

Probabilistic Use of High Frequency Words Helps Language Acquisition

Background

Recent studies suggest that language learning may benefit from the presence of high-frequency marker words¹ which may act as anchors that speech segmentation can occur around². Marker words may also assist grammatical categorisation³, possibly *while* learners are using them to segment speech⁴.

We can consider how this may work with the sentence “*you eat the cheese yet you drink the wine*”. The high frequency words leave only “*cheeseyet*” unsegmented. Further, *you* reliably precedes verbs, while *the* precedes nouns.

In a recent study⁴, we trained participants on an artificial language in which target words were reliably preceded by markers. Crucially, marker words indicated target words’ membership to one of two otherwise indistinguishable categories. Marker words helped participants to identify targets from speech, while also assisting grammatical categorisation.

However, targets and markers appeared together with perfect reliability. Language is much noisier than this, and this noise may benefit learning^{5, 6}.

So, how does varying the use of marker words influence performance on these tasks?

Method

Participants

- 72 adults (56 females, 16 males, mean age = 20.25 years).

The language

- Eight bisyllabic *target* words, arbitrarily split into two categories.
e.g. A: *noli*, *kapu*, *fede*, *samu* B: *tero*, *buza*, *vegi*, *tore*
- Two monosyllabic *marker* words (one per category).
e.g. A: *zu* B: *ni*

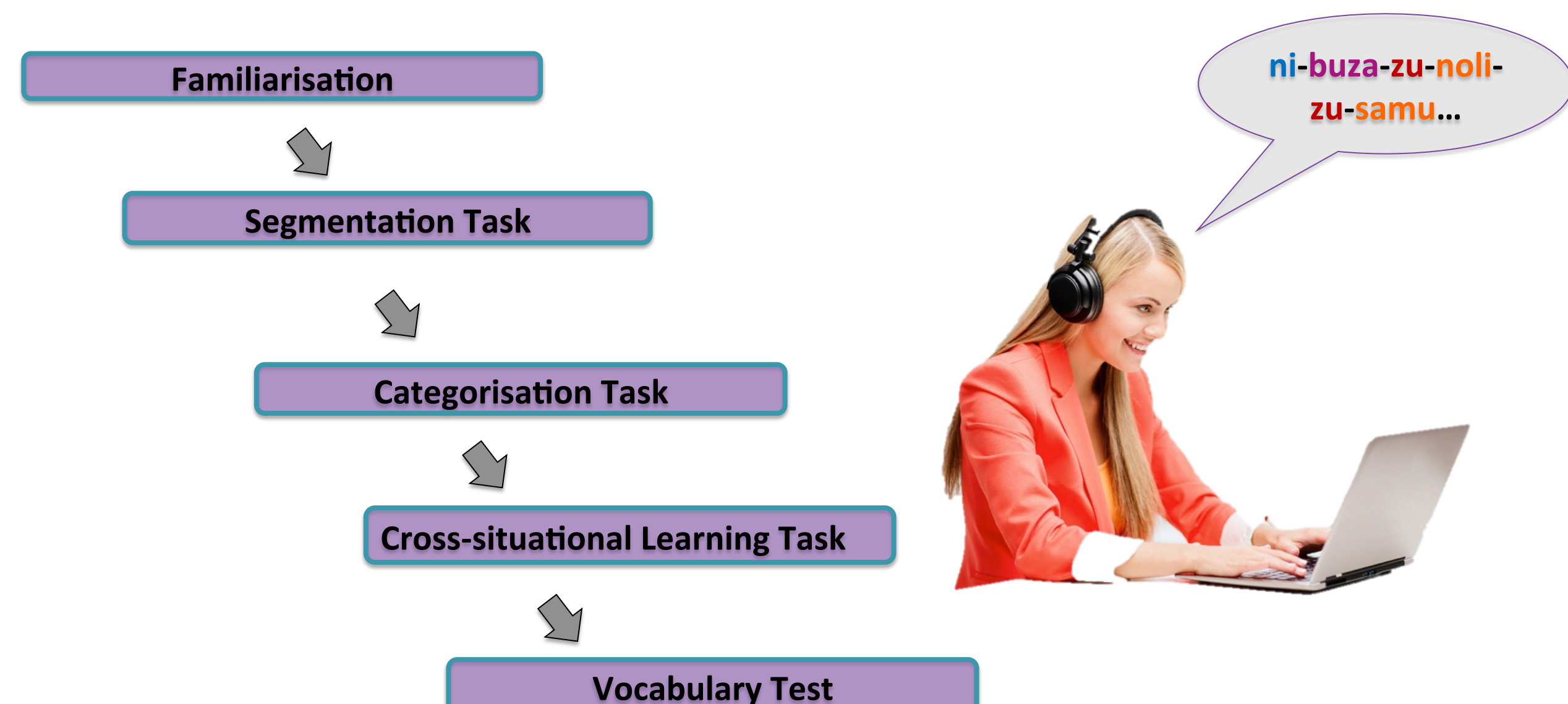
Conditions of variability

100% e.g. *zu-noli-ni-tero-zu-kapu-zu-fede-ni-buza-zu-samu-ni-tore-zu-noli-ni-vegi*

67% e.g. *zu-noli-tero-zu-kapu-fede-ni-buza-zu-samu-ni-tore-noli-ni-vegi*

33% e.g. *-noli-tero-zu-kapu-fede-ni-buza-samu-tore-noli-ni-vegi*

Procedure



Method cont.

Familiarisation

- Continuous speech stream (5-7 mins).

Segmentation Task

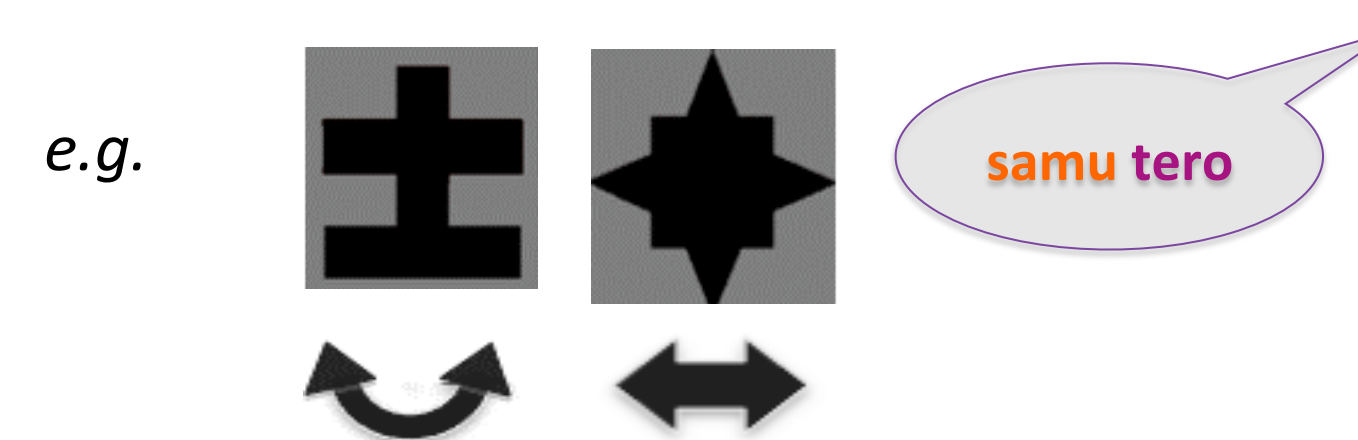
- Eight test-pairs, word vs. part-word comparisons (2AFC).
e.g. */fede/* vs. *de/sa*

Categorisation Task

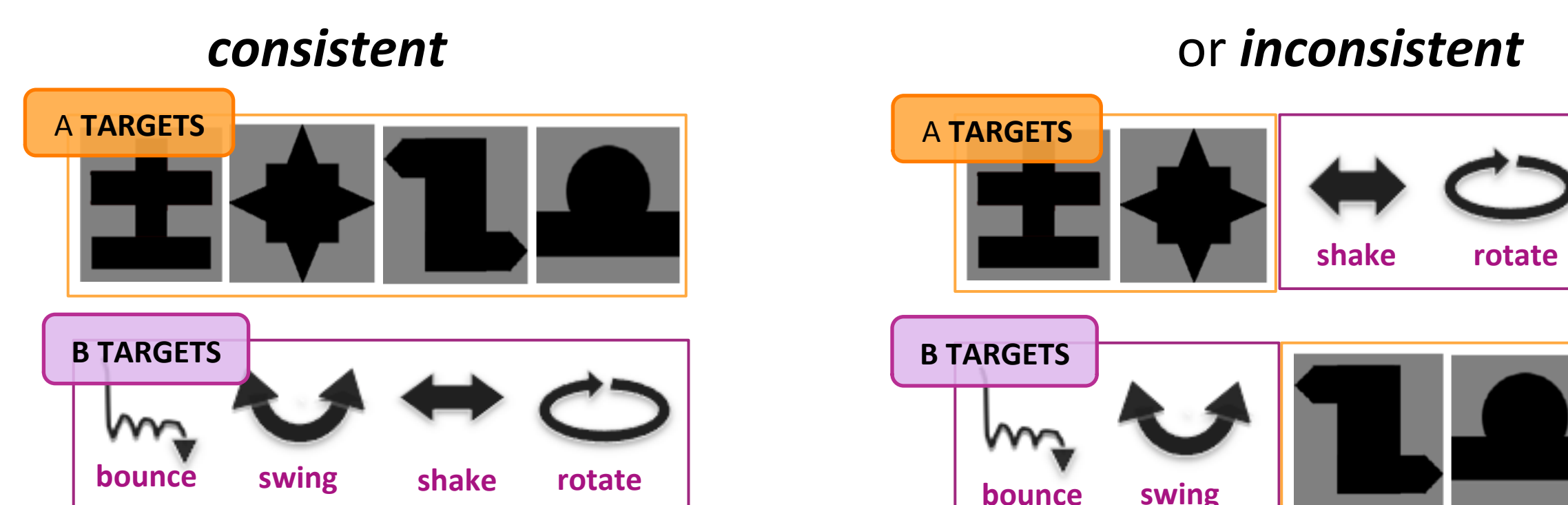
- 24 test-pairs: 12 contained two words from the same category, and 12 contained one word from each category.
e.g. *noli samu*, or *noli tero*
- Participants rated the similarity of roles in speech using a likert scale (1-6).

Cross-situational Learning Task

- We introduced four objects and four actions.
- Participants heard a sentence and saw a scene containing two objects, each undertaking a different action. Participants stated which object/action pairing the sentence described (no feedback).



Critically, pairing of A/B words with actions/objects was either:



- Six blocks, each containing eight scenes.
- Participants learnt word-action/object mappings using cross situational statistics. Inconsistent mappings should be harder to learn.

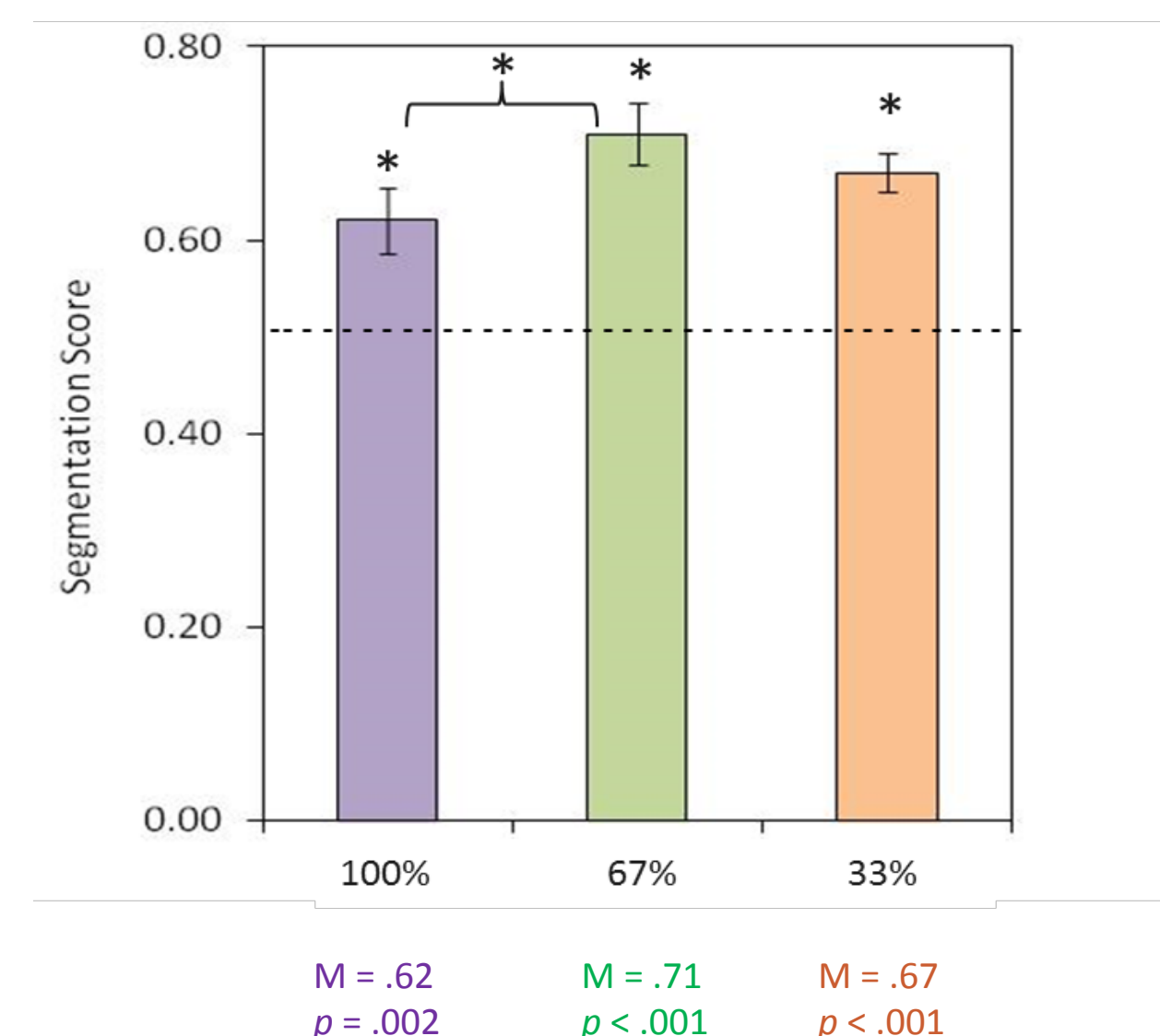
Vocabulary Test

- 16 2AFC trials. Participants selected the correct label for an action/object.

Results

Segmentation

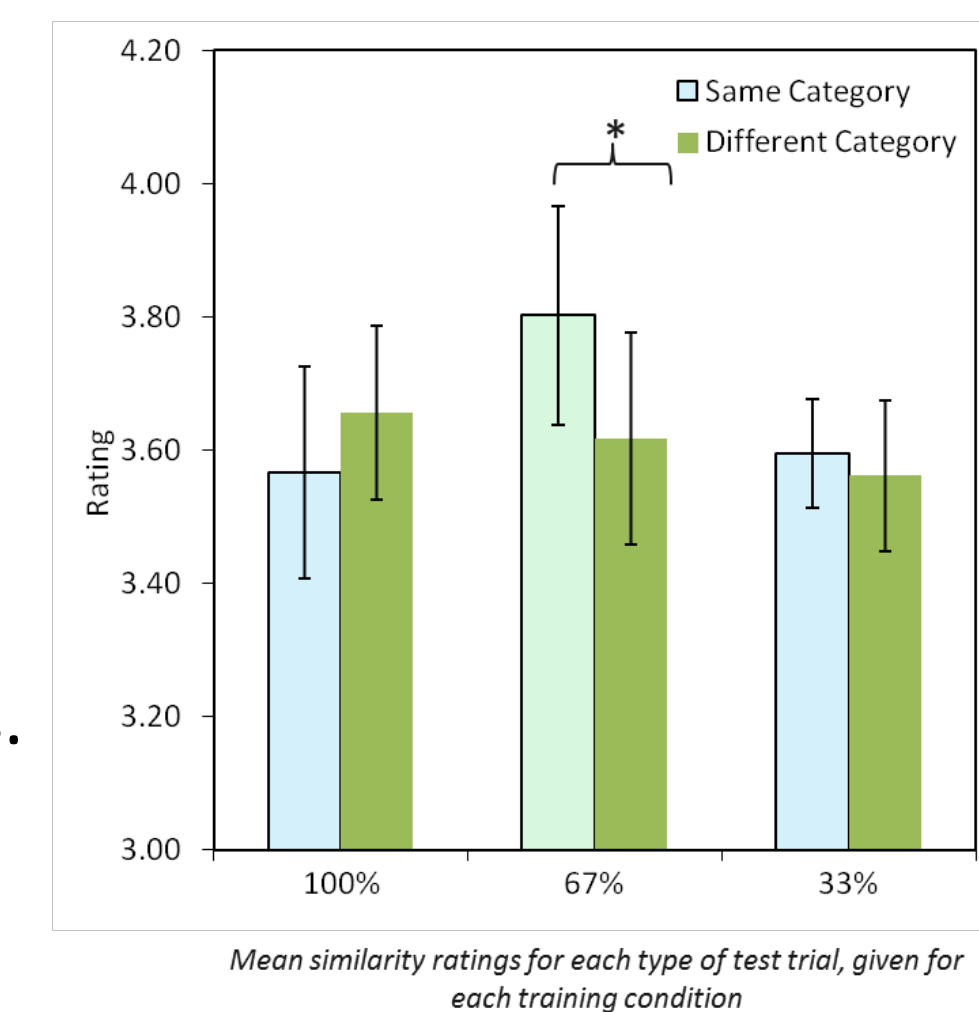
- Performance was significantly above chance for all conditions.
- Significant effect of condition $F(2, 54) = 3.647, p = .033, \eta^2 = .119$
- The 67% group outperformed the 100% group, $p = .010$; other comparisons $p > .118$



Results cont.

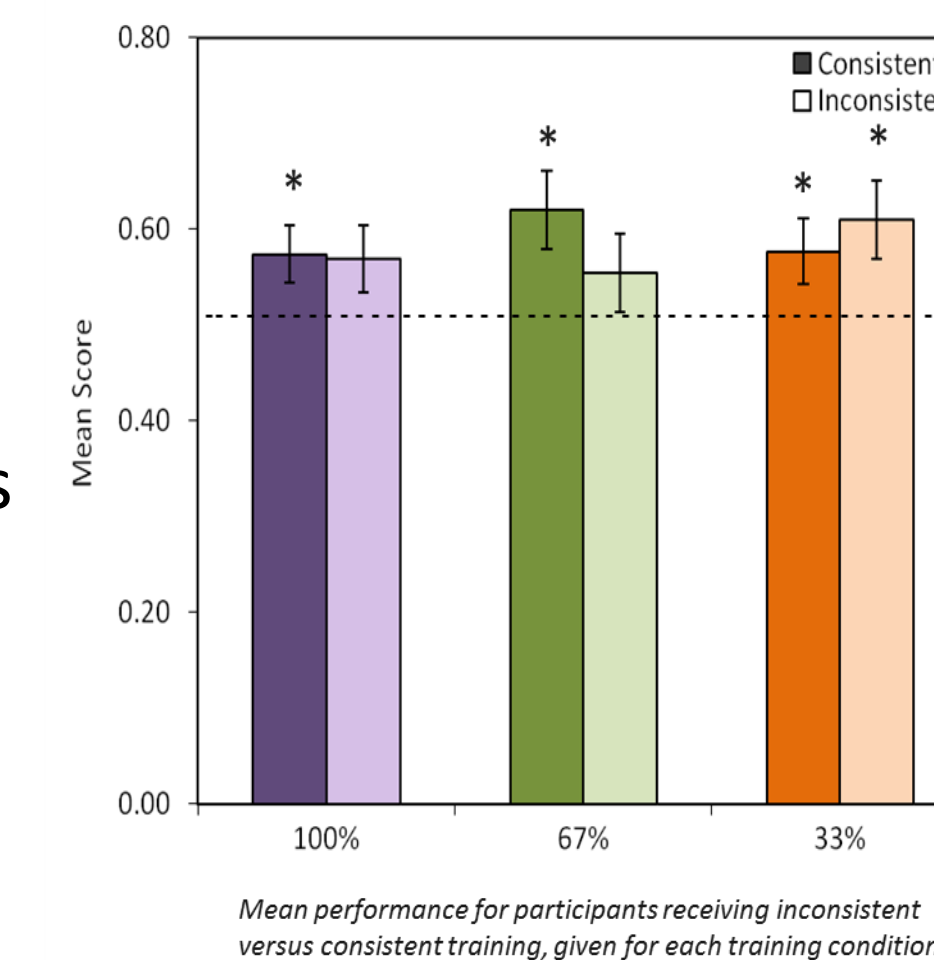
Categorisation

- No significant effect of test-pair type or variability condition.
- No significant test-pair type*condition interaction, $F(2, 54) = 1.913, p = .157$
- But participants in the **67% group** gave significantly higher ratings to pairs containing items from the same ($M = 3.802, SE = .165$) vs. different ($M = 3.617, SE = .159$) categories, $F(23) = 2.194, p = .039$.



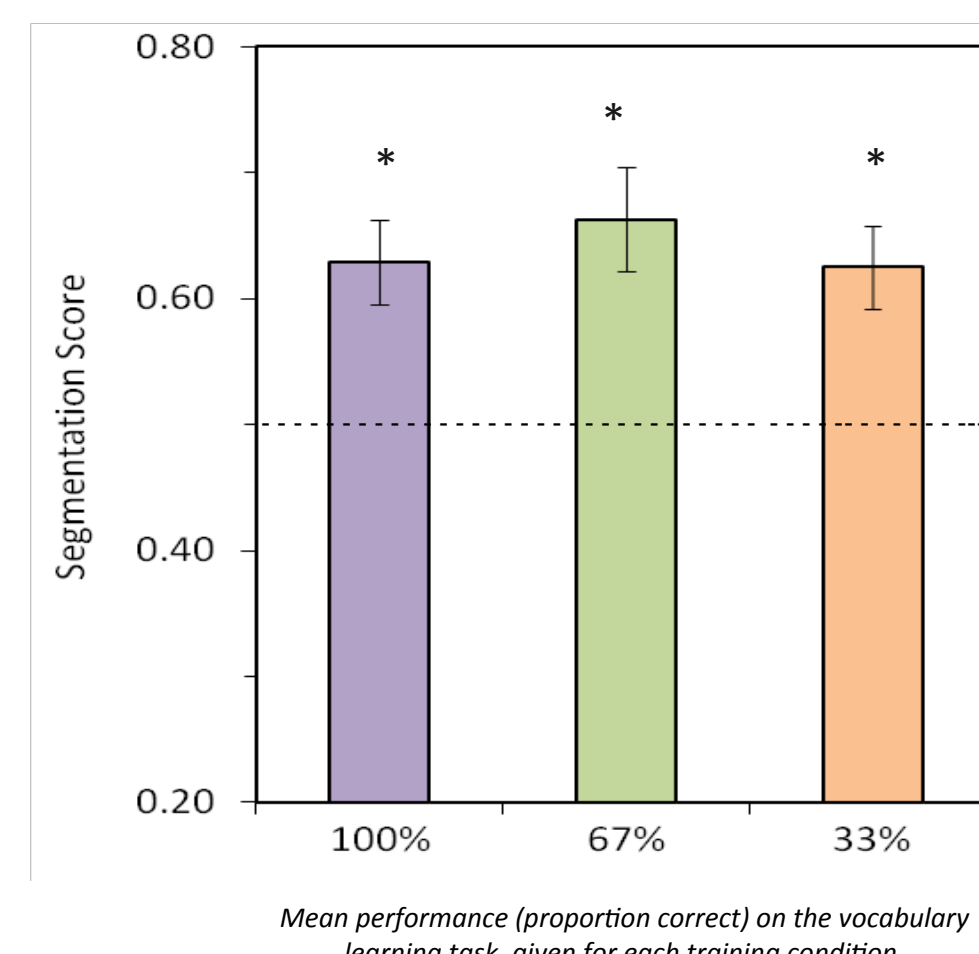
Cross-situational Learning Task

- No significant effects or interactions.
- For the 100% and 67% groups, only those receiving **consistent labelling** performed significantly above chance.
- Largest difference between consistency groups was seen for the 67% condition (but not sig).
- The 33% group performed above chance regardless of whether labelling was consistent or inconsistent.



Vocabulary Task

- All conditions performed above chance.
- Nouns > Verbs.
- Only one group learnt the verbs; those in the **67% group** who received **consistent labelling** ($M = 658, p = .017$).
- Learning of verbs approached significance in the 100% consistent group ($M = .605, p = .05$).



Conclusion

- Probabilistic co-occurrence of markers and targets led to the best segmentation performance, and the strongest demonstration of distributional categorisation.
- Distributional cues may have influenced verb learning in particular.

Data support claims that learners develop rule-like linguistic regularities while they are learning to segment speech^{5,7}.

Benefit of variability on performance; marker words may benefit language learning (segmentation, categorisation, and word learning) most when they appear in speech often, rather than always. Probabilistic co-occurrence may aid isolation of targets from markers, and lead learners toward other sources of information in speech.

References

- Bortfeld, H., Morgan, J.L., Golinkoff, R.M., & Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychological Science*, 16, 298-304.
- Monaghan, P. & Christiansen, M. H. (2010). Words in puddles of sound: modelling psycholinguistic effects in speech segmentation. *Journal of Child Language*, 37, 545-564.
- Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological distribution coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, 55, 259-305.
- Frost, R. L. A., Monaghan, P., & Christiansen, M. H. (2016). Using statistics to learn words and grammatical categories: How high frequency words assist language acquisition. *Proc. 38th Cognitive Science Society Conference*.
- Monaghan, P. (2017). *Topics in Cognitive Science*, 9, 21-34.
- Monaghan, P., Brand, J., Taylor, G., & Frost, R. L. A. (2017). *Proc 39th Cognitive Science Society Conference*.
- Frost, R. L. A. & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, 147, 70-74.