



This is a repository copy of *How well are primary and secondary meanings of L2 words acquired?*.

White Rose Research Online URL for this paper:  
<https://eprints.whiterose.ac.uk/209826/>

Version: Supplemental Material

---

**Article:**

González-Fernández, B. [orcid.org/0000-0002-4370-1822](https://orcid.org/0000-0002-4370-1822) and Webb, S. (Accepted: 2024)  
*How well are primary and secondary meanings of L2 words acquired?* *Studies in Second Language Acquisition*. ISSN 0272-2631 (In Press)

---

This article has been accepted for publication in a revised form in *Studies in Second Language Acquisition* <https://www.cambridge.org/core/journals/studies-in-second-language-acquisition>. This version is free to view and download for private research and study only. Not for re-distribution, re-sale or use in derivative works. © The Author(s) 2024.

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

## APPENDIX 3 – ADDITIONAL ANALYSES

### 1. *Preliminary analyses*

Before examining the effect of the different types of meaning on the learning of the form-meaning connections of L2 words, the two groups (treatment and control) were compared to ensure that they did not differ significantly. Table 1 presents the descriptive statistics by group for prior vocabulary knowledge in English, and knowledge of the familiar meanings and target new meanings in the pretest. It shows that both learner groups could recall virtually all primary meanings of familiar words, while they knew, on average, only 6% of the secondary target meanings in recognition and less than 1% in recall. Because real words were used in this study, correct responses were expected for some items on the pretest. For that reason, relative gains were employed for analyses.

To make sure that the two groups did not differ significantly, they were compared statistically in their pretest results, uVLT scores, and age. A Mann-Whitney test (data non-normally distributed) showed no significant difference between the two groups in the pretest scores assessing a) recall of the primary meaning of the familiar words ( $U = 605.00$ ,  $z = .950$ ,  $p = .342$ ,  $\eta^2 = .014$ ) and b) knowledge of the secondary target meanings at either recognition level ( $U = 524.50$ ,  $z = -.204$ ,  $p = .838$ ,  $\eta^2 = 6.4e-4$ ) or recall level ( $U = 561.50$ ,  $z = .489$ ,  $p = .625$ ,  $\eta^2 = .004$ ). No statistically significant differences were found either between the groups for the uVLT total scores ( $U = 537.50$ ,  $z = -.032$ ,  $p = .974$ ,  $\eta^2 = 1.6e-5$ ) or any of their individual sections ( $p$  between .256 and .816), or for age ( $U = 458.50$ ,  $z = -1.108$ ,  $p = .268$ ,  $\eta^2 = .019$ ). These results show that the two groups were matched well in terms of their age, overall prior vocabulary level, their recall knowledge of the primary meanings of the target familiar words and their lack of recognition and recall knowledge of the new, target meanings.

**TABLE 1** Comparison of groups ( $N = 66$ ) (Mean in percentages)

	<b><u>Treatment</u></b>		<b><u>Control</u></b>	
	<b><u>group (n=36)</u></b>		<b><u>group (n=30)</u></b>	
	<b><i>M</i>(<i>SD</i>)</b>	<b>Range</b>	<b><i>M</i>(<i>SD</i>)</b>	<b>Range</b>
Gender	F = 35; M = 1		F = 25; M = 5	
Age	22.22(1.84); 20-28		22.2(3.12); 20-35	
1 <sup>st</sup> Meaning Recall	97.5(3.9)	85-100	96.83(3.8)	85-100
New Meaning Recog (pretest)	6.94(6.1)	0-21.8	6.88(5.3)	0-21.8
New Meaning Recall (pretest)	0.65(1.7)	0-6.6	0.44(1.4)	0-6.6
uVLT Total	55.5(16.6)	24.7-97.3	54.58(12.3)	32.7-74.7
uVLT 1K	84.54(11.9)	56.7-100	86.11(9.9)	66.7-100
uVLT 2K	57.59(21.3)	20-100	54.78(17.6)	10-86.7
uVLT 3K	50.1(19.2)	10-93.3	55.44(14.5)	26.7-83.3
uVLT 4K	43.52(18.1)	16.7-93.3	39.11(15.2)	13.3-76.7
uVLT 5K	41.76(19.9)	10-100	37.44(16.4)	3.3-66.7

## 2. Model comparison

To examine the potential predictors that influence the acquisition of meaning recognition and recall, the nested models were explicitly compared using *Chi-Square* test for significant differences between them, as well as the AIC and BIC scores, to ensure the best model was chosen (see Table 2 for the recognition gains' models and Table 3 for the recall gains' models).

The analysis showed that excluding one-by-one insignificant covariates (i.e., length and frequency) did not significantly affect the model fit. In all cases (immediate/delayed recognition and recall knowledge), the best-fitting model was Model D, which included fixed

effects of group, meaning type, the interaction between these two variables, and uVLT, plus immediate test scores in the delayed tests. There was no statistically significant difference between the models with one or two interactions (Models D and E), but Model D was selected for reasons of parsimony, as the additional interaction between meaning type and uVLT in Model E was not significant in any case, and because Model D presented better AIC and BIC values. Thus, the more parsimonious Model D is the one presented and discussed in the Main Document to test the effect of type of meaning on the acquisition of new meanings of L2 words.

**TABLE 2** Recognition knowledge - Model comparison

Model	<u>Recognition Immediate posttest</u>					Model	<u>Recognition Delayed posttest</u>				
	AIC	BIC	Compared	Chisq (df)	p		AIC	BIC	Compared	Chisq (df)	p
<b>A</b> <i>Core</i>	1088.8	1133.5	A-B	3.99 (3)	.26	<b>A</b> <i>Core</i>	1334.0	1378.7	A-B (B better)	145.64 (4)	<.001* **
<b>B</b> <i>All predictors</i>	1090.8	1152.3	B-C	0.07 (1)	.79	<b>B</b> <i>All predictors</i>	1196.3	1263.4	B-C	0.04 (1)	.84
<b>C</b> <i>-Frequency</i>	1088.9	1144.8	C-D	0.23 (1)	.63	<b>C</b> <i>-Frequency</i>	1194.4	1255.9	C-D	0.05 (1)	.83
<b>D</b> <i>-Length</i>	1087.1	1137.5	D-A (D better)	3.68 (1)	.05*	<b>D</b> <i>-Length</i>	1192.4	1248.3	D-A (D better)	145.6 (2)	<.001* **
<b>E</b> <i>+Meaning* VLT</i>	1088.5	1150.0	D-E	2.61 (2)	.27	<b>E</b> <i>+Meaning* VLT</i>	1194.8	1261.8	D-E	1.66 (2)	.44

Model A = core model (fixed effects of *group*, *meaning type* and their interaction);  
 Model B = core + all potential predictors in it (*length*, *frequency* and *uVLT*, plus *immediate test scores* in the delayed test model);  
 Model C = Model B with one insignificant predictor excluded; Model D = Model C with one insignificant predictor removed;  
 Model E = Model D plus the additional interaction of meaning type and uVLT.

**TABLE 3** Recall knowledge - Model comparison

Model	<u>Recall Immediate posttest</u>					<u>Recall Delayed posttest</u>					
	AIC	BIC	Compared	Chisq (df)	p	Model	AIC	BIC	Compare d	Chisq (df)	p
<b>A</b> <i>Core</i>	1136.0	1180.8	A-B (B better)	14.42 (3)	.006**	<b>A</b> <i>Core</i>	1392.2	1436.9	A-B (B better)	66.68 (4)	<.001* **
<b>B</b> <i>All predictors</i>	1129.6	1191.1	B-C	0.03 (1)	.87	<b>B</b> <i>All predictors</i>	1333.5	1400.6	B-C	0.01 (1)	.92
<b>C</b> <i>-Length</i>	1127.7	1183.6	C-D	0.01 (1)	.95	<b>C</b> <i>-Frequency</i>	1331.5	1393.0	C-D	1.45 (1)	.23
<b>D</b> <i>-Frequency</i>	1125.7	1176.0	D-A (D better)	12.39 (1)	<.001* **	<b>D</b> <i>-Length</i>	1331.0	1386.9	D-A (D better)	65.22 (2)	<.001* **
<b>E</b> <i>+Meaning* VLT</i>	1129.1	1190.6	D-E	0.54 (2)	.77	<b>E</b> <i>+Meaning* VLT</i>	1333.6	1400.7	D-E	1.38 (2)	.50

The model structure (formulae) of the best-fitting models is included below. All the models followed a generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod'] and a binomial (logit) distribution.

- MeaningRecognitionIMMEDIATE ~ MeaningType \* Group + uVLTTTotal + (1 | Participant) + (1 | Item)
- MeaningRecognitionDELAYED ~ MeaningType \* Group + uVLTTTotal + MeaningRecognitionIMMEDIATE + (1 | Participant) + (1 | Item)
- MeaningRecallIMMEDIATE ~ MeaningType \* Group + uVLTTTotal + (1 | Participant) + (1 | Item)
- MeaningRecallDELAYED ~ MeaningType \* Group + uVLTTTotal + MeaningRecallIMMEDIATE + (1 | Participant) + (1 | Item)

### 3. Contrast analyses for the interaction between meaning type and group

To further explore the effect of type of meaning on each group independently, post-hoc contrast analyses were conducted. The interaction between meaning type and group was fully decompose using the *emmeans* function (Lenth, 2022). Family-wise error rates were controlled with Tukey-adjusted *p*-values within each set of comparisons (i.e., effects of the three meaning types on the learning of new meanings by each of the two groups). Model probability estimates (odds ratio scale) for the two groups are presented in Table 4 for recognition knowledge, and in Table 5 for recall knowledge. The results are interpreted in the Main Document.

**TABLE 4** Post-hoc contrast analyses of effect of meaning type in each group – Recognition knowledge

Group	Fixed effect	<u>Recognition Immediate</u>				<u>Recognition Delayed</u>			
		Probability	SE	<i>z</i>	<i>p</i>	Probability	SE	<i>z</i>	<i>p</i>
<i>Treatment</i>	Polysemy/Homonymy	1.194	0.42	0.50	.87	0.971	0.40	-0.07	.99
	Polysemy/Primary	1.211	0.43	0.54	.85	0.489	0.21	-1.69	.21
	Homonymy/Primary	1.014	0.36	0.04	.99	0.475	0.20	-1.76	.19
<i>Control</i>	Polysemy/Homonymy	7.820	5.34	3.01	.01**	3.915	2.25	2.31	.06
	Polysemy/Primary	5.951	3.69	2.88	.01**	4.235	2.44	2.51	.03*
	Homonymy/Primary	0.761	0.62	-0.33	.94	1.139	0.75	0.19	.98

*Note:* Tests are performed on the log odd ratio scale. *P*-values were corrected for multiple comparisons using the Tukey method. \*\*\* *p* <.001, \*\* *p* <.01, \* *p* <.05.

**TABLE 5** Post-hoc contrast analyses of effect of meaning type in each group – Recall knowledge

<b>Group</b>	<b>Fixed effect</b>	<b><u>Recall Immediate</u></b>				<b><u>Recall Delayed</u></b>			
		<b>Probability</b>	<b>SE</b>	<b>z</b>	<b>p</b>	<b>Probability</b>	<b>SE</b>	<b>z</b>	<b>p</b>
<i>Treatment</i>	Polysemy/Homonymy	2.9e+00	1.00e+00	2.18	.08	2.140	0.98	1.66	.22
	Polysemy/Primary	2.0e+00	1.00e+00	1.41	.34	1.061	0.48	0.13	.99
	Homonymy/Primary	7.0e-01	0.00e+00	-0.78	.72	0.496	0.23	-1.54	.28
<i>Control</i>	Polysemy/Homonymy	6.4e+05	4.44e+08	0.02	.99	8.425	9.61	7.87	.15
	Polysemy/Primary	6.0e-01	1.00e+00	-0.42	.91	2.326	1.74	1.13	.50
	Homonymy/Primary	0.0e+00	0.00e+00	-0.02	.99	0.276	0.33	-1.07	.54

## REFERENCES

Lenth, R. V. (2022). Emmeans: Estimated marginal means, aka least-squares means. In *R Package Version 1.8.3*. <https://CRAN.R-project.org/package=emmeans>.