Critical update on the manuscript "Distributional Effects of Frequency in Pronoun Processing"

Sol Lago, Wing-Yee Chow, Ewan Dunbar and Colin Phillips
Department of Linguistics, University of Maryland, 1401 Marie Mount Hall, College Park, MD 20742, USA
April 8, 2013

This is a critical update to an unpublished manuscript on a study that we reported at the 24th annual CUNY Human Sentence Processing Conference titled "Distributional Effects of Frequency in Pronoun Processing". In the manuscript we reported that reading times for pronouns are impacted by the frequency of their antecedents: pronouns with infrequent antecedents were read more slowly, suggesting that pronoun processing involves reaccess to the lexical representation of the antecedent. We showed this in two experiments, using both common noun and proper name antecedents, and we also showed that the parallel frequency effect extended to the distribution of reading times in ex- Gaussian analyses. This result was surprising in light of earlier findings by van Gompel and Majid [van Gompel, R.P.G. & Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition*, 90, 255-264], which had argued that pronoun processing showed an inverse frequency effect (a *prominence effect* in their terms). However, as we tried to further understand the discrepancy between our findings and those of van Gompel and Majid (vG&M), we discovered an error in the script that we had used to process the eyetracking data. This error specifically affected reading times beyond the first line of text in multi-line materials, especially with the configuration that we had used in our studies, following vG&M. After correcting this error, the reanalysis of our two previous experiments and a new experiment that consisted of a full replication of vG&M showed that the problem had systematically affected measurements in critical regions, and that the key claim of our earlier manuscript was incorrect: **we find no clear evidence for antecedent frequency effects on pronoun processing**. We sincerely apologize for any inconvenience that our error has caused. Meanwhile, since only the reading times in the second line of our two-line experimental items were affected, the results we reported for the antecedent region (which was always located in the first line) are unchanged. Here we summarize briefly the actual findings from our three experiments. In brief:

- Frequency effects at the antecedent nouns were consistent across all three experiments.
- Effects of antecedent frequency following the pronoun were inconsistent across the three experiments, sometimes showing no effect, sometimes a frequency effect, and sometimes a prominence/inverse-frequency effect. We do not consistently find the frequency effect that we previously reported. Nor do we find consistent evidence for the prominence effect that vG&M reported. This includes the new Experiment 3, a direct replication of their study.
- All effects of frequency on fixation distributions at the antecedent are the same as we originally reported.

In what follows we describe the specific results and statistical analyses in more detail.

Results Summary

Previous results (van Gompel and Majid, 2004 Cognition)

In a previous paper in 2004, van Gompel and Majid reported an inverse effect of antecedent lexical frequency in pronoun resolution: pronouns referring to infrequent antecedents had shorter reading times than pronouns referring to frequent antecedents, and this effect was found in first fixation and first pass times at the post-anaphor region. Following their terminology we refer to this effect as a *prominence effect*.

In contrast, in two eyetracking experiments we reported a rather different pattern. We observed a direct frequency effect: pronouns referring to infrequent antecedents elicited longer reading times than pronouns referring to frequent antecedents. This effect was observed in the post-anaphor region in first fixation, first-pass and total times. In what follows, we will refer to this effect as an *antecedent frequency effect*. We observed this pattern in an experiment that used an extended dataset including vG&M's original materials (*Experiment 1*), and in a follow-up experiment which manipulated antecedent type, and contained both common noun and proper name antecedents (*Experiment 2*). We also performed distributional analysis of first-fixation times in Experiment 1, and reported that the effect of antecedent lexical frequency was only observed on the tail (τ) of the fixation distributions, and that this pattern was strikingly similar in the antecedent and post-anaphor regions.

Re-analyzed results from previous manuscript

In addition to re-analyzing our two previous experiments, we conducted a full replication of the experiment by vG&M, using the same items and fillers. Below is the summary of the reanalyses and the full replication.

In the re-analyzed results from *Experiment 1* we found an effect of lexical frequency at the antecedent region (significant in first-fixation, first-pass and total times), consistently with our manuscript and the results of vG&M. However, there were no effects of lexical frequency in either the anaphor or the post-anaphor region. Therefore, we did not observe either an antecedent frequency effect or the prominence effect that was reported by vG&M. Table 1 shows the mean reading times by condition, and Table 2 shows the statistics calculated using linear mixed effects modeling. To be consistent with our manuscript, we split the experimental items into those taken from vG&M, the new items we created, and the results from the first 10 items in each condition, to rule out effects of having a large number of trials per condition (n=40). Lastly, we also re-did the distributional analyses of first-fixation times at the antecedent and post-anaphor region. Since only the reading times in the second line of our 2-line experimental items were affected by the script error, the distributional results we reported for the antecedent region (which was always located in the first line) are unchanged: lexical frequency only affects the tail of the first-fixation distributions. The results in the post-anaphor region, however, were different from our previous report – we did not observe an effect of antecedent frequency either in the non-parametric vincentiles (Figure 1) or the parameters resulting from fitting the ex-Gaussian distribution (Table 3).

In *Experiment 2* we used a 2 × 2 design with antecedent type (common noun vs. proper name) and antecedent frequency (high vs. low) as within-subjects factors. Each frequency pair of common nouns used in Experiment 1 was paired with a pair of proper names. After re-analyzing our results, we found a lexical frequency effect at the antecedent, both in the common noun and proper name conditions, and it was again significant in the three measures of interest. Unlike our previous report, we now found a prominence effect consistent with vG&M in the post-anaphor region: the region was read more quickly when the antecedent was an infrequent word. This effect was only significant in first-fixation times and was obtained both in the combined analysis of the proper noun and common name conditions and also when they were analyzed separately. To be consistent with our manuscript, we also analyzed reading times in the pronoun+2 region, and this revealed a marginally significant antecedent frequency effect consistent with our previous results. This effect, however, was only present in first-fixation times in the proper name condition. Therefore, there was a contrast between the effects we observed in the immediate post-anaphor region (where we found a prominence effect) and the pronoun+2 region (where we found a frequency effect). The means for Experiment 2 are shown in Table 4, and the statistical analysis in Table 5.

Finally, we conducted a full replication of vG&M's original experiment (n=32). (We are grateful to Roger van Gompel for sharing the materials from his study with us.) Following vG&M's study, we used a 2 × 2 design with antecedent frequency (high vs. low) and anaphor type (repeated NP vs. pronoun) as within-subjects factors. At the antecedent region, infrequent words were read more slowly than frequent words, and this was significant in the three measures of interest. Also, the results of this experiment replicated vG&M's previous findings in the repeated noun conditions, where we found clear frequency effects in the anaphor and post-anaphor regions. This was significant in first-pass and total times (although the effect was only marginal in first-pass in the post-anaphor region); consistent with vG&M's previous findings, the repeated name frequency effect was smaller than the antecedent frequency effect. In the pronoun conditions in the post-anaphor region, however, there were no effects of frequency in any of the three measures, contrary to vG&M's main finding. Therefore, the results of the third experiment showed neither a prominence effect nor an antecedent frequency effect. The only effect found in the pronoun conditions was obtained at the pronoun region itself, and showed the direction of an antecedent frequency effect: pronouns with infrequent antecedents had longer first-fixation times than pronouns with frequent antecedents. The effect, however, was only marginally significant. The means for Experiment 3 are shown in Table 6, and the statistical analysis in Table 7.

To summarize: (1) the reanalyzed results do not change the previously described effects of noun antecedent frequency while reading sentences, and the new results from the full replication are consistent qualitatively and quantitatively. (2) The reanalyzed and new results on the effect of antecedent frequency on pronoun reading times show little consistency with either our previous reports or with vG&M's, and do not clearly point to consistent effects of antecedent frequency on reading times during coreference resolution: only one of the three experiments shows a prominence-type effect similar to vG&M, while a different experiment shows a marginal antecedent frequency effect similar to our previous report. (3) Finally, we found that repeated NPs show attenuated frequency effects in their reading times, consistent with vG&M.

Updated tables and figures

		First-fixation duration		*First-pass duration*		*Total time*	
	Region	Low	High	Low	High	Low	High
All items	Antecedent	244	229	329	270	542	413
	Pronoun	221	229	247	259	335	338
	Pronoun+1	217	222	252	262	358	362
VG&M items	Antecedent	249	224	339	256	583	405
	Pronoun	217	223	239	254	325	337
	Pronoun+1	216	224	244	257	359	362
New items	Antecedent	241	233	323	280	512	420
	Pronoun	223	233	252	264	342	340
	Pronoun+1	218	221	259	267	358	362
First 10 items	Antecedent	247	228	356	279	631	477
	Pronoun	222	232	254	274	356	376
	Pronoun+1	217	217	258	259	376	409

Table 1. Averages in first fixation duration, first-pass duration, and total time across critical regions in **Experiment 1.**

		First-fixation duration			*First-pass duration*			*Total time*		
	Region	β (ms)	CI (ms)	Pr(β<0)	β (ms)	CI (ms)	Pr(β<0)	β (ms)	CI (ms)	Pr(β<0)
All items	Antecedent	16	[9,23]	**<.01**	61	[46,76]	**<.01**	137	[110,163]	**<.01**
	Pronoun	-9	[-18,0]	0.97	-14	[-28,0]	0.97	-3	[-24,18]	0.61
	Pronoun+1	-6	[-12,1]	0.96	-12	[-26,1]	0.96	-4	[-25,18]	0.65
vG&M items	Antecedent	26	[14,37]	**<.01**	81	[59,105]	**<.01**	179	[137,220]	**<.01**
	Pronoun	-7	[-20,6]	0.86	-17	[-37,5]	0.94	-10	[-41,23]	0.73
	Pronoun+1	-8	[-18,2]	0.94	-12	[-29,4]	0.93	-1	[-34,31]	0.53
New items	Antecedent	9	[-1,18]	**<.05**	48	[27,69]	**<.01**	105	[69,138]	**<.01**
	Pronoun	-10	[-22,4]	0.92	-12	[-31,7]	0.89	3	[-25,30]	0.41
	Pronoun+1	-4	[-13,4]	0.83	-12	[-32,7]	0.89	-7	[-36,22]	0.69
First 10 items	Antecedent	18	[3,33]	**<.01**	79	[47,113]	**<.01**	170	[107,220]	**<.01**
	Pronoun	-11	[-30,6]	0.91	-24	[-52,5]	0.95	-19	[-66,28]	0.80
	Pronoun+1	1	[-12,12]	0.47	3	[-26,29]	0.47	-29	[-79,16]	0.90

Table 2. Statistics for first fixation, first-pass and total time measures in **Experiment 1.** Reliable effects are in bold font, marginal effects with .05< *p* <.07 are in bold and italic font.

Region	*Variable*	Mean	μ	σ	τ
Antecedent	Low	244	171	28	76
	High	229	180	35	50
	Frequency effect	***15***	*-9*	*-7*	***26***
Pronoun	Low	221	162	27	57
	High	229	162	25	58
	Frequency effect	*-8*	*0*	*2*	*-1*
Pronoun+1	Low	217	166	32	50
	High	222	175	35	37
	Frequency effect	*5*	*9*	*3*	*-13*

Table 3. Fitted ex-Gaussian parameter values in **Experiment 1**, along with the sample means for comparison. Reliable differences are in bold.

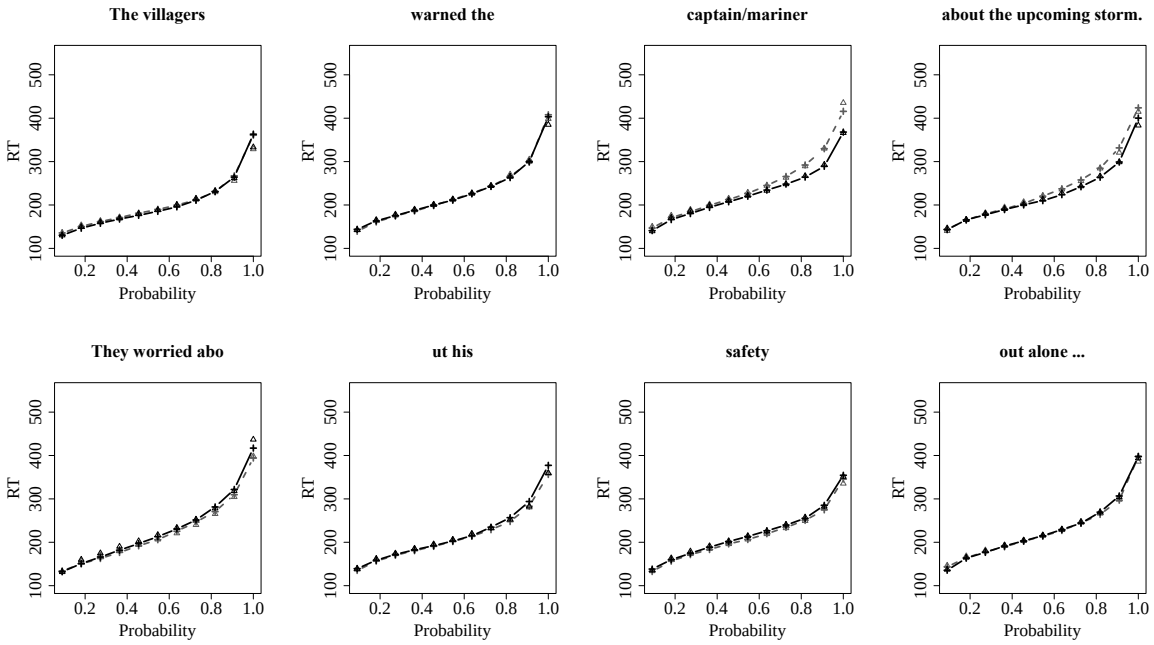

Figure 1. Predicted ex-Gaussian density plots based on parameter values averaged across participants from **Experiment 1.** The high frequency condition is plotted as a solid line, and the low frequency condition is plotted as a dashed line. The regions of interest are the antecedent (*captain/mariner*), anaphor (*ut his*) and post-anaphor (*safety*) regions.

	First fixation duration		*First-pass duration*		*Total time*	
Region	Low	High	Low	High	Low	High
Common noun						
Antecedent	265	239	345	278	612	459
Pronoun	253	260	281	295	375	389
Pronoun+1	235	247	275	278	382	368
Pronoun+2	260	254	302	305	420	432
Proper name						
Antecedent	260	240	289	257	480	399
Pronoun	254	257	286	290	401	401
Pronoun+1	234	244	265	273	400	400
Pronoun+2	259	249	309	300	449	450

Table 4. Averages in first fixation duration, first-pass duration, and total time across critical regions in **Experiment 2.**

	First fixation duration			*First-pass duration*			*Total time*		
Region	β (ms)	CI (ms)	Pr(β<0)	β (ms)	CI (ms)	Pr(β<0)	β (ms)	CI (ms)	Pr(β<0)
Common noun									
Antecedent	26	[13,38]	**<.01**	67	[47,88]	**<.01**	157	[114,200]	**<.01**
Pronoun	-5	[-21,10]	0.26	-11	[-32,11]	0.15	-11	[-45,22]	0.26
Pronoun+1	-10	[-22,1]	**<.05**	-1	[-21,18]	0.46	11	[-19,43]	0.24
Pronoun+2	5	[-8,18]	0.22	-4	[-21,14]	0.34	-12	[-36,21]	0.25
Proper name									
Antecedent	21	[8,34]	**<.01**	33	[16,49]	**<.01**	87	[49,123]	**<.01**
Pronoun	-2	[-16,13]	0.40	-3	[-22,18]	0.42	0	[-34,34]	0.49
Pronoun+1	-8	[-19,1]	**<.05**	-6	[-23,12]	0.25	4	[-34,41]	0.43
Pronoun+2	10	[-3,22]	***0.06***	11	[-9,31]	0.14	-3	[-45,35]	0.43

Table 5. Statistics for first fixation, first-pass and total time measures in **Experiment 2.** Reliable effects are in bold font, marginal effects with .05< *p* <.07 are in bold and italic font.

	Region		
	Antecedent	Anaphor	Post-anaphor
First fixation times			
Pronoun- frequent antecedent	241 (6)	247 (5)	233 (5)
Pronoun- infrequent antecedent	272 (7)	255 (7)	240 (5)
Repeated NP- frequent antecedent	249 (6)	234 (5)	245 (5)
Repeated NP- infrequent antecedent	278 (8)	239 (5)	252 (5)
First pass times			
Pronoun- frequent antecedent	293 (9)	272 (7)	269 (8)
Pronoun- infrequent antecedent	374 (13)	289 (10)	268 (8)
Repeated NP- frequent antecedent	294 (9)	307 (9)	275 (8)
Repeated NP- infrequent antecedent	381 (15)	328 (10)	290 (10)
Total times			
Pronoun- frequent antecedent	488 (21)	411 (18)	394 (18)
Pronoun- infrequent antecedent	637 (22)	419 (19)	408 (17)
Repeated NP- frequent antecedent	486 (21)	464 (21)	390 (15)
Repeated NP- infrequent antecedent	645 (25)	515 (22)	428 (24)

Table 6. Averages in first fixation duration, first-pass duration, and total time across critical regions in **Experiment 3.**

	First fixation duration			*First-pass duration*			*Total time*		
Region	β (ms)	CI (ms)	Pr(β<0)	β (ms)	CI (ms)	Pr(β<0)	β (ms)	CI (ms)	Pr(β<0)
Pronoun									
Antecedent	31	[13,48]	**<.01**	80	[50,110]	**<.01**	149	[94,203]	**<.01**
Pronoun	8	[-7,23]	0.16	17	[-7,40]	***<.08***	3	[-41,48]	0.44
Post-anaphor	7	[-8,21]	0.18	0	[-21,20]	0.51	15	[-28,59]	0.25
Repeated NP									
Antecedent	28	[10,47]	**<.01**	87	[55,117]	**<.01**	161	[101,221]	**<.01**
Pronoun	5	[-7,17]	0.20	23	[-1,46]	**<.05**	52	[4,102]	**<.05**
Post-anaphor	6	[-7,20]	0.18	18	[-5,39]	***<.07***	44	[-4,90]	**<.05**

Table 7. Statistics for first fixation, first-pass and total time measures in **Experiment 3.** Reliable effects are in bold font, marginal effects with .05< *p* <.07 are in bold and italic font.