A Simplest Systematics for the Organization of Turn-taking for Conversation

Harvey Sacks, Emanuel A. Schegloff, Gail Jefferson

Francesca Capuano

University of Saarland

15/12/2015

2 components

- **TURN-CONSTRUCTIONAL COMPONENT**

  - A turn can be a sentential, clausal, phrasal or simply lexical unit

  - Projection of the unit → Transition Relevance Place (TRP)

- **TURN-ALLOCATION COMPONENT**

  - Current speaker selecting next speaker

  - Self-selection

Set of rules

Rule 1

(a) If the current speaker has identified, or selected, a particular next speaker, then that speaker should take a turn at that place.

(b) If no such selection has been made, then any next speaker may (but need not) selfselect at that point. If self-selection occurs, then first speaker has the right to the turn.

(c) If no next speaker has been selected, then alternatively the current speaker may, but need not, continue talking with another turn-constructional unit, unless another speaker has self-selected, in which case that speaker gains the right to the turn.

Rule 2

Whichever option has operated, then rules 1a-c come into play again for the next transition-relevance place.

How the system accounts for the facts

(1) Speaker-change recurs, or at least occurs

- speaker-change is not automatic → rule 1c

How the system accounts for the facts

(2) Overwhelmingly, one party talks at a time

- exclusive right to talk
- overlaps can only occur at TRP

How the system accounts for the facts

(3) Occurrences of more than one speaker at a time are common, but brief

- multiple self-selection → first to self-select speaks
- wrong prediction of TRP
- addition of optional elements at the end of the sentence
- overlaps can only occur at TRP

How the system accounts for the facts

(4) Transitions (from one turn to a next) with no gap and no overlap are common. Together with transitions characterized by slight gap or slight overlap, they make up the vast majority of transitions

How the system accounts for the facts

(5) Turn order is not fixed, but varies

- last speaker as next speaker bias

- possibility of local monitoring for hearing, understanding,

agreement

How the system accounts for the facts

(6) Turn size is not fixed, but varies

- speaker in charge can choose any unit type

- rule 1c allows self-reselection → continuation of the same turn

How the system accounts for the facts

(7) Length of conversation is not specified in advance

- the model does not predetermine length of conversation

How the system accounts for the facts

(8) What parties say is not specified in advance

- ceremonies, interviews, etc.
- first turns → greetings
- content of previous turn
- adjancy pairs first parts bias

Constraints depend on systems other that the turn-taking one, like the organization of the types of sequences.

How the system accounts for the facts

(9) Relative distribution of turns is not specified in advance

- the set of rules allows for any possible selection of speaker
- turn distribution manipulations
- turn-order bias

How the system accounts for the facts

(10) Number of parties can vary

- current and next speaker → mechanisms for entry/exit
- small number of participants bias (last as next bias)

Number of parties is connected to relevance of variabilities:

2 parties → differential **turn size**

3 parties → differential **distribution of turns**

4 parties → number of turn-taking systems → **schism**

How the system accounts for the facts

(11) Talk can be continuous or discontinuous

- rules 1a, 1b and 1c are not applied → enlarged gap = lapse

- rules 1b and 1c are available for continuation

- rule 1a is applied → enlarged gap = pause before turn-beginning

How the system accounts for the facts

(12) Turn-allocation techniques are obviously used. A current speaker may select a next speaker (as when he addresses a question to another party) or parties may self-select in starting to talk

Turn-allocation techniques(curret-selects-next vs. self-selection):

- first pair-parts of adjancy pairs + term of address/gaze
- repair techniques (repetitions, 'what?') → turn-order bias
- recompleters (tag questions) → exit techniques
- social identities
- pre-starters (well, but, so) → analysability of the unit
- subsequent starters techniques

How the system accounts for the facts

(13) Various 'turn-constructional units' are employed; e.g. turns can be projectedly 'one word long', or they can be sentential in length

- **syntactic completeness**
- **intonation**
- **3-part structure of turns**

How the system accounts for the facts

(14) Repair mechanisms exist for dealing with turn-taking errors and violations; e.g. if two parties find themselves talking at the same time, one of them will stop prematurely, thus repairing the trouble

- turns as repairs
- rules as repairs

The type of model this is

- **LOCAL MANAGEMENT SYSTEM**

  with respect to turn-order:

a) it deals with a single transition at a time

b) the single turn it allocates on each occasion is the 'next turn'

c) it deals with transitions:

- comprehensively
- exclusively
- serially

The type of model this is

- **LOCAL MANAGEMENT SYSTEM**

  with respect to turn-size:

the determination of turn-size is accomplished locally, in the developmental course of each turn, under constraints imposed by a next turn, and by an orientation to a next turn in the current one

The type of model this is

- **LOCAL MANAGEMENT SYSTEM**

  - party-administered

  - turn-size and turn-order are interdependent

The type of model this is

- **INTERACTIONALLY MANAGED SYSTEM**

  with respect to turn-order:

every rule is contingent on higher options not having been exercised and constrained by the prospective operation of lower-order options

The type of model this is

- **INTERACTIONALLY MANAGED SYSTEM**

  with respect to turn-size:

Turn-size not exclusively determined by the speaker → distribution of tasks:

- speaker allows projection and use of a TRP

- another speaker's beginning of talk, if properly placed, can determine where he ought to stop

The type of model this is

LOCALLY MANAGED

+

INTERACTIONALLY CONTROLLED

=

general principle of

RECIPIENT DESIGN

orientation and sensitivity to particular co-participants → contex-sensitive

Some consequences of the model

- An intristic **motivation for listening** is identifiable (other than interest or politeness)
- Turn-taking organization at least partially controls the **understanding** of utterances
- The turn-taking system has a proof procedure for the **analyses of turns**

The place of conversation among speech-exchange systems

- **one party talking at a time while speaker change**
- **other parameters change:**

conversation → one-turn-allocation at a time

debates → ordering of turns is pre-allocated

meetings with chair-persons → turns are partially pre-allocated

The place of conversation among speech-exchange systems

Linear array of turn-taking systems

with respect to allocational arrangement

Local allocation → **Pre-allocation**

Conversation

- maximization of set of potential speakers in each next turn
- no methodical equalization of turns
- Increasing internal complexity of units

Meetings

Debates

- minimization of set of potential speakers in each next turn
- possible methodical equalization of turns among speakers
- Increasing turn-size

Timing in turn-taking and its implications for processing models of language

Stephen C. Levinson and Francisco Torreira

Francesca Capuano

University of Saarland

15/12/2015

Sacks et. Al (1974)

2 major issues:

1)what counts as a turn, and how participants can recognize a unit as complete

- any word or phrase may in context constitute a turn, syntactic units can be nested or conjoined indefinitely

- Sacks et al. (1974) "some understanding of sound production (i.e., phonology, intonation, etc.) is also very important to turn-taking organization." BUT:

Ford and Thompson (1996, p. 151):

K: Vera (.) was talking §on the phone §to her mom?

C: mm hm

K: And uh she got off §the phone §and she was incredibly upset?

C:Mm hm

In addition to syntactic and prosodic completeness, pragmatic completeness may be required to terminate a turn (Ford and Thompson, 1996; Levinson, 2013)

Sacks et. Al (1974)

2 major issues:

2)'projection' or predictive language understanding

- It is not at all clear how this works, given the flexibility and extendibility of most syntactic units

turn-completion by the other (Lerner, 1991, 2002; Hayashi, 2013) → bi-clausal structure (*If..then..* or *Whenever. . ., X. . .*) → sometimes exactly the same words do occur in overlap

Recipients accurately predict the content of the second clause

Alternative proposals: Duncan (1972,1974)

- turn-handing-over signals: prosodic (type of final intonation, final syllable duration, final drop in pitch, or loudness), gestural (end of a gesture), and lexical/syntactic (tag, clause end, etc.)
- turn-maintaining signals: continuing gesture or a gaze switch away

the turn- taking system is entirely **under the control of the current speaker**

vs

CA model: speaker transition is contingently achieved by one speaker coming to the end of a unit and another starting

Alternative proposals: Duncan (1972,1974)

Signaling view largely superseded, but:

- **importance of visual cues**

  Kendon (1967)

  Rossano (2013) → gaze actually oriented to larger units of conversation (sequences), which it may serve to open and close

- **the coincidence of turn transitions with a number of features of turn construction, prosody, gesture, etc**

Alternative proposals: Heldner and Edlund (2010)

- **Speakers don't aim at no-gap-no-overlap**

- Actual zero gaps (under 10 ms) represent less than 1% of transitions and overlaps average 40% of transitions in their corpora

BUT

- 10ms precision may not be realistic of human performance:

- voiceless stops in English average between 60 and 80 ms (Crystal and House, 1988; Byrd, 1993).

- perceptual "no gap" was always estimated by conversation analysts to be of the order of 150–250 ms (Schegloff, 2000; Heldner, 2010)

- some overlaps are not heard as intrusions (*hmhm*), some are expectable (competing first starts)

- interruption is a sanctionable breach of social mores (one speaker tends to rapidly drop out)

Distribution of gaps

- **Although mean values vary, the factors affecting response times are uniform across cultures:**

  - Task-oriented interaction shows similar patterns (Weilhammer and Rabold, 2003)

  - Responses to wh- questions are slower than polar (Stivers et al., 2009)

  - Response times change to match new interlocutor

  - Heldner and Edlung (2010) → average for speaker transition at 200ms (short gap)

Distribution of overlaps

Switchboard Corpus of English telephone conversations (Godfrey et al., 1992)

- **Overlaps:**

- common (30% of transitions) → 'occurrences of more than one speaker at a time are common, but brief'

- short duration (less than 5% of speech signal, between overlaps mean = 275ms) → 'overwhelmingly, one party speaks at a time'

- occur largely in principled places (between-overlaps, possible completions, simultaneous turn-start) → Sacks predicted signs of overlap avoidance

- mostly involve backchannels (which do not constitute full turns) → backchannels pass up the opportunity to take a turn, and are therefore principled intrusions (Schegloff, 2010)

'Proto-conversation' and Turn Taking in Human Development

1970s → interest in childrens' aquisition of turn taking abilities

- **Trevarthen (1977) and Bruner (1983) → "protoconversation"**
- **Bateson (1975) → average turn transitions are about 1.5 s at 3 months → turn-taking may have an instinctive basis**
- **Jasnow and Feldstein (1986); Beebe et al. (1988) → gap reduces in the following pre-linguistic months to around 800 ms**
- **Garvey and Berninger (1981) → gap duration increases toward a second and a half in toddlers → cognitive difficulties of language production → remains at around a second for 5-year-olds**

'Proto-conversation' and Turn Taking in Human Development

Renewed interest

- Tice and Henetz (2011); Casillas and Frank (2013); Keitel et al. (2013) → eye-tracking → 3-year-olds can anticipate speaker transitions
- Casillas and Frank (2013) → children under 3 are better in the prosody-only condition (with words filtered out) than in the words-only condition (with prosody filtered) → early advantage for prosody (adults only showed an advantage for words + prosody)

Predictive language comprehension

Chomsky (1969) → probability and prediction have no possible role in a scientific theory of language

BUT listeners use different sorts of linguistic information (i.e., semantic, morphosyntactic, prosodic) in order to predict the content of an incoming utterance

- eye-movement studies → listeners predict upcoming entities from likely collocations
- EEG → the morphosyntactic frame is used to predict upcoming material
- Ito and Speer (2008) → participants could anticipate referents on a screen on the basis of the location of contrastive pitch accents
- Magyari et al.(2014) → predictable turns show a very early EEG signature of preparation to respond about half way through the turn (c. 1200 ms before the end)

Quite long-range prediction is normally involved in understanding language in a conversational mode

Latencies in language production

- Wheeldon and Levelt (1995) → pre-articulation processes run 3 or 4 times faster than actual articulation

- Levelt (1989) → average reaction from seeing a picture to beginning the naming is 600 ms

- Schnur et al. (2006) → for multiword utterances the effect is not linear → naming two nouns takes 740–800 ms before output begins

- Draper et al. (1960) → 140–320 ms inhalation latency

- McFarland (2001) → inhalation time in spontaneous dialog is typically over 500 ms long

- Torreira et al. (2015) → short responses to questions are made on residual lung air, longer responses require planned inhalation → the trigger for inhalation, during the last few hundred ms of the interlocutor's turn, is based on a prediction that the current speaker will imminently end his/her turn

- Schaeffler et al. (2014) → tongue movements start before the acoustic signal (120 and 180 ms)

Language production involves latencies of well over half a second, and a multi-word utterance is likely to involve a second or more of processing before articulation begins

Experimental studies of turn-taking

- De Ruiter et al. (2006) → accuracy of turn-end anticipation is preserved under No Pitch, but significantly lost under No Words

- Bögels and Torreira (2015) → cross-spliced sentences with different pitch and syllable length

Speakers do use prosodic cues to judge turn-ending, but they need to be integrated with the lexical/syntactic information

Bögels and Torreira (2015) → neural signature 500ms after answer becomes available

Speakers begin planning their response as soon as they can, up to a second or more before the incoming turn ends

The core psycholinguistic puzzle

in spite of the long latencies involved in language production (600–1500 ms or more), participants often manage to achieve smooth turn transitions (with the most typical gaps as little as 100–300 ms)

comprehension is even more predictive than is currently thought

extraction of the speech of the incoming utterance

planning and encoding of the response

The core psycholinguistic puzzle

conversation involves constant double tasking

Segaert et al. (2011) → both comprehension and production use much of the same neural circuitry → rapid task switching

Pickering and Garrod (2013) → full comprehension and production processes running simultaneously + two fast prediction systems (one for self, one for other)

More likely the real production system may be involved minus the phonological and phonetic encoding

Toward an adequate psycholinguistic model

A recipient's tasks:

1) identify or predict the speech act being carried

2) if a floor exchange relevant or due, production planning begins

3) production proceeds through conceptualization, lemma retrieval, phonological retrieval, and phonetic encoding, extending 600–1200 ms or more before articulation

4) actual articulation held in a buffer

5) turn-final cues are detected → comprehension system signals an imminent completion of the incoming turn

6) articulation is launched

Toward an adequate psycholinguistic model

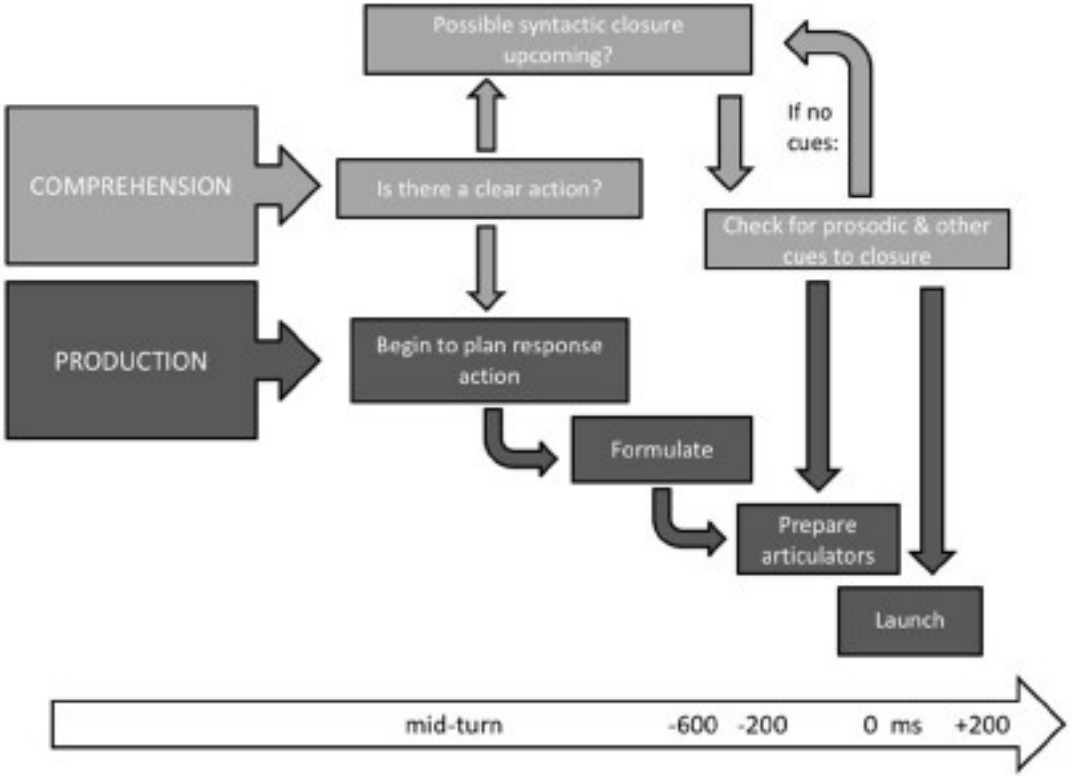

FIGURE 3 | Sketch of the interleaving of comprehension and production in the recipient of an incoming turn.

Toward an adequate psycholinguistic model

What this model crucially accounts for:

(a) short latencies in responses despite long latencies of the production system

(b) modal response with positive offsets of around 100–300 ms → reaction time to the turn-final prosodic cues in the incoming turn

Toward an adequate psycholinguistic model

What accounts for overlaps and long gaps?

- speakers may decide to launch articulation without waiting to identify turn-final cues

- speakers may not have been able to plan the initial stages of their turn early enough

Thank you!