Lessons from Computational Modelling of Reference Production in Mandarin and English (Supplementary Materials)

Guanyi Chen and **Kees van Deemter**
Department of Information and Computing Sciences
Utrecht University
{g.chen, c.j.vandeemter}@uu.nl

A Annotating MTUNA

Following and adapting from **?**, the attributes and corresponding values we used in our annotation (annotating the MTUNA and re-annotating the ETUNA) are shown in Table 1 and 2 for furniture and people domain respectively. An example of the annotated RE sample in MTUNA is shown in Figure 1.

Attribute	Possible Values	Freq.
TYPE	*chair, sofa, desk, fan*	347
COLOUR	*blue, red, green, grey*	326
ORIENTATION	*front, back, left, right*	185
SIZE	*large, small*	141
X-DIMENSION	1, 2, 3, 4, 5	5
Y-DIMENSION	1, 2, 3	5
OTHER	-	10

Table 1: Attributes and their values for REs in furniture domain.

```
{ "sno": "Object",
  "subject_id": "2",
  "object": [
    { "attributes": [
        {"name": "COLOUR",
         "value": "dark"
        },
        {"name": "TYPE",
         "value": "table"
        }]
    }],
  "trial_id": "1",
  "utt": "灰桌子"
}
```

Figure 1: An example annotated data sample from MTUNA, for the RE 灰桌子 (*grey table*).

Attribute	Possible Values	Freq.
TYPE	*person*	278
AGE	*young, old*	74
ORIENTATION	*front, left, right*	6
hasBread	0, 1	169
breadColour	*dark, light*	95
hasHair	0, 1	116
hairColour	*dark, light*	100
hasGlasses	0, 1	165
hasShirt	0, 1	14
hasTie	0, 1	26
hasSuit	0, 1	60
X-DIMENSION	1, 2, 3, 4, 5	6
Y-DIMENSION	1, 2, 3	6
OTHER	-	151

Table 2: Attributes and their values for REs in people domain.

B Results for Annotating Over-/Under-specifications

Table 3 listed all trails we used for analysing the use of over- and under-specifications in this study, each of which we annotated how many alternative minimal descriptions and numerical over-specifications are there. The raw data of the MTUNA corpus (and trial ID) can be found at: homepages.abdn.ac.uk/k.vdeemter-/pages/mtuna-webpage/,

and those of ETUNA are available at: https://www.abdn.ac.uk/ncs/departments/computing-science/tuna-318.php.

MTUNA	ETUNA	#mini.	#num.
1	1	1	0
2	2	1	0
3	3	1	0
4	4	1	0
5	5	1	0
6	6	1	0
7	7	1	0
8	-	1	1
9	-	2	0
10	-	3	0
21	21	1	1
22	22	1	0
23	23	1	0
24	24	1	0
25	25	2	1
26	26	1	2
27	-	1	2
28	-	1	2
29	-	2	1
30	-	1	3

Table 3: The number of possible minimal descriptions (**mini.** and numerical over-specifications (**num.**) for each trial. The first two colomns are the trial IDs in MTUNA and ETUNA.