The Golden Age of Speech and Language Science

Mark Liberman
University of Pennsylvania

http://ling.upenn.edu/~myl
Abstract:

From the perspective of a linguist, today's vast archives of digital text and speech, along with new analysis techniques and inexpensive computation, look like a wonderful new scientific instrument, a modern equivalent of the 17th-century invention of the telescope and microscope. We can now observe linguistic patterns in space, time, and cultural context, on a scale three to six orders of magnitude greater than in the past, and simultaneously in much greater detail than before. Scientific use of these new instruments remains mainly potential, but the next decade is likely to be a new "golden age" of research. This talk will discuss some of the barriers to be overcome, present some successful examples, and speculate about future directions.
According to the National Academy of Sciences:

We see that the computer has opened up to linguists a host of challenges, partial insights, and potentialities. We believe these can be aptly compared with the challenges, problems, and insights of particle physics. Certainly, language is second to no phenomenon in importance. And the tools of computational linguistics are considerably less costly than the multibillion-volt accelerators of particle physics. The new linguistics presents an attractive as well as an extremely important challenge.

There is every reason to believe that facing up to this challenge will ultimately lead to important contributions in many fields.

Language and Machines: Computers in Translation and Linguistics
Report by the Automatic Language Processing Advisory Committee (ALPAC), National Academy of Sciences
That’s what they all say . . .

Progress in any science depends on a combination of improved observation, measurement, and techniques. The cheap computing of the past two decades means there has been a tremendous increase in the availability of economic data and huge strides in econometric techniques. As a result, economics stands at the verge of a golden age of discovery.

Two wrinkles

(1) ALPAC ’s main recommendation was to de-fund Machine Translation research.

... wait, what?

(2) And, the ALPAC report came out in 1966 (!) so 44 years later, where’s the QCD of linguistics?
The plan vs. the reality

• ALPAC ‘s idea:
  1. computers → new language science
  2. language science → language engineering

• What actually happened:
  1. computers → new language engineering
  2. engineering → new language science (???)
Parenthesis: What is linguistics?

The term *linguistics* is ambiguous:

1. “rational inquiry into questions of speech and language”
2. “the institutions of academic linguistics”
3. “what people identified as linguists do”

In sense (1), most linguistics today is not done by linguists, but rather by computer scientists, electrical engineers, psychologists, neurologists, anthropologists, biologists, lexicographers, classicists, etc.

Still, . . .
Speech and language technology
1966-2010

• Despite the “funding winter” that followed ALPAC (and Pierce 1969), there has been an extraordinary flowering of technologies for dealing with digital speech and text.

• As more and more of the world’s intellectual and social life is mediated by digital networks, we can expect this to continue.
The science of speech 1966-2010

• Plenty of computer use
  – minicomputers in the 1960s
  – micro- and super-computers in the 1980s
  – ubiquitous laptops today

• Applications:
  – replaced tape splicing
  – replaced sound spectrograph
  – easier pitch tracking, formant tracking
  – more convenient statistical analysis
  – and so on

• BUT…
No phonetic quantum mechanics

• Great speech science by smart people
• But surprisingly little change
  – in style and scale of research 1966-2009
  – in scientific questions about speech
  – in the rate of progress compared to 1946-1966 (the first golden age of phonetics)
    ...at least on the acoustic analysis side.
• Peterson & Barney 1951
  – data is still relevant
  – many contemporary publications are similar in style and scale
The science of language 1966-2010

• Despite ALPAC’s 1966 predictions, computers have not had a large impact on “the new linguistics” (in the sense of what “linguists” do).

• This has begun to change over the past decade, at least in some sub-disciplines.

• But for the most part, the effects are still in the future tense.
What went wrong?

• There are still many unmet challenges,
  – not enough new insights,
  – and the scientific potentialities of 1966 are still mostly potential.

• Is this just cultural conservatism?

• No. (1966-era) Computers were not enough: we also need
  – adequate accessible digital data
  – tools for large-scale automated analysis
  – applicable research paradigms

• Now: we have (at least) two out of three . . .
Why 2010 is like 1610

... at least a little ...

• Telescope: invented 1608
  Galileo 1609, Kepler 1611, Newton 1668

• Microscope: invented 1590
  Hooke 1665, Leeuwenhoek 1674

Instruments that opened new worlds to view
Why 2010 is like 1610

Science needs theory -- but

“Sometimes you can observe a lot just by watching”

-Yogi Berra
Breakfast experiments

• Our “telescope and microscope” are
  – Easily available collections of speech and text
  – Computer algorithms for
    • analyzing speech and text
    • aligning speech and text
    • collecting, displaying, analyzing statistics
• When we point these new instruments in almost any direction, we see interesting new things
• This is so easy and fast that we can often do an “experiment” on a laptop over breakfast.
These quick looks are not a substitute for serious research.

But they illustrate the power of our new tools, and allow us to explore interesting new directions quickly.

(All of the cited Breakfast Experiments™ were published in Language Log)
Six One-Hour Explorations

• Do Japanese speakers show more gender polarization in pitch than American speakers?
• Do American women talk more (and faster) than men?
• How does word duration vary with phrase position?
• How does local speaking rate vary in the course of a conversation?
• How does disfluency vary with sex and age?
• “you know”/”I mean” ratio over the lifespan
One-hour exploration #1

• Gender polarization in conversational speech
• Question: are Japanese men and women more polarized (more different) in pitch than Americans or Europeans?
• Method:
  – Pitch-track published telephone conversations
  – LDC “Call Home” publications for Japanese, U.S. English, German
    • about 100 conversations per language
  – Compare quantiles of pooled values
    (about 2 million numbers per sex/culture combination)
• Answer: yes, apparently so.
F0 quantiles for Japanese (red), English (blue), German (black)
Male (M) & Female (F) speakers

Data from CallHome M/F conversations; about 1M F0 values per category.

11/12/2010  Groningen: The Golden Age
As usual, more questions:

• Other cultures and languages
• Effects of speaker’s age
• Effects of relationship between speakers, nature of discussion
• Formal vs. conversational speech
• Effects of social class, region
One-hour exploration #2a

• Sex differences in conversational word counts
• Question: Do women talk more than men?
• Method: Count words in “Fisher” transcripts
  – Conversational telephone speech
    • Collected by LDC in 2003
    • 5,850 ten-minute conversations
      – 2,368 between two women
      – 1,910 one woman, one man
      – 1,572 between two men

• Answer: No.
Female vs. Male Word Counts, Fisher 2003
(all conversations)

proportion of speakers

0.0000  0.0004  0.0008  0.0012

0  500  1000  1500  2000

total words (bins of 100)
Female vs. Male Word Counts, Fisher 2003
(mixed-sex conversations only)
One-hour experiment #2b

• Sex differences in conversational speaking rates
• Question: Do women talk faster than men?
• Method: Words and speaking times in Fisher 2003
• Answer: No.
Speech rates in Fisher English 2003

(11,700 conversational sides; mean speaking rate=173 wpm, sd=27)
(Male mean 174.3, female 172.6: difference 1.7, effect size d=0.06)
One-Hour Experiment #3

• Phrasal modulation of speaking rate
  – “final lengthening” a well-established effect
  – first observed by Abbé J.-P. Rousselot ~1870
  – other phrase-position effects are less clear

• What is a “phrase”? 
  – A syntactic unit?
  – A unit of information structure?
  – A unit of speech production?

• Method: word duration by position in “pause group” (stretch of speech without internal silence >100 msec)

• Data: Switchboard corpus

• Result: Amazingly regular (average) pattern
Data from Switchboard; phrases defined by silent pauses
(Yuan, Liberman & Cieri, ICSLP 2006)
One-hour experiment #4

• How does speaking rate reflect the ebb and flow of a conversation?

• Method: word- or syllable-count in moving window over time-aligned transcripts

• Result: suggestive pictures
One-hour experiment #5

• How does disfluency vary with sex and age?
• Method: count “filled pauses” in transcripts of U.S. English conversations by demographic categories of speakers
• Result: systematic but unexpected interaction
'Uh' by sex and age

Speaker Age

uh/the ratio

20-39 40-59 60-69

M M M

F F F
'Um' by sex and age

![Graph showing the frequency of the word 'Um' by sex and age.]
One-hour experiment #6

The News Editor at Psychology Today wrote to me:

Sometimes I wonder if there are underlying personality differences between people who punctuate (litter?) their speech with "you know" versus those who use "I mean" more frequently. Any hunch on that?

I didn’t have any hunches, and there didn’t seem to be anything relevant in the literature. But I did have access to an indexed copy of the the 14,137 conversations (26,151,602 words) in the LDC’s English-language conversational speech corpora.

(...and so do you!)
And there’s demographic data for (almost) all speakers. So I checked:

<table>
<thead>
<tr>
<th>Age Group</th>
<th>&quot;you know&quot;</th>
<th>&quot;I mean&quot;</th>
<th>&quot;you know&quot;/&quot;I mean&quot; ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-39</td>
<td>58,364</td>
<td>24,478</td>
<td>2.38</td>
</tr>
<tr>
<td>40-59</td>
<td>278,099</td>
<td>73,211</td>
<td>3.80</td>
</tr>
<tr>
<td>60+</td>
<td>33,477</td>
<td>7,518</td>
<td>4.45</td>
</tr>
</tbody>
</table>

Elapsed time: 6 queries + 3 ratio calculations = 5 minutes
What about the effect of years of education?

<table>
<thead>
<tr>
<th></th>
<th>&quot;you know&quot;</th>
<th>&quot;I mean&quot;</th>
<th>&quot;you know&quot;/&quot;I mean&quot; ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>High school</td>
<td>2,608</td>
<td>408</td>
<td>6.39</td>
</tr>
<tr>
<td>College</td>
<td>191,088</td>
<td>51,143</td>
<td>3.72</td>
</tr>
<tr>
<td>Post-graduate</td>
<td>167,893</td>
<td>51,389</td>
<td>3.27</td>
</tr>
</tbody>
</table>

(Caveat: High-school-only group was small, and perhaps mainly older...)
Sex differences?

<table>
<thead>
<tr>
<th></th>
<th>&quot;you know&quot;</th>
<th>&quot;I mean&quot;</th>
<th>&quot;you know&quot;/&quot;I mean&quot; ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women</td>
<td>198,086</td>
<td>51,689</td>
<td>3.83</td>
</tr>
<tr>
<td>Men</td>
<td>173,321</td>
<td>53,892</td>
<td>3.22</td>
</tr>
</tbody>
</table>

Elapsed time:
15 minutes for queries, 45 minutes to write it up

("I mean, you know", Language Log, 8/19/2007)
My conclusion

Maybe greater use of "I mean" means greater involvement with self as opposed to others, and that age makes people less self-involved, but education makes them more self-involved, and men are somewhat more self-involved than women.

But this is even more tenuous than such explanations generally are, since the demographic variables in this collection of conversations are not orthogonal.

So you'd want to do some sort of hierarchical regression, and it would take a day or too to get the data and run it.

But still . . .
Serious speech science

• Transcribed speech is available in very large quantities
• By applying
  – forced alignment
  – pronunciation modeling
  – automated measurements
  – multilevel regression
  we see a new universe of speech data, on a scale 4-5 orders of magnitude greater than the laboratory recordings of the past.

• And interesting patterns are everywhere!
Interdisciplinary opportunities

• These techniques will have rich applications in other fields
  – Clinical diagnosis and evaluation
  – Educational assessment
  – Social science survey methods
  – Studies of performance style
  – . . . and so on . . .

• Wherever speech and language are relevant!
Even in classical scholarship!

The early years of the twenty-first century have seen a heroic age for intellectual life. Ideas have poured across the world and new minds have joined the professionalized academics and authors in grappling with the heritage of humanity. [...] 

No field of study is poised to benefit more than those of us who study the ancient Greco-Roman world and especially the texts in Greek and Latin to which philologists for more than two thousand years have dedicated their lives. [...] 

The terms eWissenschaft and ePhilology, like their counterparts eScience and eResearch, point towards those elements that distinguish the practices of intellectual life in this emergent digital environment from print-based practices. Terms such as eWissenschaft and ePhilology do not define those differences but assert that those differences are qualitative. We cannot simply extrapolate from past practice to anticipate the future.

-- Gregory Crane et al., “Cyberinfrastructure for Classical Philology”,
*Digital Humanities Quarterly*, Winter 2009
An historic opportunity:

• Take an interesting problem, and add
  – a little linguistics and phonetics
  – a little psychology
  – a little signal processing
  – a little statistics and machine learning
  – a little computer science
  – your curiosity and initiative

• And the future is yours!
Thank you!