Sentence-final particle *de* in Mandarin as an informativity maximizer[1]

Jun CHEN — *University of Stuttgart*
Sean PAPAY — *University of Stuttgart*

Abstract. We propose that the Mandarin sentence-final particle *de* is attached to answers with maximal utility. We quantify this utility as informativity by drawing on a cross-entropy model, and explore the plausibility of applying cross-entropy methods (as well as related Kullback-Leibler divergence-based methods) across languages, as a precise model of understanding the subtle pragmatic meanings of sentence-final particles.

Keywords: game-theoretic pragmatics, sentence-final particle, speaker-oriented meaning, cross-entropy, rational speech act model

1. Introduction

East Asian sentence-final particles (SFPs) express a range of subtle, speaker-oriented meanings that pertain to the way the speaker conveys her belief states to the listener (e.g. Chu, 1998; Simpson, 2014; Constant, 2014). The aim of this study is to motivate employing information-theoretic pragmatic methods as a way to obtain a fine-grained understanding of the non-literal meaning contributions expressed by these particles. Specifically, we present a formal characterization of the Mandarin Chinese sentence-final particle *de*, based on cross-entropy as a part of the Rational Speech Act model (Shannon, 1948; Jäger, 2007; Goodman and Stuhlmüller, 2013). In a nutshell, we entertain the novel idea that *de* is added to the end of an aswer, when the speaker signals that the answer is maximally useful/informative, in response to the question under discussion in the immediate discourse.

The remainder of the paper is structured as follows. We first present data establishing the constraints that *de*-answers are subject to. We argue that these data are captured if *de*-answers are the most informative in the context. This is followed by a formal implementation of information load in terms of negated cross-entropy under a simplified Rational Speech Act model. Afterwards we discuss potential alternatives to the RSA approach, compare our notion of informative partial answer with other definitions of partial answerhood, and suggest applications of similar probabilistic approaches to other particles cross-linguistically.

2. Data

We focus on the distribution of *de* in narrow focus answers that address an immediate prior *wh*-question. We assume conversation is recognized as a signaling game (van Rooy, 2004; Merin, 2011). Our central empirical claim is thus that in a cooperative game, the speaker adds the particle *de* to a propositional answer to signal that she conveys the most informative answer to

[1]We are indebted to the anonymous abstract reviewers for the Sinn und Bedeutung conference 2021 for their valuable input. We have also benefited from discussions with Daniel Hole, Judith Tonhauser, Fabian Bross, Sebastian Padó and Swantje Tönnis, as well as audiences at the 2021 ESSLLI student session and the linguistic colloquium at University of Stuttgart. Needless to say, all the remaining errors are our own.

© 2022, Jun Chen and Sean Papay. In: Daniel Gutzmann and Sophie Repp (eds.): Proceedings of Sinn und Bedeutung 26, pp. 225–238. University of Cologne.

her knowledge. As far as we are concerned, this observation has not been made in the literature.

Example (1) illustrates the constraint *de* imposes upon possible answers. Imagine it is in the knowledge of B that teacher Cai takes charge of reimbursement on Monday, and teacher Wang takes charge of reimbursement on the other weekdays (Tuesday till Friday). Now this information is not part of mutual knowledge, i.e. it is not already known by A. Given this, B may provide information about teacher Cai within a partial answer as in (1a). In contrast, however, it is less felicitous for B to utter a *de*-answer as in (1b).[2] Here the judgment has to be elicited based on the specific prior context. An *out-of-the-blue* utterance does not yield a similar contrast of acceptability. The same applies to subsequent reported judgments.

(1) QUD (Speaker A): Who should I find if I want to get a reimbursement? B answers:

a. Zhouyi shi cai laoshi fuze baoxiao.
Monday SHI Cai teacher in.charge reimbursement
'On Monday, it is teacher Cai who is in charge of reimbursement.'

b. ?? Zhouyi shi cai laoshi fuze baoxiao **de**.
Monday SHI Cai teacher in.charge reimbursement DE
Intended: 'On Monday, it is teacher Cai who is in charge of reimbursement.'

Compared to (1b), the attachment of *de* to the more informative answer about teacher Wang gives rise to no infelicity, as in (2).

(2) QUD is the same as in (1)

Cong zhouer dao zhouwu dou shi wang laoshi fuze baoxiao **de**.
from Tuesday till Friday DOU COP Wang teacher in.charge reimbursement DE
'From Tuesday until Friday, it is teacher Wang who is in charge of reimbursement.'

The information about teacher Cai and teacher Wang may be provided in two parallel partial answers juxtaposed with each other. In such case, we entertain four possibilities where the *de* particle could be attached. In these permutations, listed in (3), the pattern is that *de* prefers the more informative answer about teacher Wang, and not the less informative answer about teacher Cai: ($3B_1$) and ($3B_2$) are judged better compared with ($3B_3$) and ($3B_4$).[3]

(3) QUD (Speaker A): Who should I find if I want to get a reimbursement?

B_1: Cong zhouer dao zhouwu, dou shi Wang laoshi fuze baoxiao **de**.
from Tuesday until Friday DOU COP Wang teacher in.charge reimburse DE
Zhouyi, shi cai laoshi fuze baoxiao.
Monday, COP Cai teacher in.charge reimburse
'From Tuesday until Friday teacher Wang is in charge of reimbursement. On Monday, teacher Cai is in charge of reimbursement. '

[2]Both answers (1a) and (1b) are not intuitively speaking the most informative one, given that the information about teacher Wang better addresses the questioner's goal of identifying the most likely way to get the reimbursement done. (1a) is typically used as a secondary partial answer (for the sake of providing complete answers) alongside another partial answer providing teacher Wang's information.

[3]($3B_3$) and ($3B_4$) may be accepted, but only if the utterer wants to emphasize Monday as particularly relevant to the question under discussion under a different knowledge state. We return to this issue shortly.

B$_2$: $^?$Zhouyi, shi cai laoshi fuze baoxiao. Cong zhouer dao zhouwu,
Monday, COP Cai teacher in.charge reimburse from Tuesday until Friday
dou shi Wang laoshi fuze baoxiao **de**.
DOU COP Wang teacher in.charge reimburse DE
'On Monday, teacher Cai is in charge of reimbursement. From Tuesday until Friday teacher Wang is in charge of reimbursement.'

B$_3$: $^{??}$Cong zhouer dao zhouwu, dou shi Wang laoshi fuze baoxiao.
from Tuesday until Friday DOU COP Wang teacher in.charge reimburse
Zhouyi, shi cai laoshi fuze baoxiao **de**.
Monday, COP Cai teacher in.charge reimburse DE
'From Tuesday until Friday teacher Wang is in charge of reimbursement. On Monday, teacher Cai is in charge of reimbursement.'

B$_4$: $^{??}$Zhouyi, shi cai laoshi fuze baoxiao **de**. Cong zhouer dao zhouwu,
Monday, COP Cai teacher in.charge reimburse DE from Tuesday until Friday
dou shi Wang laoshi fuze baoxiao.
DOU COP Wang teacher in.charge reimburse
'On Monday, teacher Cai is in charge of reimbursement. From Tuesday until Friday teacher Wang is in charge of reimbursement.'

Table 1 summarizes the pattern of answers in (3).[4] The generalization is maintained regardless of the order of the two partial answers, excluding the possibility that the above contrast is due to the preference for uttering the more informative partial answer earlier.

Example numbers	Prejacents & DE & Continuations	Judgments
(B$_1$)	Tue–Fri DE < Mon	✔
(B$_2$)	Mon < Tue–Fri DE	?
(B$_3$)	Tue–Fri < Mon DE	??
(B$_4$)	Mon DE < Tue–Fri	??

Table 1: Pattern of judgment based on permutations of partial answers with placement of *de*

We suggest that linear order might factor into the rather mild differences between the examples (3B$_1$) and (3B$_2$). Specifically, with the novel information structured into two information chunks/units, by holding back the more informative chunk (hence uttering first the less informative chunk), the maxim of Quantity is violated. The maxim of Quantity is maintained if the more informative partial answer is uttered preceding a less informative one (Afterwards, what's left is now the most informative chunk available, so that no violation is incurred).[5] In this way,

[4]Our proposal has treated the placement of the *de* particle as a pragmatic phenomenon. The pattern we are dealing with here is obviously very subtle. To make sure that the empirical basis of the paper is as solid as possible, and particularly because we are arguing for a pattern of the particle that has not been pointed out before, we have conducted a pilot acceptability judgement task (please see our description and results here). We want to note that the positioning of *de* in less informative partial answers still yields scores tending towards being acceptable. This, we believe, points to the strong cancellability and tendency for accommodation that are characteristic of pragmatic-based constraints, especially those involving speaker-oriented meaning, and hence could be made compatible with a game-theoretic pragmatic characterization.

[5]It is possible to have less informative partial answer in front of the more informative one, if some order of

the observation that ($3B_2$) is rated as more natural than ($3B_1$) can be explained from independent Gricean grounds. Crucially, since both ($3B_3$) and ($3B_4$) are rated worse than ($3B_2$), the infelicity with *de*-attachment in the former two examples cannot be reduced to the precedence of a less informative answer over a more informative one.

Again, bare, *de*-less narrow focus answers receive no degraded judgment. Without *de* attachment, it is fine to juxtapose a less informative answer with its more informative continuation, evidenced in (4).

(4) Zhouyi shi cai laoshi fuze baoxiao, zhouer zhi zhouwu shi wang
Monday SHI Cai teacher in.charge reimbursement, Tuesday until Friday SHI Wang
laoshi fuze baoxiao.
teacher in.charge reimbursement
'On Monday, teacher Cai is in charge of reimbursement. On Tuesday until Friday, teacher Wang is in charge of reimbursement.'

The above pattern holds for other ways to partition possible partial answers, as far as there is a difference in informativity among the answers. Thus, in (5), *de* cannot attach to a proposition expressing quantification over a small number of events, when this partial answer is contrasted against another answer expressing quantification over a majority of events. The infelicity disappears if *de* is attached to the 'majority-event' answer.

(5) Context: A: Who should I find if I want to get a reimbursement? B answers:

Ou'er shi cai laoshi fuze baoxiao ($^{??}$**de**). Tongchang shi wang
occasionally SHI Cai teacher in.charge reimbursement (de) normally SHI Wang
laoshi fuze baoxiao.
teacher in.charge reimbursement

Intended: 'Occasionally it is teacher Cai who is in charge of reimbursement. Normally it is teacher Wang who is in charge of reimbursement.'

From the above discussions, we can understand a *de*-answer as providing the most informative answer to the listener, according to the knowledge of the speaker. The function of *de* as an information maximizer would immediately explain the fact that *de* tends to attach to exhaustive answers (Hole, 2011). This is because by excluding other alternatives than the one that corresponds to the focus value denoted by the prior question's *wh*-part, the exhaustive answer is always the most informative answer a speaker can provide.

(6) Context: A: Who should I find if I want to get a reimbursement? B answers:

Shi cai laoshi fuze baoxiao ($^{??}$**de**).
SHI Cai teacher in.charge reimbursement (de)

'It is teacher Cai who is in charge of reimbursement (not the others).'

---

listing is available. This could be the case with a conventional temporal/numerical order, e.g. the tendency to list weekday/month following the ascending order of integers (1, 2, 3, ...). What we are postulating is that the maxim of Quantity and the conventional listing order could enter into competition, with the speaker opting for one principle for bringing out information and temporarily suspending the priority of another.

Finally, under the situation in (7), *de* shifts to the 'Monday-teacher Cai' answer. The utterance here is in a situation where teacher Wang handles reimbursement by default. The speaker and the listener expect Wang every day, and it is information about the rather exceptional case of teacher Cai's handling reimbursement that carries the higher information load.

(7) Context: There is prior knowledge that teacher Wang is much more likely than teacher Cai to handle refunds (e.g., Cai only handles refunds when Wang is on vacation)

Zhouyi, shi cai laoshi fuze baoxiao **de**. Cong zhouer dao zhouwu, dou
Monday, COP Cai teacher in.charge reimburse DE from Tuesday until Friday DOU
shi Wang laoshi fuze baoxiao.
COP Wang teacher in.charge reimburse

'On Monday, teacher Cai is in charge of reimbursement. From Tuesday until Friday teacher Wang is in charge of reimbursement.'

3. Analysis

3.1. Formal implementation

In the above we have argued that the *de* particle encodes an informativity optimizer, to the effect that the proposition it attaches to offers the maximal amount of useful information that the speaker wants to convey to the listener. Our goal now is to provide a precise characterization of the notion of informativity that underpins our analysis. To achieve this goal we adopt a framework coupling information theory with Bayesian statistics. Specifically, we build our analysis upon the Rational Speech Act (RSA) model (Goodman and Stuhlmüller, 2013; Spector, 2017), with which we measure the amount of information contained in a given proposition and compare it with that of minimally different propositions. In mathematical terms, informativity can be defined in terms of the remaining uncertainty (as measurable in i.e. bits) for the hearer about the state of the world. We follow the assumptions that the number of bits of information still missing to the hearer is finite (as long as the speaker is not lying) and calculable.

RSA, as a formally precise way to quantify and compare the informativity across different utterances, has found fruitful applications in topics that seek to capture the interlocutors' presumptions about their beliefs, and to predict which utterances or actions they may prefer (Frank and Goodman, 2012; Goodman and Stuhlmüller, 2013). A standard case is classic scalar implicature, e.g. uttering *some P* invites the inference that *not all P* (Horn, 2005). The ability to cancel such *not-all* inference is subject to the increasing knowledge given a situation. In general, the extent to which an implicature can be cancelled depends on what the listener knows about the speaker's knowledge state. Consider a scenario with three apples in total. The speaker and the listener have the conversational goal of conveying the information about the number of red apples in the simplest way. In the case where the speaker has complete knowledge of the number of red apples, and the listener knows that the speaker has complete knowledge, then the listener is very likely to infer 'not all apples are red', upon hearing the speaker's utterance 'some apples are red'. Consider now an alternative scenario where the speaker has partial access to information about the apples, e.g. he only knows if one apple is red, and the listener knows the speaker does not have complete knowledge. Then the listener is less likely to draw a *not-all* inference upon hearing 'some apples are red', compared to the previous scenario. Thus, the implicature is 'canceled.'

Under an RSA-based approach, incremental cancellation is achieved by resorting to expected utility (informativity) – a quantity that depends on the speaker's belief distribution (Frank & Goodman 2012, Goodman & Stuhlmüller 2012). The RSA model predicts an interaction between (shared) knowledge and captures how the speaker's belief states could influence a listener's interpretation. We similarly propose an analysis of the *de*-particle in an RSA-based Bayesian framework, in which the knowledge states of the speaker and the listener are represented probabilistically. We start with a description of our model and how it characterizes the *de* particle. We then demonstrate that alternative characterizations of informativity based on entailment facs challenges to capture the same data, offering additional motivation for our use of probabilisic machinery. We will see later in Section 4 that our approach additionally allows us to represent phenomena such as imperfect speaker knowledge, or arbitrary common-knowledge priors.[6]

In our model, *de* is characterized as an informativity maximizer in information-theoretic terms, by measuring the **cross-entropy** encoded in the prejacent of the particle. To do so, the following components are required:

- a set T of all possible worlds (or equivalence classes of worlds, where two worlds are equivalent if they address the QUD in the same way);
- a speaker, who holds a belief state $S(t)$ over the worlds $t \in T$ that she would like to convey;
- a listener, who forms a belief state $L(t|m)$ over worlds $t \in T$, dependent on some message m.

Informativity thus can be measured as the amount of information provided about the speaker's belief state, representable as the negation of the cross entropy H (Shannon, 1948) between the speaker's and the listener's belief states after the speaker communicates message m, i.e.

(8) $-H(S(\cdot), L(\cdot|m)) = \sum_{t \in T} S(t) \log L(t|m).$

Utterances which bring the listener's belief state closer to the speaker's have a higher informativity, while utterances which contradict worlds deemed possible by the speaker have a negative infinite informativity.

Consider now the following example for illustration.

[6] Additional advantages enjoyed by such game-theoretic modeling of beliefs, compared to the single-step calculation of expected utility, include that it enables us to optimize the speaker's distribution given incomplete information access (hence capturing the epistemic effects, Franke (2005)) and allows for multiple levels of recursion (*I think that you think, that I think that...*) so that the influence of incremental shared knowledge on the listener's interpretation can be seen clearly.

(9) QUD: Who should I find if I want to get a refund this week?

{Zhouer yizhi dao zhouwu}/ {tongchang}, shi Wang laoshi fuze
{Tuesday all.the.way until Friday}/ {usually} COP Wang teacher handle
baoxiao (**de**); {zhouyi}/{ou'er}, shi cai laoshi fuze baoxiao ($^{??}$**de**).
refund DE; Monday/occasionally, COP Cai teacher handle refund DE

'From Tuesday until Friday/Usually, teacher Wang handles refunds; on Monday/occasionally, teacher Cai handles refunds.'

Let T be the set of all who's-in-charge assignments, each such assignment mapping one day of the workweek to one of two people who handle refunds. Let $S(t) = \delta_{t,s}$, where $s \in T$ is the unique world described in example (9), i.e. the speaker has perfect knowledge of world s. Finally, let $L(t|m)$ be the uniform distribution over all worlds consistent with the literal meaning of m, i.e. a literal listener with a uniform prior over worlds. From this, the informativity of a message m is $\log L(s|m)$. If m is compatible with s, this is the negative logarithm of the number of worlds compatible with m. As logarithms are monotonic, informativity can be measured by counting worlds: utterances with more compatible worlds are less informative.

The response featuring *from Tuesday to Friday* is compatible with 2 worlds, and *on Monday* with 16 worlds (thus more uncertain and less informative). In the case of quantifier *usually*, assuming it represents *more than half of the weekdays*, the proposition is compatible with 16 worlds. As for *occasionally*, assuming a meaning of *at least one weekday*, the second proposition is compatible with 31 worlds, and is thus less informative (as it only excludes one possibility).

Hence we can formally capture the intuition that example (9)'s first proposition is more informative than the second. The constraint that the *de* particle must attach to the more informative proposition can be formulated accordingly.

Now we postulate the following pragmatic constraint of *de*, in which the usefulness of an utterance is characterized in terms of informativity.

1. Maximal expected informativity condition
   de $(p)(w)$ is felicitous iff the informativity $I_{\max}(u|w)$, given the utterance u that corresponds to the proposition containing *de* and its prejacent p.
2. Definition of $I_{\max}$
   Given alternatives to u at w, where $Alt(u) = \{u_1, u_2, \ldots u_i\}$, $I_{\max}(u|w)$ iff $I(u|w) > I(u_i|w)$.

3.2. Comparison with other proposals

Our above notion of utterance informativity differs from the understanding of informative answers in terms of the entailment relation. Entailment-based characterizations have been employed to capture exhaustive answers. To see this, we illustrate with one particular approach that treats partial answers as elements of partially ordered domain (Szabolcsi and Zwart, 1993; Champollion, 2017). That is, partial answers encoded in the proposition could be mapped to the join semilattice.

A partial ordering is a reflexive, transitive, antisymmetric relation, defined in (10).

(10) a. Reflexivity: Everything is part of itself.
b. Transitivity: Any part of any part of a thing is itself part of that thing.
c. Antisymmetry: Two distinct things cannot both be part of each other.
d. Unique sum: Every nonempty set has a unique sum.

A free join semilattice could be described by its property of reflexivity, transitivity, antisymmetry, taken together with the uniqueness of sums.[7] A typical model is shown in Figure 1.

(11) A binary sum $x \oplus y$ is defined as:
The sum of two things is the thing which contains both of them and whose parts each overlap with one of them.

Suppose the relevant context contains a QUD *who will come*? and a domain consisting of individuals {Zhangsan, Lisi, Wangwu}, we arrive at a set of propositions based on atomic individuals: {$^\wedge$Zhangsan will come, $^\wedge$Lisi will come, $^\wedge$Wangwu will come}. These propositions form a free join semi-lattice based on the entailment relation, from which we have all the possible Hamblin answers. Using *a*, *b*, *c* to stand for Zhangsan, Lisi and Wangwu, respectively, and using *K* to refer to the extension of *will come*, the set of Hamblin answers are then represented as $\{K(a), K(b), K(c), K(a+b), K(a+c), K(b+c), K(a+b+c)\}$. The exhaustive (actual) answer then corresponds to the highest level of summation, based on the mapping of the summations given the partial answers. The answer at the highest level asymmetrically entails the other Hamblin answers. In this sense, the exhaustive answer is the most informative among alternative answers.

However, the entailment-based characterizations are ill-suited for informative answers under our definition of expected utility. We illustrate with the example (12).[8]

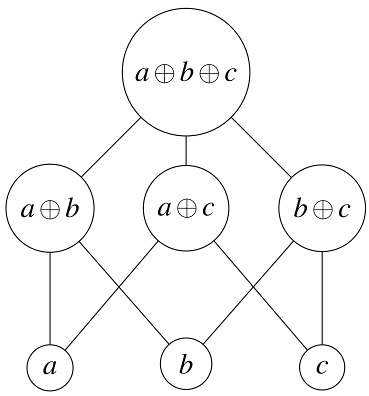

Figure 1: An example for a free join semilattice with binary sums

[7]The freedom property means that whenever two pairs of elements are distinct, their unions are distinct (Szabolcsi & Zwarts 1993).

(12) Context: A: When can I get a reimbursement? B answers:

Shi zhouyi baoxiao (??**de**).
SHI Monday reimbursement (de)

'It is on Monday that reimbursement is processed.'

In the current example, the information that varies in the two partial answers (Tuesday until Wednesday versus Monday) could be recognized as two elements on the join semilattice. One may propose that the information maximizer *de* is attached to the highest level of summation. However, this solution faces the problem that the information conveyed by contrasting partial answers is not comparable on a join semilattice. *Monday* is a basic element at the bottom level of the semilattice (e.g. a in Figure 1). It is considered as an atom of the lattice, since it has no further proper part in the current context. On the other hand, *Tuesday until Wednesday* corresponds to a join of the atom *Tuesday* (b) and the atom *Wednesday* (c). This join could be represented by $b \oplus c$ in Figure 1. Though it seems that *Tuesday until Wednesday* (i.e. $b \oplus c$) is positioned higher than the atom *Monday* (a) on the semilattice, they are incomparable since a is not a proper part of $b \oplus c$ and vice versa. In other words, *Monday* is not a part of *Tuesday until Wednesday* that is distinct from *Tuesday until Wednesday* and vice versa (*Tuesday* would be a proper part of *Tuesday until Wednesday*). Thus, in the current case, a lattice-theoretic formulation cannot predict that *Tuesday until Wednesday* is more informative than *Monday*.

Aside from the incomparability problem with elements that do not stand in a part-of relation, the entailment-based approach also fails to cover cases where the dynamic update during the speaker-listener interaction leads to the change in the relative informativity of individual messages. For example, an atomic answer that encodes an unlikely event (represented in the prior knowledge states) may receive a higher expected utility, which cannot be reflected in terms of the entailment relation (see section 4.1 for details). As another example, the presentational order of information affects the epistemic knowledge about the speaker belief state: The information presented at the beginning would imply the current knowledge state of the speaker. Consider a setting where the listener expects a conventional temporal/numeral order. If the speaker starts with a partial answer about Tuesday (instead of Monday), the listener may realize that information about Monday is unexpectedly not provided. This could be accounted for by the first-level recursive RSA model. The first-level listener, upon hearing the partial answers presented in an unconventional temporal ordering, now assigns a lower utility value (or a higher cost, depending on the details of the setting) to the partial answer with Monday. Intuitively, the listener would infer that either the speaker does not know about Monday, or anticipates more important information about Monday in the later conversation.

To sum up, the entailment-based account faces problems in the following two aspects: 1) the case where two partial answers are not in a part-of relation (e.g. *Monday* vs. *Tuesday until Wednesday*); 2) the case involving backward inference and the update over speaker's prior knowledge state. A game-theoretic pragmatic approach thus enjoys the advantage of measuring differences in expected utility in mutually non-entailing propositions, as well as keeping track of the dynamic update of (shared) knowledge.

[8] Here we use a simpler *wh*-question for demonstration, since in our previously used weekday-teacher examples, the underlying QUD is broken down into pair-list questions where weekday-teacher pairs yield the atomic answers. We want to avoid such complication.

4. Extensions

Our model, with modifications under the information theoretic framework, promises to expand beyond the use of the *de* particle and apply to more cases involving speaker-listener interaction in which the respective belief states are crucial to our understanding. Specifically, we show that the current model is able to capture the case where the speaker and the listener have an arbitrary common-knowledge prior. Additionally, we discuss the potential settings for other kinds of partial answers (i.e. mention-some vs. mention-all). Finally, we show that with a minimal amount of changes, the model could be expanded to capture a range of particles cross-linguistically.

4.1. Cases with enriched prior knowledge

In the current setting of the model, we assume a uniform common knowledge prior for the listener and speaker. However, this could be changed so that more data could be explained. For example, in example (7, repeated here as in 13) we already see that *de* attaches to the Monday-teacher Cai pair, given a situation where the speaker and the listener both share a common knowledge that teacher Cai only handles refunds when teacher Wang is on vacation.

(13) Context: A great proportion of the faculty is on vacation during August. A asked: 'Who should I find if I want to get a refund this week?' B:

zhouyi shi Cai laoshi fuze baoxiao **de**. zhouer yizhi dao zhouwu,
Monday COP Cai teacher in.charge reimburse **DE** Tuesday all.the.way until Friday
shi Wang laoshi fuze baoxiao.
COP Wang teacher in.charge reimburse

'On Monday, teacher Cai is in charge of reimbursement DE. From Tuesday until Friday, teacher Wang is in charge of reimbursement. '

In other words, the speaker and listener both hold non-uniform priors, with higher probabilities assigned to worlds in which teacher Wang is in charge a majority of the days except the days when she is on vacation. In addition, the speaker knows that the listener doesn't know when teacher Cai is on vacation. With the additional assumptions that teacher Cai has a 1 percent chance of being on vacation on a given day, and that each day is independent from each other, we can calculate informativity using cross-entropy as we have done previously.[9]

Assume a situation where teacher Wang is out for vacation on Monday, and is back to work starting Tuesday. This yields a work schedule with teacher Cai in charge on Monday, and teacher Wang on charge from Tuesday to Friday – we notate this world state s. By our priors, the probability of s is $0.01*(1- 0.01)^4$ (given that the probability that teacher Wang is on vacation is 0.01, and there are five working days in a week). After hearing the utterance m_1 'On Monday, teacher Cai handles refunds,' the conditional probability of s given the utterance

[9] We acknowledge that here the example setting is unrealistic. We use these oversimplified assumptions merely to demonstrate how informativity can be calculated numerically in cases with non-uniform priors.

$(L(w|u))$ is $0.01 * \sum_{k=0}^{4} \binom{4}{k} * 0.01^k * 0.99^{4-k}/0.01$, which simplifies to 0.99^4. Similarly, given the proposition m_2 'from Tuesday until Friday, teacher Wang handles refunds', the corresponding conditional probability is 0.01. From this, the informativity of the corresponding proposition could be calculated by $\log L(s|m)$. With base-10 logarithms, this yields an informativity of approximately -0.0174 for m_1, and exactly -2 for m_2.

Therefore, we predict that the *de* particle prefers to attaching to the utterance *On Monday teacher Cai handles refunds*, since it is more informative in this situation.

In summary, the informativities of our utterances are affected by both the number of events addressed, and the probabilities of those events. As we see in our example, an utterance addressing a single unlikely event can be more informative than one addressing many likely events.

4.2. Other types of partial answers

Our notion makes use of expected utility, which bears resemblance to analyses of mention-some answers in terms of expected utility (e.g. van Rooy, 2004). It has long been observed (e.g. Hamblin, 1973) that *wh*-questions can be addressed by partial answers that only mention *some* positive instances. The question in (14), in a context where a tourist is new to the town, is more appropriately answered by mentioning just one place where he can successfully get newspaper (a partial answer), instead of mentioning all relevant places exhaustively that the listener know (a complete answer). That is, mentioning just one element of the set of alternative 'equally best' places suffice to resolve the question.

(14) Where can I buy newspaper?

A more recent work from van Rooy (2004) argues that whether the mention-some reading suffices to resolve the question depends on the expected utility of the answers. He crucially applies the *answer rules*, which are rules that determine which answer will be given in which worlds. Assuming asking the question is cost free, together with the *answer rules* and an 'empty' context, it is possible to calculate the expected utility of the *wh*-question, which have both readings. Due to the property of mention-some and mention-all questions denoting two partitions that have subset relations, s.t. $Q_{\text{some}} \sqsubseteq Q_{\text{all}}$, the average utility of the mention-some reading of the question can never be higher than the utility on the corresponding mention-all reading. Importantly, there are cases where the expected utilities coincide: If the mention-some answer is known to be equally useful as the mention-all answer, and it needs less effort or is shorter, the *wh*-question receives the mention-some interpretation.

Van Rooy's RSA-based analysis connects to ours in the sense that expected utility underlies the calculation that determines the condition where a mention-some answer or a *de*-answer is provided. The two conditions differ, however, in that *de* requires that its prejacent has a higher expected utility than its potential sister partial answers. This way, *de* may attach to a partial answer even if it is not as useful as the mention-all answer. It just needs to be more useful than other partial answers available (Tuesday to Friday versus Monday). The requirements of *de* also allow it to attach to an exhaustive answer, as such answer vacuously satisfy the condition of maximal utility (there is no sister partial answer).

4.3. Other particles

The cross-entropy approach developed for *de* promises to be expanded to capture a broad range of sentence-final particles across languages that are shown to compare the interlocutors' belief states (e.g. Cantonese *ge*, Japanese *no*, etc.). In the remainder we mention briefly applying similar mechanisms to Japanese sentence-final particles and German discourse particles.

The particle *no* in Japanese is argued to indicate that the speaker signals that *no*'s prejacent is incompatible with what is in the pre-utterance belief states of the listener (Cook, 1990).

(15) Context: Speaker A saw a butter on the table and complained that she doesn't like the salted butter. B answers:
B: kore-ga oshio-o haitte nai batta na **no**.
this-SBJ salt-ACC add-PROG NEG butter PRT NO
'This is the butter without salt inside.'

Information theoretically, the difference between the listener's prior and posterior belief state could be measured by Kullback-Leibler (KL) divergence (Baldi, 2002), i.e. for given a message m, $H(L(\cdot|m), L(\cdot)) - H(L(\cdot|m))$. As KL divergence differs from cross entropy only by the posterior entropy $H(L(\cdot|m))$, this treatment follows as a natural extension of our account of informativity used earlier. While the informativity of an utterance measures how much it brings a listener's belief state towards the speaker's, the KL divergence here measures how much a listener's belief state has changed, by keeping track of the listener's prior and posterior belief states with regard to an utterance.

In example (15), the context indicates a low-entropy prior belief state, where A believes with high probability that *the butter is salted*. For simplicity, we assume butter-saltedness to be a binary variable, with the prior belief distribution highly skewed in the direction of it *being salted*. Since the utterance m contradicts the listener's expectation, after the utterance the listener's posterior belief state is flipped towards certain belief in *unsalted* butter. This drastic difference between prior and posterior belief states should be reflected by a high KL divergence. More explicitly, we could quantify how 'far' the listener's belief state has shifted upon hearing this utterance, based on $S(m) = -\log(L(\text{unsalted}))$, where L is the listener's prior.

Such treatment is theoretically desirable, as the distance between the prior and posterior belief states of the listener can be directly measured by interpretable units such as bits, if concrete numbers are given to the listener's prior and posterior probability. This allows us to predict the gradience in the felicity of the particles when occurring with the utterance.

Relatedly, note that the very rich literature on the role played by German discourse particles in modulating speaker's epistemic belief states (cf. Zimmermann, 2011; Grosz, 2014, 2022) could benefit from the current probabilistic treatment. We briefly illustrate with the particles *ja* and *doch* within the setting of our model. In German, *ja* and *doch* form a pair of opposing elements encoding the state of speaker belief, as in (16) and (17) (Grosz, 2022: p9).

(16) Gesüßter Mate-Tee schmeckt **ja** scheußlich.
sweetened mate-tea tastes ja disgusting
'Sweetened mate tea [JA] tastes disgusting. '

(17) A: Gesüßter Mate-Tee schmeckt gut.
sweetened mate-tea tastes good
'Sweetened mate tea tastes good. '

B: Gesüßter Mate-Tee schmeckt **doch** scheußlich.
sweetened mate-tea tastes doch disgusting
'Sweetened mate tea [DOCH] tastes disgusting. '

Their meaning contributions can be described in (18), respectively, given a proposition p.

(18) a. $[\![ja]\!](p)$= p is true and speaker believes p uncontroversial.
b. $[\![doch]\!](p)$ presupposes that somebody believes a $\neg p$-entailing proposition q.

As *doch* presupposes some disagreement about p, an analysis of *doch* could look similar to that of Japanese *no*, where we must find a **high** KL divergence between two belief states about p (possibly those of speaker and listener). Conversely, *ja*, which means that p is uncontroversial, could be analyzed to require a **low** KL divergence between two belief states about p. We leave the details to future research.

5. Conclusion

In this study, we claim that the Mandarin sentence-final particle *de* is used to indicate that the speaker signals that an answer is maximally informative, in response to the question under discussion in the immediate discourse. We then show that cross-entropy allows for a formally precise model of the pragmatic notion of informativity. We argue that our case study opens the way for the wider applicability of cross-entropy-based methods (as well as related KL-divergence-based methods) as part of RSA-based approaches to cover the speaker meanings of discourse particles across languages (East Asian sentence-final particles, German discourse particles, etc.).

Note that the probabilistic reasoning we have assumed is encoded in the meaning of the discourse particles. In doing so we deviate from the more traditional view where truth-conditional denotations are fed to a probabilistic/Bayesian pragmatics. The current approach thus is in line with recent developments towards a complex interface approach to meaning, encompassing the realm of probabilistic semantics/pragmatics (e.g. Champollion et al., 2019), as well as other facets such as the speech act theory (e.g. Krifka, 2017, 2019).

References

Baldi, P. (2002). A computational theory of surprise. In B. Farrell and H. van Tilborg (Eds.), *Information, Coding, and Mathematics*, pp. 1–25. MA: Norwell: Norwell.

Champollion, L. (2017). *Parts of a whole: Distributivity as a bridge between aspect and measurement*. Oxford: Oxford University Press.

Champollion, L., A. Alsop, and I. Grosu (2019). Free choice disjunction as a rational speech act. In K. Blake, F. Davis, K. Lamp, and J. Rhyne (Eds.), *Semantics and Linguistic Theory (SALT)*, Volume 29, pp. 238–257.

Chu, C. (1998). *A Discourse Grammar of Mandarin Chinese*. Berlin: Peter Lang.

Constant, N. (2014). *Contrastive topic: Meanings and Realizations*. Ph. D. thesis, University of Massachusetts at Amherst.

Cook, H. M. (1990). An indexical account of the japanese sentence-final particle *No*. *Discourse Processes 13*, 401–439.

Frank, M. and N. Goodman (2012). Predicting pragmatic reasoning in language games. *Science 336*(6084), 998–1008.

Franke, M. (2005). *Signal to act: Game theory in pragmatics*. Ph. D. thesis, Amsterdam: Institute for Logic, Language and Computation.

Goodman, N. and A. Stuhlmüller (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science 5*(1), 173–184.

Grosz, P. (2014). German *doch*: An element that triggers a contrast presupposition. In R. Baglini, T. Grinsell, J. Keane, A. Singerman, and J. Thomas (Eds.), *Proceedings of the Chicago Linguistic Society 46*, pp. 163–178.

Grosz, P. (2022). Discourse particles. In D. Gutzmann, L. Matthewson, C. Meier, and H. Rullmann (Eds.), *The Companion to Semantics*, pp. 619–652. Oxford: Wiley.

Hamblin, L. (1973). Questions in Montague English. *Foundations of Language 10*, 41–53.

Hole, D. (2011). The deconstruction of Chinese *shi...de* clefts revisited. *Lingua 121*(11), 1707–1733.

Horn, L. (2005). The border wars: a neo-Gricean perspective. In K. von Heusinger and K. Turner (Eds.), *Where Semantics meets Pragmatics*, pp. 21–48. Oxford: Oxford University Press.

Jäger, G. (2007). Game dynamics connects semantics and pragmatics. In A.-V. Pietarinen (Ed.), *Game Theory and Linguistic Meaning*, pp. 89–102. Amsterdam/New York: Elsevier.

Krifka, M. (2017). Epistemic, evidential and discourse modalities in commitment space semantics. In *Not-At-Issue-Meaning and Information Structure Workshop*, Volume 1, pp. 1–13.

Krifka, M. (2019). Layers of assertive clauses: propositions, judgements, commitments, acts. In J. Hartmann and A. Wöllstein (Eds.), *Propositional arguments in cross-linguistic research: Theoretical and empirical issues*, pp. 1–45. Tübingen: Gunter Narr Verlag.

Merin, A. (2011). Information, relevance, and social decision making: Some principles and results of decision-theoretic semantics. In J. Ginzburg, L. Moss, and M. de Rijke (Eds.), *Logic, Language, and Information*, pp. 179–221. Standford, California: CSLI Publications.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal 27*(3), 379–423.

Simpson, A. (2014). Sentence-final particles. In C.-T. J. Huang, Y.-H. A. Li, and A. Simpson (Eds.), *The handbook of Chinese linguistics*, pp. 156–179.

Spector, B. (2017). The pragmatics of plural predication: Homogeneity and non-maximality within the rational speech act model. In A. Cremers, T. van Gessel, and F. Roelofsen (Eds.), *Proceedings of the 21st Amsterdam Colloquium*, pp. 435–444.

Szabolcsi, A. and F. Zwart (1993). Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics 1*(3), 235–284.

van Rooy, R. (2004). Utility of mention-some questions. *Research on Language and Computation 2*(3), 401–416.

Zimmermann, M. (2011). Discourse particles. In C. Maienborn, K. von Heusinger, and P. Portner (Eds.), *Semantics: an international handbook of natural language meaning*, pp. 2011–2038. Berlin: Mouton de Gruyter.