Modelling Various Kinds of Specifications

Guanyi Chen (Utrecht University), Kees van Deemter (Utrecht University) and Chenghua Lin (University of Aberdeen)*

We introduce a new approach to defining and sub-categorising over-specification, which offers a better analysis than commonly used metrics such as DICE. Next, we report on a primary analysis of the Chinese MTuna corpus [1] based on the new approach.

1 Introduction

Referring Expression Generation (REG) algorithms focus on producing referring expressions (REs) that enable the hearer to identify the entity in the given context [2]. REG has received a considerable amount of attention in cognitive sciences [3], where some key concepts were developed, such as the notions of over- and under-specification [4, 5]. These concepts are important for understanding references yet, as we shall argue, our understanding of these concepts is still imperfect: many existing definitions lack formal precision, while also failing to make some important distinctions of different types of over- and under-specification.

Consequently, when REG algorithms are evaluated, they are seldom compared in terms of over- and under-specification, but are more often in terms of the DICE metric [6, 7], which is based on a simple comparison between sets (i.e., a set of properties expressed). Yet DICE bears no direct relation to the notions of over- and under-specification, e.g., a referring expression can have a low DICE score even though it manages to single out the referent without any over- or under-specification [7, 8]. We are developing new ways of thinking about over- and under-specification, which are fine-grained and formally precise (like DICE) and also theoretically motivated (like over- and under-specification), and which we are employing to compare different REG algorithms and different languages.

2 Re-defining Specifications

We use notations similar to that of [3]. A description D can be represented by a bag (i.e., multi-set) of properties $\mathcal{D} = \{P_1, ..., P_n\}$. For instance, $\{A, B, A\}$ is the same as $\{B, A, A\}$ but different from $\{A, B\}$. We define a description D containing properties $P_1..P_n$ as a *distinguishing description* if it can single out the referent r, i.e., $[\![P_1]\!] \cap ... \cap [\![P_n]\!] = \{r\}$. Note that $[\![P]\!]$ is a set of elements that share a property P (the *denotation* or *extension* of P).

Previous studies [4, 9, 5] motivate their understanding of over-specification on the basis of the Gricean Maxim of Quantity [10]: (1) the speakers should make their contributions as informative as required; (2) sufficient information should be included to allow an addressee to identify an intended referent. In contrast, other studies [11, 12, 13] defined *Minimal Descriptions* (MD) [14, 15] to be REs that mention as few attributes as possible. Sharing similar spirits to [11, 12, 13], we define MD as a referring expression $\mathcal{D}$ containing n properties $\{P_1, ..., P_n\}$, where there is no distinguishing description $\mathcal{D}' = \{P_1, ..., P_m\}$ such that $m < n$ (that is, $|\mathcal{D}'| < |\mathcal{D}|$).

Following the second principle of Gricean Maxim of Quantity, some researchers [4, 9, 5] defined an RE over-specified if there is more informative than necessary for successful communication. This clearly covers situations in which a RE includes non-required attributes while managing to identify the referent. However, it is less clear whether it also covers situations where a RE includes non-required attributes while failing to identify the referent. Furthermore, it is often unclear whether over-specification includes violations of Local Brevity (LB) [16], a situation where no attribute is superfluous yet it is possible to produce a shorter RE by replacing a set of existing attributes by single new attribute. We use the term *numerical over-specification*, calling a referring expression $\mathcal{D} = \{P_1, ..., P_n\}$ numerically over-specified if there is no $P \in \mathcal{D}$ such that $\bigcap_{P_j \in \mathcal{D}-\{P\}} [\![P_j]\!] = \{r\}$, but the number of attributes n is higher than that of the MD.

English requires REs have a head noun, therefore key REG algorithms such as the Incremental Algorithm (IA) [11] added a special provision to ensure that each generated RE contains a `TYPE`. However, some languages, including Dutch and Chinese among others, do not have such requirements [9]. We therefore define a separate type of over-specification called *nominal over-specification*. Specifically, a *nominal over-specification* is a description $\mathcal{D}$ in which any $P \in \mathcal{D}$ that causes $\bigcap_{P_j \in \mathcal{D}-\{P\}} [\![P_j]\!] = \{r\}$ is `TYPE`; in other words, only `TYPE` attributes are superfluous, no other attributes is superfluous. Some subjects regularly produce REs in which certain attribute (e.g., related to the difficult-to-express Attribute `ORIENTATION`) is expressed more than once. Therefore, we defined a type of over-specification called *duplicate-attribute over-specification* where a RE contains multiple attributes whose values are semantically equivalent. Finally, a description $\mathcal{D}$ is defined as a *real over-specification* if at least one of the $P \in \mathcal{D}$ is

*Email Addresses: {g.chen, c.j.vandeemter}@uu.nl, chenghua.lin@abdn.ac.uk

	total	minimal	real	nom	num	dup	ordinal	under	wrong
furniture	415	46	135	128	24	0	5	66	11
people	409	17	246	68	14	1	5	54	4

Table 1: The table indicates the number of descriptions for each category of under-/over-specification, in which **nom** represents nominal over-specification, **num** is numerical over-specification and **dup** is the duplicate-attribute over-specification. **wrong** means the number of descriptions that refers to wrong referent and **ordinal** is the number of descriptions that contains locative information/attributes.

such that $P \neq$ TYPE and $\bigcap_{P_j \in \mathcal{D}-\{P\}} [\![P_j]\!] = \{r\}$. Re-defining *under-specification* turned out to be easier: we propose $\bigcap_{P_j \in \mathcal{D}_u} [\![P_j]\!] = A$ and $\{r\}$ is a real subset of A.

With the new tagging schema, some simple hypotheses were formulated. For example, since the previous research suggests that reference in the furniture corpus was "easier" than the people corpus [17], with descriptions in the latter being more varied, and harder to model computationally than the ones in the former [7], we expected there were more over-specifications, and fewer minimal descriptions, in the people corpus than in the furniture corpus.

3 Primary Findings in MTuna

We applied our tagging schema to a portion of the M(andarin)Tuna corpus [1] that has one singular referent (as opposed to a set of two). We report here some preliminary results; further analysis is in progress.

Although there was no significant difference between the overall amounts of over-specification between the two corpora, the new tagging scheme allows us to see some striking differences (Table 1). For example, there was much less *real* over-specifications in the furniture corpus, and more minimal descriptions ($p < .01$). At the same time, the furniture corpus has more nominal over-specifications, which may be caused by two factors: 1) the furniture corpus is easier; 2) `TYPE` attribute in furniture corpus has multiple values, such as chair, table, and sofa. Compared to other languages, the proportion of under-specification in MTuna was 16.1%, which is much higher than the 5% reported in [9] on D(utch)Tuna, a figure that was in line with earlier studies of West European languages as spoken by adults.

Because our definitions of specification are mutually exclusive (e.g., a description cannot be really over-specified and nominally over-specified at the same time), a description classified as real over-specification may contain superfluous `TYPE` attributes or duplicate attributes. If we look into *real over-specification*, which is the most common over-specification type, we find that together with nominal over-specifications, 77.24% of the descriptions in MTuna have superfluous `TYPE` attributes, which is much lower than the 92.25% for E(nglish)Tuna reported in [15].

Finally, MTuna differentiates between the subject and object position of a referring expressions [1]. We expected that position would not affect over-specification, but we found that subject position descriptions have much more over-specifications and much less under-specifications. Definiteness – typically associated with the subject position – may play a role here, which based on the fact that the subject position and other preverbal positions in Mandarin favour definiteness [18], but this is a matter for further research.

4 Future Work

We hope to extend our work in the following ways: 1) We will apply the tagging schema to the whole MTuna corpus, including reference to sets for example, and carry out more detailed analysis. One interesting challenge here is aggregation. Consider the sentence "the red chair and the red table". Using *syntactic* aggregation we obtain "the red chair and table"; a further step of *semantic* aggregation would result in "the red furniture". We currently classify a lack of syntactic aggregation (though not a lack of semantic aggregation) as a type of over-specification. Furthermore, 2) we will apply the tagging schema to different corpora in order to further investigate to what extent (speakers of) different languages use reference differently. Our starting point will be four different Tuna-like corpora, for English [7], Mandarin [1], Dutch [9] and German [19]. Finally, 3) we will investigate using the tagging schema for a automatic evaluation metric with the help of a sequence tagger used for picking out the attributes used in each description. The new tagging scheme could be used for conducting improved semantic evaluation on classical REG algorithms with knowledge bases as input and on REG algorithms from the computer vision community.

References

[1] Kees van Deemter, Le Sun, Rint Sybesma, Xiao Li, Chen Bo, and Muyun Yang. Investigating the content and form of referring expressions in mandarin: introducing the mtuna corpus. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 213–217, 2017.

[2] Ehud Reiter and Robert Dale. *Building natural language generation systems*. Cambridge university press, 2000.

[3] Kees Van Deemter. *Computational models of referring: a study in cognitive science*. MIT Press, 2016.

[4] Paul E Engelhardt, Karl GD Bailey, and Fernanda Ferreira. Do speakers and listeners observe the gricean maxim of quantity? *Journal of Memory and Language*, 54(4):554–573, 2006.

[5] Paul E Engelhardt, Ş Barış Demiral, and Fernanda Ferreira. Over-specified referring expressions impair comprehension: An erp study. *Brain and cognition*, 77(2):304–314, 2011.

[6] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[7] Kees van Deemter, Albert Gatt, Ielka van der Sluis, and Richard Power. Generation of referring expressions: Assessing the incremental algorithm. *Cognitive science*, 36(5):799–836, 2012.

[8] Anja Belz and Albert Gatt. Intrinsic vs. extrinsic evaluation measures for referring expression generation. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pages 197–200. Association for Computational Linguistics, 2008.

[9] Ruud Koolen, Albert Gatt, Martijn Goudbeek, and Emiel Krahmer. Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, 43(13):3231–3250, 2011.

[10] H Paul Grice. Logic and conversation. *1975*, pages 41–58, 1975.

[11] Robert Dale and Ehud Reiter. Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive science*, 19(2):233–263, 1995.

[12] Robert Dale. Cooking up referring expressions. In *Proceedings of the 27th annual meeting on Association for Computational Linguistics*, pages 68–75. Association for Computational Linguistics, 1989.

[13] Robert Dale. *Generating referring expressions: Constructing descriptions in a domain of objects and processes*. Bradford Books, 1992.

[14] Albert Gatt, Ielka Van Der Sluis, and Kees Van Deemter. Evaluating algorithms for the generation of referring expressions using a balanced corpus. In *Proceedings of the Eleventh European Workshop on Natural Language Generation*, pages 49–56. Association for Computational Linguistics, 2007.

[15] Ielka Van der Sluis, Albert Gatt, and Kees Van Deemter. Evaluating algorithms for the generation of referring expressions: Going beyond toy domains. In *Proceedings of RANLP*, volume 7, 2007.

[16] Ehud Reiter. The computational complexity of avoiding conversational implicatures. In *Proceedings of the 28th annual meeting on Association for Computational Linguistics*, pages 97–104. Association for Computational Linguistics, 1990.

[17] Kees van Deemter, Ielka van der Sluis, and Albert Gatt. Building a semantically transparent corpus for the generation of referring expressions. In *Proceedings of the Fourth International Natural Language Generation Conference*, pages 130–132. Association for Computational Linguistics, 2006.

[18] Yuen Ren Chao. *A grammar of spoken Chinese*. Univ of California Press, 1965.

[19] David Howcroft, Jorrig Vogels, and Vera Demberg. G-tuna: a corpus of referring expressions in german, including duration information. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 149–153, 2017.