

Are Interlocutors as Sensitive to Over-informativeness as they are to Under-informativeness?

Catherine Davies (cnd24@cam.ac.uk)

Research Centre for English and Applied Linguistics, University of Cambridge, 9 West Road, Cambridge, CB3 9DP, UK

Napoleon Katsos (nk248@cam.ac.uk)

Research Centre for English and Applied Linguistics, University of Cambridge, 9 West Road, Cambridge, CB3 9DP, UK

Abstract

Research by Engelhardt, Bailey and Ferreira (2006) suggests that speakers avoid under-informativeness and listeners penalise it, but that neither of these behaviours apply to over-informativeness. We argue that what may appear as lack of sensitivity to over-informativeness is in fact preference for information that, even though it is not contributing to the unique identification of a referent, it is nevertheless addressee-oriented and pragmatically motivated because it minimizes the risk of ambiguity. In experiments 1A and 1B, where we remove the addressee-oriented motivation for over-informing that was implicit in Engelhardt et al's experiments, we document that speakers avoid and listeners penalise over-informativeness (though less than under-informativeness). This supports the Gricean prediction of some leniency towards over-informativeness relative to under-informativeness but nevertheless sensitivity to both. In a further experiment (2), increased visual salience of an attribute, typically understood as a non-Gricean, non-addressee oriented factor, was also found to contribute to increased rates of over-informativeness in production. We conclude that a careful study of the factors that affect the production and comprehension of referring expressions is needed to identify cases where over-informing is in fact the optimal choice for speakers, or for listeners, or for both. Computational modeling of the relevant constraints could lead to further testable predictions.

Keywords: Pragmatics; Informativeness; Referential Expressions; Speaker- and Hearer-oriented factors.

Research Background

Production Studies

Experimental approaches using the referential communication paradigm have documented that when speakers produce referring expressions, they frequently over-inform by describing entities in more detail than is necessary for unique identification (e.g. Deutsch & Pechmann, 1982; Maes, Arts & Noordman, 2004; Nadig & Sedivy, 2002; Pechmann, 1989).

In an early study in this domain, Deutsch & Pechmann (1982) found that 28% of adults' referring expressions in a task comprised redundant initial descriptions, i.e. contained the attribute(s) of the intended referent that were necessary for unique identification, plus at least one further attribute that was superfluous. Engelhardt, Bailey, & Ferreira (2006)

found that 30% of the referring expressions in their production study were over-informative, and Nadig & Sedivy (2002) found that adults overspecified by providing unnecessary adjectival modification in 50% of their elicited productions.

Evidently, such studies elicited a wide range of rates of overspecification, which may have been due to a number of factors, e.g. the use of rather different versions of the referential communication paradigm, and/or the interplay of several discourse-pragmatic contextual factors which impact on the level of detail provided in referring expressions. From the results to date, it is not possible to assert a conclusive rate of overspecification in production, but at the very least, we can say that humans commonly give more information than is minimally required. It is much less common for speakers to under-inform, i.e. to give less information than is required to identify an entity (Engelhardt et al; experiment 1).

The findings from both over-and under-informativeness accord with the Gricean account (1989) which concedes that a violation of the second quantity maxim (of *over-informativeness* 'Do not make your contribution more informative than is required') may not be as serious as a violation of the first quantity maxim (of *under-informativeness* 'Make your contribution as informative as is required'), and might be 'merely a waste of time'. Intuitively, saying too much is a relatively minor infraction – the message is conveyed, even though the hearer might have to filter out extraneous information, or could be misled about the point of the inclusion of extra information (e.g. generate a contrastive inference). In any case, referent identification can still go ahead, whereas if speakers under-inform, communication is severely threatened since no single referent has been identified.

Comprehension Studies

Recent studies from the comprehension literature have examined how listeners treat informativeness violations. It has been found that under-informative utterances (henceforth *U-utterances*) are predominantly rejected by adults at rates that range from 60% (Noveck, 2001) to over 85% (Papafragou & Musolino, 2003) depending on the task. Conversely, with regards to over-informative utterances (henceforth *O-utterances*), a single study up to now has found that listeners do not rate them as any worse than optimal expressions (Engelhardt, Bailey & Ferreira,

2006, experiments 2a and 2b); this lack of penalisation is even more striking given that participants did rate U-utterances worse than their optimal counterparts. The authors conclude that listeners are ‘only moderately Gricean’ in the sense that they are sensitive to under- but not to over-informativeness. This conclusion is hard to reconcile with the Gricean account which predicts some leniency but not full tolerance or insensitivity towards over-informativeness.

However, results of Engelhardt et al’s task (experiment 2, either version *a* or *b*) may not be replicable, since the utterances that the authors considered to be over-informative were nevertheless pragmatically felicitous. For example, in a critical condition, there was a display with an apple on a towel and another towel without an object on it. Participants heard the instruction ‘Put the apple on the towel on the other towel’ and were asked to rate how good the instruction was for bringing about a situation which was visually presented to the participants depicting the apple having been moved to the empty towel. The authors consider the first prepositional phrase ‘on the towel’ in the instruction to be over-informative (since there was only one apple in the display) and hence pragmatically infelicitous. But note that in contexts of more than one identical item (in this case the source and destination container), coupled with the temporary syntactic ambiguity inherent in the first PP (as reduced relative clause of the source or as the destination location), over-informative reference has a clarificatory addressee-oriented function: it serves the Gricean maxim of *manner*, which enjoins interlocutors to minimise the risk of ambiguity for the hearer. Additionally, in line with the principle of distant responsibility (Clark & Wilkes-Gibbs, 1986) speakers’ estimation of the hearer’s potential for misunderstanding may influence the amount of information provided, with uncertainty on the speaker’s side invoking overspecification. This may also go some way in explaining the ‘surprisingly common’ rate (30%) of over-descriptions observed in Engelhardt et al’s (1996) production experiment¹. Taking these factors into account and then given the likelihood that listeners give favourable ratings to speakers who over-inform in conditions when the risk of ambiguity is high, it is not clear whether the lack of a penalty for the O-utterances obtained by Engelhardt et al reflects a lack of sensitivity to over-informativeness (as the authors claim) or actual preference for over-informativeness when it minimises the risk of communicative failure. Experiment 1a was designed to address this issue.

¹ Additionally, there may be a syntactic priming effect at work: as hearers frequently encounter a PP of the form ‘on/in-the-X’ as goal component, they may come to expect and prefer such a construction for modifying the referent, in spite of its over-informativeness.

Experiment 1A: Comprehenders’ Ratings of Truly Over-informative Utterances

Pretesting for Default Descriptions

A pre-test was performed on 90 tokens of proposed stimulus materials. Items showing attributes from several dimensions (scalar and absolute; e.g. size, length, material, cleanliness, state of repair) were presented in isolation and described by native English speaking participants ($n=29$). Items which were referred to using an unmodified noun in more than 80% of elicited referring expressions in the pre-test were added to the stimulus lists for subsequent use. This was done to control for the possibility that a certain attribute is mentioned simply because it is particularly marked. Thus, the stimuli which appeared in our studies are highly likely to have an unmodified default description, unlike those items which showed some marked state, e.g. ‘a dirty cup’ which was modified as such in more than 80% of relevant referring expressions in the pretest and not used in subsequent experiments.

Method

To address the concerns arising from Engelhardt, Bailey & Ferreira’s (2006) methodology in their experiments 2a and 2b, we ran a ratings study using a referential communication paradigm which did not involve a movement-to-destination component. This was a 2 x 2 within-subjects design, where two factors were manipulated: presence of a *contrast set* in the visual array, and presence of *adjectival modification* in the heard utterance, creating 4 conditions: under- and over-informative, and two corresponding optimal conditions (illustrated in figs. 1-4). As there is no source/destination ambiguity, and there was no other entity that shared the adjectival property, adjectival modification in our over-informative condition (fig. 2) is straightforwardly redundant and does not serve a safe-guarding function, unlike in Engelhardt et al’s study (2006).

A static laptop display showed two characters, one of which had four items in her vicinity. Native English speaking university students ($n=21$) heard one interlocutor asking the other to *pass me the [referential expression]* and then rated how natural the instruction was using a magnitude estimation scale (Bard, Robertson & Sorace, 1996). Magnitude estimation scale ratings are particularly recommended for investigating participants’ sensitivity to moderately ‘mild’ violations (i.e. violations that are not overriding a core grammatical rule).

After giving a rating on the scale for each item, participants were also asked to make a categorical judgment as to whether the instruction they heard was exactly what they would have said themselves in the same context. Giving a negative judgment in this binary measure (e.g. ‘I would *not* have said it this way myself’) can be interpreted as evidence that a violation is not just warranting penalisation on the magnitude estimation scale, but it is also so severe that it warrants the downright rejection of the utterance. As such, the results from the binary measure can

reveal qualitative difference between kinds of violations (i.e. whether the violation warrants a downright rejection or not), while the results from the magnitude scale can reveal fine quantitative differences (in terms of degree of penalisation). Furthermore, when participants gave a negative judgment, they were asked to state exactly how they would have asked for the item themselves, which provides evidence about which exact aspect of the critical utterance they objected to.

There were 40 critical items, 10 in each condition, plus 20 syntactically infelicitous fillers (see appendix 1). Every target item appeared in each of the four conditions between participants, i.e. a Latin square design was used to counterbalance any item effects.

Experiment 1: Example Arrays

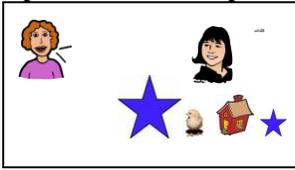


Fig 1. Under-informative:
'Pass me the star'

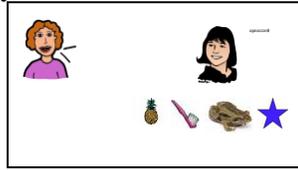


Fig 2. Over-informative:
'Pass me the small star'



Fig 3. Optimal-1:
'Pass me the star'

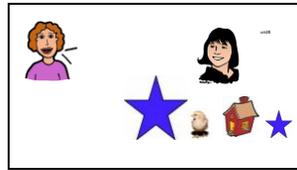


Fig 4. Optimal-2:
'Pass me the small star'

Results

The normalized data (by means of *z*-scores) from the magnitude estimation ratings are presented in tables 1 and 2. With regards to the magnitude ratings, a 2 x 2 repeated measures ANOVA for the four critical conditions reveals no main effect of contrast ($F(1, 20) = 2.46$, n.s.) and no main effect of modification ($F(1, 20) = .29$, n.s.). Instead, there was a significant interaction between contrast set and modification ($F(1, 20) = 19.46$, $p < .001$). Further pair-wise comparisons by means of *t*-tests show that, contra Engelhardt et al (2006), U-utterances (contrast & no adjective) and O-utterances (no contrast & adjective) were both rated lower than their corresponding optimal utterances, i.e. Optimal-1 (no contrast & no adjective) and Optimal-2 (contrast & adjective) respectively (both $p < .005$). They were also rated lower than the corresponding visual display (Optimal-2 for U-utterances and Optimal-1 for O-utterances, both $p < .005$). Even though O-utterances were rated numerically higher than U-utterances, this difference did not reach statistical significance ($t(20) = -1.16$, n.s.).

With regards to the binary judgments, a Friedman's ANOVA for non-parametric data revealed a significant difference between conditions ($\chi^2(3) = 36.69$, $p < .001$).

Further pair-wise comparisons by means of Wilcoxon's signed rank tests reveal that U- and O-utterances are rated worse than their optimal counterparts (i.e. Optimal-1 and Optimal-2, respectively; both $p < .001$), corroborating the ME results. Note that the two optimal conditions were significantly different from each other with Optimal-2 receiving more rejections ($p < .005$). An analysis of the rephrasings that followed the rejections for Optimal-2 reveals that participants did not object to the presence of adjectival modification per se, but rather to the specific lexical choice (e.g. rejecting and rephrasing the instruction to *pass me the old-fashioned phone* in favour of *pass me the rotary phone*). As such, this difference does not relate to informativeness per se, although a difference that is related to informativeness is that that U-utterances received more rejections and rephrasings than O-utterances ($p < .05$). This is despite the fact that some O-utterances were rejected not on the basis that they violated informativeness but on the basis of the specific lexical choice (analyses of the data excluding these cases are under way).

Table 1: Magnitude estimation scores (z-transformed) mean

Utterance type	Mean (SE)
Under-informative	0.24 (0.09)
Over-informative	0.36 (0.06)
Optimal-1 (no contrast set)	0.77 (0.05)
Optimal-2 (contrast set)	0.71 (0.05)
Ungrammatical fillers	-1.04 (0.03)

Table 2: Mean rejections and rephrasings*

Utterance type	Mean (SE)
Under-informative	7.38 (0.78)
Over-informative	5.71 (0.73)
Optimal-1 (no contrast set)	0.62 (0.48)
Optimal-2 (contrast set)	1.71 (0.53)
Ungrammatical fillers	20

*maximum = 10 for critical conditions, 20 for fillers

Discussion

Overall, the results from Experiment 1A magnitude estimation scale and binary judgments show that in contrast to Engelhardt, Bailey & Ferreira (2006), both over- and under-informative utterances are penalised more than their felicitous counterparts. Further, the data from the binary judgment measure also reveal that listeners are less severe towards overspecified utterances than under-specified utterances (this pattern was also obtained numerically in the magnitude estimation scale). This set of results supports the Gricean account of some leniency towards over-informativeness compared to under-informativeness, but a clear penalisation for both in this task where there was no pragmatic advantage for the additional information.

Experiment 1B: Production of over-informative utterances

In this experiment we investigated whether the rates of over-informativeness in *production* would fall when participants were tested with a design where over-informativeness is not pragmatically motivated.

Method

Experiment 1B used the same 40 visual stimulus arrays and the same experimental design (minus the fillers) as Experiment 1A, with an arrow added to highlight the target item. There were twenty 1-referent arrays and twenty 2-referent arrays. Twenty-four English-speaking university students took part in this study. Participants were required to state the best way of asking for the highlighted item.

Results

Quantitative Analysis The proportions of under-, optimal, and over-informative descriptions for 1-referent and 2-referent arrays are presented in table 3. With regards to the 1-referent displays, where only two types of output are possible (optimal and over-informative), a Wilcoxon's signed rank test for non-parametric data revealed a significant difference between optimal and overspecified descriptions ($p < 0.001$). With regards to the 2-referent displays, where under-, optimal-, and over-informative expressions were elicited, a Friedman's ANOVA for non-parametric data revealed a significant difference between conditions ($\chi^2(2) = 28.74, p < 0.001$). Further pair-wise comparisons by means of Wilcoxon's signed rank tests revealed significant differences between under- and optimally-informative utterances ($p < 0.05$); optimal and over-informative utterances ($p < 0.001$); and under- and over-informative utterances ($p < 0.005$). When comparing the 1- and 2-referent conditions, the rates of over-informativeness remain stable ($Z = -1.01, n.s.$).

Table 3: Under-, optimal-, and over-informative utterances in 1- and 2-referent displays (mean rates shown as %)

	Under-	Optimal-	Over-
1-ref	0.00	92.50	7.50
2-ref	26.67	67.92	5.42

Thus, when encountering displays without a contrast set, speakers used a minimally contrastive referring strategy, i.e. were overwhelmingly optimal. When referring to targets which formed part of a contrast set, speakers were also largely optimal in their referring strategies, though relatively high rates of under-informativeness were documented. Crucially, the rates of over-informativeness are the same in

the 1- and 2-referent displays, and in both cases much lower than the rates of 30% reported by Engelhardt et al.

Qualitative Analysis The 56 tokens of overspecified reference that were elicited from 960 referring expressions were also analysed qualitatively. Of these overspecified tokens, the attribute most frequently provided redundantly was colour (48%). Modified nouns such as *handheld phone* and *desktop computer* were classified as 'type' modified with 14% of overspecified referring expressions. The contents of items such as *glass of water* was provided in 13% of overspecified referring expressions. Other attributes provided in less than 5% of overspecified referring expressions were (in descending order) size, age, pattern, shape, state (e.g. *empty*), material, speaker attitude (e.g. *nice*) and location. Our results regarding the colour over-modification accord with previous work which suggests that colour attributes are often used gratuitously rather than discriminatively (Mangold & Pobel, 1988), underscored by Sedivy (2002) who found that unpredictable colour modifiers are frequently encoded in default descriptions².

Discussion of Experiments 1A and 1B:

In line with the rating study in Experiment 1A where participants did penalise over-informativeness, speakers over-informed in Experiment 1B at a mean of 6.5%, which is far lower than the rates observed by Engelhardt et al. Overall, after removing the addressee-oriented pragmatic imperative of avoiding ambiguity which was likely to have motivated over-informativeness in Engelhardt et al's study, we find that both speakers and listeners are sensitive to over-informativeness; that is to say interlocutors are fully Gricean in two senses: they are sensitive to both under- and over-informativeness and also both in production and in listener judgments.

In light of our findings, we revisited other studies documenting high incidence of over-informativeness in production. Recall our claim that avoiding source/destination ambiguity is an addressee-oriented factor that licenses over-informativeness (as accounted by the Gricean maxim of manner, and/or Clark & Wilkes-Gibbs' principle of distant responsibility). In addition to this, there is evidence in the literature that other addressee-oriented factors such as politeness (Küntay & Bahtiyar, 2008) and lexical entrainment (Brennan & Clark, 1996) can additionally increase overspecified reference.

Moreover, a non-addressee oriented factor, the salience of the double instantiation of an item makes that item simply more likely to be mentioned. This mechanism could be at work in Nadig & Sedivy's (2002) materials. In their study, the speaker could see a contrast set comprising e.g. a tall and a short glass, although one of the pair was hidden from the hearer's view. The experimental design was such that in

² Relatedly, in comprehension, colour modification does not trigger contrastive inference, at least for objects with predictable colour modification (Sedivy 2003).

this study, the ‘secret’ contrast-mate was in the speaker’s direct line of sight and was thus perceptually salient and potentially difficult to suppress. It is feasible then that the visual salience (together with its ‘special’ feature of being the only obscured item) contributed to speakers including redundant modification in relation to the target object in 50% of their utterances.

The importance of visual salience for the speaker’s perspective is also directly supported by Carbary & Tanenhaus (2008), who found that the presence of a shared attribute in an array increases the incidence of overspecified references to targets from an 11% baseline rate to 25% in contexts which contained another item with a common attribute, even though it was not necessary to mention this attribute for unique identification of the target. Further, there is ample data from the production literature to show that speakers indeed provide more modification than is needed for unique identification when more than one entity of the same kind appears in an array (Mangold & Pobel, 1988).

Furthermore, in other studies where high rates of over-informativeness have been elicited, there are systematic preferences in the use of redundant information. In Pechmann’s 1989 study, 98% of overspecified utterances mentioned colour. As discussed above, colour has been repeatedly found in default descriptions so can perhaps be classified as a special case of overspecification.

This very brief outlook adds weight to the argument that people are in fact Gricean in that in many studies where information was provided over and above that required for minimal contrastiveness this may have been so because of a communicative addressee-oriented reason (maxim of manner, of distant responsibility, lexical entrainment, politeness). This is not to exclude other non-Gricean factors contributing to rates of over-informing such as visual salience or nature of attribute (colour or not).

Experiment 2: Production with increased salience

In a second set of studies (in progress) we aim to empirically document that visual salience is implicated in rates of overspecified referring expressions.

Background

If one of the item-types in an array appears more frequently than other item-types, it is reasonable to assume that its salience is increased. Further, if one of the items is clearly different to its array-mates, for example by appearing in a container, its salience may also increase. In Engelhardt, Bailey & Ferreira’s (2006) study, target items were indeed ‘special’ on both of these counts, and so arguably were made more salient than the other items. We argue that such configurations may play a major causal factor in overspecification from the speaker’s perspective, as has been demonstrated by Carbary & Tanenhaus (2008) i.a. It is

not unfeasible to further predict that in these conditions, hearers also expect superfluity of detail in referring expressions.

Materials

To test that visual salience of an attribute is implicated in rates of over-informativeness in production, we ran a study using arrays which increased the salience of a particular attribute possessed by a target referent. There were two conditions from Experiment 1B: Condition (1) *contrasting target* (fig. 5) where adjectival modification is required for unique identification; and condition (2) *non-contrasting target* (fig. 6) where adjectival modification is redundant and visual salience of the attribute is at baseline levels. Increasing the salience of a particular attribute was achieved by including two new conditions: condition (3) the *contrasting comparison* condition (see fig. 7, which contained a non-target contrast set differing on the same dimension as an attribute possessed by the target, e.g. an open box and a closed box next to an open bag target and an unrelated distracter item; and also condition (4) a *noncontrasting comparison* condition (fig. 8) which contained one item of a different kind which shared an attribute of the target, e.g. an open box next to an open bag target, and two unrelated distracter items (mirroring Carbary & Tanenhaus’s 2008 materials).

We predict that the effect of increasing the visual salience of an attribute by including these two new conditions (3) and (4) could be reflected in higher rates of over-informing for the two increased-salience conditions relative to baseline condition (2), the non-contrastive target, where the visual salience of the attribute is low.

Of the 16 trials in each condition with a contrast, 8 contained a scalar comparison contrast (e.g. big vs. small) and 8 an absolute comparison contrast (e.g. square vs. round). The two new comparison conditions were included alongside the original 2-referent display (16 items) and the original 1-referent display (16 items).

Experiment 2: Example Arrays

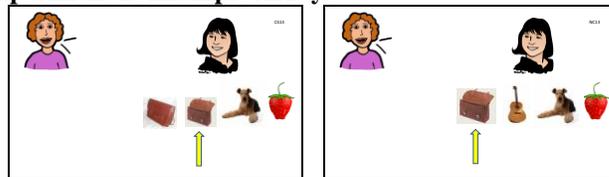


Fig. 5. Condition 1:
Contrasting target

Fig. 6. Condition 2:
No contrast

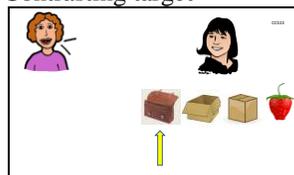


Fig. 7. Condition 3.
Contrasting comparison

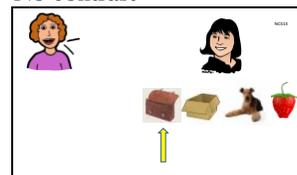


Fig. 8. Condition 4:
Noncontrasting comparison

Results

The rates of utterance-types for Experiment 2 are presented in Table 4. A significant difference was obtained between rates of over-informativeness in the four conditions, when the rates of over-informativeness for the scalar attributes and the absolute attributes were analysed together (Friedman test for non-normally distributed data, $\chi^2(3)=12.07$, $p<.01$). Moreover, pairwise comparisons by means of Wilcoxon sign ranked test reveal that the rates of over-informativeness were lower for condition 1 (where one modifier is already required) than for other three conditions ($p<.01$). However, the other three conditions did not differ from each other (all t -values < -1.1). Thus, inclusion of an item or contrast set sharing an attribute with the target in condition 3 and 4 did not significantly increase the production of modifiers relative to the baseline condition 2.

Table 4: Experiment 2;
mean % of utterance-types across conditions (all items).

Condition ³	Scalar & Absolute Items		
	Under-	Optimal	Over-
1	10.12	88.69	1.19
2	0.00	92.26	7.74
3	0.00	90.18	9.82
4	0.00	88.69	11.31

However, this null result may have been affected by the scalar and absolute nature of contrasts in the experimental trials. Scalar attributes may not have been perceptually apparent when target items did not have a contrast mate, rendering half of the items in the noncontrasting comparison condition perceptually identical to the no contrast condition. Carbery & Tanenhaus's (2008) arrays only included absolute contrasts, which may account for the discrepancy in results: they found a rise from 11% to 25% when a comparison mate was present; we found a negligible rise from 8% to 11%. In a second analysis, rates of over-informativeness were computed for just the absolute adjectives in our materials and are presented in table 5.

Table 5: Experiment 2; mean % of
utterance-types across conditions (absolute items only).

Condition	Absolute Items only		
	Under-	Optimal	Over-
1	16.67	80.95	2.38
2	0.00	85.71	14.29
3	0.00	83.93	16.07
4	0.00	79.76	20.24

³ Key to conditions: 1) contrasting target - modifier required; 2) no contrast - modifier *not* required; 3) contrasting comparison - modifier *not* required; 4) noncontrasting comparison - modifier *not* required

Again, a significant difference was obtained between rates of over-informativeness in the four conditions (Friedman test for non-normally distributed data, $\chi^2(3)=10.55$, $p<.05$). Moreover, pair-wise comparisons by means of Wilcoxon sign ranked test reveal that the rates of over-informativeness were lower for condition 1 (where one modifier is already required) than for other three conditions ($p<.01$). However, the other three conditions did not differ from each other (all t -values < -1.2). Thus, while again there were no significant differences between conditions 2, 3 and 4, when looking at absolute attributes only, there was nevertheless a numerical difference towards increased over-informativeness, climbing from 14% in the baseline condition (2), to 20% in the noncontrasting comparison condition (4). This trend, coupled with Carbery & Tanenhaus's (2008) findings tentatively suggest that there is something in the visual context which may encourage speakers to mention a non-discriminating adjective.

Discussion

These results tentatively suggest that speakers over-modify their referring expressions to a non-negligible extent when targets with absolute attributes are made saliently spotty, wooden etc by the presence of the same attribute elsewhere in the array. This finding may help account for the relatively high incidence of overspecification found in Engelhardt, Bailey & Ferreira's (2006) production study as being due to increased salience of specific containers in the paradigm (due to multiple instantiations in arrays, as well as the double-nature of the target-and-container items), in addition to the pragmatic addressee-oriented motivation that we already discussed. Ongoing work will aim to consolidate the findings of Experiment 2 and test the prediction that increasing the salience of an attribute will lead to more favourable ratings for overspecified referring expressions in the presence of a comparison set in comprehension studies.

General Discussion

From Experiment 1A and 1B we conclude that the high rates (30%) of O-utterances in production found by Engelhardt, Bailey & Ferreira (2006) and the lack of penalisation of O-utterances in their listener rating study can be (at least partially) attributed to addressee-oriented pragmatic motivation: participants were observing the maxim of manner and/or of distant responsibility in avoiding potential ambiguity. When the pragmatic motivation for O-utterances was removed, O-utterances were severely penalised (Experiment 1A), though at lower rates than under-informative utterances. Moreover, the rates of over-informativeness in production dropped to around 6% (Experiment 1B). This documents that speakers and listeners are fully Gricean in that they respectively avoid and penalise both under- and over-informativeness.

Moreover, in Experiment 2 we showed that when an attribute's visual salience (which is traditionally considered a speaker-oriented non-Gricean factor) was increased, the

rates of over-informativeness in production rose again to 20% from a baseline of 14%. Such results inform our work in that they highlight a principled relationship between aspects of the linguistic- or extra-linguistic context, and the incidence of over-informativeness. From this, we approach a clearer distinction between overspecification and redundancy in that certain types of overspecified expressions are not truly redundant but rather they obey addressee- and speaker-oriented constraints. This is in accordance with a review of the psycholinguistic literature, which suggests several reasons why interlocutors may apparently be acting over-informatively but do not in fact violate pragmatic constraints.

Future work is planned along two dimensions: First, we need to explore whether pragmatic (addressee-oriented) and non-pragmatic speaker-oriented factors can be disentangled. In the present work we differentiated between addressee-oriented factors, such as avoiding the source/destination ambiguity, and speaker-oriented factors such as attribute visual salience. Note however, that with regards to the former, it is not fully possible to isolate the effect of the addressee-oriented factors, neither in our current design, nor in Engelhardt et al's studies, since the need for over-informing to avoid potential ambiguity is introduced in exactly the same conditions where attribute salience is also increased. Conversely, while visual salience is indeed a factor that affects the speaker's choice, we do not think it is unlikely that listeners also expect that visually salient attributes will be mentioned and might penalise lack of such mention. We are now in the process of testing this prediction by running the corresponding listener version of Experiment 2, where we predict that increased visual salience does not only affect speaker production but also leads to less severe penalisation from listeners. Moreover, we are currently exploring experimental paradigms where the relative contribution of speaker- and addressee-oriented factors can be measured independently.

Second, Engelhardt et al (2006) also found an *online* cost (as fixation delays) for the comprehension of O-utterances. Since, in their understanding, they had already established that speakers and listeners are not sensitive to over-informativeness, they attributed this online cost to syntactic disambiguation processes. To further complement the ratings and production studies reported in this paper, in future work we aim to address the online comprehension of under-, over-, and optimally-informative utterances, using the visual world paradigm and the stimulus materials from Experiment 1A, which remove the motivation for O-utterances. On the basis of our rating and production results, we expect to find a temporal penalty in referent resolution for truly over-informative utterances, this time straightforwardly attributable to violation of Grice's second quantity maxim.

Finally, research in experimental pragmatics which tracks the incidence and nature of apparent violations of optimal informativeness can inform and be informed by a variety of neighboring disciplines and in particular computational

work on the creation of algorithms (e.g. Viethen & Dale, 2006) for the generation of referring expressions. Such work, which aims to replicate the characteristic informativeness found in human speech processes, looks to pragmatic concepts for guidance and reinterprets pragmatic theory appropriate to its own aims (Dale & Reiter, 1995; 1996, Frederking, 1995). Our present research helps build of a profile of overspecification in human communication, and although much of the descriptive work on over- and under-informativeness is well under way, there remains much work to be done in identifying and exploring the specific constraints on levels of informativeness above (and below) that required for unique identification of a referent. It is hoped that such a research programme can make valuable contributions to computational linguistics. Likewise, computational modeling of the weighting of the constraints which combine to produce specific levels of informativeness would lead to psycholinguistically testable predictions.

Acknowledgments

The authors would like to thank John Williams and Henriette Hendriks for their guidance on this project. Catherine Davies was supported by a Domestic Research Scholarship from the University of Cambridge.

References

- Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, 72(1), 32–68.
- Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, 46, 391–418.
- Carbary, K. & Tanenhaus, M. K. (2008) Syntactic priming and message encoding in unscripted dialogue. Presented at *AMLaP*. Cambridge, Sept. 2008.
- Clark, H.H. & Wilkes-Gibbs, D. (1986) Referring as a collaborative process. *Cognition*, 22, 1-39
- Dale, R. & Reiter, E. (1995) Computational Interpretations of the Gricean Maxims in the generation of referring expressions. *Cognitive Science*, 19, 233-263.
- Dale, R. & Reiter, E. (1996). The Role of the Gricean Maxims in the Generation of Referring Expressions. Pages 16-20 in *Working Notes for the AAAI Spring Symposium on Computational Implicature: Computational Approaches to Interpreting and Generating Conversational Implicature*, Stanford, March 25-27
- Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, 54(4), 554-573.
- Frederking, R.E. (1996) Grice's Maxims: "Do the Right Thing". Presented at the *Computational Implicature*

- workshop at the AAAI-96 Spring Symposium Series, Stanford.*
- Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press
- Küntay, A.C. & Bahtiyar, S. (2008). Integration of Communicative Partner's Visual Perspective in Patterns of Referential Requests. *Journal of Child Language*.
- Maes, A., Arts, A., & Noordman, L. (2004). Reference management in instructive discourse. *Discourse Processes*, 37(2), 117-144.
- Mangold, R. & Pobel, R. (1988). Informativeness and instrumentality in referential communication. *Journal of Language and Social Psychology*, 7, 181-191.
- Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, 13(4), 329-336.
- Noveck, I. (2001). When children are more logical than adults. *Cognition*, 86, 253–282.
- Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics/pragmatics interface. *Cognition*, 86, 253–282.
- Pechmann, T. (1989). Incremental Speech Production and Referential Overspecification. *Linguistics*, 27(1), 89-110.
- Sedivy, J. C. (2002). Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory and Language*, 46, 2, 341-370
- Sedivy, J.C. (2003) Pragmatic Versus Form-Based Accounts of Referential Contrast: Evidence for Effects of Informativity Expectations. *Journal of Psycholinguistic Research*, 32 (1): 3-23.
- Viethen, J. & Dale, R. (2006) Algorithms for generating referring expressions: Do they do what people do?, *Proceedings of the Fourth International Natural Language Generation Conference* (Sydney, Australia), July 2006, pp. 63–70.

Appendix 1

20 syntactically infelicitous filler items were randomly distributed throughout the critical trials, and comprised the following forms:

- i. Clefted, e.g. *The kettle, pass me*
- ii. Adjective-verb inversion, e.g. *Pass me the jug tall*
- iii. Scrambled, e.g. *Me car pass big the*
- iv. Article deletion, e.g. *Pass me pineapple*