Supporting materials: Quantitative linguistic predictors of infants' learning of specific English words

These materials give outcomes for three additional analyses of infants' word understanding. These analyses are alternatives to the cumulative link models described in the article. The first and second show regression analyses for 12-month-olds and 15-month-olds separately, whereas the third examines 15-month-olds but includes only words that children were *not* said to understand at 12 months.

1 Supplementary analysis 1: word understanding at 12 months

This analysis examines word understanding at 12 months. The reported model is the outcome of a series of logistic regression models evaluated by eliminating predictors that did not improve model fits. The overall analysis strategy was as in the paper's analysis of words said at 15 months. Children s2 and f1 were excluded because s2 had no CDI and f1 was only reported to know two words. The remaining six children knew between 19 and 135 words (mean, 82; sd, 48).

Table 1:

```
glmer( understands.12months ~ total.freq.log.c + isol.freq.log.c +
   concrete.c + word.category + MLU.c +
   total.freq.log.c:word.category + isol.freq.log.c:word.category +
   (1+word.category|mom) + (1|word))
```

predictor	coef.	(SE coef.)	z	p
Intercept	-2.9110	(1.1043)	-2.6359	0.0084
log total frequency.c	0.5843	(0.0897)	6.5162	*0.0000
log isolated frequency.c	0.4102	(0.2646)	1.5505	0.1210
median MLU.c	-0.0526	(0.0511)	-1.0289	0.3035
concreteness	0.4727	(0.1656)	2.8550	*0.0043
class(closed)	0.3442	(1.0904)	0.3157	0.7522
class(predicate)	-0.4025	(0.4419)	-0.9108	0.3624
class(closed):total.freq	-0.8342	(0.2355)	-3.5428	*0.0004
class(predicate):total.freq	-0.1777	(0.1514)	-1.1741	0.2404
class(closed):isol.freq	0.6955	(0.4010)	1.7345	+0.0828
class(predicate):isol.freq	-0.6380	(0.3484)	-1.8310	+0.0671
Observations	1540			
Log Likelihood	−680.5			

The final model included total (log) frequency, isolated (log) frequency, their interactions with word type, and concreteness. MLU was left in though not significant (z=-1.05, p=.29) for comparison with other models (removing MLU has only a negligible impact on the remaining coefficients).

In this analysis, total frequency was the strongest predictor, though this was strong for nouns (the reference level) and somewhat weaker for predicates. At 12 months, closed-class words were not aided by frequency (the beta at baseline plus the interaction =-.25).

Isolated frequency was not quite reliable for nouns (beta=.41, p=0.12). However, isolated frequency was a significant predictor among closed-class words (the beta for nouns plus the beta for the interaction =.75). The large negative beta for the interaction with predicates nullifies the positive coefficient for nouns (the base level), meaning that being heard in isolation did not predict understanding of predicates. This result was also found in the cumulative link model.

The phonotactic predictors, phonetic duration ratio, and MLU did not predict understanding significantly. With the possible exception of MLU (marginal in the longitudinal cumulative link model), this analysis of 12-month-olds alone revealed little evidence that predictors of learning at or before 12 months differed from predictors of learning through to 15 months as revealed in the cumulative link model.

2 Supplementary analysis 2: word understanding at 15 months

This analysis examines word understanding at 15 months. All eight children were included. They were reported to know between 17 and 260 words (mean, 175; sd, 81). As in the analysis of 12 months olds, the models were logistic regression analyses, here predicting the binary outcome of whether children knew a word at 15 months (1) or not (0).

The final model included total (log) frequency, isolated (log) frequency, concreteness, and word type as the most important effects, and MLU retained for comparison with other models (z=-1.5).

Table 2:

```
glmer( understands.15months ~ total.freq.log.c + isol.freq.log.c +
  concrete.c + word.category + MLU.c +
  (1+word.category|mom) + (1|word))
```

predictor	coef.	(SE coef.)	z	p
Intercept	0.1230	(0.5675)	0.2168	0.8284
log total frequency.c	0.5004	(0.0603)	8.3037	* 0.0000
log isolated frequency.c	0.2912	(0.1533)	1.8992	+ 0.0575
median MLU.c	-0.0560	(0.0362)	-1.5451	0.1223
concreteness	0.4443	(0.1540)	2.8852	* 0.0039
class(closed)	-1.1344	(0.9192)	-1.2341	0.2172
class(predicate)	-0.0626	(0.7277)	-0.0861	0.9314
Observations	2712			
Log Likelihood	-1303.5			

Total frequency was once again the strongest predictor, an effect that did not interact with word type. Isolated frequency only approached significance, as in the analysis of 12 month olds above. Comparing the 12-month-old and 15-month-old analyses, the results are quite similar. A version of the 12-month analysis without the frequency interactions resulted in coefficients that matched the 15-month analysis fairly closely (MLU, -.039; total frequency, .500; isolated frequency, .304; concreteness, .597).

3 Supplementary analysis 3: gains in understanding from 12 to 15 months

This analysis examines 15 month olds' "understands" scores, considering only the words that were not reported as known by individual children at 12 months. The question this addresses is: which linguistic features predict that a word will go from not known, to known, for a given child, as he or she grows from 12 to 15 months? This analysis could allow us to see what 12–15 month olds find most useful in word learning, and whether this is different from what is most effective for children up to 12 months. However, caution is justified in making this interpretation. It could also be that the words that children have failed to learn by 12 months are unusual, as a set, such that different predictors support their acquisition.

The number of words children gained between 12 and 15 months ranged from 16 to 202 (mean, 109; sd, 73).

The final model included word category (noun, predicate, closed-class), (log) total frequency, (log) isolated frequency, concreteness, and the interaction of concreteness and word category. Among these only total frequency was statistically significant. Isolated frequency did not come close (but is left in here because it is part of the point of the analysis. If removed, the beta coefficient for MLU becomes slightly more negative (-.082) and significant (z=-2.81, p=.029)). Concreteness was not signficant as a predictor but models lacking it and its interactions with word class were inferior by chi-square test.

One interpretation of this model is that isolated frequency and concreteness are more important for younger children (up to 12 months) than for somewhat older children (12–15 months). These variables were significant in the previous analysis of understanding at 15 months, but not here with the words already known by 12 months excluded. However, as stated above, the present "gains" analysis is not really an experiment in the right way; 15 month olds could be quite similar to 12 month olds in what they benefit from the most, but the words with those properties were already "taken" (i.e., learned) a few months earlier.

Table 3:

```
glmer( understands.15months ~ total.freq.log.c + isol.freq.log.c +
  concrete.c + word.category + concrete.c:word.category +
  (1+word.category|mom) + (1|word))
```

predictor	coef.	(SE coef.)	z	p
Intercept	-1.1597	(1.5973)	-0.7260	0.4678
log total frequency.c	0.4331	(0.0632)	6.8550	*0.0000
log isolated frequency.c	0.2204	(0.1885)	1.1690	0.2424
MLU.c	-0.0745	(0.0382)	-1.9504	+0.0511
concreteness	0.1791	(0.2547)	0.7031	0.4820
class(closed)	-1.4452	(1.9345)	-0.7471	0.4550
class(predicate)	-2.9374	(1.8955)	-1.5496	0.1212
concreteness:predicate	0.6497	(0.3393)	1.9146	+0.0555
concreteness:closed-cl.	-0.0413	(0.3518)	-0.1175	0.9065
Observations	1874			
Log Likelihood	-900.1			