Entropy Converges Between Dialogue Participants: Explanations from an Information-Theoretic Perspective

Yang Xu and **David Reitter**
College of Information Sciences and Technology
The Pennsylvania State University
University Park, PA 16802, USA
`yang.xu@psu.edu, reitter@psu.edu`

Abstract

The applicability of entropy rate constancy to dialogue is examined on two spoken dialogue corpora. The principle is found to hold; however, new entropy change patterns within the topic episodes of dialogue are described, which are different from written text. Speaker's dynamic roles as *topic initiators* and *topic responders* are associated with decreasing and increasing entropy, respectively, which results in local convergence between these speakers in each topic episode. This implies that the sentence entropy in dialogue is conditioned on different contexts determined by the speaker's roles. Explanations from the perspectives of grounding theory and interactive alignment are discussed, resulting in a novel, unified information-theoretic approach of dialogue.

1 Introduction

Information in written text and speech is strategically distributed. It has been claimed to be ordered such that the rate of information is not only close to the channel capacity, but also approximately constant (Genzel and Charniak, 2002, 2003; Jaeger, 2010); these results were developed within the framework of Information Theory (Shannon, 1948). In these studies, the per-word cross-entropy of a sentence is used to model the amount of information transmitted. Language is treated as a series of random variables of words.

Most existing work examined written text as opposed to speech. Spoken dialogue is different from written text in many ways. For example, dialogue contains more irregular or ungrammatical components, such as incomplete utterances, disfluencies etc. (Jurafsky and Martin, 2014, ch 12), which are "theoretically uninterested complexities that are unwanted" (Pickering and Garrod, 2004). Dialogue is also different from written text in high level discourse structure. The paragraphs in written text, which function as relatively standalone topic units, are constructed under the guidance of one consistent author. On the other hand, the constitution and transformation of topics in dialogue are more dynamic processes, which are the result of the joint activity from multiple speakers (Linell, 1998). In nature, written text is a monologue, while dialogue is a joint activity (Clark, 1996).

From the application perspective, investigating entropy in dialogue can help us better understand which speaker contributes the most information, and thus may potentially benefit tasks such as conversational roles identification (Traum, 2003) etc. From the theoretical perspective, we believe that such investigation will reveal some unique features of the formation of higher level discourse structure in dialogue that are different from written text, e.g., *topic episode* shifts, because previous studies have found the correlation between entropy decrease and potential topic shift in written text (Qian and Jaeger, 2011). Finally, entropy is closely related to predictability and processing demands, which has implications for cognitive aspects of communication.

The main purpose of this study is to characterize how lexical entropy changes in spoken language. We will focus on spontaneous dialogue of two speakers and carry out two steps of investigation. First, we examine the overall entropy patterns within dialogue as a whole context that does not differentiate speakers. Second, we zoom in to *topic episodes* within dialogue and explore how each of the two speakers' entropy develops. The goal of the second step is to account the complexity of topic shifts within spoken dialogues and to reach a more detailed understanding

of human communication from an information-theoretic perspective. If topic shifts in dialogue do correlate with changes in entropy, how do they affect the two speakers, only one of whom typically initiates the topic shift, while another follows along? To answer this question, we use the transcribed text data from two well-developed corpora.

2 Related Work

2.1 The principle of entropy rate constancy

The *constancy rate principle* governing language generation in human communication was first proposed by Genzel and Charniak (2002). Inspired by ideas from Information Theory (Shannon, 1948), this principle asserts that people communicate (written or spoken) in a way that keeps the rate of information being transmitted approximately constant.

Genzel and Charniak (2002) provide evidence to support this principle by formulating the problem into Equation 1. They treat text as a sequence of random variables X_i, and X_i corresponds to the i th word in the corpus. They focus on the entropy of a word conditioned on its context, i.e., $X_i | X_1 = w_1, \ldots, X_{i-1} = w_{i-1}$, and decompose the context into two parts: the global context C_i that refers to all the words from preceding sentences, and the local context L_i that refers to all the preceding words within the same sentence as X_i. Thus, the conditioned entropy of X_i is also decomposed into two terms (see the right side of Equation 1): the local measure of entropy (first term), and the mutual information between the word and global context (second term).

$$H(X_i|C_i, L_i) = H(X_i|L_i) - I(X_i, C_i|L_i) \quad (1)$$

The constancy rate principle predicts that the left side of Equation 1 should be constant as i increases. Because $H(X_i|C_i, L_i)$ itself is difficult to estimate (because it is hard to define C_i mathematically), and that the mutual information turn $I(X_i, C_i|L_i)$ is known to increase with i, the whole problem becomes examining whether the local measure of entropy $H(X_i|L_i)$ also increases with i. Genzel and Charniak (2002) have confirmed this prediction by showing that $H(X_i|L_i)$ does increase with i within multiple genres of written text of different languages.

The constancy rate principle also leads to an interesting prediction about the relationship between entropy change and topic shift in text. Generally, a sentence that initiate a shift in topic will have lower mutual information between its context, because the previous context provides little information to the new topic. Thus, a topic shift corresponds to the drop of the mutual information term $I(X_i, C_i|L_i)$. Then in order to keep constancy of the left term as predicted by the principle, the entropy term needs to decrease when a topic shift happens. Genzel and Charniak (2003) verified this prediction by showing that paragraph-starting sentences have lower entropy than non-paragraph-starting ones, with the assumption that a new paragraph often indicates a topic shift in written text. More recently, latent topic modeling (Qian and Jaeger, 2011) showed that lower sentence entropy was associated with topic shifts.

Genzel and Charniak's work has been extended to integrate non-linguistic information into the principle. Doyle and Frank (2015) leveraged Twitter data to find further support to the constancy rate principle: the entropy of message gradually increases as the context builds up, and it sharply goes down when there is a sudden change in the non-linguistic context (Baseball world series news reports, Doyle and Frank, 2015). *Uniform Information Density* (UID) (Jaeger and Levy, 2006) extends the principle in a framework that governs how people manage the amount of information in language production, from lexical levels to all levels of linguistic representations, e.g., syntactic levels. Its core idea is that people avoid salient changes in the density of information (i.e., amount of information per amount of linguistic signal) by making specific linguistic choices under certain contexts (Jaeger, 2010).

2.2 Topic shift in dialogues

As a conversation unfolds, topic changes naturally happen when a current topic is exhausted or a new one occurs, which is referred to as *topic shift* in the field of *Conversation Analysis* (CA) (Ng and Bradac, 1993; Linell, 1998). In CA, the basic unit of topical structure analysis in dialogue is *episode*, which refers to a sequence of speech events that are "about" something specific in the world (Linell, 1998, ch 10, p 187). Here, to be precise, we use the term *topic episode*.

According to related theories in CA, the for-

Table 1: Basic statistics of corpora

Statistics	Switchboard	BNC
# of dialogues.	1126	1346
Avg # of turns in dialogue	109	52
Avg # of sentences in dialogue	141	70

mation of topic episode is a joint accomplishment from two speakers and a product of initiatives and responses (Linell, 1990). When establishing a new topic jointly, one speaker first produces an initiatory contribution that introduce a "candidate" topic, and the other speaker makes a response that shares his perspective on that (Linell, 1998). From the information theoretic point of view, the *initiator* of a new topic plays a role of introducing *novelty* or *surprisal* into the context, while the other speaker, the *responder*, is more of a *commenter* or *evaluator* of information, who does not contribute as much in terms of novelty.

Since previous studies have shown that the decrease of sentence entropy is correlated with topic shifts in written text (Genzel and Charniak, 2003; Qian and Jaeger, 2011), it is reasonable to expect the same effect to be present at the boundaries of topic episodes in dialogue. Furthermore, considering the *initiator* vs. *responder* discrepancy in speaker roles, we expect their entropy change patterns also to be different.

3 Overall Trend of Entropy in Dialogue

In this section we examine whether the overall entropy increase trend is present in dialogue text.

3.1 Corpus data

The Switchboard corpus (Godfrey et al., 1992) and the British National Corpus (BNC) (BNC, 2007) are used in this study. Switchboard contains 1126 dialogues by telephone between two native North-American English speakers in each dialogue. We use only a subset of BNC (spoken part) that contain spoken conversations with exactly two participants, so that the dialogue structures are consistent with Switchboard.

3.2 Computing Entropy of One Sentence

We use language model to estimate the sentence entropy, which is similar to Genzel and Charniak (2003)'s method. A sentence is considered as a sequence of words, $W = \{w_1, w_2, \ldots, w_n\}$, and its per-word entropy is estimated by:

$$H(w_1 \ldots w_n) = -\frac{1}{n} \sum_{w_i \in W} \log P(w_i|w_1 \ldots w_{i-1})$$

where $P(w_i|w_1 \ldots w_{i-1})$ is estimated using a trigram language model. The model is trained using Katz backoff (Katz, 1987) and Lidstone smoothing (Chen and Goodman, 1996).

For the two corpora respectively, we extract the first 100 sentences from each conversation, and apply a 10-fold cross-validation, i.e., dividing all the data into 10 folds. Then we choose each fold as the testing set, and compute the entropy of each sentence in it, using the language model trained against the rest of the folds.

3.3 Eliminating sentence length effects

Intuitively, longer sentences tend to convey more information than short ones. Thus, the per-word entropy of a sentence should be correlated with the sentence length, i.e., the number of words. This correlation is confirmed in our data by calculating the Pearson correlation between the per-word entropy and sentence length: For Switchboard, $r = 0.258, p < 0.001$; for BNC, $r = 0.088, p < 0.001$.

Sentence length is found to vary with its relative position in text (Keller, 2004). Thus, in order to truly examine the variation pattern of sentence entropy within dialogue, we need to eliminate the effect of sentence length from it. We calculate a *normalized entropy* that is independent of sentence length in the following way. (This method is used by Genzel and Charniak (2003) to get the length-independent *tree depth* and *branching factor* of sentence.) First, we compute $\bar{e}(n)$, the average per-word entropy of sentences of the same length n, for all lengths ($n = 1, 2, \ldots$) that have occurred

$$\bar{e}(n) = \frac{1}{|L(n)|} \sum_{s \in L(n)} e(s)$$

where $e : S \to \mathbb{R}$ is the original per-word entropy of a sentence s, and $L(n) = s|l(s) = n$ is the set of sentences of length n. Then we compute the sentence-length adjusted entropy measure that we want by

$$e'(s) = \frac{e(s)}{\bar{e}(n)}$$

This normalized entropy measure sums up to 1, and is not sensitive to sentence length. In later part

of this paper, we demonstrate our results in both entropy and normalized entropy because the former is the direct measure of information content.

3.4 Results

We plot the per-word entropy and normalized entropy of sentence against its *global position*, which is the sentence position from the beginning of the dialogue (Figure 1). It can be seen that both measures increase with global position. BNC shows larger slope than Switchboard, and the latter has a flatter curve but sharper increase at the early stage of conversations.

To test the reliability of the observed increasing trend, we fit linear mixed-effect models using entropy and normalized entropy as response variables, and the global position of sentence as predictor (fixed effect), with a random intercept grouped by distinct dialogues. The `lme4` package in `R` is used (Bates et al., 2014). The results show that the fixed effects of global position are significant for both measures in both corpora: Entropy in Switchboard, $\beta = 4.2 \times 10^{-3}$, $p < 0.001$; normalized entropy in Switchboard, $\beta = 5.9 \times 10^{-4}$, $p < 0.001$; entropy in BNC, $\beta = 1.5 \times 10^{-2}$, $p < 0.001$; normalized entropy in BNC, $\beta = 1.4 \times 10^{-3}$, $p < 0.001$).

In particular, since the curves of Switchboard seem flat after a boost in the early phase (between 0 to 5 in global position), we fit extra models to examine whether the entropy increase for global positions larger than 10 is significant. The long-term changes are reliable, too: Entropy, $\beta = 3.4 \times 10^{-3}$, $p < 0.001$; normalized entropy, $\beta = 5.1 \times 10^{-4}$, $p < 0.001$.

In sum, we find increasing entropy over the course of the whole dialogue. These findings are consistent with previous findings on written text.

4 Topic Shift and Speaker Roles

Since the topic structure of dialogue differs from written text, it is our interest to investigate how this difference affects the sentence entropy patterns. First, we identify the boundaries of topic episodes, and examine the presence of entropy drop effect at the boundaries. Second, we differentiate the speakers' roles in initiating the topic episode, i.e., *initiator* vs. *responder*, and compare their entropy change patterns within the episode.

4.1 Topic segmentation

There are multiple computational frameworks for topic segmentation, such as the Bayesian model (Eisenstein and Barzilay, 2008), Hidden Markov model (Blei and Moreno, 2001), latent topic model (Blei et al., 2003) etc. Considering that performance is not the prior requirement in our task, and also to avoid being confounded by segmentation method that utilize entropy measure per se, we use a less sophisticated cohesion-based TextTiling algorithm (Hearst, 1997) to carry out topic segmentation.

TextTiling algorithm inserts boundaries into dialogue as a sequence of sentences. We treat the segments between those boundaries as topic episodes. For each episode within a dialogue, we assign it a unique *episode index*, indicating its relative position in the dialogue (e.g., from 1 to N for a dialogue that contains N episodes). For each sentence, we assign it a *within-episode position*, indicating its relative position within the topic episode.

In Figure 2 we plot the entropy (and normalized) of sentence against the within-episode positions, grouped by episode index. Due to the space limit, we only present the first 6 topic episodes and the first 10 sentences in each episode. It can be seen that entropy drops at the beginning of topic episode, and then increases within the episode.

To examine the reliability of the entropy increase within topic episodes, we fit linear mixed effect models using entropy (and normalized) as response variables, and the within-episode position of sentence as predictor (fixed effect), with a random intercept grouped by the unique episode index of each topic episode. We find a significant fixed effect of within-episode position on both measures for both corpora: Entropy in Switchboard, $\beta = 5.9 \times 10^{-4}$, $p < 0.001$; normalized entropy in Switchboard, $\beta = 4.5 \times 10^{-3}$, $p < 0.001$; entropy in BNC, $\beta = 2.5 \times 10^{-2}$, $p < 0.001$; normalized entropy in BNC, $\beta = 3.0 \times 10^{-3}$, $p < 0.001$.

Our results show that when we treat the sentences in dialogue indiscriminately, their entropy change patterns at topic boundaries are consistent with previous findings on written text.

4.2 Identifying topic initiating utterances

Having dialogue segmented into topic episodes, our next step is to identify each speaker's role in initiating the topic. According to the theories
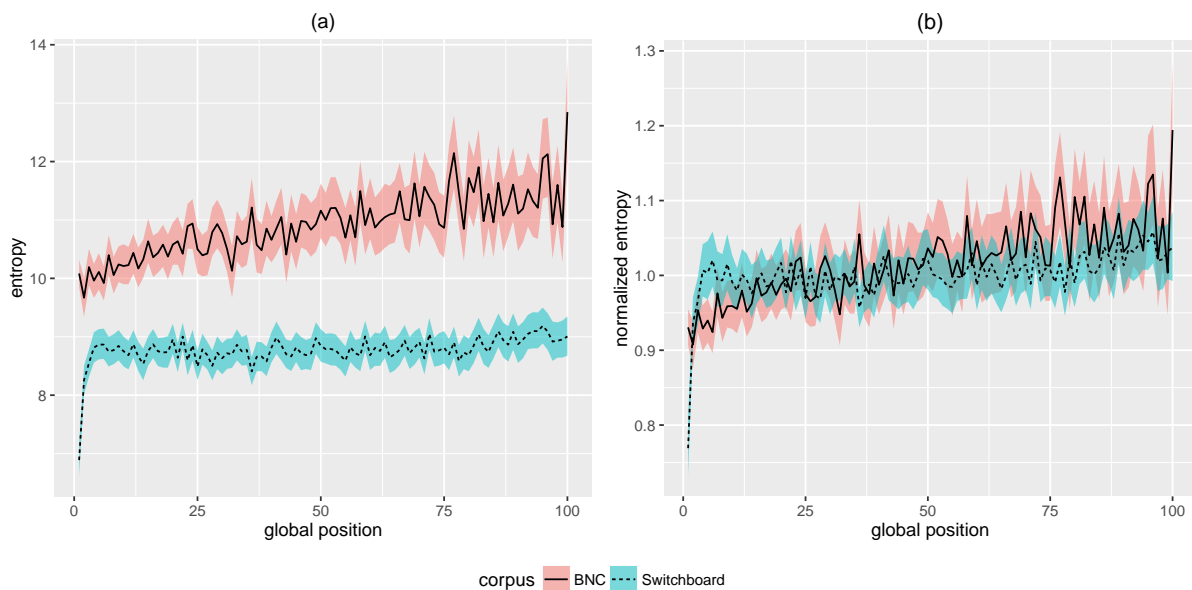
Figure 1: Entropy (a) and normalized entropy (b) against global position of sentences (from 1 to 100). Shadow area indicates 95% bootstrapped Confidence Interval.

reviewed in Section 2.2, the key to identify the speaker roles is to identify who produces the initiatory "candidate" topic. To be convenient, we use the term *topic initiating utterance* (TIU) to refer to the very first utterance produced by the *initiator* to bring up the new topic. Here, we give an empirical operational definition of TIU.

Since we treat dialogue as a series of sentences, and apply the TextTiling algorithm to insert topic boundaries indiscriminately (without differentiating whether adjacent sentences are from the same speaker or not), it results in two types of topic boundaries: *Within-turn* boundaries, the ones located in the middle of a turn (i.e., from one speaker). *Between-turn* boundaries, the ones located at the gap between two different turns (i.e., from two speakers). Our survey shows that in Switchboard 27.2% of the topic boundaries are within turns, and 72.8% are between turns. For BNC the two proportions are 41.2% and 58.8% respectively.

Intuitively, a within-turn topic boundary suggests that the speaker of the current turn is initiating the topic shift. On the other hand, a between-turn boundary suggests that the following speaker who first gives *substantial* contribution is more likely to be the initiator of the next topic. Following this intuition, for within-turn boundaries, we define TIU as the rest part of current turn after the boundary. For between-turn boundaries, we define

TIU as the whole body of the next *relatively long* turn after the boundary, whose length is larger than N words. Note that the determination of threshold N is totally empirical, because our goal is to identify the most *probable* TIU, based on the intuition that longer sentences *tend* to contain more information, and thus are more likely to initiate a new topic. For the results shown later in this paper, we use $N = 5$, and our experiments draw similar results for $N \geq 5$. The operational definition of TIU is demonstrated in Figure 3.

4.3 The effect of topic initiator vs. responder

Based on the operational definition of topic initiating utterance (TIU), we distinguish the two speakers' roles in each topic segment: the author of TIU is the *initiator* of the current topic, while the other speaker is the *responder*.

Again, we plot the sentence entropy (and normalized) against the within-episode position respectively, this time, grouped by speaker roles (*initiator* vs. *responder*) in Figure 4. It can be seen that at the beginning of a topic, initiators have significantly higher entropy than responders. As the topic develops, the initiators' entropy decreases (Figure 4a) or stays relatively steady (Figure 4b), and the responder's entropy increases. Together they form a convergence trend within topic episode.

We use standard linear mixed models to examine the convergence trend observed, i.e., to test

541

(a) Entropy vs. within-episode position

(b) Normalized entropy vs. within-episode position

Figure 2: Entropy (a) and normalized entropy (b) against within-episode position grouped by episode index. The x-axis in each block indicates the within-episode position of sentence. The number 1 to 6 on top of the blocks are episode indexes. Shadow area indicates 95% bootstrapped Confidence Interval
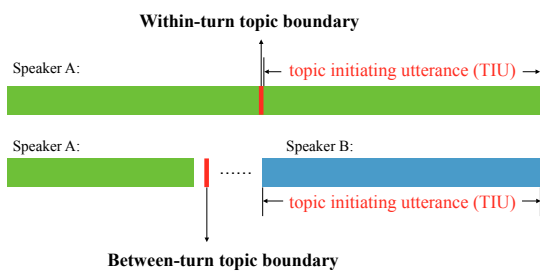

Figure 3: Operational definition of topic initiating utterances (TIUs). The red vertical bars indicate the topic boundaries placed using TextTiling. A complete horizontal bar of one color represents a turn from one speaker (green for speaker A and blue for speaker B). The upper line shows the case of within-turn topic boundary, and the lower line shows the case of between-turn topic boundary.

whether the initiators' entropy reliably decreases and whether the responders' entropy reliably increases. Models are fitted for initiators and responders respectively, using the entropy (and normalized) as response variables, and the within-episode position as predictor (fixed effect), with a random intercept grouped by the unique episode index. Our models show that for the entropy measure, the fixed effect of within-episode position is reliably negative for initiators (Switchboard, $\beta = -3.6 \times 10^{-2}$, $p < 0.001$; BNC, $\beta = -2.9 \times 10^{-2}$, $p < 0.05$) and reliably positive for responders (Switchboard, $\beta = 3.3 \times 10^{-1}$, $p < 0.001$; BNC, $\beta = 1.4 \times 10^{-1}$, $p < 0.001$). For the normalized entropy measure, the fixed effect of within-episode position is insignificant for initiators, which means there is neither increase nor decrease, and is reliably positive for responders (Switchboard, $\beta = 1.4 \times 10^{-2}$, $p < 0.001$; BNC, $\beta = 1.2 \times 10^{-2}$, $p < 0.001$). Thus, the convergence trend is confirmed.

The entropy change patterns of topic *initiators* (decrease or remain constant within topic episode) are inconsistent with previous findings that assert an entropy increase in written text (Genzel and
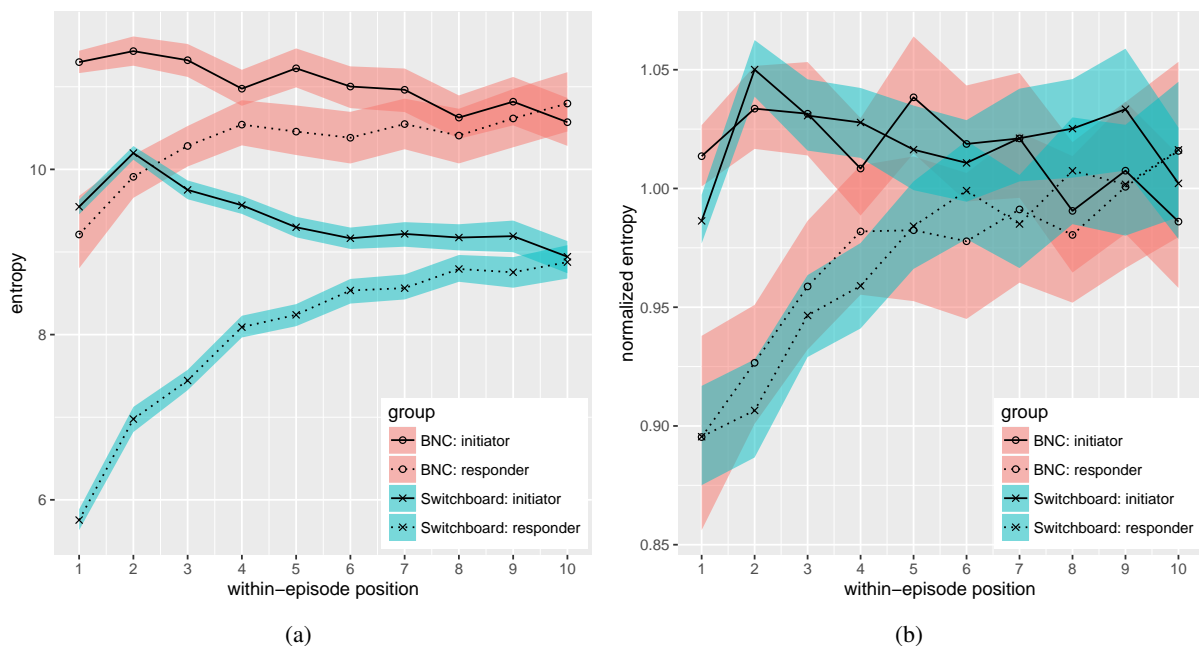
542

Figure 4: Entropy (a) and normalized entropy (b) against within-episode position grouped by speaker roles (*topic initiator* vs. *topic responder*)

Charniak, 2002, 2003), which will be discussed in the next section.

5 Discussion

5.1 Summary

Our main contribution is that we find new entropy change patterns in dialogues that are different from those in written text. Specifically, when distinguishing the speakers' roles by topic *initiator* vs. *responder*, we see that the initiator's entropy decreases (or remain steady) whilst the responder's increases within a topic episode, and together they form a convergence pattern. The partial trend of entropy decrease in topic initiators seems to be contrary to the principle of entropy rate constancy, but as we will discuss next, it is actually an effect of the unique topic shift mechanism of dialogues that is different from written text, which does not violate the principle.

From an information theoretic perspective, we view dialogue as a process of information exchange, in which the interlocutors play the roles of information provider and receiver, interactively within each topic episode.

Beyond differences in speaker roles, we do observe that sentence entropy increases with its global position in the dialogue, which is consistent with written text data (Genzel and Charniak, 2002, 2003; Qian and Jaeger, 2011; Keller, 2004).

Thus, overall speaking, spoken dialogue do follow the general principle of entropy rate constancy.

5.2 Dialogue as a process of information exchange

By combining topic segmentation techniques and fine-grained discourse analysis, we provide a new angle to view the big picture of human communication: the perspective of how information is distributed between different speakers.

One critical difference between written text and spoken text in conversation is that there is only one direct input source of information in the former, i.e., the author of the text, but for the latter, there are multiple direct input sources, i.e., the multiple speakers. That means, when language production is treated as a process of choosing proper words (or other representations) within a context, the definition of "context" is different between the two categories of text. In written language (see Equation 1 in Section 2), C_i, the global context of a word X_i, is assumed to be all the words in preceding sentences. This is a reasonable assumption, because when one author is writing a complete piece of text, he may organize information smoothly to keep the entropy rate constant. Within a dialogue, for any upcoming utterance, all preceding utterances together can be viewed as the *shared* context for the two speakers. To help us un-

derstand the nature of this shared context, we propose the following mental experiment. Suppose we, as researchers and "super-readers", observe the transcript of a dialogue between interlocutors A and B. To us, all utterances are based upon the context of previous ones, which is why we can observe consistent entropy increase within the whole dialogue (Figure 1 in Section 3). Also, to us, a new topic episode in dialogue is just like a new paragraph in written text, within which we can observe steady entropy increase without differentiating the utterances from the two speakers. By contrast, let's look at the context used by the two speakers. They will not necessarily leverage the preceding utterances as a coherent context. A *topic initiator* introduces new information from a context outside of the dialogue. Therefore the mutual information between the initiator's current sentence and the previous context is reduced, which causes the sentence entropy to start high before decreasing. On the other side, a *topic responder* relies much on the previous shared context (because he is not an active topic influencer). The responder is dynamically updating the context as the initiator pours new information into the mix. This causes the mutual information with the previous context to be high, and thus the sentence entropy start low before increasing again.

We think that the respective cognitive load in the topic responder imposed by following the other speaker in a new topic direction may be complemented by reduced information at the language level. This is, again, compatible with a cognitive communication framework that imposes a tendency to limit or keep constant overall information levels. It is also an example of extralinguistic information that causes complementary entropy changes in a speaker's language (cf., Doyle and Frank, 2015).

5.3 Dialogue as a process of building up common ground

Our findings can also be explained by a theory of *grounding* (Clark and Brennan, 1991; Clark, 1996) of communication. Dialogue can be seen as a joint activity during which multiple speakers contribute alternatively to build *common ground* (Clark and Brennan, 1991). Common ground can be understood as the mutual knowledge shared between interlocutors.

Clark (1996) proposes that joint activities have a number of characteristics: First, participants play different roles in the activity. Second, a major activity is usually comprised of sequences of sub-activities, and the participants' role may differ from sub-activity to next. Third, to achieve the goal of the activity, it requires coordination between participants of different roles.

In our design, the local roles of *topic initiator* vs. *topic responder* correspond to roles suggested by the joint-activity theory. The initiator sets up the dominant goal of the sub-activity, i.e., developing a new *topic episode*, and the responder joins him or her in order to achieve the goal. The converging sentence entropy indicates that the mutual knowledge between them is accumulating, i.e., the common ground is being gradually built up. Once the goal is achieved, i.e., the current topic is fully developed, a new goal will emerge, and a new common ground needs to be built again, which is sometimes accompanied by a change in participants roles.

5.4 Convergence of linguistic behaviors

One mechanism that may lead to the convergence of sentence entropy may be the interactive alignment of linguistic features between speakers (Pickering and Garrod, 2004); repeating words and syntactic structure leads to increased similarity. The entropy-converging pattern also reflects the convergence of higher-level dialogical behavior, say, speakership occupancy; the discrepancy between the two speakers' roles gradually becomes smaller, i.e., the "speaker" becomes more of a "listener", and vice versa. A psychologist might treat the fragmented topic episodes in dialogues as the locus where interlocutors build temporarily shared understanding (Linell, 1998), through the process of "synchronization of two streams of consciousness" (Schutz, 1967).

6 Conclusion

In this study, we validate the principle of entropy rate constancy in spoken dialogue, using two common corpora. Besides the results that are consistent with previous findings on written text, we find new entropy change patterns unique to dialogue. Speakers that actively initiate a new topic tend to use language with higher entropy compared to the language of those who passively respond to the topic shift. These two speaker's respective entropy

levels converge as the topic develops. A model of this phenomenon may provide explanations from the perspectives of information exchange, common ground building, and the convergence of linguistic behaviors in general.

With this, we put forward what we think is a new perspective to analyzing dialogue. As much dialogue happens for the purpose of information exchange, loosely defined, it makes sense to apply information-theoretic models to the semantics as well as the form of speaker's messages. The quantitative approach taken here augments rather than supplants speech acts (Searle, 1976), identifying who leads the dialogic process by introducing topics and shifting them.

Furthermore, our approach actually provides a unified perspective of dialogue that combines Grounding theory (Clark and Brennan, 1991) and Interactive Alignment (Pickering and Garrod, 2004). These two models are often described as opposite; by applying each theory to the dialogic structure between and within topic episodes, we find both of them can explain our findings. The entropy measure of information content quantifies interlocutors' contributions to common ground and also allows us to show convergence patterns.

This unified information-theoretic perspective may eventually allow us to identify further systematic patterns of information exchange between dialogue participants. There is, of course, no reason to think that multi-party dialogue should work differently; we leave the empirical examination as an open task.

Acknowledgments

References

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823* .

David M Blei and Pedro J Moreno. 2001. Topic segmentation with an aspect hidden markov model. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New Orleans, LA, USA, pages 343–348.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *The Journal of Machine Learning Research* 3:993–1022.

BNC. 2007. The British National Corpus, version 3 (BNC XML Edition).

Stanley F Chen and Joshua Goodman. 1996. An empirical study of smoothing techniques for language modeling. In *Proceedings of the 34th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, pages 310–318.

Herbert H Clark. 1996. *Using language*. Cambridge University Press.

Herbert H Clark and Susan E Brennan. 1991. Grounding in communication. *Perspectives on socially shared cognition* 13(1991):127–149.

Gabriel Doyle and Michael C Frank. 2015. Shared common ground influences information density in microblog texts. In *Proceedings of NAACL-HLT*. Denver, CO, USA.

Jacob Eisenstein and Regina Barzilay. 2008. Bayesian unsupervised topic segmentation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Honolulu, HI, USA, pages 334–343.

Dmitriy Genzel and Eugene Charniak. 2002. Entropy rate constancy in text. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, Philadelphia, PA, USA, pages 199–206.

Dmitriy Genzel and Eugene Charniak. 2003. Variation of entropy and parse trees of sentences as a function of the sentence number. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Sapporo, Japan, pages 65–72.

John J Godfrey, Edward C Holliman, and Jane McDaniel. 1992. Switchboard: Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*. IEEE, volume 1, pages 517–520.

Marti A Hearst. 1997. Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational linguistics* 23(1):33–64.

T Florian Jaeger. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61(1):23–62.

T Florian Jaeger and Roger P Levy. 2006. Speakers optimize information density through syntactic reduction. In *Advances in Neural Information Processing Systems*. pages 849–856.

Dan Jurafsky and James H Martin. 2014. *Speech and language processing*. Pearson.

Slava M Katz. 1987. Estimation of probabilities from sparse data for the language model component of a speech recognizer. *Acoustics, Speech and Signal Processing, IEEE Transactions on* 35(3):400–401.

Frank Keller. 2004. The entropy rate principle as a predictor of processing effort: An evaluation against eye-tracking data. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Barcelona, Spain, volume 317, page 324.

Per Linell. 1990. *The power of dialogue dynamics*. Harvester Wheatsheaf.

Per Linell. 1998. *Approaching dialogue: Talk, interaction and contexts in dialogical perspectives*, volume 3. John Benjamins Publishing.

Sik Hung Ng and James J Bradac. 1993. *Power in language: Verbal communication and social influence*. Sage.

Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27(02):169–190.

Ting Qian and T Florian Jaeger. 2011. Topic shift in efficient discourse production. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Boston, MA, USA, pages 3313–3318.

Alfred Schutz. 1967. *The phenomenology of the social world*. Northwestern University Press.

John R Searle. 1976. A classification of illocutionary acts. *Language in Society* 5(01):1–23.

Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27:379–423.

David Traum. 2003. Issues in multiparty dialogues. In *Advances in Agent Communication*, Springer, pages 201–211.