Given the topic knowledge ϕ, the entity name knowledge ψ and the entity context knowledge ξ:

1. For each doc d in **D,** sample its topic distribution $\theta_d \sim Dir(\alpha)$;
2. For each of the M_d mentions m_i in doc d:
   a) Sample a topic assignment $z_i \sim Mult(\theta_d)$;
   b) Sample an entity assignment $e_i \sim Mult(\phi_{z_i})$;
   c) Sample a mention $m_i \sim Mult(\psi_{e_i})$;
3. For each of the N_d words w_i in doc d:
   a) Sample a target entity it describes from d's referent entities $a_i \sim Unif(e_{m_1}, e_{m_2}, \cdots, e_{m_d})$;
   b) Sample a describing word using a_i's context word distribution $w_i \sim Mult(\xi_{a_i})$.

Figure 2. The document generative process, with $Dir(.)$, $Mult(.)$ and $Unif(.)$ correspondingly *Dirichlet*, *Multinomial* and *Uniform distribution*

Given the entity list **E** = $\{e_1, e_2, ..., e_E\}$ in the knowledge base, the word list **V** = $\{w_1, w_2, ..., w_v\}$, the entity name list **K** = $\{n_1, n_2, ..., n_K\}$ and the global knowledge described in above, the generation process of a document collection (corpus) **D** = $\{d_1, d_2, ..., d_D\}$ is shown in Figure 2. To demonstrate the generation process, we also demonstrate how the document in Figure 1 can be generated using our model in following steps:

Step 1: The model generates the topic distribution of the document as θ_d = {*Apple Inc.*$^{0.45}$, *Operating System(OS)*$^{0.55}$};

Step 2: For the three mentions in the document:

i. According to the topic distribution θ_d, the model generates their topic assignments as z_1=*Apple Inc.*, z_2 = *Apple Inc.*, z_3 = *OS*;

ii. According to the topic knowledge $\phi_{Apple\ Inc.}$, ϕ_{OS} and the topic assignments z_1, z_2, z_3, the model generates their entity assignments as e_1 = *Apple Worldwide Developers Conference*, e_2 = *Apple Inc.*, e_3 = *Mac OS X Lion*;

iii. According to the name knowledge of the entities *Apple Worldwide Developers Conference, Apple Inc.* and *Mac OS X Lion*, our model generates the three mentions as m_1=*WWDC*, m_2 = *Apple*, m_3 = *Lion*;

Step 3: For all words in the document:

i. According to the referent entity set in document $\mathbf{e_d}$ = {*Apple Worldwide Developers Conference*, *Apple Inc.*, *Mac OS X Lion*}, the model generates the target entity they describes as a_3=*Apple Worldwide Developers Conference* and a_4=*Apple Inc.*;

ii. According to their target entity and the context knowledge of these entities, the model generates the context words in the document. For example, according to the context knowledge of the entities *Apple Worldwide Developers Conference*, the model generates its context word w_3 =*conference*, and according to the context knowledge of the entity *Apple Inc.*, the model generates its context word w_4 = *introduces*.

Through the above generative process, we can see that all entities in a document are extracted from the document's underlying topics, ensuring the topic coherence; and all words in a document are extracted from the context word distributions of its referent entities, resulting in the context compatibility. Furthermore, the generation of topics, entities, mentions and words are highly correlated, thus our model can capture the correlation between the topic coherence and the context compatibility.

2.2 Global Knowledge Generative Process

The entity-topic model relies on three types of global knowledge (including the topic knowledge, the entity name knowledge and the entity context knowledge) to generate a document. Unfortunately, all three types of global knowledge are unknown and thus need to be estimated from data. In this paper we estimate the global knowledge through Bayesian inference by also incorporating the knowledge generation process into our model.

Specifically, given the topic number T, the entity number E, the name number K and the word number V, the entity-topic model generates the global knowledge as follows:

1) $\phi|\beta \sim Dir(\beta)$

For each topic z, our model samples its entity distribution ϕ_z from an E-dimensional Dirichlet distribution with hyperparameter β.

2) $\psi|\gamma \sim Dir(\gamma)$

For each entity e, our model samples its name distribution ψ_e from a K-dimensional Dirichlet distribution with hyperparameter γ.

3) $\xi|\delta \sim Dir(\delta)$