Prosody and the brain II: Turn-taking

Sara Bögels

Max Planck Institute for Psycholinguistics
Overview

Turn-taking models

Turn-taking & non-prosodic cues
  early anticipation (EEG)
  early planning (EEG)

Turn-taking & final prosodic cues
  importance of final information
  importance of prosodic cues

Turn-taking & silence (EEG)
Turn-taking models
Turn-taking

70% of speaker changes [-250 ms, 500 ms]

(Switchboard Corpus of American English)

Levinson & Torreira (2015)
A psycholinguistic puzzle

Planning and producing language takes time, e.g.:

- picture naming: 600 ms Levelt et al. (1999)
- simple sentence (SVO) production: 1500 ms Griffin & Bock (2000)

• Planning starts in overlap with the current turn
• The turn end should be estimated precisely
A psycholinguistic puzzle

Planning and producing language takes time, e.g.:
- picture naming: 600 ms Levelt et al. (1999)
- simple sentence (SVO) production: 1500 ms Griffin & Bock (2000)

• Planning starts in overlap with the current turn
• The turn end should be estimated precisely
Model 1

1. Predict words up to end of turn
2. Estimate duration of these words
3. Launch planning to start speaking around estimated turn end

Sacks et al. (1974); De Ruiter, Mitterer & Enfield (2006)
1. Early conceptual and linguistic planning
2. Identification of turn-completion
3. Launch articulation

e.g., Levinson & Torreira (2015)
Turn-taking & non-prosodic effects
Model 1: Anticipation

1. Predict words up to end of turn
2. Estimate duration of these words
3. Launch planning to start speaking around estimated turn end

Sacks et al. (1974); De Ruiter, Mitterer & Enfield (2006)
Turn-end anticipation (1)

**Task**: Listen to isolated conversational turns
Press a button exactly at the moment when the turn ends

De Ruiter, Mitterer, & Enfield (2006)

Predictable & Unpredictable turns
Lower beta (11-18.5Hz) power for predictable turns from 1.8 s before button-press

=> Reflects early anticipation of turn end

Magyari, Bastiaansen, De Ruiter, & Levinson (2014)
Turn-end anticipation (2)

Readiness Potential (RP)

- button-press
- verbal response

=> Reflects early anticipation of turn end

Modeling Long-range Anticipation (Lexical Prediction)

Lexical Prediction

1. A’s turn
   - Gap
   - B’s production planning
   - B’s turn

2. A’s turn
   - Gap
   - Articulation
   - B’s turn
   - B’s production planning

Long-range anticipation appears possible (but what is it?)

(1) and (2) are the same interaction model.

Lexical Prediction is the first step in the model, followed by production planning and then articulation before the turn is completed.
EARLY: Which character, also called 007, appears in the famous movies?
LATE: Which character from the famous movies, is also called 007?

Bögels, Magyari, & Levinson (2015)
Which character, also called 007, appears in the famous movies?
Which character from the famous movies, is also called 007?

Response - planning

Absence/reduced in control study

=> Listeners start planning ASAP
Models

1. A’s turn → gap → B’s production planning → B’s turn

Listeners start planning ASAP

2. A’s turn → gap → articulation → B’s turn → B’s production planning
Turn-taking & Final prosodic cues
Are final cues too late?

How fast is reaction to turn-final cues?

De Ruiter, Mitterer & Enfield (2006)
Methods

Say ‘ja’ in reaction to stimulus offset ASAP

Materials (blocked presentation):

Speech-like repetitive stimulus (mamama...)

- ‘flat’: no prosodic markers
- lengthening on last syllable
- pitch change on last syllable
- pitch change on penultimate syllable
- mixed pitch block: less predictable

Torreira & Bögels (in preparation)
Results

- flat: mode later than in conversation
- prosodic cues lead to faster RTs
- participants anticipate (even though asked to react)
- mixed block: still anticipation (less)
Are final cues too late?

- Prosodic cues do not seem to occur too late: they seem to enable ‘anticipation’ of the end
- ‘Anticipation’ seems to be a default strategy if possible (even if asked to react)
- Distribution of turn transitions in conversation similar to prosodic conditions
  - If content of turn is already planned; final prosodic cues could account for distribution in conversation
- A part of turn transitions in conversation could even be a reaction to silence
Final cues not too late

Prosodic cues enable ‘anticipation’: not too late
Orientation to turn-ends? Breathing

Inbreath: related to speech timing

A’s turn

B’s production planning

B’s turn

A’s turn

B’s production planning

B’s turn
Conversational corpus with Respitrace inductive plethysmography

Torreira, Bögels, & Levinson (2015)
Answer duration & inbreaths

Question-answer pairs

=> Inbreaths related to planning
Timing relative to question end

A’s question

B’s inbreath

B’s answer

Time
Inbreath timing

[Diagram showing a timing graph with categories for Question and Answer, along with a frequency distribution histogram.]
Answer timing

Question

Inbreath

Answer

![Answer latency box plot]

- Inbreath: 
- No Inbreath:

n = 80
n = 91
Conclusions

Inbreaths are more likely to occur before longer answers

> breathing behavior can be informative about speech planning in conversation

Speech inbreaths before answers appear to be timed to the end of questions

> consistent with the idea that articulation is timed to turn ends (planning might start earlier)

> answers with an inbreath occur later
Listeners orient to turn-ends

1. A’s turn → gap → B’s production planning → B’s turn
   - Inbreath (if long answer)

2. A’s turn → gap → inbreath → articulation → B’s production planning → B’s turn
   - at least some planning
Use of final prosodic cues?

1. A’s turn → gap → B’s production planning → B’s turn

2. A’s turn → gap → articulation → B’s turn → B’s production planning

Prosodic cues
Prosody & turn-boundaries

- Turn-boundaries coincide with:
  - Intonational completion (nuclear pitch accent, boundary tone, final lengthening)
  - Lexico-syntactic completion

- Observational studies: correlation
- Offline experiments: same as online?

Online experiment

Extraction of turns from a corpus of conversational Dutch

Button-press task:

– Participants anticipate the end of turns
– Conditions:
  • Original (no manipulation)
  • No pitch (flattened)
  • No words (low-pass filtering)

De Ruiter, Mitterer & Enfield (2006)
Online experiment

Results:

No pitch = Original
No words < Original

Conclusions:

• Lexico-syntactic information is necessary and possibly sufficient for turn-end projection
• Intonation is neither necessary nor sufficient

But pitch ≠ intonation!
Participants may have used other cues to intonational phrasing in the experiment (e.g. final lengthening)

De Ruiter, Mitterer & Enfield (2006)
Prosodic cues to turn-taking

RA interviewed participants via microphones and headphones

Short and long questions from a script embedded in otherwise free interview:

  e.g.,  
  \textit{So you are a student?}
  \textit{So you are a student here at the Radboud University?}

Same lexico-syntax: end should be ambiguous

Bögels & Torreira (2015)
Response times

• Long questions: not many overlaps and none close to syntactic completion point

  So you are a student here at the Radboud University?

• Short questions: close to the end; 18% gap < 200 ms (no reaction)

  So you are a student?
Acoustic measurements

SHORT

Dus je bent  student  H%

L*

LONG

Dus je bent  student  hier op de  Radboud  Universiteit  H%

LH*

L*
Acoustic measurements

- Duration of final syllable
- Duration of final word
- Duration of preceding words

SHORT

LONG
Acoustic measurements

- Duration of final syllable
- Duration of final word
- Duration of preceding words

- F0 valleys
- F0 peaks
Findings

- No consistent differences before ‘final’ word
- Clear differences in ‘final’ word:
Experiment: Stimuli

Only 2 manipulated items per participant
Results: short stimuli

[Dus je bent student]_{IP}

[Dus je bent student...]

[Dus je bent student...]

Original

Cut Final only

Cut All

Mean RT (ms)

0 100 200 300 400 500

* n.s.
Results: long stimuli

- Original: [Dus je bent student]_IP hier op de Radboud Universiteit]_IP
  - Early point of syntactic completion
  - 0% button presses

- Spliced Final only: [Dus je bent student]_IP hier op de Radboud Universiteit]_IP
  - Early point of syntactic completion
  - 27% button presses

- Spliced All: [Dus je bent student]_IP hier op de Radboud Universiteit]_IP
  - Early point of syntactic completion
  - 32% button presses
Discussion

• Listeners use prosodic cues to determine turn-ends
  – late button-presses if no IPh boundary at the end
  – some button-presses if IPh boundary in the middle

• Final cues appear most important
  – acoustic analyses
  – same effects if only last word was manipulated
Use of final prosodic cues

1. A’s turn → gap → B’s production planning → B’s turn

2. A’s turn → gap → articulation → B’s turn

Syntactic and prosodic completion
Generalizability

1. A’s turn → gap → B’s production planning → B’s turn

2. A’s turn → gap → articulation → B’s turn → B’s production planning

Prosodic cues in all turns?
Methods

Materials

• 96 turns from spontaneous telephone calls (no overlap)
• Naïve readers identified plausible points of syntactic completion
  – See transcripts of turns word-by-word (1 s.)
  – Task: press button when see last word

=> 35 early plausible syntactic completion points:
10+/24 readers thought this was the last word
Prosodic analysis

Annotation of prosodic cues at these points:

- Sentence accent
- Salient final lengthening
- Salient phrase-final pitch movement

Cues are not independent

- When no accent: no melody & no lengthening (2)
Experiment

Auditory task

• Button-press

• All non-manipulated turns
Window [-250, 250 ms] around plausible points of syntactic completion => how many button-presses?

More button-presses if more prosodic cues are present ($p < .001$)
Conclusions

• Plausible early syntactic completion points are abundant (36%)
  – Not all also prosodically complete!

• Prosodic cues – not syntactic completion alone – predict anticipatory button-presses

• Still < 30% button-presses even with all cues
  – Silence used some of the time?
  – Anticipation + inhibition?
Prosodic cues are used

Lexical prediction

1. A’s turn → gap → B’s production planning → B’s turn

2. A’s turn → gap → B’s production planning → B’s turn

Syntactic and prosodic completion (all turns) + silence?
Conclusions turn-taking

Lexical prediction is not enough

A’s turn

Gap

B’s turn

B’s production planning

Planning starts ASAP
Early anticipation?

Turns appear to be timed based on information at the end of the previous turn

Turn-final prosodic cues are:
• not too late (given planning)
• used in questions and other turns

2
Some future directions

• Early planning in overlap with listening?
  – Preliminary evidence for trade-off between comprehension and production planning
• More natural task: button-press => answer
• Importance of different final cues: strength
• Turn-keeping cues, e.g.
  – Rush-through
  – Hesitation
Turn-taking & Silence
Long silences

Some gaps are long. Why?
  e.g., dispreferred responses

A:  *I could come to you right now?*
B:  *Sure.*  ‘accepting’ (preferred)
B’: *Well, I have to work.*  ‘declining’ (dispreferred)

(e.g., Kendrick & Torreira, 2015)
Silence & anticipation

Off-line experiments e.g., Roberts, Francis & Morgan (2006)
  – Listen to request + positive response with different gap lengths
  – Willingness to comply with a request judged smaller for longer silence

Online expectations affected by gap length?

EEG

Research Question

Do interlocutors anticipate the type of response (preferred or dispreferred) based on the duration of inter-turn silence alone?

Bögels, Kendrick & Levinson (2015)
Methods

• Turns from a Dutch spoken corpus (CGN): telephone conversations between friends and acquaintances
• 120 questions: requests, invitations, proposals, and offers
• 60 responses from elsewhere in the corpus
  – preferred responses: *ja* (‘yes’)
  – dispreferred response: *nee* (‘no’)
• Two response timings
  – 300 ms gap
  – 1000 ms gap
Procedure

Written context

Question

+ gap (recording noise): 300 or 1000 ms

+ response: ‘yes’ or ‘no’

+ Statement true false

20%
EEG Hypotheses

N400: smaller for expected words e.g., Kutas (1980)

- 300 ms gap: ‘yes’ is more expected than ‘no’ => N400 for ‘no’ vs. ‘yes’
- 1000 ms gap: N400 disappears or even flips?
Results: N400

Timing by Response type Interaction for N400

N400 for ‘no’

no N400

300-500 ms
Results: late positivity

No Timing by Response type interaction after 500 ms
Late positivity for ‘no’ irrespective of timing
Discussion N400

- After 300 ms gap: larger N400 for ‘no’ => dispreferred is less expected than preferred after a short gap
- After 1000 ms gap: same N400 for ‘yes’ and ‘no’ => expectations converge
  ⇒ Mere silence changes expectations: expectations for preferred and dispreferred response converge

Why does ‘no’ not become more expected?
1. Dispreferreds often qualified or mitigated (e.g., well, maybe) => dispreferred more expected than preferred but not a flat ‘no’
2. General normative expectation for preferred over dispreferred + expectation for dispreferred after long gap based on frequencies = no effect?
Discussion positivity

- Late positivity for ‘no’ *irrespective* of response timing
- Dispreferred responses are socially accountable actions => require explanations (Garfinkel, 1967; Heritage, 1984)
- A flat ‘no’ might require more analysis or extra processing to understand (e.g., a search for an account)
- cf. positivity to social norm violations (e.g., Leuthold et al., 2015)
  ⇒ No-responses are socially disaffiliative

Future direction
- Create positive/negative context: when is the gap too long?
References (1)


Thank you!