

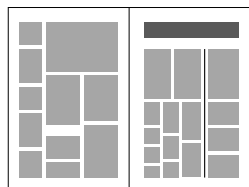
Modeling Norms of Turn-Taking in Multi-Party Conversation

Kornel Laskowski

Carnegie Mellon University
Pittsburgh PA, USA

13 July, 2010

Comparing Written Documents

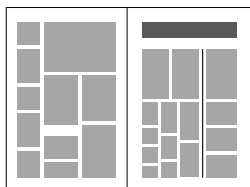
 w_1  w_2

- If have a form for a density model Θ of word sequences, and
- techniques for estimating the parameters of Θ from data, and
- techniques for estimating $P(w | \Theta)$,
- Can easily compare w_1 with w_2 , with respect to
- how far each deviates from the norms encoded in Θ

Comparing Written Documents



w_1



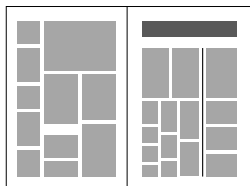
w_2

- If have a form for a density model Θ of word sequences, and
- techniques for estimating the parameters of Θ from data, and
- techniques for estimating $P(\mathbf{w} | \Theta)$,
- Can easily compare w_1 with w_2 , with respect to
- how far each deviates from the norms encoded in Θ

Comparing Written Documents



\mathbf{w}_1

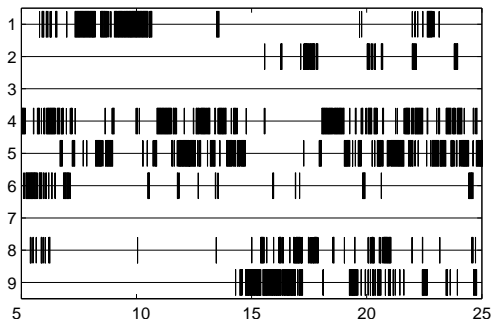


\mathbf{w}_2

- If have a form for a density model Θ of word sequences, and
- techniques for estimating the parameters of Θ from data, and
- techniques for estimating $P(\mathbf{w} | \Theta)$,
- Can easily compare \mathbf{w}_1 with \mathbf{w}_2 , with respect to
- how far each deviates from the norms encoded in Θ

Representing Spoken Documents

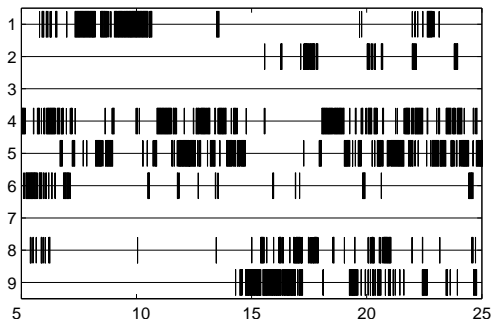
- $K > 1$ sources (participants)



- interaction chronograph (Chapple, 1939)
- aka vocal interaction record (Dabbs & Ruback, 1987)
- a **content-independent** representation

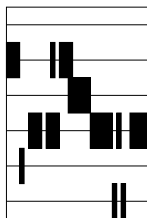
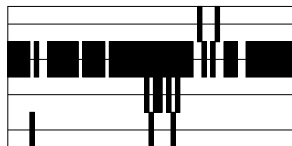
Representing Spoken Documents

- $K > 1$ sources (participants)



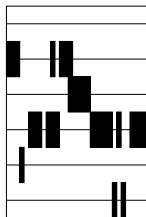
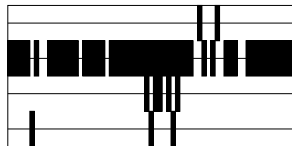
- interaction chronograph (Chapple, 1939)
- aka vocal interaction record (Dabbs & Ruback, 1987)
- a **content-independent** representation

Comparing Spoken Documents


 Q_1

 Q_2

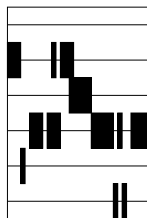
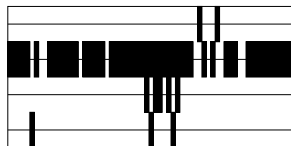
- Cannot compare spoken documents Q_1 and Q_2
 - no candidate form for model Θ ,
 - no techniques of estimating the parameters of Θ from data,
 - no techniques of estimating $P(Q|\Theta)$.
- Comparison must remain **manual** and **qualitative**.

Comparing Spoken Documents


 Q_1

 Q_2

- Cannot compare spoken documents Q_1 and Q_2
 - no candidate form for model Θ ,
 - no techniques of estimating the parameters of Θ from data,
 - no techniques of estimating $P(Q|\Theta)$.
- Comparison must remain **manual** and **qualitative**.

Comparing Spoken Documents


 Q_1

 Q_2

- Cannot compare spoken documents Q_1 and Q_2
 - no candidate form for model Θ ,
 - no techniques of estimating the parameters of Θ from data,
 - no techniques of estimating $P(Q | \Theta)$.
- Comparison must remain **manual** and **qualitative**.

Wouldn't It Be Nice if We Could ...

- Compare meetings in organizations to determine which interaction patterns correlate with successful business practice?
- Find instants within conversations where interaction management breaks down (hotspots)?
- Classify conversations according to a spectrum of interactivity?
- Contrast conversational behavior across languages and cultures?
- Assess emergent turn-taking performance in dialogue systems?

Outline of this Talk

- 1 Compositional Modeling Framework
- 2 Direct estimation in compositional models
- 3 “Extended Degree of Overlap” (EDO) Model
- 4 Experiments with Naturally Occurring Conversation
 - within-conversation prediction
 - across-conversation prediction
- 5 Summary

Turns and Talk Spurts ...

- ① **turn-taking**: generally observed phenomenon in conversation
- ② but **turn**: ?? (no generally agreed upon definition)
- ③ here: **turn** \equiv **(talk) spurt** (Norwine & Murphy, 1938)
 - prefer “speech regions uninterrupted by pauses longer than 500 ms” (Shriberg et al, 2001)
 - with a threshold $T_{\square} = 300$ ms (NIST RT Evaluations, 2002–)
 - similar to **inter-pausal unit**, $T_{\square} = 100$ ms (Koiso et al, 1998)
- ④ specific choice may have minor numerical consequences
- ⑤ but no impact on the mathematical viability of the modeling techniques of this work

Turns and Talk Spurts ...

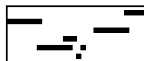
- ① **turn-taking**: generally observed phenomenon in conversation
- ② but **turn**: ?? (no generally agreed upon definition)
- ③ here: **turn** \equiv **(talk) spurt** (Norwine & Murphy, 1938)
 - prefer “speech regions uninterrupted by pauses longer than 500 ms” (Shriberg et al, 2001)
 - with a threshold $T_{\square} = 300$ ms (NIST RT Evaluations, 2002–)
 - similar to **inter-pausal unit**, $T_{\square} = 100$ ms (Koiso et al, 1998)
- ④ specific choice may have minor numerical consequences
- ⑤ but no impact on the mathematical viability of the modeling techniques of this work

Turns and Talk Spurts ...

- ① **turn-taking**: generally observed phenomenon in conversation
- ② but **turn**: ?? (no generally agreed upon definition)
- ③ here: **turn** \equiv **(talk) spurt** (Norwine & Murphy, 1938)
 - prefer “speech regions uninterrupted by pauses longer than 500 ms” (Shriberg et al, 2001)
 - with a threshold $T_{\square} = 300$ ms (NIST RT Evaluations, 2002–)
 - similar to **inter-pausal unit**, $T_{\square} = 100$ ms (Koiso et al, 1998)
- ④ specific choice may have minor numerical consequences
- ⑤ but no impact on the mathematical viability of the modeling techniques of this work

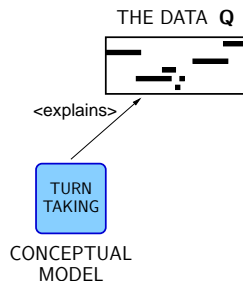
... and Models of Either or Both

THE DATA Q



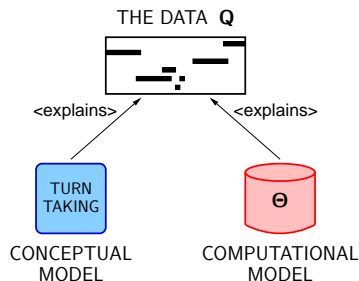
- data Q: speech activity in time and across participants
- **turn-taking**: name of a conceptual model
- **goal in this work**: propose a **computational** model

... and Models of Either or Both



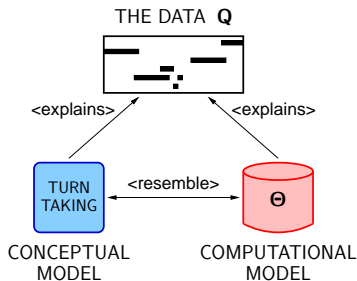
- data **Q**: speech activity in time and across participants
- **turn-taking**: name of a conceptual model
- **goal in this work**: propose a **computational model**

... and Models of Either or Both



- data Q : speech activity in time and across participants
- **turn-taking**: name of a conceptual model
- **goal in this work**: propose a **computational** model

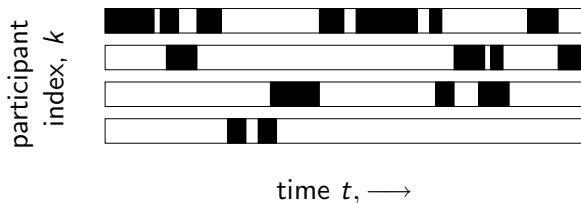
... and Models of Either or Both



- data Q : speech activity in time and across participants
- **turn-taking**: name of a conceptual model
- **goal in this work**: propose a **computational** model

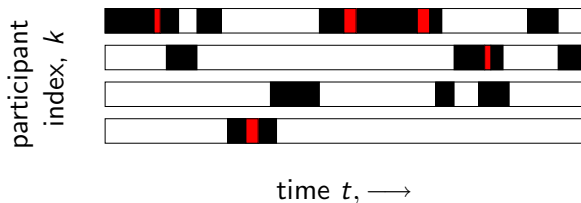
What Is Q?

- 1 time-align speech □/■ activity of all K participants
- 2 close all □ gaps shorter than $T_{\square} = 300$ ms
- 3 discretize with a frame step of 100 ms
- 4 yields a $Q \in \{\square, \blacksquare\}^{K \times T}$



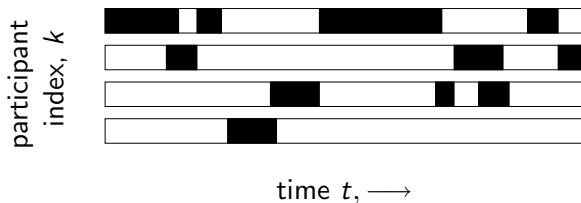
What Is Q?

- 1 time-align speech □/■ activity of all K participants
- 2 close all □ gaps shorter than $T_{\square} = 300$ ms
- 3 discretize with a frame step of 100 ms
- 4 yields a $Q \in \{\square, \blacksquare\}^{K \times T}$



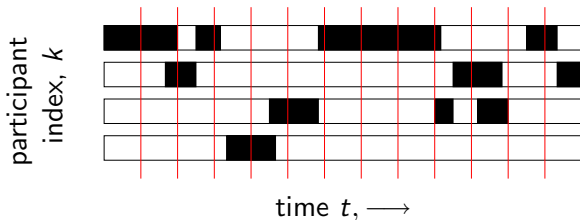
What Is Q?

- 1 time-align speech □/■ activity of all K participants
- 2 close all □ gaps shorter than $T_{\square} = 300$ ms
- 3 discretize with a frame step of 100 ms
- 4 yields a $Q \in \{\square, \blacksquare\}^{K \times T}$



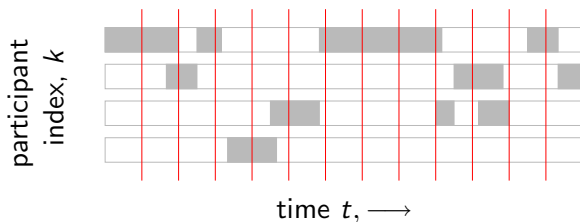
What Is Q?

- 1 time-align speech □/■ activity of all K participants
- 2 close all □ gaps shorter than $T_{\square} = 300$ ms
- 3 discretize with a frame step of 100 ms
- 4 yields a $Q \in \{\square, \blacksquare\}^{K \times T}$



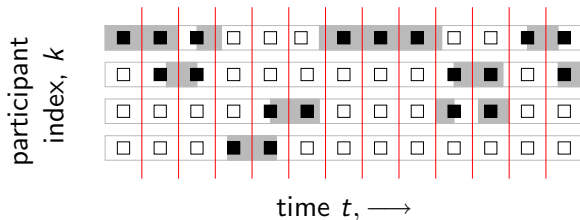
What Is Q?

- ① time-align speech □/■ activity of all K participants
- ② close all □ gaps shorter than $T_{\square} = 300$ ms
- ③ discretize with a frame step of 100 ms
- ④ yields a $Q \in \{\square, \blacksquare\}^{K \times T}$



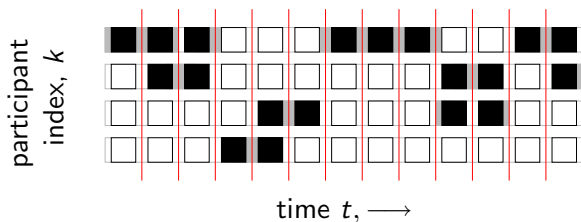
What Is Q?

- 1 time-align speech \square/\blacksquare activity of all K participants
 - 2 close all \square gaps shorter than $T_{\square} = 300$ ms
 - 3 discretize with a frame step of 100 ms
- 4 yields a $Q \in \{\square, \blacksquare\}^{K \times T}$



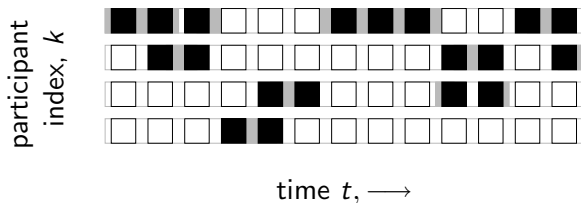
What Is Q?

- ① time-align speech □/■ activity of all K participants
- ② close all □ gaps shorter than $T_{\square} = 300$ ms
- ③ discretize with a frame step of 100 ms
- ④ yields a $Q \in \{\square, \blacksquare\}^{K \times T}$



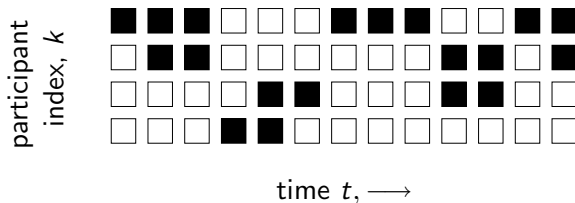
What Is Q?

- 1 time-align speech □/■ activity of all K participants
- 2 close all □ gaps shorter than $T_{\square} = 300$ ms
- 3 discretize with a frame step of 100 ms
- 4 yields a $Q \in \{\square, \blacksquare\}^{K \times T}$



What Is Q?

- 1 time-align speech □/■ activity of all K participants
- 2 close all □ gaps shorter than $T_{\square} = 300$ ms
- 3 discretize with a frame step of 100 ms
- 4 yields a $\mathbf{Q} \in \{\square, \blacksquare\}^{K \times T}$



Modeling A Vector-Valued Markov Process

- model conversation as a Markov process

$$\dots, \mathbf{q}_{t-1} = \begin{bmatrix} \square \\ \square \\ \blacksquare \\ \square \end{bmatrix}, \mathbf{q}_t = \begin{bmatrix} \blacksquare \\ \square \\ \square \\ \square \end{bmatrix}, \mathbf{q}_{t+1} = \begin{bmatrix} \blacksquare \\ \square \\ \square \\ \square \end{bmatrix}, \dots$$

- then (assuming first-order model Θ)

- easy to train, just count bigrams !

Modeling A Vector-Valued Markov Process

- model conversation as a Markov process

$$\dots, \mathbf{q}_{t-1} = \begin{bmatrix} \square \\ \square \\ \blacksquare \\ \square \end{bmatrix}, \mathbf{q}_t = \begin{bmatrix} \blacksquare \\ \square \\ \square \\ \square \end{bmatrix}, \mathbf{q}_{t+1} = \begin{bmatrix} \blacksquare \\ \square \\ \square \\ \square \end{bmatrix}, \dots$$

- then (assuming first-order model Θ)

$$\begin{aligned} P(\mathbf{Q} | \Theta) &= P_0 \prod_{t=1}^T P(\mathbf{q}_t | \mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_{t-1}, \Theta) \\ &\doteq P_0 \prod_{t=1}^T P(\mathbf{q}_t | \mathbf{q}_{t-1}, \Theta) \end{aligned}$$

- easy to train, just count bigrams !

Modeling A Vector-Valued Markov Process

- model conversation as a Markov process

$$\dots, \mathbf{q}_{t-1} = \begin{bmatrix} \square \\ \square \\ \blacksquare \\ \square \end{bmatrix}, \mathbf{q}_t = \begin{bmatrix} \blacksquare \\ \square \\ \square \\ \square \end{bmatrix}, \mathbf{q}_{t+1} = \begin{bmatrix} \blacksquare \\ \square \\ \square \\ \square \end{bmatrix}, \dots$$

- then (assuming first-order model Θ)

$$\begin{aligned} P(\mathbf{Q} | \Theta) &= P_0 \prod_{t=1}^T P(\mathbf{q}_t | \mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_{t-1}, \Theta) \\ &\doteq P_0 \prod_{t=1}^T P(\mathbf{q}_t | \mathbf{q}_{t-1}, \Theta) \end{aligned}$$

- easy to train, just count bigrams !

Some Past Work

- interaction chronography (Chapple, 1939; Chapple, 1949)
- **modeling in dialogue:** $K = 2$
 - telecommunications (Norwine & Murphy, 1938; Brady, 1969)
 - sociolinguistics (Jaffe & Feldstein, 1970)
 - psycholinguistics (Dabbs & Ruback, 1987)
 - dialogue systems (cf. Raux, 2008)
- **modeling in multi-party settings:** $K > 2$
 - psycholinguistics: GroupTalk model (Dabbs et al, 1987)
 - not quite serviceable for current task
 - pre-ASR segmentation: EDO model (Laskowski & Schultz, 2007)

Defining Turn-Taking Perplexity (PPL)

In language modeling,

w : word $\|\mathbf{w}\|$ -sequence

Θ : “language model”

$$\text{NLL} = -\frac{1}{\|\mathbf{w}\|} \log_e P(\mathbf{w} | \Theta)$$

$$\text{PPL} = 10^{\text{NLL}}$$

Here,

Q : $K \times T$ chronograph

Θ : “turn-taking model”

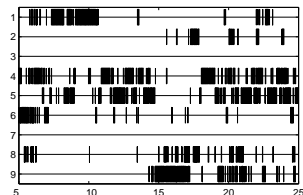
$$\text{NLL} = -\frac{1}{KT} \log_2 P(\mathbf{Q} | \Theta)$$

$$\begin{aligned} \text{PPL} &= 2^{\text{NLL}} \\ &= (P(\mathbf{Q} | \Theta))^{-1/KT} \end{aligned}$$

- Can also window negative log-likelihood (NLL) to yield a measure of local perplexity (in time).

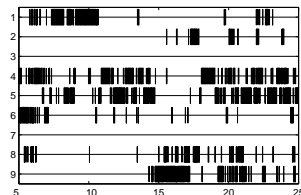
An Example of the Perplexity Trajectory in Time

- 1 obtain \mathbf{Q} for ICSI Bmr024
 - $K = 9$ participants
 - ≈ 55 minutes
- 2 train the Θ model
- 3 compute local perplexity using 60-second Hamming windows



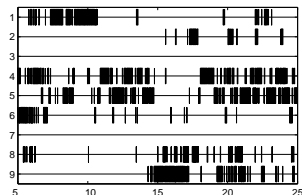
An Example of the Perplexity Trajectory in Time

- 1 obtain \mathbf{Q} for ICSI Bmr024
 - $K = 9$ participants
 - ≈ 55 minutes
- 2 train the Θ model
- 3 compute local perplexity using 60-second Hamming windows



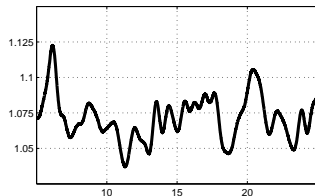
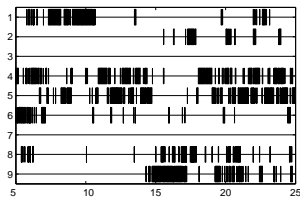
An Example of the Perplexity Trajectory in Time

- 1 obtain \mathbf{Q} for ICSI Bmr024
 - $K = 9$ participants
 - ≈ 55 minutes
- 2 train the Θ model
- 3 compute local perplexity using 60-second Hamming windows



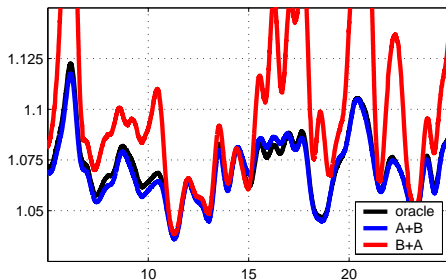
An Example of the Perplexity Trajectory in Time

- 1 obtain \mathbf{Q} for ICSI Bmr024
 - $K = 9$ participants
 - ≈ 55 minutes
- 2 train the Θ model
- 3 compute local perplexity using 60-second Hamming windows



Generalization

- **A+B**: train on first half (A) only, test on A
- **B+A**: train on second half (B) only, test on A



- a multi-participant compositional model Θ generalizes poorly
- even to other parts of the **same conversation!**

Circumscribing Model Complexity

$$P(\mathbf{Q}) \approx P_0 \cdot \prod_{t=1}^T P(\mathbf{q}_t | \mathbf{q}_{t-1}, \Theta^{CD}) \quad 2^K \cdot (2^K - 1)$$

Circumscribing Model Complexity

$$\begin{aligned}
 P(\mathbf{Q}) &\approx P_0 \cdot \prod_{t=1}^T P(\mathbf{q}_t | \mathbf{q}_{t-1}, \Theta^{CD}) && 2^K \cdot (2^K - 1) \\
 &\approx P_0 \cdot \prod_{t=1}^T \prod_{k=1}^K P(\mathbf{q}_t[k] | \mathbf{q}_{t-1}, \Theta_k^{CI}) && K \cdot 2^K
 \end{aligned}$$

Circumscribing Model Complexity

$$P(\mathbf{Q}) \approx P_0 \cdot \prod_{t=1}^T P(\mathbf{q}_t | \mathbf{q}_{t-1}, \Theta^{CD}) \quad 2^K \cdot (2^K - 1)$$

$$\approx P_0 \cdot \prod_{t=1}^T \prod_{k=1}^K P(\mathbf{q}_t[k] | \mathbf{q}_{t-1}, \Theta_k^{CI}) \quad K \cdot 2^K$$

$$\approx P_0 \cdot \prod_{t=1}^T \prod_{k=1}^K P(\mathbf{q}_t[k] | \mathbf{q}_{t-1}, \Theta_{any}^{CI}) \quad 2^K$$

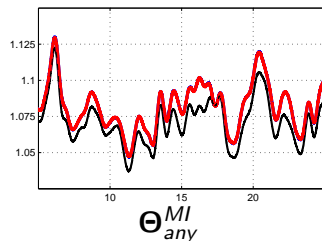
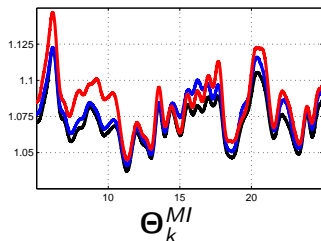
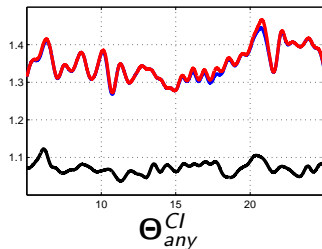
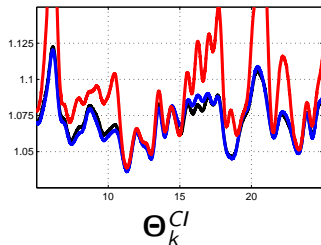
Circumscribing Model Complexity

$$\begin{aligned}
 P(\mathbf{Q}) &\approx P_0 \cdot \prod_{t=1}^T P\left(\mathbf{q}_t \mid \mathbf{q}_{t-1}, \Theta^{CD}\right) && 2^K \cdot (2^K - 1) \\
 &\approx P_0 \cdot \prod_{t=1}^T \prod_{k=1}^K P\left(\mathbf{q}_t[k] \mid \mathbf{q}_{t-1}, \Theta_k^{CI}\right) && K \cdot 2^K \\
 &\approx P_0 \cdot \prod_{t=1}^T \prod_{k=1}^K P\left(\mathbf{q}_t[k] \mid \mathbf{q}_{t-1}, \Theta_{any}^{CI}\right) && 2^K \\
 &\approx P_0 \cdot \prod_{t=1}^T \prod_{k=1}^K P\left(\mathbf{q}_t[k] \mid \mathbf{q}_{t-1}[k], \Theta_k^{MI}\right) && K \cdot 2
 \end{aligned}$$

Circumscribing Model Complexity

$$\begin{aligned}
 P(\mathbf{Q}) &\approx P_0 \cdot \prod_{t=1}^T P\left(\mathbf{q}_t \mid \mathbf{q}_{t-1}, \Theta^{CD}\right) && 2^K \cdot (2^K - 1) \\
 &\approx P_0 \cdot \prod_{t=1}^T \prod_{k=1}^K P\left(\mathbf{q}_t[k] \mid \mathbf{q}_{t-1}, \Theta_k^{CI}\right) && K \cdot 2^K \\
 &\approx P_0 \cdot \prod_{t=1}^T \prod_{k=1}^K P\left(\mathbf{q}_t[k] \mid \mathbf{q}_{t-1}, \Theta_{any}^{CI}\right) && 2^K \\
 &\approx P_0 \cdot \prod_{t=1}^T \prod_{k=1}^K P\left(\mathbf{q}_t[k] \mid \mathbf{q}_{t-1}[k], \Theta_k^{MI}\right) && K \cdot 2 \\
 &\approx P_0 \cdot \prod_{t=1}^T \prod_{k=1}^K P\left(\mathbf{q}_t[k] \mid \mathbf{q}_{t-1}[k], \Theta_{any}^{MI}\right) && 2
 \end{aligned}$$

Circumscribing Model Complexity: Perplexity Trajectories



Directly Estimated Compositional Model Limitations

- Direct compositional models either:
 - 1 Periodically underperform — grossly — due to overfitting, or
 - 2 Do not model interaction (Θ_k^{MI} , Θ_{any}^{MI}).
- Variants which **do** model interaction (Θ^{CD} , Θ_k^{CI} , Θ_{any}^{CI}):
 - ❑ Fail to exhibit K -independence.
 - the number and identity of states is a function of K
 - cannot be trained on conversations with K participants, and applied to conversations with $K' \neq K$ participants
 - ❑ Fail to exhibit N -independence.
 - sensitive to participant index assignment
 - perplexities differ if Ω is rotated by arbitrary rotation R
 - exhaustive rotation during training has complexity $K!$

Directly Estimated Compositional Model Limitations

- Direct compositional models either:
 - ① Periodically underperform — grossly — due to overfitting, or
 - ② Do not model interaction (Θ_k^{MI} , Θ_{any}^{MI}).

- Variants which **do** model interaction (Θ^{CD} , Θ_k^{CI} , Θ_{any}^{CI}):
 - ① Fail to exhibit *K-independence*.
 - the number and identity of states is a function of K
 - cannot be trained on conversations with K participants, and applied to conversations with $K' \neq K$ participants
 - ② Fail to exhibit *R-independence*.
 - sensitive to participant index assignment
 - perplexities differ if \mathbf{Q} is rotated by arbitrary rotation \mathbf{R}
 - exhaustive rotation during training has complexity $K!$
 - ③ Insufficiently parsimonious \rightarrow theoretically vacuous.

Directly Estimated Compositional Model Limitations

- Direct compositional models either:
 - ① Periodically underperform — grossly — due to overfitting, or
 - ② Do not model interaction (Θ_k^{MI} , Θ_{any}^{MI}).

- Variants which **do** model interaction (Θ^{CD} , Θ_k^{CI} , Θ_{any}^{CI}):
 - ① Fail to exhibit *K-independence*.
 - the number and identity of states is a function of K
 - cannot be trained on conversations with K participants, and applied to conversations with $K' \neq K$ participants
 - ② Fail to exhibit **R-independence**.
 - sensitive to participant index assignment
 - perplexities differ if \mathbf{Q} is rotated by arbitrary rotation \mathbf{R}
 - exhaustive rotation during training has complexity $K!$
 - ③ Insufficiently parsimonious → theoretically vacuous.

Directly Estimated Compositional Model Limitations

- Direct compositional models either:
 - ① Periodically underperform — grossly — due to overfitting, or
 - ② Do not model interaction (Θ_k^{MI} , Θ_{any}^{MI}).

- Variants which **do** model interaction (Θ^{CD} , Θ_k^{CI} , Θ_{any}^{CI}):
 - ① Fail to exhibit *K-independence*.
 - the number and identity of states is a function of K
 - cannot be trained on conversations with K participants, and applied to conversations with $K' \neq K$ participants
 - ② Fail to exhibit **R-independence**.
 - sensitive to participant index assignment
 - perplexities differ if \mathbf{Q} is rotated by arbitrary rotation \mathbf{R}
 - exhaustive rotation during training has complexity $K!$
 - ③ Insufficiently parsimonious \rightarrow theoretically vacuous.

Degree-of-Overlap (DO) Model

Replace the probability of transition between **compositional states** by the probability of transition between **the number of participants speaking simultaneously** in them:

$$P(\mathbf{q}_t | \mathbf{q}_{t-1}, \Theta^{CD}) \doteq \alpha P(\|\mathbf{q}_t\| | \|\mathbf{q}_{t-1}\|, \Theta^{DO})$$

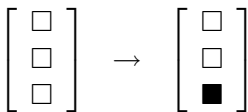
where

$$\|\mathbf{q}\| = \sum_{k=1}^K \delta(\mathbf{q}[k], \blacksquare)$$

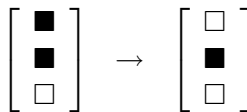
$$\in \{0, 1, \dots, K\}$$

Model contains only $K \cdot (K + 1)$ free parameters.

DO Model Examples

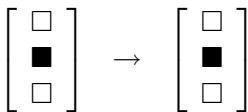


0 → 1

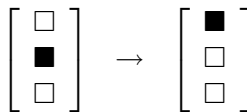


2 → 1

But unfortunately,



1 → 1



1 → 1

Extended-Degree-of-Overlap Model (Laskowski & Schultz, 2007)

Extend the “to” state,

- to a 2-element vector, with the
- **number of participants speaking in both \mathbf{q}_{t-1} and \mathbf{q}_t :**

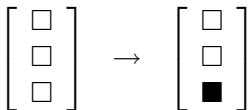
$$P\left(\mathbf{q}_t \mid \mathbf{q}_{t-1}, \Theta^{CD}\right) \doteq \alpha P\left(\|\mathbf{q}_t\|, \|\mathbf{q}_t \cdot \mathbf{q}_{t-1}\| \mid \|\mathbf{q}_{t-1}\|, \Theta^{EDO}\right)$$

where

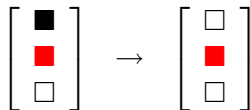
$$(\mathbf{q} \cdot \mathbf{q}') [k] = \begin{cases} \blacksquare & \text{if } \mathbf{q} [k] = \blacksquare \text{ and } \mathbf{q}' [k] = \blacksquare \\ \square & \text{otherwise} \end{cases}$$

Also easy to train: just count the bigrams!

EDO Model Examples

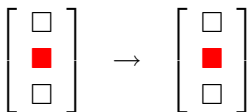


$$0 \rightarrow [0, 1]$$

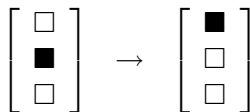


$$2 \rightarrow [1, 1]$$

And as desired,



$$1 \rightarrow [1, 1]$$



$$1 \rightarrow [0, 1]$$

EDO Desiderata Scorecard

- 1 The EDO Model achieves **R**-invariance:
 - $\| \cdot \|$ is a sum
 - commutative \rightarrow rotation-independent
 - **results same regardless of participant index assignment**
- 2 The EDO Model achieves *K*-invariance:
 - sums performed over ■-state participants only
 - remaining participants, in □, ignored
 - **can apply to any K , with $K_{train} \neq K_{test}$**
- 3 The EDO state space is tractably small:
 - parameters are individually meaningful
 - can be further constrained by mapping all $\| \cdot \| > K_{max}$ to K_{max}

EDO Desiderata Scorecard

- 1 The EDO Model achieves **R**-invariance:
 - $\| \cdot \|$ is a sum
 - commutative \rightarrow rotation-independent
 - **results same regardless of participant index assignment**

- 2 The EDO Model achieves **K**-invariance:
 - sums performed over \blacksquare -state participants only
 - remaining participants, in \square , ignored
 - **can apply to any K , with $K_{train} \neq K_{test}$**

- 3 The EDO state space is tractably small:
 - parameters are individually meaningful
 - can be further constrained by mapping all $\| \cdot \| > K_{max}$ to K_{max}

EDO Desiderata Scorecard

- 1 The EDO Model achieves **R**-invariance:
 - $\| \cdot \|$ is a sum
 - commutative \rightarrow rotation-independent
 - **results same regardless of participant index assignment**

- 2 The EDO Model achieves **K**-invariance:
 - sums performed over \blacksquare -state participants only
 - remaining participants, in \square , ignored
 - **can apply to any K , with $K_{train} \neq K_{test}$**

- 3 The EDO state space is tractably small:
 - parameters are individually meaningful
 - can be further constrained by mapping all $\| \cdot \| > K_{max}$ to K_{max}

Data

ICSI Meeting Corpus (Janin et al, 2003; Shriberg et al, 2004):

- 75 meetings
- would have occurred even if they had not been recorded
- approximately 1 hour long
- 3–9 participants each
- forced-alignment-mediated ■/□ references available

Same-Conversation Training

- iterate over all meetings:
 - ① split meeting into halves A and B
 - ② A+B condition: { train A, score A } & { train B, score B }
 - ③ B+A condition: { train A, score B } & { train B, score A }
- scoring intervals of the **same** conversation
 - number of participants K invariable
 - participant index assignment \mathbf{R} invariable
- assess
 - independent-participant model Θ_{any}^{MI}
 - compositional models: Θ^{CD} , Θ_k^{CI} , Θ_k^{MI}
 - EDO model, with $K_{max} = K$

Same-Conversation Results

Model	PP, A+B		PP, B+A	
	“all”	“sub”	“all”	“sub”
oracle	1.0905	1.6444	1.0905	1.6444
Θ^{CD}	1.0905	1.6444	1.1225	1.8395
$\{\Theta_k^{CI}\}$	1.0915	1.6576	1.1156	1.7809
$\{\Theta_k^{MI}\}$	1.0978	1.7236	1.1086	1.7950
Θ^{MI}	1.1046	1.8047	1.1047	1.8059
Θ^{EDO}	1.0977	1.7257	1.0985	1.7323

- on unseen same-conversation data, EDO model outperforms all compositional models

Same-Conversation Results

Model	PP, A+B		PP, B+A	
	“all”	“sub”	“all”	“sub”
oracle	1.0905	1.6444	1.0905	1.6444
Θ^{CD}	1.0905	1.6444	1.1225	1.8395
$\{\Theta_k^{CI}\}$	1.0915	1.6576	1.1156	1.7809
$\{\Theta_k^{MI}\}$	1.0978	1.7236	1.1086	1.7950
Θ^{MI}	1.1046	1.8047	1.1047	1.8059
Θ^{EDO}	1.0977	1.7257	1.0985	1.7323

- on unseen same-conversation data, EDO model outperforms all compositional models

Other-Conversation Training

- iterate over all meetings:
 - 1 train on remaining 74 meetings
 - 2 score held out meeting
- scoring **different** conversations
 - number of participants K **variable**
 - participant index assignment **R unknown**
- assess
 - independent-participant model Θ_{any}^{MI}
 - EDO model, over a range of K_{max}

Other-Conversation Results

Model	PP		Δ PP (%)	
	"all"	"sub"	"all"	"sub"
oracle	1.0921	1.6616	-100	-100
Θ^{MI}	1.1051	1.8170	0	0
Θ^{EDO} (6)	1.0992	1.7405	-45	-49
Θ^{EDO} (5)	1.0968	1.7127	-64	-67
Θ^{EDO} (4)	1.0953	1.6947	-75	-79
Θ^{EDO} (3)	1.1082	1.8502	+24	+21

Conclusions

- 1 **The Extended-Degree-of-Overlap (EDO) model:**
 - can be used as a density estimator for *any* conversation;
 - *any* conversation can be used to infer its parameters.
- 2 The EDO model vastly outperforms standard single-participant alternatives,
 - e.g., those used in speech activity detection,
 - by 75%rel from the oracle.
- 3 Participant behavior is (in measurable part) predicted by interlocutor behavior.

Conclusions

- 1 The Extended-Degree-of-Overlap (EDO) model:**
 - can be used as a density estimator for *any* conversation;
 - *any* conversation can be used to infer its parameters.
- 2 The EDO model vastly outperforms standard single-participant alternatives,**
 - e.g., those used in speech activity detection,
 - by 75%rel from the oracle.
- 3 Participant behavior is (in measurable part) predicted by interlocutor behavior.**

Conclusions

- 1 **The Extended-Degree-of-Overlap (EDO) model:**
 - can be used as a density estimator for *any* conversation;
 - *any* conversation can be used to infer its parameters.
- 2 **The EDO model vastly outperforms standard single-participant alternatives,**
 - e.g., those used in speech activity detection,
 - by 75%rel from the oracle.
- 3 **Participant behavior is (in measurable part) predicted by interlocutor behavior.**

Contributions

- 1 A **framework for computing perplexity** in ■/□ interaction chronographs;
- 2 Evidence of the **unsuitability of directly estimated compositional models**;
- 3 Evidence of the **suitability of the EDO model** as a baseline for future research; and
- 4 Empirical assessment of EDO performance on one of the **largest multi-party corpora of naturally occurring conversation**.

Contributions

- 1 A **framework for computing perplexity** in ■/□ interaction chronographs;
- 2 Evidence of the **unsuitability of directly estimated compositional models**;
- 3 Evidence of the **suitability of the EDO model** as a baseline for future research; and
- 4 Empirical assessment of EDO performance on one of the **largest multi-party corpora of naturally occurring conversation**.

Contributions

- 1 A **framework for computing perplexity** in ■/□ interaction chronographs;
- 2 Evidence of the **unsuitability of directly estimated compositional models**;
- 3 Evidence of the **suitability of the EDO model** as a baseline for future research; and
- 4 Empirical assessment of EDO performance on one of the **largest multi-party corpora of naturally occurring conversation**.

Contributions

- 1 A **framework for computing perplexity** in ■/□ interaction chronographs;
- 2 Evidence of the **unsuitability of directly estimated compositional models**;
- 3 Evidence of the **suitability of the EDO model** as a baseline for future research; and
- 4 Empirical assessment of EDO performance on one of the **largest multi-party corpora of naturally occurring conversation**.

Impact/Recommendations

- 1 The precise EDO model formulation complements and possibly supersedes the (heretofore usefully) imprecise notion of taking turns.
 - both account for the distribution of speech in time and across participants
- 2 The EDO model and PPL measure provide a computational means for corroborating the findings of conversation analysis (in particular) on vastly larger collections of conversation than have been analyzed to date.
- 3 The EDO model and PPL measure provide an unambiguous measure of spoken document similarity; can now easily:
 - compare turn-taking across conversations,
 - find where turn-taking deviates from norms (“hotspots”), and
 - assess turn-taking appropriateness in dialogue systems.

Impact/Recommendations

- 1 The precise EDO model formulation complements and possibly supersedes the (heretofore usefully) imprecise notion of taking turns.
 - both account for the distribution of speech in time and across participants
- 2 The EDO model and PPL measure provide a computational means for corroborating the findings of conversation analysis (in particular) on vastly larger collections of conversation than have been analyzed to date.
- 3 The EDO model and PPL measure provide an unambiguous measure of spoken document similarity; can now easily:
 - compare turn-taking across conversations,
 - find where turn-taking deviates from norms (“hotspots”), and
 - assess turn-taking appropriateness in dialogue systems.

Impact/Recommendations

- 1 The precise EDO model formulation complements and possibly supersedes the (heretofore usefully) imprecise notion of taking turns.
 - both account for the distribution of speech in time and across participants
- 2 The EDO model and PPL measure provide a computational means for corroborating the findings of conversation analysis (in particular) on vastly larger collections of conversation than have been analyzed to date.
- 3 The EDO model and PPL measure provide an unambiguous measure of spoken document similarity; can now easily:
 - compare turn-taking across conversations,
 - find where turn-taking deviates from norms (“hotspots”), and
 - assess turn-taking appropriateness in dialogue systems.

Thank You!