

Heather Turnbull¹, Rebecca Frost² & Jill Lany¹
¹The University of Liverpool, ²Edge Hill University

Background

Infants can track statistics e.g., transitional probabilities (TPs) in artificial languages (Saffran et al., 1996 among others)

The ability to track TPs in artificial language is correlated with vocabulary size (Frost et al., 2020)

Very little is known about whether infants track TPs in their native language

Evidence that they can is limited (Ngon et al., 2013)

Research Aim

To test whether toddlers are sensitive to native language TPs.

At this age, language is still being acquired and natural language measures can be obtained

1) If toddlers track native language TPs, they should be better at repeating sequences that are consistent with native language TPs than sequences containing TPs of 0

2) The older they get, as exposure to the native language increases, the more sensitive they should be to native language TPs.

- This may result in improved repetition of consistent TP sequences and worse repetition of TP-0 sequences as age increases
- As sensitivity to native language TPs increases, so should vocabulary size

Planned Methods

Develop a nonword repetition task suitable for British-English speaking toddlers aged 2-4

- This is a common task to assess repetition at a young age and can be used as a measure of sensitivity to statistical structure (Stark et al., 2022)
- Create CVCV nonwords with high transitional probabilities (HTPs) and TPs of 0

MOT: <well why don't you play inside> [>] ?
 DAD: <daddy's coming back in , sweetheart> [<] .
 DAD: daddy's just going to the garage to get his drill .
 DAD: and he's coming straight back in .
 MOT: let's &-um play with some jigsaws or something , Lara .
 MOT: bricks ?
 (ENG-UK corpora, CHILDES database: MacWhinney, 2000)

1. Identifying high TP sequences in English speech

- 1 million utterances and 4.5 million total words.
- Calculated frequency and forward TP (FTP) of all bigrams
- Used a script to auto syllabify text, resulted in some mis-segmentation.
- Did not include those chunks in our stimuli so they did not contribute to the TPs used

syll1	syll2	bigram_freq	syll1_freq	syll2_freq	ftp	btp
you	go	10902	214287	52599	0.050876	0.207266
you	do	8784	214287	62752	0.040992	0.13998
you	want	8263	214287	12536	0.03856	0.659142
you	can	7952	214287	29661	0.037109	0.268096
you	wan	7510	214287	10888	0.035046	0.68975
you	like	7227	214287	24022	0.033726	0.300849

2. Choosing CVCV stimuli

Top syllables and bigrams = word sequences, and highly frequent function words

- Extracted highly frequent CV syllable sequences
- Paired this list exhaustively to make new bigrams
- Looked at TPs in range of .1 - .01.
- TPs imperfect but represent those in naturalistic speech.

Nonword	bigram_freq	syll1_freq	syll2_freq	ftp	btp
meto	2240	16466	82687	0.14	0.02709
caco	64	613	2661	0.1	0.024051
nago	2034	22049	52599	0.09	0.03867
nado	1981	22049	62752	0.09	0.031569
tago	1097	15003	52599	0.07	0.020856
mato	526	7868	82687	0.07	0.006361

SET 1 Mean TP=.04

TP	BIGRAM FREQUENCY	HTP NONWORD
0.1	64	CACO
0.07	1097	RADI
0.03	126	FARI
0.02	36	LIGA
0.02	78	MUTI
0.01	4	TAGO
TP	BIGRAM FREQUENCY	TP-0 NONWORD
0	0	DOLO
0	0	MARO
0	0	BOTO
0	0	NIBE
0	0	POME
0	0	NALE

SET 2 Mean TP=.04

HTP NONWORD	TP	BIGRAM FREQUENCY
NADO	0.09	1981
MATO	0.07	526
BOLO	0.05	41
RONI	0.03	20
POBE	0.01	16
LEME	0.01	14
TP-0 NONWORD	TP	BIGRAM FREQUENCY
RIGO	0	0
TIRA	0	0
COTA	0	0
FALI	0	0
MUCA	0	0
GADI	0	0

3. Deciding on the stimuli set

- Two sets of 6xHTP and 6xTP-0 nonwords
- Matched for phonotactic probability
- Some real words but not interpretable as English multi-word sequences

HTP nonwords recombined to make TP-0 nonwords for opposite set- syllables unique across sets

Planned Procedure



NWR task

- Set 1 or Set 2 repeated twice
- Add CVCV real words at random e.g., baby, dada, doggy

Naturalistic Play

- 20 min with caregiver
- Generate MLU and explore link to productive language



BPVS

- Explore link to vocabulary size

Planned Analysis

- To test if toddlers are sensitive to native language statistics we will:
 - Look for effect of nonword type (HTP vs TP-0) on correct repetition
- To investigate whether toddlers become more sensitive to native language statistics with age we will:
 - Look for effect of age on repetition of HTP vs TP-0 nonwords
- To examine the link between sensitivity to native language statistics and vocabulary development we will:
 - Explore whether repetition of HTP nonwords correlates with vocabulary size (receptive and productive)

Acknowledgements

With thanks to Andrew Jessop for his script and advice on TP calculations and Jada Mattis for her help with creating the stimuli

References

Frost, R.L.A et al. (2020). Non-adjacent dependency learning in infancy, and its link to language development. *Cognitive Psychology*, 120 / MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates. | Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., & Peperkamp, S. (2013). (Non) words, (non) words, (non) words: evidence for a protollexicon during the first year of life. *Developmental Science*, 16(1), 24-34. | Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, 274(5294), 1926-1928. | Stärk, K., Kidd, E., & Frost, R. L. (2022). The effect of children's prior knowledge and language abilities on their statistical learning. *Applied Psycholinguistics*, 43(5), 1045-1071.