

# SITS: A Hierarchical Nonparametric Model using Speaker Identity for Topic Segmentation in Multiparty Conversations

## Supplementary Material

**Viet-An Nguyen**  
Department of Computer Science  
and UMIACS  
University of Maryland  
College Park, MD  
vietan@cs.umd.edu

**Jordan Boyd-Graber**  
iSchool  
and UMIACS  
University of Maryland  
College Park, MD  
jbg@umiacs.umd.edu

**Philip Resnik**  
Department of Linguistics  
and UMIACS  
University of Maryland  
College Park, MD  
resnik@umd.edu

In this supplementary document, we derive Gibbs sampling equations for for both parametric and nonparametric SITS.

## 1 Nonparametric SITS

In this section, we describe the general Gibbs sampler for our nonparametric model. The state space of our chain is the topic indices assigned to all tokens  $\mathbf{z} = \{z_{c,t,n}\}$  and topic shifts assigned to all turns  $\mathbf{l} = \{l_{c,t}\}$ . In order to obtain  $z_{c,t,n}$  we need to know the *path* assigned for token  $w_{c,t,n}$  through the hierarchy which includes  $k_{c,t,n}^T$ ,  $k_{c,s,j}^S$  and  $k_{c,i}^C$ . For ease of reference, the meaning of these symbols (and others used in this appendix) are listed in Table 1. Figure 1c shows the relationship among the latent variables in our model. As shown, once we know the three seating assignments  $k_{c,t,n}^T$ ,  $k_{c,s,j}^S$  and  $k_{c,i}^C$ ,  $z_{c,t,n}$  can obtained by

$$z_{c,t,n} \equiv k_{c,k_{c,t,n}^T, k_{c,s,j}^S, k_{c,i}^C}^C \quad (1)$$

To perform inference, we marginalize over all other latent variables and alternate between sampling paths  $k$  and the sampling topic shifts  $l$ .

**Sampling topic assignments** Before deriving the sampling equations, let us use  $f_k^{-c,t,n}(w_{c,t,n})$  to denote the conditional density of token  $w_{c,t,n}$  under topic  $k$  given all other items except  $w_{c,t,n}$

$$f_k^{-c,t,n}(w_{c,t,n}) = \frac{\int_{\phi_k} P(\mathbf{w} | \phi_k) P(\phi_k | \lambda) d\phi_k}{\int_{\phi_k} P(\mathbf{w}_{-c,t,n} | \phi_k) P(\phi_k | \lambda) d\phi_k} = \begin{cases} \frac{TW_{k,w_{c,t,n}} + \lambda}{TW_{k,\cdot} + V\lambda}, & \text{if } k \text{ exists;} \\ \frac{1}{V}, & \text{if } k \text{ is new.} \end{cases} \quad (2)$$

To sample the path, we take a similar approach to the the first sampling method described in Section 5.1 of [Teh et al., 2006]. We first sample the table assignment for each customer. A customer can sit at an existing table  $j$  or create a new one  $j^{new}$ . If a new table is created, an item in the restaurant’s local menu will be sampled. Again, the table can be served with an existing local menu item  $i$  or a new one  $i^{new}$ . If a new local menu item is sampled, it has to be assigned a global item. An existing global dish  $k$  or a new one  $k^{new}$  can be sampled. The details of each step is as follows:

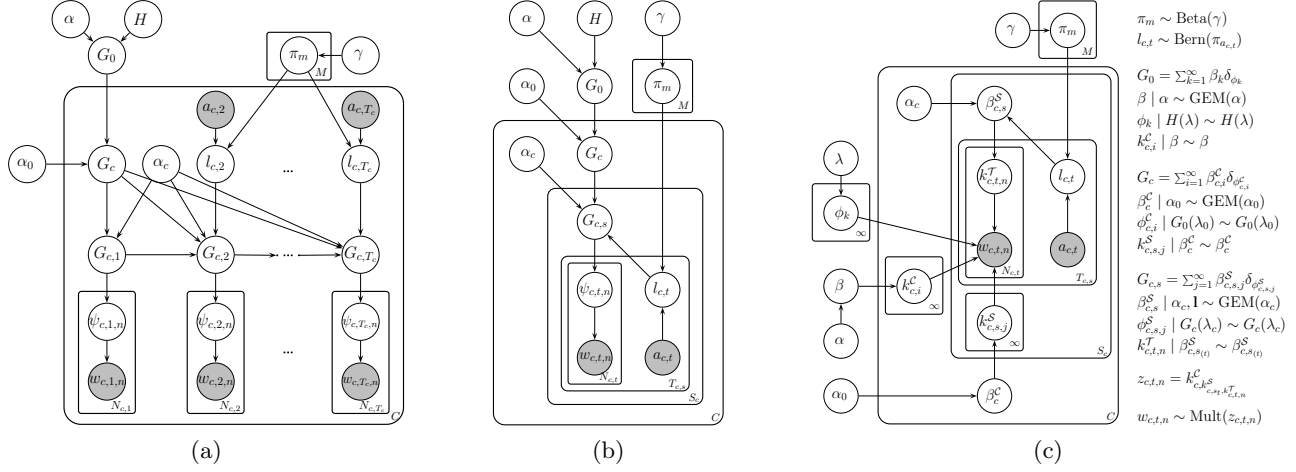


Figure 1: Graphical model representations of our nonparametric model. (a) the original representation according to the generative process; (b) the representation using the notion of *segment*. We use  $S_c$  to denote the number of segments in conversation  $c$  and  $T_{c,t}$  to denote the number of turns in segment  $s$  of conversation  $c$ ; (c) the representation where explicit path assignments are shown.

**Sampling table assignment to customers  $k_{c,t,n}^T$**  We start sampling the path by sampling the table assignment to each customer

$$\begin{aligned}
 &P(k_{c,t,n}^T = j \mid \mathbf{k}_{-c,t,n}^T, \mathbf{k}^S, \mathbf{k}^C, \mathbf{l}, \mathbf{w}, *) \\
 &\propto P(k_{c,t,n}^T = j \mid \mathbf{k}_{-c,t,n}^T) P(w_{c,t,n} \mid k_{c,t,n}^T, \mathbf{k}_{-c,t,n}^T, \mathbf{w}_{-c,t,n}, \mathbf{k}^S, \mathbf{k}^C, \mathbf{l}, *) \\
 &= \begin{cases} \frac{N_{c,st,j}^S}{N_{c,st,\cdot}^S + \alpha_c} f_{k_{c,st,j}^S}^{-c,t,n}(w_{c,t,n}), & \text{if } j \text{ exists;} \\ \frac{\alpha_c}{N_{c,st,\cdot}^S + \alpha_c} P(w_{c,t,n} \mid k_{c,t,n}^T = j^{\text{new}}, \mathbf{k}_{-c,t,n}^T, \mathbf{w}_{-c,t,n}, \mathbf{k}^S, \mathbf{k}^C, \mathbf{l}, *), & \text{if } j \text{ is new.} \end{cases} \quad (3)
 \end{aligned}$$

Marginalize out all assignments of to  $k_{c,st,j^{\text{new}}}^S$  (i.e. all possible local items  $i$ 's including a new item  $i^{\text{new}}$  of restaurant  $c$ ), we have

$$\begin{aligned}
 &P(w_{c,t,n} \mid k_{c,t,n}^T = j^{\text{new}}, \mathbf{k}_{-c,t,n}^T, \mathbf{w}_{-c,t,n}, \mathbf{k}^S, \mathbf{k}^C, \mathbf{l}, *) \\
 &= \sum_{i=1}^{I_c^C} \frac{N_{c,i}^C}{N_{c,\cdot}^C + \alpha_0} f_{k_{c,i}^C}^{-c,t,n}(w_{c,t,n}) + \frac{\alpha_0}{N_{c,\cdot}^C + \alpha_0} P(w_{c,t,n} \mid k_{c,st,j^{\text{new}}}^S = i^{\text{new}}, \mathbf{k}_{-c,st,j^{\text{new}}}^S, \mathbf{w}_{-c,t,n}, \mathbf{k}^C, \mathbf{l}, *) \quad (4)
 \end{aligned}$$

Marginalize out all possible values that can be used to assign to  $k_{c,i^{\text{new}}}^C$  (i.e. all possible global dishes  $k$ 's including a new  $k^{\text{new}}$ ), we have

$$\begin{aligned}
 &P(w_{c,t,n} \mid k_{c,st,j^{\text{new}}}^S = i^{\text{new}}, \mathbf{k}_{-c,st,j^{\text{new}}}^S, \mathbf{w}_{-c,t,n}, \mathbf{k}^C, \mathbf{l}, *) \\
 &= \sum_{k=1}^K \frac{N_k}{N_{\cdot} + \alpha} f_k^{-c,t,n}(w_{c,t,n}) + \frac{\alpha}{N_{\cdot} + \alpha} f_{k^{\text{new}}}^{-c,t,n}(w_{c,t,n}) \quad (5)
 \end{aligned}$$

**Sampling local menu item for tables  $k_{c,s,j}^S$**  When a new table  $j^{\text{new}}$  is created after sampling  $k_{c,t,n}^T$ , we need to assign an item from the local menu to the table. That is

Notation	Descriptions
$H$	the base probability measure over topics, a symmetric Dirichlet
$G_0$	the probability measure drawn from $H$ and is shared across all conversations
$G_c$	the probability measure drawn from $G_0$ for each conversation $c$ and is shared across all turns in $c$
$G_{c,t}$	the probability measure drawn from $G_c$ for turn $t$ in conversation $c$
$G_{c,s}$	the probability measure drawn from $G_c$ for segment $s$ in conversation $c$
$\phi_k$	the $k^{th}$ multinomial distribution over words (i.e. the $k^{th}$ topic)
$\psi_{c,t,n}$	the multinomial distribution over words generating token $w_{c,t,n}$ (i.e. $\psi_{c,t,n} = \phi_{z_{c,t,n}}$ )
$l_{c,t}$	the topic shift assigned to turn $t$ of conversation $c$
$\pi_m$	the topic shift tendency of speaker $m$
$w_{c,t,n}$	the $n^{th}$ token of turn $t$ of conversation $c$
$a_{c,t}$	the speaker of turn $t$ of conversation $c$
$s_t$	The segment index of turn $t$
$N_{c,t}$	Number of tokens in turn $t$ of conversations $c$
$T_{c,s}$	Number of turns in segment $s$ of conversation $c$
$T_c$	Number of turns in conversation $c$
$S_c$	Number of segments in conversation $c$
$C$	Number of conversations
$N_{c,s,j}^S$	Number of customers sitting at table $j$ during segment $s$ at restaurant $c$
$N_{c,i}^C$	Number of tables in restaurant $c$ serving local dish item $i$ over time
$N_k$	Number of local menu items across all restaurants assigned the global dish $k$
$k_{c,t,n}^T$	Index of the table assigned to customer $n$ on day $t$ at restaurant $c$
$k_{c,s,j}^S$	Index of the item on the local menu assigned to table $j$ during segment $s$ of restaurant $c$
$k_{c,i}^C$	Index of the global dish assigned to item $i$ on the local menu of restaurant $c$
$J_{c,s}^S$	Number of tables during segment $s$ of restaurant $c$
$I_c^C$	Number of items on the local menu of restaurant $c$
$K$	Number of global dishes
$*$	The set of all hyperparameters
$\cdot$	This notation is used to denote the marginal count

Table 1: Notations used and their descriptions

$$\begin{aligned}
& P(k_{c,st,j^{new}}^S = i \mid \mathbf{k}_{-c,st,j^{new}}^S, \mathbf{k}^C, \mathbf{w}, \mathbf{l}, *) \\
& \propto P(k_{c,st,j^{new}}^S = i \mid \mathbf{k}_{-c,st,j^{new}}^S) P(w_{c,t,n} \mid k_{c,st,j^{new}}^S = i^{new}, \mathbf{k}_{-c,st,j^{new}}^S, \mathbf{w}_{-c,t,n}, \mathbf{k}^C, \mathbf{l}, *) \\
& = \begin{cases} \frac{N_{c,i}^C}{N_{c,\cdot}^C + \alpha_0} f_{k_{c,i}^C}^{-c,t,n}(w_{c,t,n}), & \text{if } i \text{ exists;} \\ \frac{\alpha_0}{N_{c,\cdot}^C + \alpha_0} P(w_{c,t,n} \mid k_{c,st,j^{new}}^S = i^{new}, \mathbf{k}_{-c,st,j^{new}}^S, \mathbf{w}_{-c,t,n}, \mathbf{k}^C, \mathbf{l}, *), & \text{if } i \text{ is new.} \end{cases} \quad (6)
\end{aligned}$$

The value of  $P(w_{c,t,n} \mid k_{c,st,j^{new}}^S = i^{new}, \mathbf{k}_{-c,st,j^{new}}^S, \mathbf{w}_{-c,t,n}, \mathbf{k}^C, \mathbf{l}, *)$  can be obtained in Equation 5.

**Sampling global dish for local menu item  $k_{c,i}^C$**  When a new local menu item  $i^{new}$  is added, a global dish is sampled from

$$P(k_{c,i^{new}}^C \mid \mathbf{k}_{-c,i^{new}}^C, \mathbf{w}, \mathbf{l}, *) \propto \begin{cases} \frac{N_k}{N_{\cdot} + \alpha} f_k^{-c,t,n}(w_{c,t,n}), & \text{if } k \text{ exists;} \\ \frac{\alpha}{N_{\cdot} + \alpha} f_{k^{new}}^{-c,t,n}(w_{c,t,n}), & \text{if } k \text{ is new.} \end{cases} \quad (7)$$

By having all the seating assignments, we obtain the topic assignment  $z_{c,t,n}$  for every token  $w_{c,t,n}$  by using Equation 1.

**Sampling topic shifts** Given all the path assignments for all token  $w_{c,t,n}$ , we will sample the topic shift  $l_{c,t}$  for every turn  $t$  of conversation  $c$ .

$$P(l_{c,t} \mid \mathbf{l}_{-c,t}, \mathbf{k}^\mathcal{T}, \mathbf{w}, \mathbf{a}, *) \propto P(l_{c,t} \mid \mathbf{l}_{-c,t}, \mathbf{a}, *) \cdot P(\mathbf{k}_{c,t}^\mathcal{T} \mid \mathbf{k}_{-c,t}^\mathcal{T}, l_{c,t}, \mathbf{l}_{-c,t}, *) \quad (8)$$

We now compute two factors in the RHS of Equation 8.

**Computing  $P(l_{c,t} \mid \mathbf{l}_{-c,t}, \mathbf{a}, *)$**  Recall that the topic shifts  $\mathbf{l} = \{l_{c,t}\}$  are drawn from a Bernoulli distribution parameterized by the topic shift tendency  $\pi$ , which is drawn from a conjugate prior  $\text{Beta}(\gamma)$ . Marginalizing out  $\pi$  we have

$$P(\mathbf{l}) = \int_0^1 P(\mathbf{l} \mid \pi) P(\pi; \gamma) d\pi = \prod_{m=1}^M \frac{\Gamma(2\gamma) \Gamma(S_{m,1} + \gamma) \Gamma(S_{m,0} + \gamma)}{\Gamma(\gamma)^2 \Gamma(S_{m,\cdot} + 2\gamma)} \quad (9)$$

When  $l_{c,t} = 0$ , the counts of number of times being assigned topic shift of value 1 for all speakers will remain unchanged. Similarly, for the case of  $l_{c,t} = 1$ . Thus, we have  $P(l_{c,t} \mid \mathbf{l}_{-c,t}, \mathbf{a}, *)$  for the two cases as follow [Resnik and Hardisty, 2010]

$$P(l_{c,t} \mid \mathbf{l}_{-c,t}, \mathbf{a}, *) = \frac{P(\mathbf{l} \mid \mathbf{a}, *)}{P(\mathbf{l}_{-c,t} \mid \mathbf{a}, *)} \propto \begin{cases} \frac{S_{a,c,t,0}^- + \gamma}{S_{a,c,t,\cdot}^- + 2\gamma}, & \text{if } l_{c,t} = 0 \\ \frac{S_{a,c,t,1}^- + \gamma}{S_{a,c,t,\cdot}^- + 2\gamma}, & \text{if } l_{c,t} = 1 \end{cases} \quad (10)$$

where  $S_{a,x}^{-c,t}$  denotes the number of times speaker  $a$  is assigned topic shift of value  $x \in \{0, 1\}$  excluding  $l_{c,t}$ .

**Computing  $P(\mathbf{k}_{c,t}^\mathcal{T} \mid \mathbf{k}_{-c,t}^\mathcal{T}, l_{c,t}, \mathbf{l}_{-c,t}, *)$**

$$P(\mathbf{k}_{c,t}^\mathcal{T} \mid \mathbf{k}_{-c,t}^\mathcal{T}, l_{c,t}, \mathbf{l}_{-c,t}, *) \propto \frac{P(\mathbf{k}_{c,t}^\mathcal{T}, \mathbf{k}_{-c,t}^\mathcal{T} \mid l_{c,t}, \mathbf{l}_{-c,t}, *)}{P(\mathbf{k}_{-c,t}^\mathcal{T} \mid l_{c,t}, \mathbf{l}_{-c,t}, *)} = \frac{P(\mathbf{k}^\mathcal{T} \mid l_{c,t}, \mathbf{l}_{-c,t}, *)}{P(\mathbf{k}_{-c,t}^\mathcal{T} \mid l_{c,t}, \mathbf{l}_{-c,t}, *)} \quad (11)$$

Given all the customers assigned to all tables, the joint probability of all tables [Gershman and Blei, 2012] can be computed as follow

$$P(\mathbf{k}^\mathcal{T} \mid \mathbf{l}) = \prod_{c=1}^C \prod_{s=1}^{S_c} \frac{\alpha_c^{J_{c,s}^\mathcal{S}} \prod_{j=1}^{J_{c,s}^\mathcal{S}} (N_{c,s,j}^\mathcal{S} - 1)!}{\prod_{x=1}^{N_{c,s,\cdot}^\mathcal{S}} (x - 1 + \alpha_c)} \quad (12)$$

where recall that  $S_c$  denote the number of segments in restaurant  $c$ .

Applying Equation 12 to Equation 11, we have

$$P(\mathbf{k}_{c,t}^\mathcal{T} \mid l_{c,t}, \mathbf{l}_{-c,t}, \mathbf{k}_{-c,t}^\mathcal{T}) \propto \begin{cases} \frac{\alpha_c^{J_{c,s_t}^{S,0}} \prod_{j=1}^{J_{c,s_t}^{S,0}} (N_{c,s_t,j}^{S,0} - 1)!}{\prod_{x=1}^{N_{c,s_t,\cdot}^{S,0}} (x - 1 + \alpha_c)}, & \text{if } l_{c,t} = 0 \\ \frac{\alpha_c^{J_{c,s_t-1}^{S,1}} \prod_{j=1}^{J_{c,s_t-1}^{S,1}} (N_{c,s_t-1,j}^{S,1} - 1)!}{\prod_{x=1}^{N_{c,s_t-1,\cdot}^{S,1}} (x - 1 + \alpha_c)} \frac{\alpha_c^{J_{c,s_t}^{S,1}} \prod_{j=1}^{J_{c,s_t}^{S,1}} (N_{c,s_t,j}^{S,1} - 1)!}{\prod_{x=1}^{N_{c,s_t,\cdot}^{S,1}} (x - 1 + \alpha_c)}, & \text{if } l_{c,t} = 1 \end{cases} \quad (13)$$

where

- $J_{c,s_t}^{S,x}$  denotes the number of tables during segment  $s$  of restaurant  $c$  if  $l_{c,t} = x$ .
- $N_{c,s_t,j}^{S,x}$  denotes the number of customers sitting at table  $j$  during segment  $s$  of restaurant  $c$  if  $l_{c,t} = x$ . The marginal count  $N_{c,s_t}^{S,x}$  denotes the total number of customers during segment  $s$  of restaurant  $c$ .

Combining Equations 10 and 13, we have the sampling equation for topic shift,  $P(l_{c,t} \mid \mathbf{l}_{-c,t}, \mathbf{k}^\mathcal{T}, \mathbf{w}, \mathbf{a}, *)$

$$\propto \begin{cases} \frac{S_{a_{c,t},0}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma} \frac{\alpha_c^{J_{c,s_t}^{S,0}} \prod_{j=1}^{J_{c,s_t}^{S,0}} (N_{c,s_t,j}^{S,0} - 1)!}{\prod_{x=1}^{N_{c,s_t}^{S,0}} (x - 1 + \alpha_c)}, & \text{if } l_{c,t} = 0 \\ \frac{S_{a_{c,t},1}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma} \frac{\alpha_c^{J_{c,s_t}^{S,1}} \prod_{j=1}^{J_{c,s_t}^{S,1}} (N_{c,s_t,j}^{S,1} - 1)!}{\prod_{x=1}^{N_{c,s_t}^{S,1}} (x - 1 + \alpha_c)}, & \text{if } l_{c,t} = 1 \end{cases} \quad (14)$$

Thus, using Equations 3, 6 and 7 to obtain the topic assignment  $z_{c,t,n}$  as in Equation 1 and using Equation 14 to obtain the topic shift  $l_{c,t}$ , we complete the derivation of sampling equations for our Gibbs samplers.

## 2 Parametric SITS

Similar to the nonparametric version, the state space of the Markov chain includes the topic indices assigned to all tokens  $\mathbf{z} = \{z_{c,t,n}\}$  and topic shifts assigned to all turns  $\mathbf{l} = \{l_{c,t}\}$ . Here, we present the sampling equations for both variables.

### Sampling topic assignments

$$\begin{aligned} P(z_{c,t,n} = y \mid \mathbf{z}_{-c,t,n}, \mathbf{l}, \mathbf{w}, *) &= \frac{P(z_{c,t,n} = y, \mathbf{z}_{-c,t,n}, \mathbf{l}, \mathbf{w})}{P(\mathbf{z}_{-c,t,n}, \mathbf{l}, \mathbf{w})} \\ &= \frac{P(\mathbf{w} \mid \mathbf{z}) P(\mathbf{z} \mid \mathbf{l}) P(\mathbf{l})}{P(\mathbf{w} \mid \mathbf{z}_{-c,t,n}) P(\mathbf{z}_{-c,t,n} \mid \mathbf{l}) P(\mathbf{l})} = \frac{P(\mathbf{w} \mid \mathbf{z})}{P(\mathbf{w} \mid \mathbf{z}_{-c,t,n})} \frac{P(\mathbf{z} \mid \mathbf{l})}{P(\mathbf{z}_{-c,t,n} \mid \mathbf{l})} \\ &\propto \frac{TW_{y,w_{c,t,n}}^{-c,t,n} + \beta}{TW_{y,\cdot}^{-c,t,n} + V\beta} \cdot \frac{ST_{c,s_t,y}^{-c,t,n} + \alpha}{ST_{c,s_t,\cdot}^{-c,t,n} + K\alpha} \end{aligned} \quad (15)$$

where

- $TW_{k,w}$  denotes the number of times that topic  $k$  is assigned to token  $w$  in the vocabulary.
- $ST_{c,s,k}$  denotes the number of times topic  $k$  is assigned to segment  $s$  of conversation  $c$ .
- $V$  denotes the size of the vocabulary.
- $K$  denotes the number of predefined topics.

**Sampling topic shifts**  $P(l_{c,t} \mid \mathbf{l}_{-c,t}, \mathbf{z}, \mathbf{w}, \mathbf{a}, *)$

$$\propto \begin{cases} \frac{S_{a_{c,t},0}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma} \cdot \frac{\prod_{k=1}^K \Gamma(ST_{c,s_t,k}^0 + \alpha)}{\Gamma(ST_{c,s_t,\cdot}^0 + K\alpha)}, & \text{if } l_{c,t} = 0 \\ \frac{S_{a_{c,t},1}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma} \cdot \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \frac{\prod_{k=1}^K \Gamma(ST_{c,s_t-1,k}^1 + \alpha)}{\Gamma(ST_{c,s_t-1,\cdot}^1 + K\alpha)} \frac{\prod_{k=1}^K \Gamma(ST_{c,s_t,k}^1 + \alpha)}{\Gamma(ST_{c,s_t,\cdot}^1 + K\alpha)}, & \text{if } l_{c,t} = 1 \end{cases} \quad (16)$$

where

- $S_{m,x}$  denotes the number of times that topic shift of value  $x$  is assigned to speaker  $m$ .
- $ST_{c,t,k}^x$  denotes the number of times that topic  $k$  is assigned to segment  $s$  of conversation  $c$  if  $l_{c,t} = x$ .

## References

- [Gershman and Blei, 2012] Gershman, S. J. and Blei, D. M. (2012). A tutorial on bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1):1 – 12.
- [Resnik and Hardisty, 2010] Resnik, P. and Hardisty, E. (2010). Gibbs sampling for the uninitiated. Technical Report UMIACS-TR-2010-04, University of Maryland. <http://www.lib.umd.edu/drum/handle/1903/10058>.
- [Teh et al., 2006] Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.