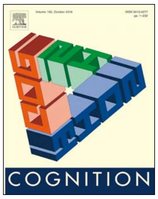
Cognition 170 (2018) 147–163

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit



Original Articles

Information density converges in dialogue: Towards an information-theoretic model

Yang Xu, David Reitter*

College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA 16802, USA

A R T I C L E I N F O

Keywords:
Dialogue
Grounding
Interactive alignment
Uniform information density
Entropy
Information

A B S T R A C T

The principle of entropy rate constancy (ERC) states that language users distribute information such that words tend to be equally predictable given previous contexts. We examine the applicability of this principle to spoken dialogue, as previous findings primarily rest on written text. The study takes into account the joint-activity nature of dialogue and the topic shift mechanisms that are different from monologue. It examines how the information contributions from the two dialogue partners interactively evolve as the discourse develops. The increase of local sentence-level information density (predicted by ERC) is shown to apply to dialogue overall. However, when the different roles of interlocutors in introducing new topics are identified, their contribution in information content displays a new converging pattern. We draw explanations to this pattern from multiple perspectives: Casting dialogue as an information exchange system would mean that the pattern is the result of two interlocutors maintaining their own context rather than sharing one. Second, we present some empirical evidence that a model of Interactive Alignment may include information density to explain the effect. Third, we argue that building common ground is a process analogous to information convergence. Thus, we put forward an information-theoretic view of dialogue, under which some existing theories of human dialogue may eventually be unified.

1. Introduction

Dialogue can be understood from multiple theoretical perspectives. It is a *joint activity* in which language plays a prominent role, a dynamic process in which the common goal and the mutual understanding between speakers are achieved through *grounding* (Clark, 1996). It can also be viewed as a two-way communication system where information flow follows general regularities, e.g., the rule of optimizing the rate of information transmission (Genzel & Charniak, 2002, 2003; Shannon, 1948), and the tendency to distribute material such that the density of information remains constant (*Uniform Information Density* hypothesis, UID, Jaeger, 2010; Jaeger & Levy, 2006; Temperley & Gildea, 2015). Dialogue is also interpreted as a process in which alignment between interlocutors occurs at multiple levels of linguistic representation, as a result of primitive priming mechanisms (a.k.a, the *Interactive Alignment Model*, IAM, Pickering & Garrod, 2004), which can result in reduced surprisal in terms of the language that each dialogue partner can observe. Whether alignment should be seen in an integrated theory as the cause, an epiphenomenon, or an additional process, is unclear; however, in this paper we explore the possibility of an account of dialogue that integrates these theories through observable, regular patterns of information distribution.

The motivation of this paper comes from two directions. First, in the work on the entropy rate constancy (ERC) principle and later in the UID framework, most of the existing studies rely on empirical evidence from corpora of written-form natural language. Only some preliminary work exists on transcripts of dialogue (Vega & Ward, 2009). It is unknown whether the inherent differences between dialogue and written text will bring up issues concerning the applicability of the theories. For example, questions can be asked such as "does the ERC/UID principle apply to the language from individual speaker alone, or to the dyad of interlocutors as a system where multiple inputs of information merge?"

Second, the other two theories mentioned above, IAM and grounding, have rich connections, but are not fully reconciled in terms of their subjects and scopes. For example, Pickering and Garrod's (2004) IAM proposes that the alignment of *situation models*, i.e., the multi-dimensional representation of the situation under discussion, is central to a successful dialogue, and that interlocutors proceed towards this higher level alignment by aligning on what they termed an *implicit common ground*. This concept analogizes to the notion of *common ground* by Clark and Brennan (1991), but whether they refer to the same process remains debatable.

---

* Corresponding author.
E-mail address: reitter@psu.edu (D. Reitter).

http://dx.doi.org/10.1016/j.cognition.2017.09.018
Received 11 October 2016; Received in revised form 6 September 2017; Accepted 27 September 2017

Therefore, by answering the call for a better examination on how the information centered theories (ERC and UID) apply to the language production in dialogue, we propose a novel information-theoretic perspective of dialogue, which captures the basic needs of successful communication – effective and efficient information exchange, and also potentially reconciles the IAM and grounding theory. From a hierarchical view of language use in dialogue, the IAM covers a full range of linguistic representation levels, but most of the empirical work only examines alignment at lower levels, such as phonemes (Pardo, 2006), lexicon (Garrod & Anderson, 1987) and syntactic structures (Branigan, Pickering, & Cleland, 2000; Pickering & Branigan, 1998; Reitter & Moore, 2014). On the other hand, grounding theory (Clark, 1996) provides mostly qualitative descriptions at higher levels by viewing dialogue as indistinctive example of joint activities in general sense. The approach taken in this study investigates dialogue at a level somewhere between the IAM and grounding perspectives, which starts with quantifying the information density at sentence level, and then takes into account the unique discourse structure in spoken conversations.

The first step of our work is to quantify the sentence information (the reason for focusing on sentence is explained in Section 2.1) in dialogue, and to examine whether it demonstrates the same overall increasing pattern that has been discovered in written language. The purpose of this step is to confirm whether the shared context between interlocutors consistently accumulates in dialogue. In the second step, we zoom into the level of the topic episodes (The reason of doing so is discussed in Section 2.2), which are delineated using a topic segmentation technique. We focus on how information density changes near the boundaries of topic episodes. The goal is to relate the information interchange between interlocutors to the process of building common ground, as well as the alignment approach. Third, we turn to alignment, hypothesizing that it is either cause or consequence of the informational exchange that requires coordination between speakers.

The remainder of this paper is organized into seven sections. After an introduction of background and raising the research questions in Section 1, some previous studies of related issues are reviewed in Section 2. In Section 3 we examine the overall trend of information density in dialogues. In Sections 4 and 5, we examine the effect of topic shifts on sentence information, and demonstrate the information converging patterns between interlocutors of different roles within the scope of topic episodes. Then in Section 6, we demonstrate how lexical alignment can be used to explain the convergence pattern. Finally, the implications of these observations are discussed in Section 7.

2. Related work

2.1. The principle of entropy rate constancy

Human communication systems such as written text and speech have been claimed to be optimized in that the rate of information being transmitted keeps constant and is close to the channel capacity, a.k.a., following principle of *entropy rate constancy* (ERC) (Genzel & Charniak, 2002, 2003; Qian & Jaeger, 2011). This line of work was inspired by Information Theory (Shannon, 1948), which relates uncertainty about the next signal to the amount of information that can be transmitted. The observation that communicators tend to distribute information evenly may have much to do with the idea that information is a measure of cognitive load: much information, or at least much surprisal at the information conveyed, is a model of how difficult it is to process (Hale, 2001).

In Genzel and Charniak's (2002) original work, the amount of information conveyed in natural language is estimated by treating the word as a random variable, and then a bulk of text becomes a sequence of random variables X_i, where X_i corresponds to the i-th word in the text. ERC predicts that the information amount of the conditional random variable, $X_i|X_1 = w_1,...,X_{i-1} = w_{i-1}$, should remain constant as i increases. Because natural language organizes messages into sentences, the context formed by all previous words $X_1 = w_1,...,X_{i-1} = w_{i-1}$ can be decomposed into two parts: The *global context* C_i, all the words from preceding sentences, and the *local context* L_i, all the preceding words within the same sentence as X_i. Then the variable that remains constant can be written as the left term in Eq. (1), where H refers to entropy. According to Information Theory, this term can be further decomposed into the two terms on the right side: $H(X_i|L_i)$, the entropy of X_i conditioned on local context L_i, and $I(X_i,C_i|L_i)$, the conditioned mutual information between X_i and its global context C_i.

$$H(X_i|C_i,L_i) = H(X_i|L_i)-I(X_i,C_i|L_i) \tag{1}$$

Intuitively, Eq. (1) says that knowing about the global context (i.e., having positive mutual information between X_i and C_i), will make the words under the current local context more predictable (hence, having lower entropy). Another important fact is that as i increases, the mutual information term $I(X_i,C_i|L_i)$ will also increase, because knowing more about the context makes it easier to predict the upcoming content. Since the whole right side of Eq. (1) also needs to remain constant (according to Information Theory), then the entropy conditioned on local context, $H(X_i|L_i)$ must also increase with i. Therefore, the increase of locally-conditioned entropy $H(X_i|L_i)$ is an indicator of the ERC principle in natural language.

Theoretically, the entropy of a random variable is defined as the expected value of the information conveyed, and the amount of information is measured by the negative logarithm of the probability of the event (Shannon, 1948). For a discrete random variable X, which has n possible outcomes, its entropy is:

$$E[-\log(P(X))] = -\sum_{i=1}^{n} P(x_i)\log P(x_i) \tag{2}$$

where the unit of entropy is *bit* when the base 2 logarithm is computed.

Applying these concepts to natural language, the random variable of our interest here is the next word given its preceding context (words), a.k.a., an N-gram. For example, to estimate the entropy of the third word X after the bigram context *this is*, then ideally, we need to enumerate all possible words that follow the context, e.g., *good*, *bad*, *great*, etc., and estimate their probabilities, $P(good|thisis)$,$P(bad|thisis)$,$P(great|thisis)$. The conditional entropy of X,$E[P(X|thisis)]$, is computed in the same way as Eq. (2). However, this method is impractical because it is nearly impossible to enumerate all the possible outcomes of X and estimate their probabilities accurately.

Therefore, alternative methods need to be used to properly estimate the entropy of words. Genzel and Charniak (2002) provide such a method by averaging the negative logarithm of probabilities of all the trigrams in a sentence, and use this per-word information to approximate the average entropy of words in the sentence (see Section 3.1 for details). Thus, Genzel and Charniak (2002) compute the $H(X_i|L_i)$ term in Eq. (1) at the sentence level, and they have confirmed that this variable increases with i, which now indicates the position of a sentence within the text.

We adopt Genzel and Charniak's method in this study and examines how the estimated information of sentence is distributed and evolves in dialogue. Genzel and Charniak (2002, 2003) in their original work and Keller (2004) in a following study all referred to the per-word average information as *entropy*, although this measure does not follow the definition of Shannon's entropy (Eq. (2)). Because *entropy* is a measure of information that can be potentially conveyed over a channel before seeing the actual transmission, as opposed to the actual *information* conveyed by actual signals, which Genzel and Charniak's approximation quantifies. Through the remainder part of this paper, we use the term *sentence information* instead of entropy to account for the post hoc nature of the approximation of $H(X_i|L_i)$.

More recently, the idea of *uniform information density* (UID, e.g., Jaeger & Levy, 2006) has extended ERC into a broader framework that