NM: UNM Press.

Nagy, N. 2009. *Heritage Language Variation and Chan*ge. http://projects.chass.utoronto.ca/ngn/HLVC.

Nagy, N., N. Aghdasi, D. Denis, & A. Motut. 2011. Pro-drop in Heritage Languages: A cross-linguistic study of contact-induced change. *Penn Working Papers in Linguistics* 17.2.

Nichols, J. 2008. Universals and Diachrony: Some Observations. In J. Good, ed. *Linguistic Universals and Language Change*. Oxford.

Pustovalova, E. 2011. Null Subject Variation in the Russian spoken language (based on the materials of the Russian National Corpus). National Research University – Higher School of Economics *ms*.

Thomason, S. G. 2001. *Language contact: An introduction.* Edinburgh University Press.