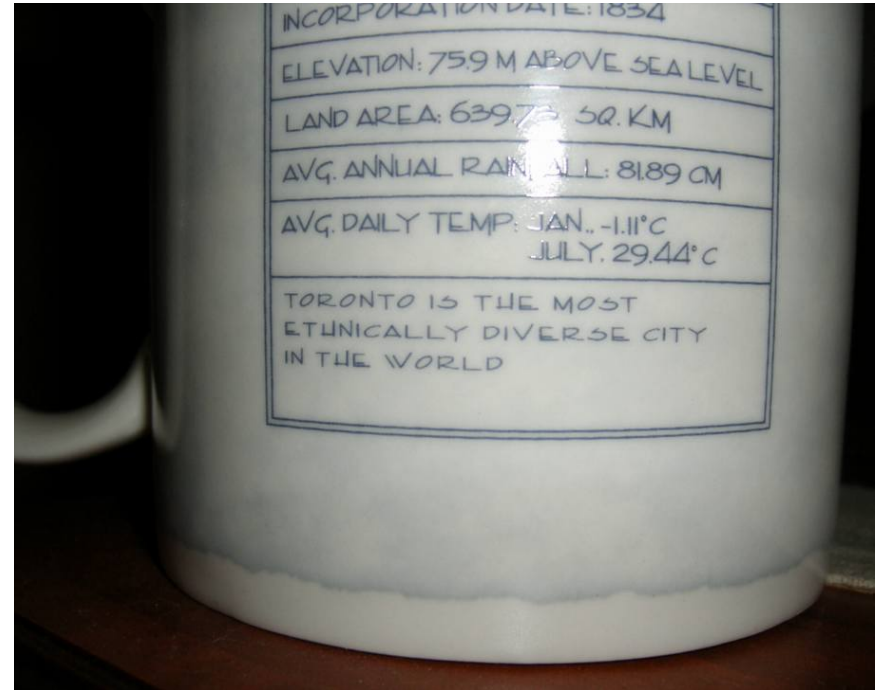


# HLVC in TO: A New Project

Naomi Nagy, Yoonjung Kang,  
Alexei Kochetov, James Walker



Heritage Language Workshop  
May 7, 2009

# Innovations

- Many languages, 1 Methodology, 1 City
- “Special” context: a city where multilingualism is the norm
  - Does this mean less pressure to switch?
  - Does the complexity mean more pressure to switch?
  - Does this mean more languages for contact?
- Sharing data across languages, fields, the public

# The language quilt

By Catherine Farley and Damian Lister/TORONTO STAR

English is still, by far, the first language across Greater Toronto. But strip away that blanket of dominance and a colourful patchwork emerges, showing where newcomers from around the world chose to settle. The map shows the most prevalent mother tongue after English in more than 1,000 neighbourhoods across the GTA, as revealed by a Star analysis of 2006 census data

## Some discoveries

- English is the second language in **47** of the GTA's **1,076** census tracts
- English is in third place in **7** tracts in Agincourt, on Toronto's northern edge
- In **57** tracts, **70%** or more of the population has a non-English mother tongue
- The preponderance of English as a mother tongue is **90%** or more in **42** census tracts
- In **200** tracts, more than **30** distinct mother tongues are spoken by 15 or more people — the minimum number of speakers required for a language to count in the census
- In **13** tracts, there are more than **40** mother tongues

**56%** of the 5.4 million GTA residents count English as their mother tongue. Here are the next most prevalent:

## Top 10 mother tongues (after English)



## Behind the mosaic

**Tight knit:** The Greek community, comprising 47,305 speakers, is one of the GTA's most compact with Greek placing second behind English in 20 census tracts in central Toronto. Estonian, with 4,200 speakers, is one of the smallest communities, but it ranks No. 2 in one rural community in North Durham, ahead of the German and French contingents.

**Loose knit:** Mother tongues with wide distribution rarely get into second place. Although a sizeable community in numbers (33,831), Hindi has no dominant census tracts and Hungarian (21,325) has just one. In contrast, Vietnamese, with 39,740 speakers,

## Mother tongue mix



## KEY TO MOTHER TONGUES

Number of speakers in the GTA

### Western Europe, Americas

- English 3,029,955 (56%)
- Italian 189,775 (3.5%)
- Portuguese 190,255 (3.2%)
- Spanish 190,225 (3.2%)
- French 65,435 (1.2%)
- Greek 47,305 (0.9%)
- German 42,955 (0.8%)
- Dutch 13,460 (0.2%)

### East Asia

- Chinese (no language specified) 375,405 (6.9%)
- Cantonese 167,305 (3.1%) (Hong Kong)
- Tagalog 100,485 (1.9%) (Philippines)
- Mandarin 63,140 (1.2%) (China)
- Korean 48,385 (0.9%)
- Vietnamese 45,725 (0.8%)

### South Asia

- Punjabi 133,190 (2.5%) (Sikhs in Pakistan, India)
- Urdu 98,300 (1.8%) (Pakistan)
- Sanskrit 84,580 (1.6%) (South India, Sri Lanka)
- Gujarati 54,710 (1.0%) (West India)
- Bengali 25,240 (0.5%) (Bangladesh)

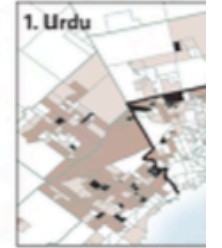
### Eastern Europe & Middle East

- Polish 83,950 (1.6%)
- Russian 45,730 (0.8%)
- Farsi 65,025 (1.2%) (Iran)
- Arabic 56,875 (1.0%) (Middle East, N. Africa)
- Ukrainian 29,040 (0.5%)

No data in 8 census tracts

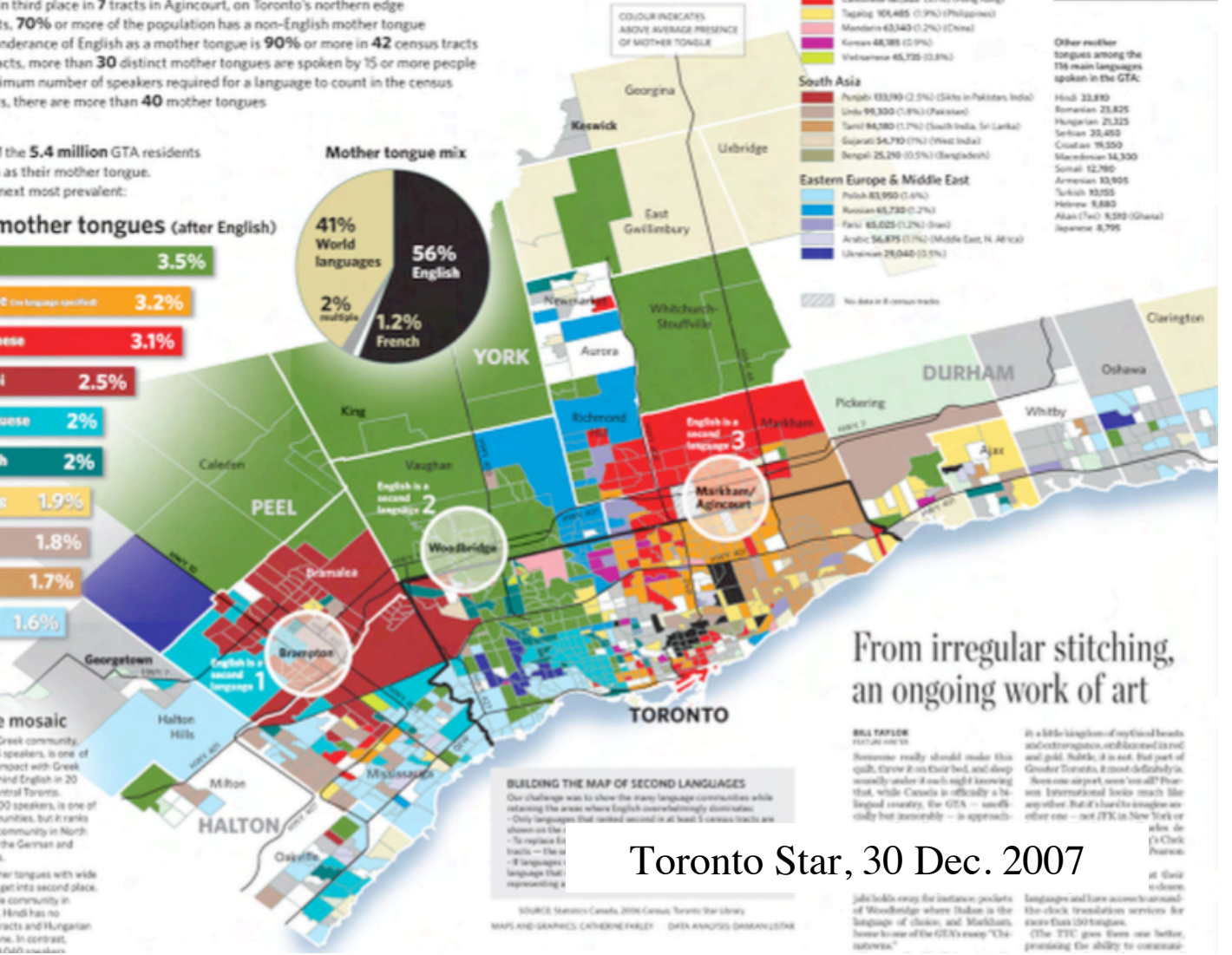
## Top 8 new voices

The new languages that will shape it show where the 452,000 immigrants



Other mother tongues among the 116 main languages spoken in the GTA:

- Hindi 33,830
- Romanian 23,825
- Hungarian 21,325
- Serbian 20,480
- Croatian 19,500
- Macedonian 14,330
- Somali 12,790
- Armenian 10,905
- Turkish 10,155
- Hebrew 9,880
- Albanian 9,310 (Ghana)
- Japanese 8,795



## From irregular stitching, an ongoing work of art

BILL TAYLOR  
PHOTOGRAPH

Someone really should make this quilt. Cover it in their bed, and sleep soundly under it each night knowing that, while Canada is officially a bilingual country, the GTA — unofficially but intensely — is approaching

it's a little kingdom of mythical beasts and extraterrestrials, and located in red and gold. Subtle, it is not. But part of Greater Toronto, it most definitely is. One one side, you've got Pearson International which is much like any other. But it's hard to imagine another one — not JFK in New York or even de Gaulle in Paris.

at their

languages and have access to around-the-clock translation services for more than 100 languages. (The TTC gives them one button, providing the ability to communicate

## BUILDING THE MAP OF SECOND LANGUAGES

Our challenge was to show the many language communities while retaining the areas where English overwhelmingly dominates.

- Only languages that ranked second in at least 5 census tracts are shown on the map.
- To replace English in the map — the map is a language quilt.
- Language that is representing a

SOURCE: Statistics Canada, 2006 Census; Toronto Star library  
MAPS AND GRAPHICS: CATHERINE FARLEY DATA ANALYSIS: DAMIAN LISTER

Toronto Star, 30 Dec. 2007

# Languages we'll look at

- Korean
- Russian
- Italian
- Faetar
- Greek? Ukrainian? Urdu? Cantonese Serbo-Croatian? Estonian? Tagalog?

# Theoretical Questions

- LINGUISTIC: Are cross-linguistic generalizations possible about the types of features, structures, rules or constraints that are borrowed earlier and more often in this type of contact situation? If so, what do they include? How are social factors relevant?
- SOCIAL: Do the same (types of) speakers lead changes in their HL and in English? Is the propensity to be a “leader” in language change an inherent trait, or do speakers choose to use one language or the other for this social “work”?

# Methods

1. Establish communities of interest
2. Interview & record speakers
3. Transcribe.
4. Analyze variables within each language
5. Compare trends across languages
6. Develop a generalized framework

# 1. Establish communities of interest

- Language selection
  - Different histories
  - Different language types
  - Our linguistic expertise
- Speaker selection
  - Use fieldworkers' networks
  - Distribution by:
    - age
    - generation
    - class/education
    - sex
    - fluency ?

# The Fluency Issue

- to use for selecting a range of speakers?
- or just to index them later?
- Either way we need a metric.
  - To avoid circularity
  - does speech rate work?
  - what predictions do we have?



## 2. Interview & Record

- Digital recording
- Conversational Interview
- Ethnicity and Language Questionnaire
- Linguistically controlled elicitations

# Ethnic Orientation Questionnaire

(excerpt)

## Ethnicity Questionnaire

Ethnic identification:

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?
4. Are the people you work with Italian?
5. When you were growing up, were the kids in your school Italian? Were your friends? The kids in your neighbourhood?

Language:

1. Do you speak Italian? How well? How often?  
If no: Can you understand Italian?
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?

Do you read Italian magazines and newspapers? — Which ones?

All responses  
to be  
quantified for  
comparisons

# Ethnic Orientation Questionnaire

## Topics

- Ethnic Identification
- Language
- Language Choice
- Cultural Heritage
- Parents
- Partner
- Culture
- Discrimination

# Enclave Status

- [+ Enclave]
  - Oriented toward ethnic social networks, language, community activities
- [- Enclave]
  - Fewer/weaker ties to ethnic group
  - More diverse social networks

### 3. Transcription

ELAN  
makes  
time-  
aligned  
transcript  
ions

-->

Transcrip  
tion can  
be broad

The screenshot displays the ELAN software interface with the following components labeled:

- Menu Bar:** File, Edit, Annotation, Tier, Type, Search, View, Options, Window, Help.
- Grid, Text, Subtitles, Controls:** A row of buttons for switching between different annotation views.
- Media Player controls:** Includes Volume and Rate sliders.
- Video Viewer:** A window showing a video of two people in a room.
- Waveform Viewer:** A view showing the audio waveform.
- Annotation Density Viewer:** A view showing the density of annotations over time.
- Timeline Viewer:** A view showing the transcription tiers, including K-Spch, W-Spch, W-Words, W-POS, W-IPA, W-RGU, W-RGph, and W-RGMe.

The interface also shows a timeline with time markers (e.g., 00:00:11.683, 00:00:12.000, 00:00:13.000) and a selection bar indicating the current time range (00:00:10.060 - 00:00:10.460).

# ELAN

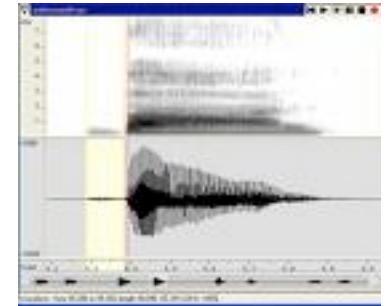
“ELAN is a professional tool for  
the creation of complex  
annotations on video and audio  
resources”

[www.lat-mpi.eu/tools/elan/](http://www.lat-mpi.eu/tools/elan/)

## 4. Analysis - Language internal

- Choose some variables
- Mark/code what we need in more detail
- Look at correlations to social characteristics of groups of speakers
- after first “factoring out” linguistic factor effects

# Korean



[www.iub.edu/.../manual/  
wsman157/wsman08.htm](http://www.iub.edu/.../manual/wsman157/wsman08.htm)

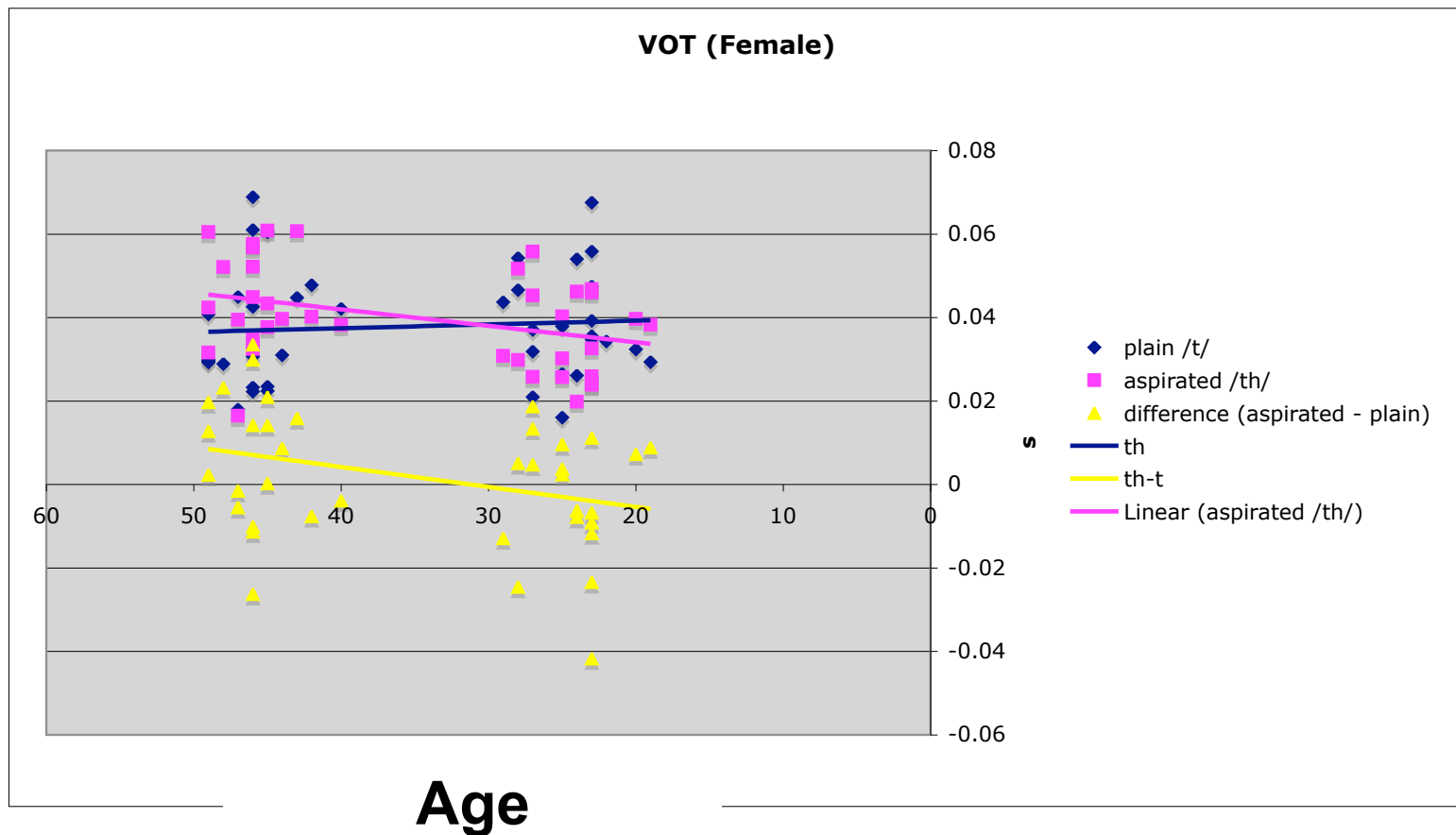
- Variable
  - Laryngeal contrast in obstruents
    - Recent shift in Korea: merger of VOT between plain and aspirated stops phrase-initially --> f0 became the primary cue (cf. Silva 2006, Wright 2006, Park & Iverson 2008)
    - Does Heritage Korean
      - show a parallel development?
      - show an independent development (due to bilingualism with English, cf. Kang & Guion 2006) or not?
- Method
  - VOT values sensitive to prosodic structure (Jun 1993, 1995): Prosody should be controlled to allow for direct comparison
    - sentence reading task (for those who can read)
    - match NIKL data base
      - 120 Seoul speakers balanced for gender & age
      - reading passages from novels



# NIKL database (2003)

120 Seoul speakers balanced for gender & age, reading passages from novels  
(930 sentences in all)

(Thanks to Pheba Ninan for her assistance with the acoustic analysis)



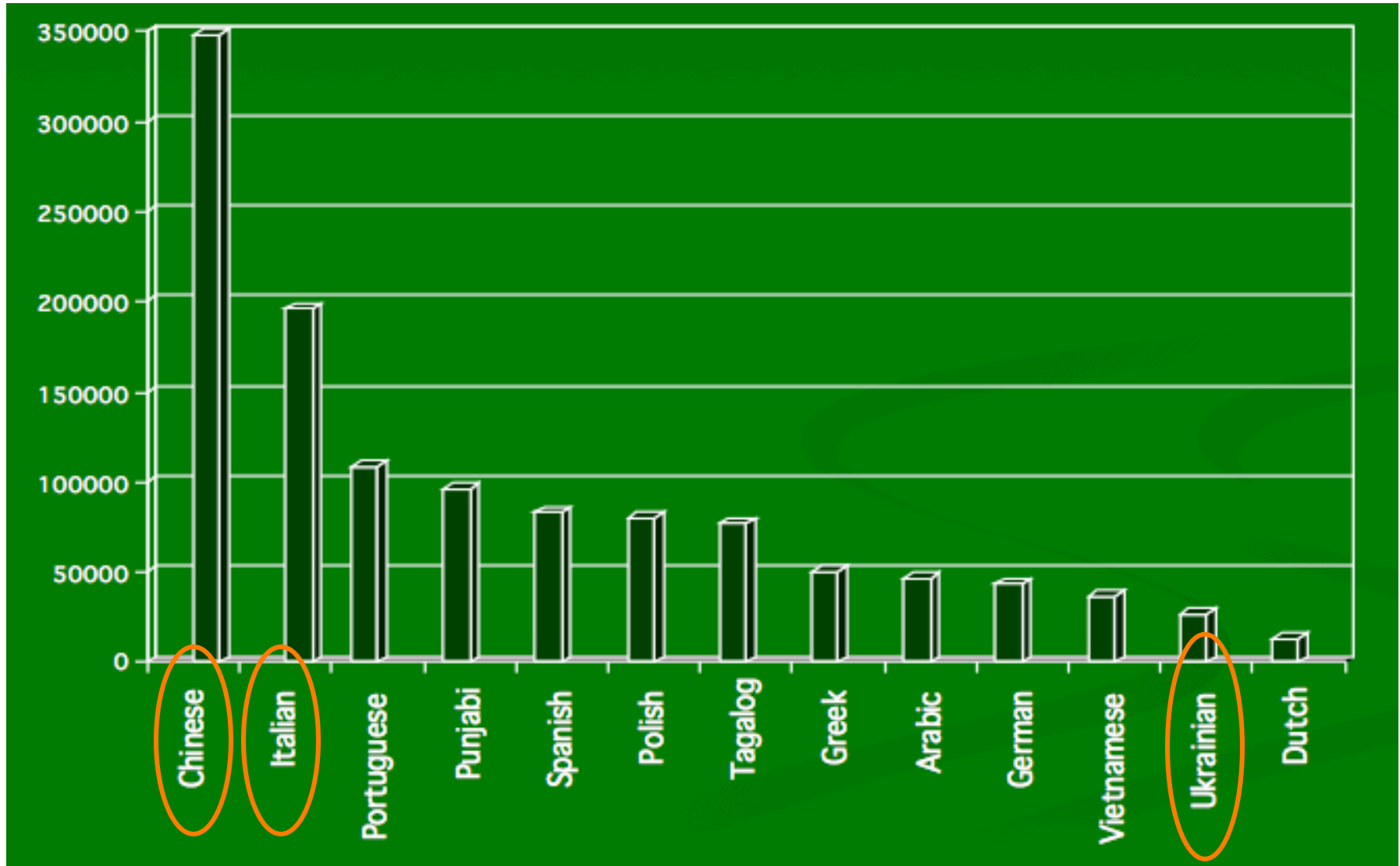
# Korean (continued)

- Morpho-phonological alternations
  - Verb paradigms in American Korean: “” based on “intimate”-level ending (Choi 2003, Kang 2006)
  - Nouns in GTA Korean (Kang and Park, next session)
- Honorifics: Morphological & Lexical
  - Elaborate honorific system (cf. Sohn 2001)
    - E.g., different verb endings for 6 speech levels: deferential, polite, blunt, familiar, intimate, plain
  - Simplification reported in homeland Korean
  - Likely more so in Heritage Korean
- Case marker drop, misplacement, doubling etc.
- Anything else?

# Variables: Cross-linguistic

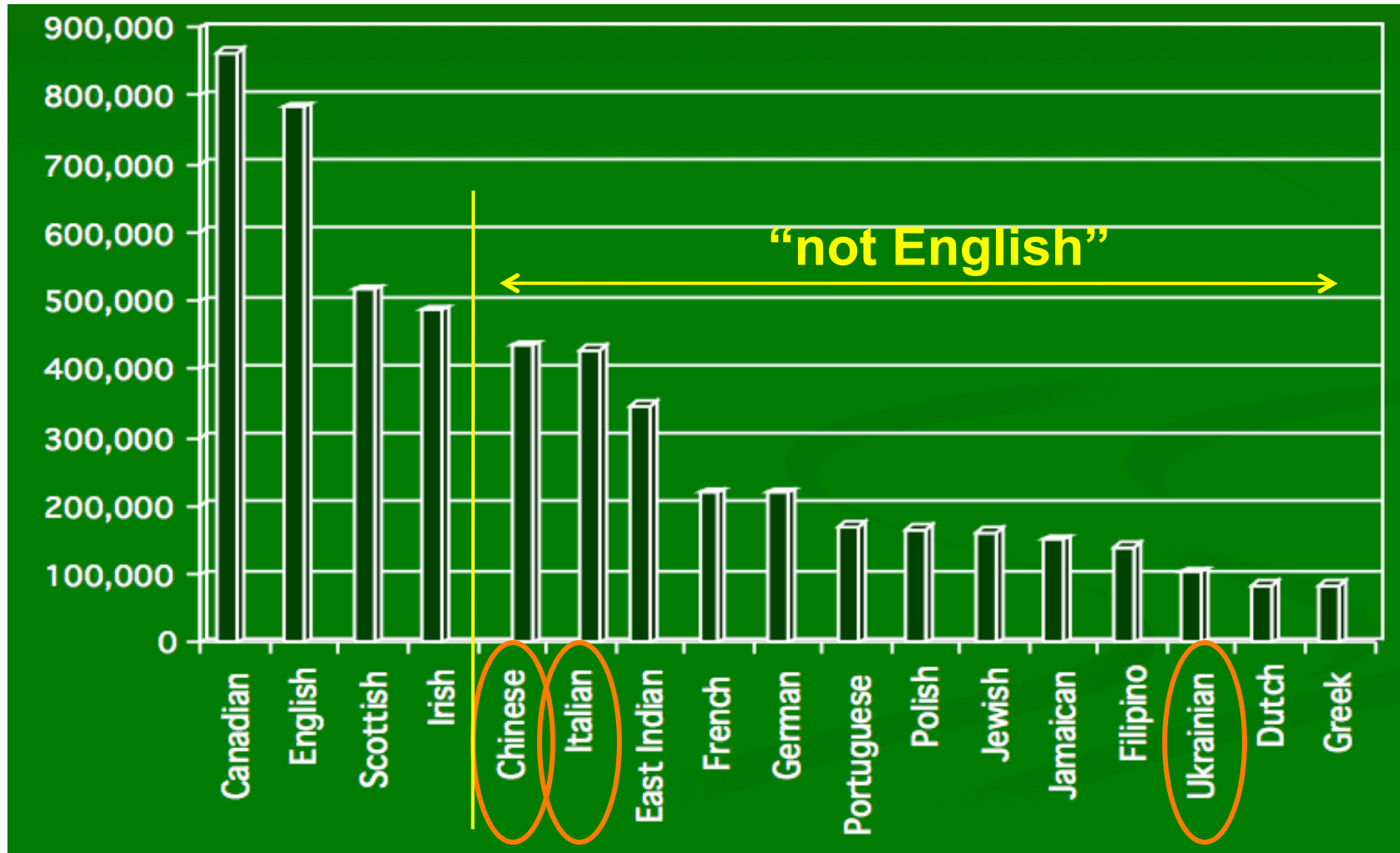
- Looking for possibility of English-contact influence across all languages
- Some ideas are:
  - Phonetic
    - VOT
  - Phonological
    - Word-final C deletion - universal? Compare conditioning factors to Home Country variety.
  - Morphological
    - Case and gender marking
    - Pro-drop (Variable null subject presence)
  - Syntactic
    - Word order
  - Lexical
    - *tu/vous* (deference/solidarity) distinctions
    - Use of borrowings from English, other
    - Use of archaic (in Home country variety) words
  - Other ideas?

# Toronto Population by Mother Tongue (Non-Official Languages) Census of Canada 2001



# Toronto by Ethnic Origin

(Census of Canada 2001)



# Retention of MT as Home Language in Ontario

(King 1998: 407)

