Sociolinguistics and TalkBank

Brian MacWhinney

CMU - Psychology, Modern Languages, LTI,
SDU - IFKI

http://talkbank.org/socio.ppt

CHILDES and TalkBank

	CHILDES	TalkBank
Age	26 years	10 years
Words	44 million	8 + 55 million
Media	2 TB	.5 TB
Languages	33	18
Publications	3500+	130
Users	3200	600

Lots of Banks

- CHILDES
- AphasiaBank
- PhonBank (link to sociophonetics)
- SLABank
- BilingBank
- ClassBank
- SCOTUS
- AAC, Gesture, Fluency, TBI, Dementia, Tutoring

Where is sociolinguistics?

- Lots of CA corpora
- CallFriend courtesy Chris Cieri
- SBCSAE from TalkBank
- SLX from Labov

- But .....

What data types?

- Written or spoken?
- Corpus or Interaction?
- Phone call or face-to-face?
- Audio or video?

- Answer: we need all of the above
- Data-sharing mandate vs. the "IRB"
- IRB is not the real problem

The Rise of Corpus Studies

Across the last ten years of LLBA citations, there has been a 50% drop in citations of *Chomsky* and a 100% rise in citations of *corpus*.

But language change occurs in spoken interactions in the moment. So our corpora must include these components.

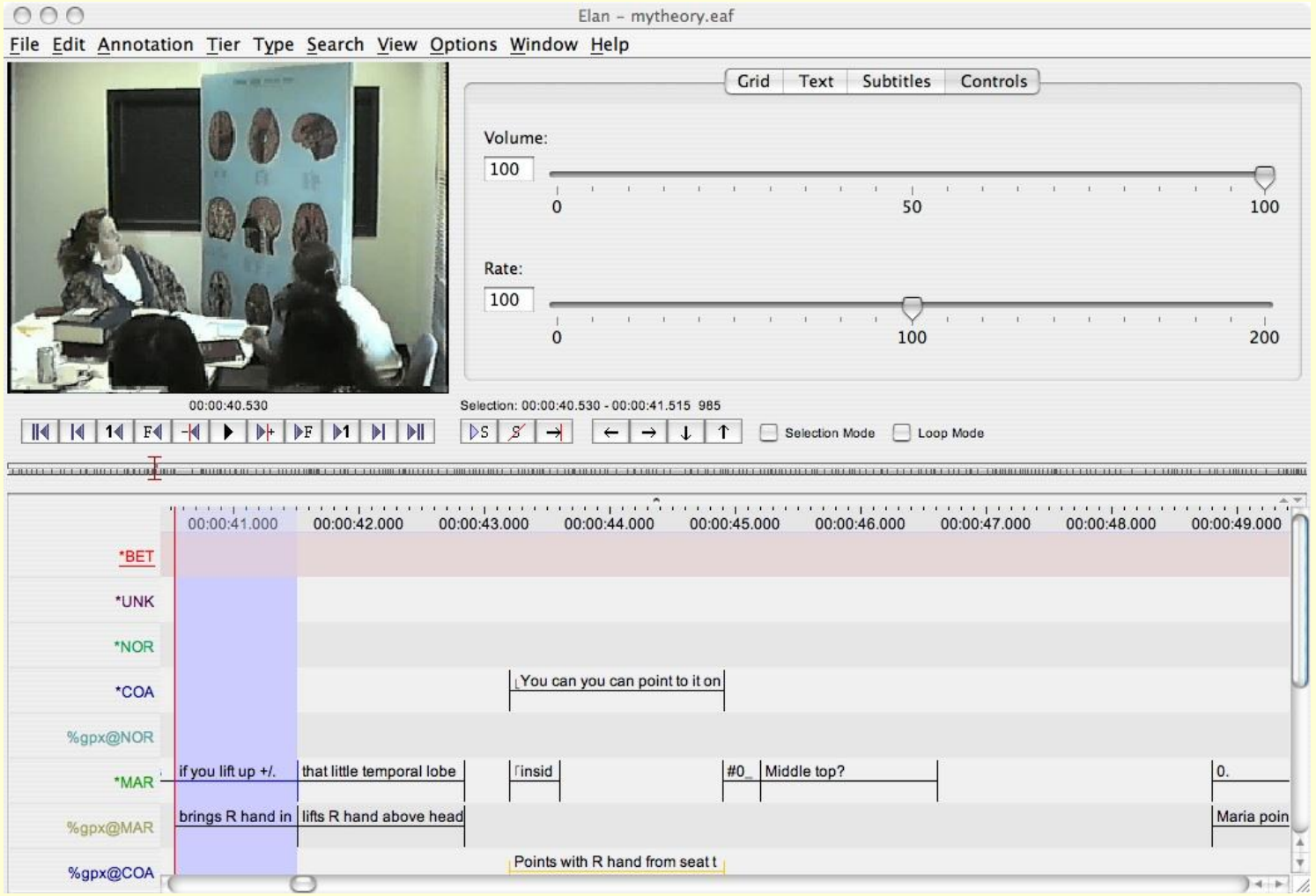
A sample moment:

Transcript linked to video

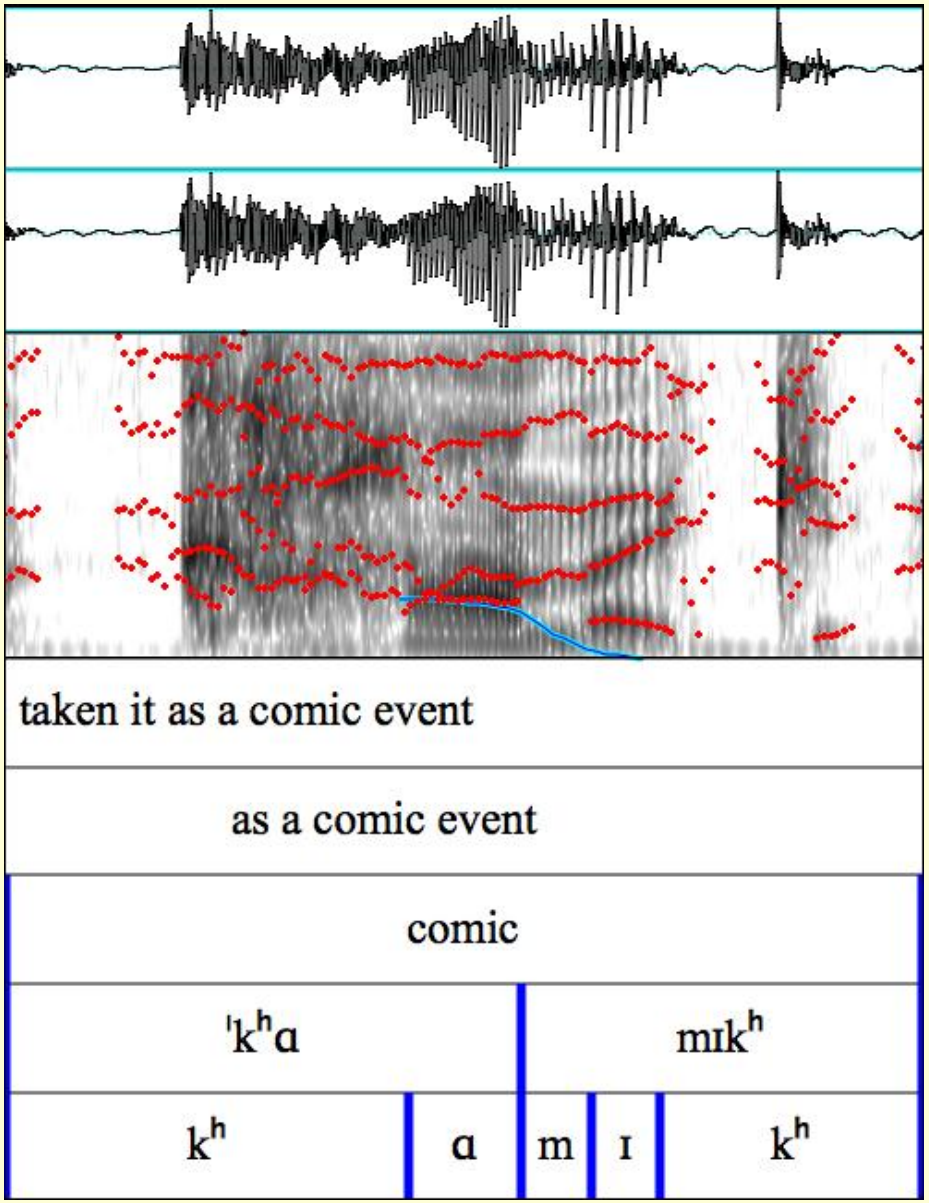
Coyote:demo:MyTheory.ca

```
1     @Begin
2     @Transcriber:   Tim. Koschmann. Last revision 8.1.2000 Johannes Wagner
3     @Participants:  Be Betty, No Norman, Co Coach, Mar Maria, May, Jen Jenny,
4                     Lill, ? unidentified Person, Ps Pauses
5     @Dependent:     ges
6     @Filename:      MyTheory.ca. Moviefile  MyTheory.mov
7     @Time:          6 minutes
8     @Contents:      fragment of tutor-group discssion
9     @Comment:       numbering is by TCUs and pauses, not lines
10
11    Be:     See what it said in here (.) in- my theory (hhh) •
12            (0.4)
13    ?:      khu- [(.hhh)
14    Be:           [about this amnesic (.) dysnomic aphasia, •
15            (0.3)
16    Be:     u:hm (it) says the cause of lesion is usually deep in
17                    temporal lobe just like Kathy was saying ↑presumably
18                    interrupting connections of sensory speech areas with the
CLAN [E][CA]   1
```

Other views


Elan – mytheory.eaf
File Edit Annotation Tier Type Search View Options Window Help
Grid Text Subtitles Controls
Volume:
100
0
50
100
Rate:
100
0
100
200
00:00:40.530
Selection: 00:00:40.530 - 00:00:41.515 985
Selection Mode
Loop Mode
00:00:41.000
00:00:42.000
00:00:43.000
00:00:44.000
00:00:45.000
00:00:46.000
00:00:47.000
00:00:48.000
00:00:49.000
*BET
*UNK
*NOR
*COA
%gpx@NOR
*MAR
%gpx@MAR
%gpx@COA
You can you can point to it on
if you lift up +/.
that little temporal lobe
insid
#0_
Middle top?
0.
brings R hand in
lifts R hand above head
Maria poin
Points with R hand from seat t

Acoustic Views



Gestural Views

Segment	N1
Action	rests chin on hand, elbow on table, right shoulder back
Gaze	front to Deedee
Classification	Attention
Meaning	Attention

*D: ⌈så er det snart⌉ ⌈torturtid→⌉

%ges: ⌊-------D1------⌋ ⌊----D2----⌋

⌊----------N1----------⌋

%com: assimilating the pronounciation of a danish actor in a then tv show

pic * pic *

Analysis Programs

- Searching
- Coding
- MOR, GRASP
- Phon
- Fluency
- EVAL
- nothing yet for sociolinguistics

Rich Data

- For data depth, we need
  - Good recording
  - Good microanalytic methods
- For data breadth, we need
  - Sharing across projects – no navigator can map the world alone
  - This then leads to the need for data-sharing and interoperability

Data Sharing

- 42 reasons not to share data
- The reason to share: it is our responsibility

- The solutions:
  - Methods for password protection
  - Methods for anonymization
  - Credit to contributor
  - Group commitment

Interoperability

- Format Babel: 86 formats
- Program Babel: 55 programs

The solutions:

- CHAT XML
- Roundtrip Convertors for 8 formats
- Program uniformity (nice but not crucial)

Access: Multilingual Corpora

- Ad Backus summary for Moyer and Wei
- CHILDES: Bilingualism
- BilingBank
  - Multilingualism
  - Second Language Acquisition

CHILDES

- AarsenBos - Arabic, Dutch
- DeHouwer - English, Dutch
- Deuchar - English, Spanish
- FerFuLice - English, Spanish
- Genesee - English, French
- Guthrie - English L2
- Hayashi - Danish, Japanese
- Ionin - English, Russian
- Klammler - German, Italian

CHILDES

- Koroschetz: Italian, German
- Krupa: English, Polish
- MCF: Portuguese, English, Swedish
- Perez: English, Spanish
- Serra: Spanish, Catalan
- vanOosten: Dutch, Italian
- Vila: Spanish, Catalan
- YipMatthews: English, Cantonese

Multilingualism

- Bangor
- BlumSnow
- Eppler
- Gardner-Chloros
- Hatzidaki
- Køge
- Langman
- Qatar

Multilingualism - others

- Hamburg?
- LIDES?
  - Moyer
  - Housen
  - Berlin
- CALPIU
- Gardner-Chloros

SLA

- DiazRodriguez
- Dresden
- ESF
- FLLOC/TCD
- Fluency / ELI
- Langman
- PAROLE
- Reading
- SPLLOC

Analysis Methods

1. Bag of Words
2. QDA = a.k.a. Hand Coding
3. Tagging = a.k.a. Automatic Coding
4. Profiles = a.k.a. Canned Analyses
5. Group/treatment comparisons
6. CA Analysis
7. Gesture Analysis
8. Phonetic Analysis
9. Collaborative Commentary
10. Error analysis
11. Longitudinal analysis
12. Modeling

Competing Motivations

“The forms of natural languages are created, governed, constrained, acquired, and used in the service of communicative functions.”

-- MacWhinney, Bates & Kliegl (1984)

Need for a broader framework

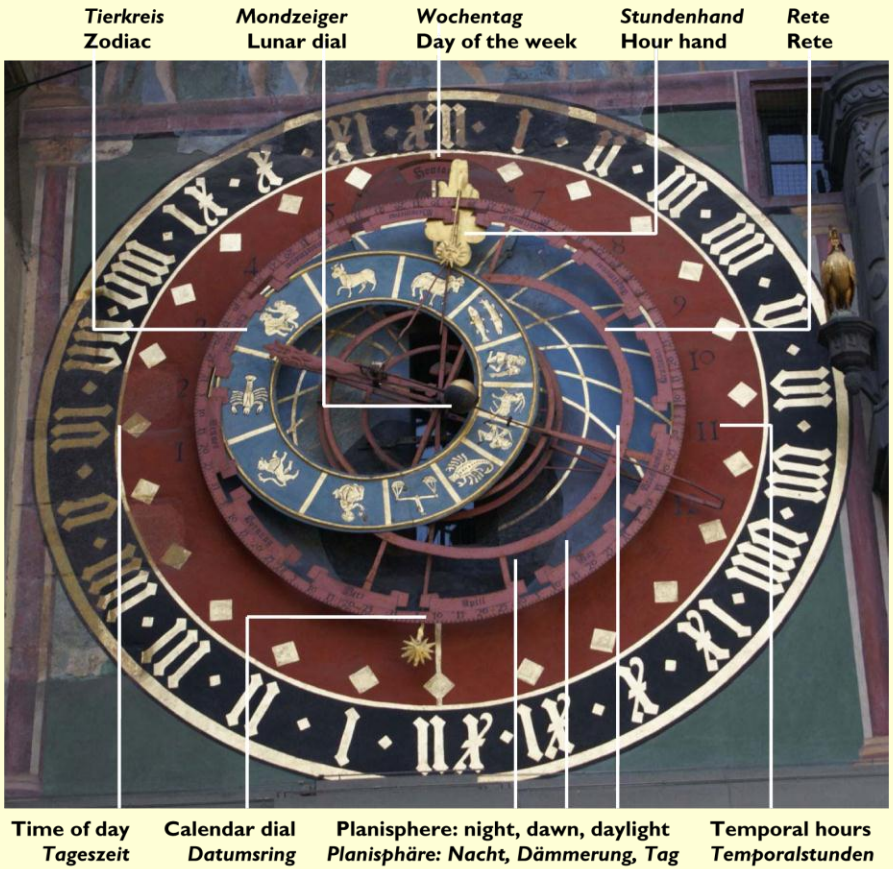
- Emergent modularity
- Revised conception of generativity
- Integrating L1 and L2 acquisition
- Grounding in social process

Interacting Processes within Timeframes

Uniformitarian Principle

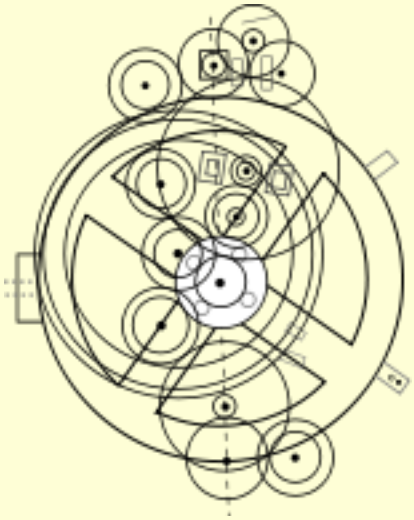
- Hutton in Geology
- Forces determining the geologic record are all observable in the present
  - erosion
  - vulcanism
  - tectonics
  - but not asteroid collisions
- Historical changes in language are based on things observable in current interactions

Meshing of space-time scales



Orloj of Prague -- 1490

The Antikythera – Greece 150BC



How do timeframes mesh?

- They mesh through processes.
  - Goodwin, Lemke, Leontiev, Bahktin
- Many processes are biological.
- Many are social.
- Social frameworks extend to artifacts with long-term permanence (books, mountains, Hungarian crown)

How do the processes mesh?

- The 8 big timeframes are each implemented by dozens of smaller process wheels
- Examples:
  - Gating of lexicon by syntax.
  - Roles configured through embodied action.
  - Licensing of conversational contributions.
  - Use of objects as long-term memories -- Goodwin
  - Graduated interval recall -- Pierre-Humbert
- Processing biases accumulate diachronically, but there can be “defining moments” as in “needs washed”, “repudiate”, and “hun”.

Generativity

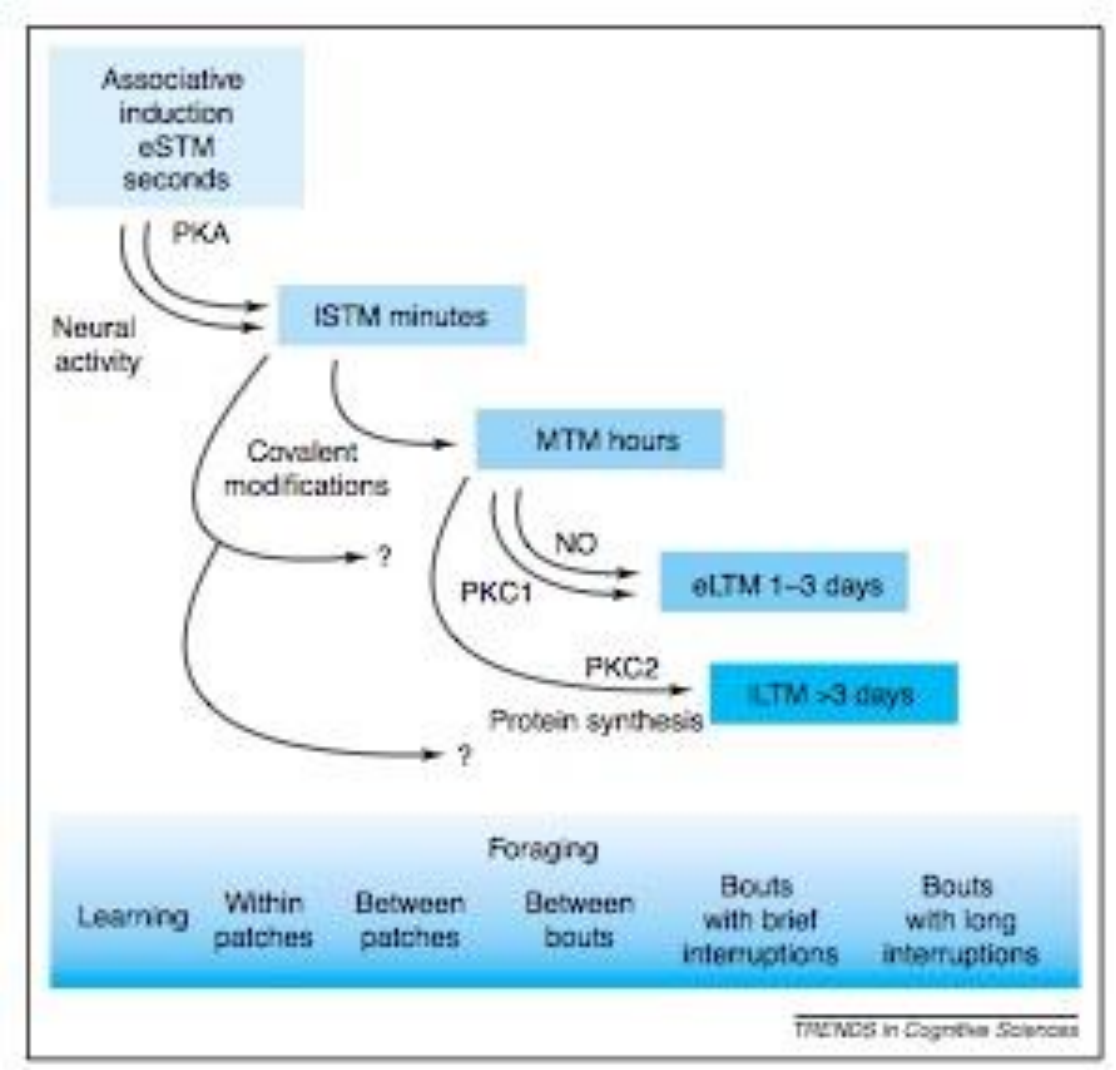
- Modular Generativity: machine that generates and describes all possible sentences (words, sounds) in the language and no impossible ones.
- Interactive Generativity: a collection of emergent processes that interact competitively to generate observed linguistic patterns in corpora.

Basic Issue

1. Language is a system for mapping functions to forms.
2. The forms come from the functions.
3. Where do the functions come from?
4. Current thesis: the functions come from multiple timeframes which integrate in the moment.
5. This suggests a new understanding of ***generativity*** and a new goal for linguistics.

2

Timeframes in Bees

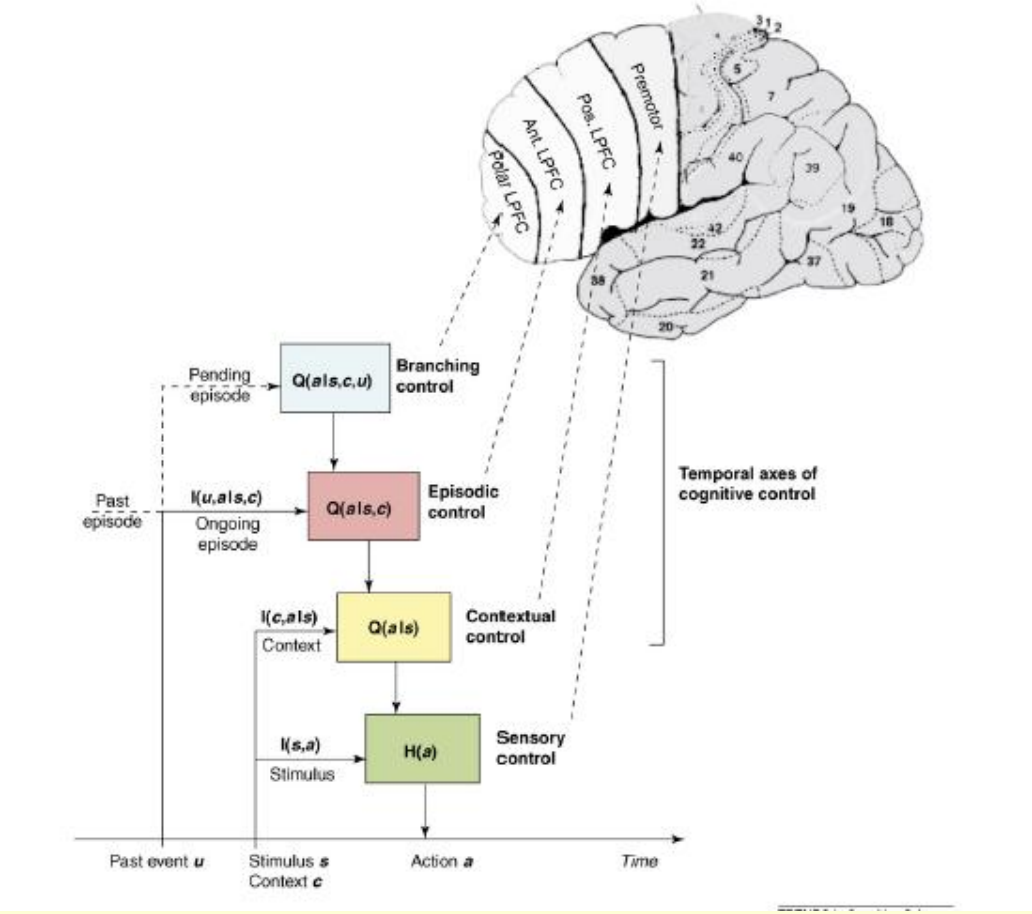


Timeframes in Humans

- Neuronal transmission
- Acoustic storage
- Gaze tracking
- Short-term storage
- Syntactic priming
- Hippocampal function
- Proceduralization
- ....
- Social role identification

Timeframes in Frontal Cortex

Koechlin & Summerfield



8 timeframe groups

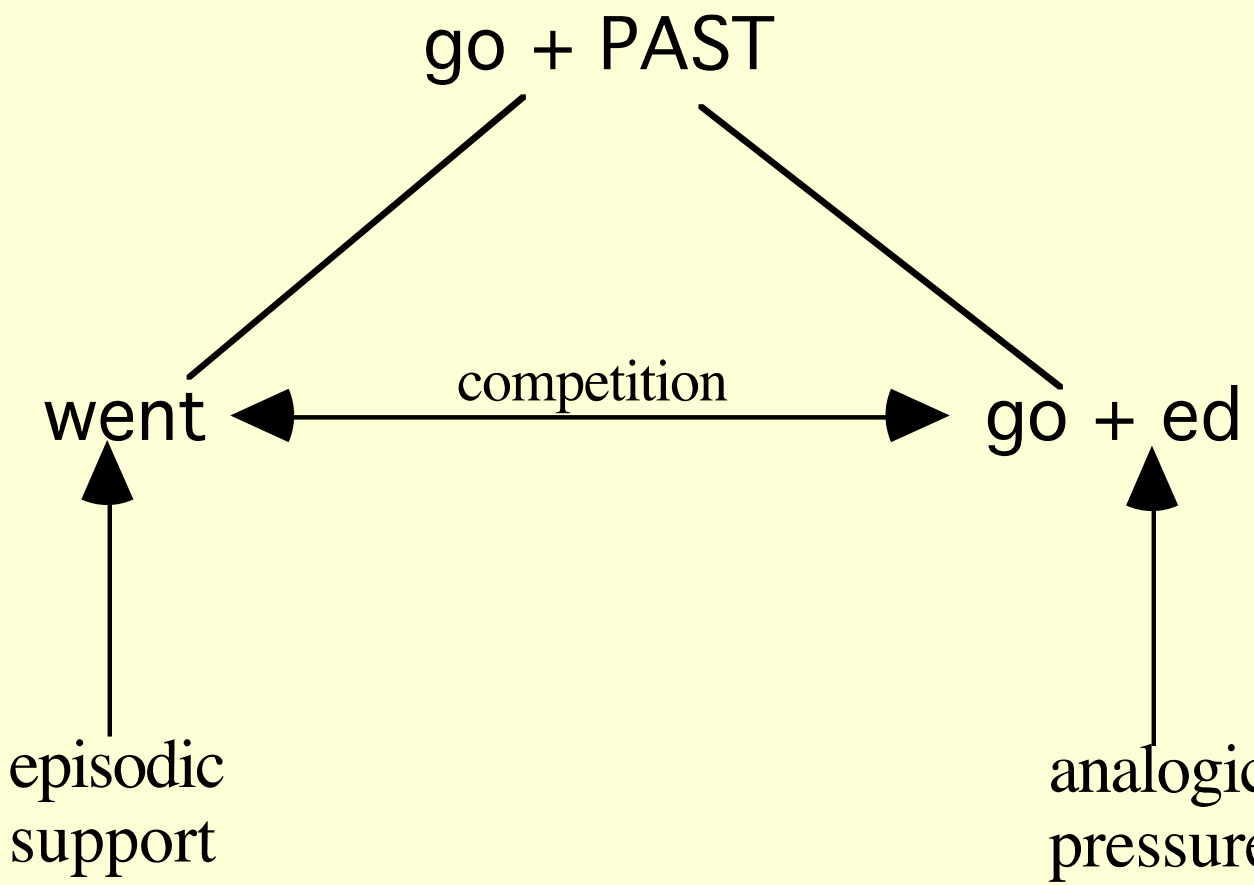
1. Comprehension [10ms - 5sec]
2. Production [10ms - 5sec]
3. Interaction [10ms - 5sec]
4. Encounters [1sec - 20min]
5. Social [days, years]
6. Developmental [days, years]
7. Diachronic [years, decades]
8. Phylogenetic [millenia]

- Interaction

1. Production Wheels

- gating of lexicon by syntax (MacWhinney)
- gesture-speech linkages (McNeill)
- phonological activation (Dell)
- gang effects (all six linguistic levels)
- rote, combination (Nathan, MacWhinney)
- perspective tracking

Dual Routes


go + PAST
went
competition
go + ed
episodic
support
analogic
pressure

2. Perceptual Wheels

- statistical learning (Aslin, Thiessen)
- attention to ends and beginnings (Slobin)
- attention to stress (Juszczyk)
- BOSS, cohorts (M-W, Dell)
- input vs output frequency (Bybee)
- parsing efficiency, attachment (Hawkins ...)
- changes in attentional biases (Rieger)

3. Interactional Wheels

- Gaze contact, posture alignment (Condon)
- Repair, correction, recast (Pfeiffer)
- Variation sets, scaffolding (Waterfall)
- Repetition, imitation, choral (Ochs)
- Turn projection, completion, overlap (CA)

4. Encounter wheels

- Alignment, affiliation, disaffiliation
- Commitment (Social Psychology)
- Mutual Plans, negotiation (Clark)
- Shared mental models (Fauconnier)
- Perspective taking (MacWhinney, Kuno)
- Frequency effects: the toothbrush problem

5. Social wheels

- Immigration (Jørgensen)
- Age group stratification (Ervin-Tripp)
- Rites of passage (Kozniol)
- Educational stratification (Hart)
- Groups: clubs, religions (Wagner)

6. Developmental Wheels

- Body: vocal tract, metabolism (Oller)
- Brain: neurogenesis, connectivity (Bates)
- Motor control: entrainment, coupling
- Learning: Entrenchment, generalization

7. Diachronic Wheels

- Uniformism – Grimm's Law
- Northern Cities shift, push-pull
- Lexical diffusion (Ota)
- Founder's effect (Kiesling)
- Long-term social-affiliation (Labov)

8. Phylogenetic Wheels

- Growth of social support (Tomasello)
- Linking of IFG to STG (Macneilage)
- Organization of dorsal frontal mechanisms
- CV frame-content (Davis-Macneilage)
- Articulatory control (FoxP2)
- Connectivity methods

Memory Reflexes of Frames

- short-term precise acoustic
- mid-term lexical
- frontal timescales
- hippocampal reentrant consolidation
- proceduralization
- …..
- like the bees, but more complex

Linking Timeframes

- Frames impact memory which then provides inputs to the competition
- Slower, marked processes must come to override initial, unmarked processes
- Competition Model: Effects of frequency, reliability, availability, detectability, conflict validity, error tagging

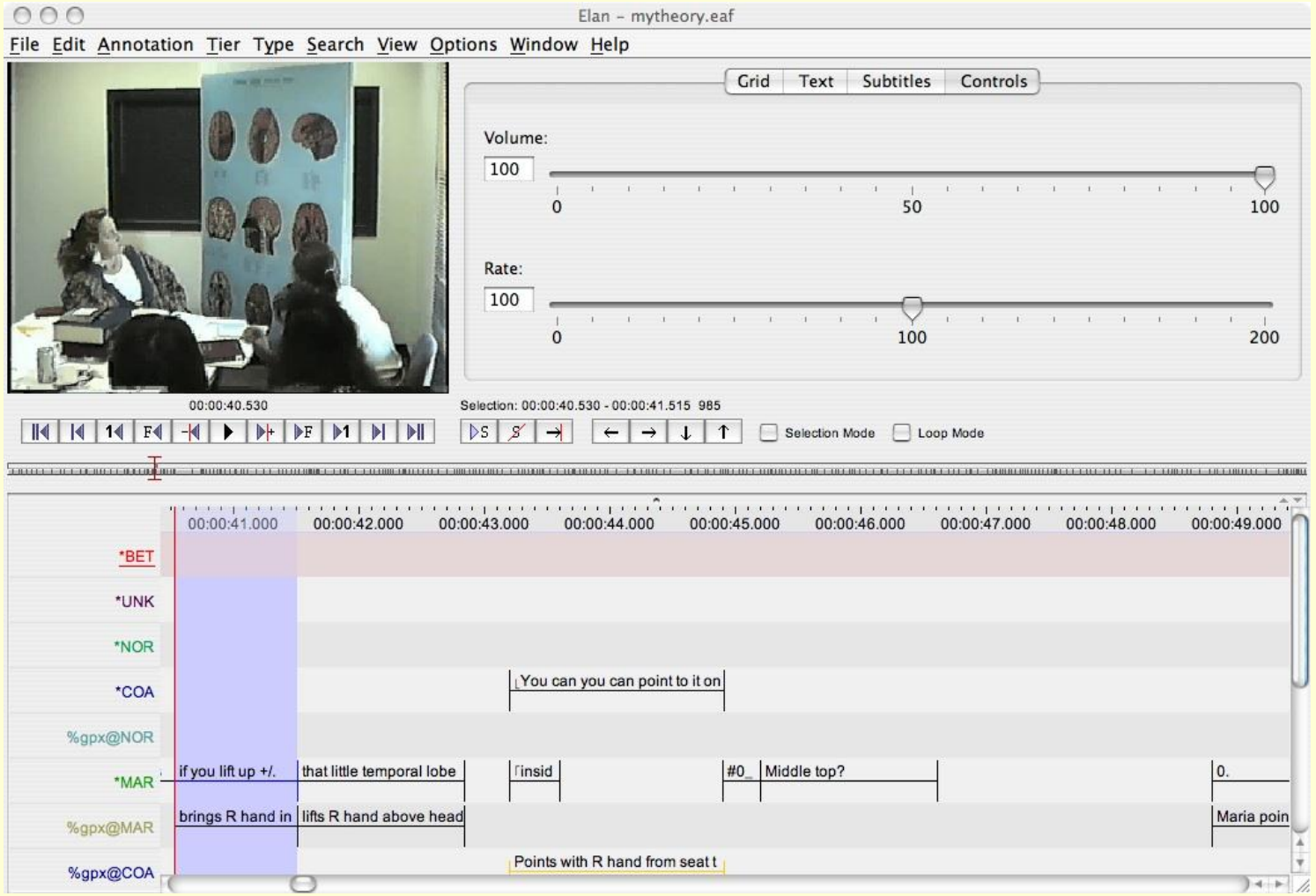
Interaction Sites

- hun - Dutch, yinz - Pittsburgh
- extraposition - Strunk, Hawkins
- self-repair - Pfeiffer
- dative alternation - Bresnan
- Conversational Examples
  - flip up that little temporal lobe - Koschmann
  - dependable -- Sfard, McCobb
  - up to your standards - MacWhinney

Data Capture

- All of the space-time frames must show their effects and be conditioned in actual moments in time and space.
- We can capture The Moment and The Place on video.
- However, we will need to compare across time and space to understand the texture of the process.

Other views


Elan – mytheory.eaf
File Edit Annotation Tier Type Search View Options Window Help
Grid Text Subtitles Controls
Volume:
100
0
50
100
Rate:
100
0
100
200
00:00:40.530
Selection: 00:00:40.530 - 00:00:41.515 985
Selection Mode
Loop Mode
00:00:41.000
00:00:42.000
00:00:43.000
00:00:44.000
00:00:45.000
00:00:46.000
00:00:47.000
00:00:48.000
00:00:49.000
*BET
*UNK
*NOR
*COA
%gpx@NOR
*MAR
%gpx@MAR
%gpx@COA
You can you can point to it on
if you lift up +/.
that little temporal lobe
insid
#0_
Middle top?
0.
brings R hand in
lifts R hand above head
Maria poin
Points with R hand from seat t

Linkage expands Science

- Scientific advance comes from adding additional constraints, considerations.
- The challenge of linking timeframes will force us to expand our view of communication.
- To do this, we must link together a wider data network

The Rise of Corpus Studies

During the last ten years of LLBA citations, there was a 50% drop in citations of Chomsky and a 100% rise in citations of “corpus”.

What changes?

- Fundamental methods do not change
  - Linguistic tests, comparisons
  - VARBRUL, Competition Model, stats
  - eye movement, ERP
  - corpora, video, transcriprts
- What changes is the new focus on the interlocking of processes
  - wider sampling of data
  - more generalization across findings

Conclusion

- Competition is central, to be sure ...
- But to really understand how forms are used, we will need to ask where functions come from
- This requires use to look at
  - processes
  - timeframes
  - meshing

http://talkbank.org/timeframes.ppt