University of Pittsburgh

VA | U.S. Department of Veterans Affairs
Veterans Health Administration
VA Pittsburgh Healthcare System

# Within-trial evidence for the importance of semantic-feature generation during SFA

**Robert Cavanaugh[1,2*], Alexander Swiderski[1,2,3], Emily Goldberg[1,2,3], William D. Hula[2,1], William S. Evans[1], Michael Walsh Dickey[1,2,3]**

[1] Communication Sciences and Disorders, University of Pittsburgh,
[2] Speech Pathology & Audiology, VA Pittsburgh Healthcare System,
[3] Center for Neural Basis of Cognition, Carnegie Mellon University

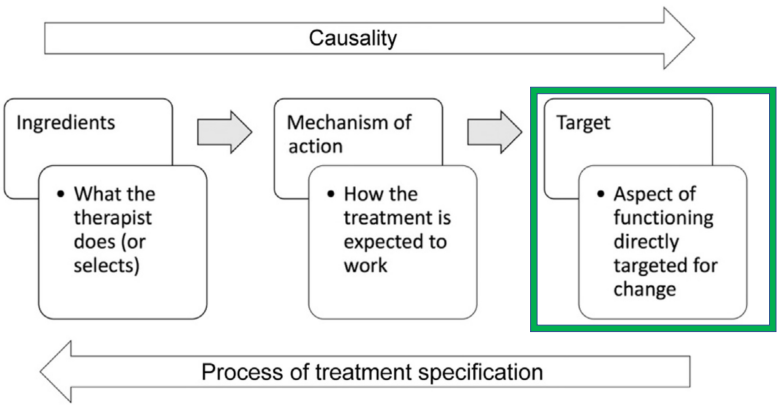



**Fig 1** Tripartite structure of treatment theory.

Van Stan et al., 2018

University of Pittsburgh

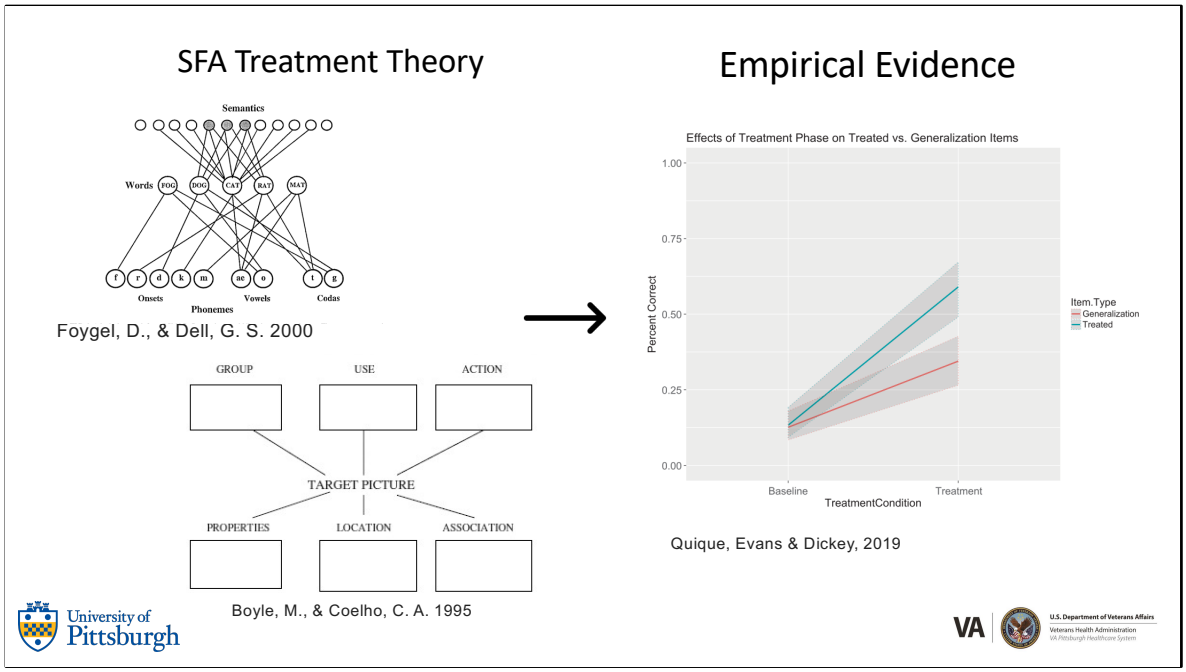
VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

I'd like to start with a figure that I hope many of us are increasingly familiar with. The Rehabilitation Treatment Specification System developed by John Whyte and colleagues provides a framework for specifying complex behavioral interventions which are common in aphasia.

The figure shows the central component of the RTSS. In essence, our treatments can be specified by considering one or more treatment targets, which change through a hypothesized mechanism of action, engendered through the active ingredients of treatment.

NEXT

For the most part, we have supported our theories about the ways in which our treatments work by evaluating treatment outcomes. That is - based on our theory of the treatment mechanism and the ingredients necessary to engage that mechanism, we hypothesize that we will see a specific pattern of change.




Let's walk through a familiar example.

In Semantic Feature Analysis treatment for aphasia, theory suggests that production of features that are semantically-related to a target word will increase the spread of and/or resting activation for words within a semantic category.

The expectation is that word retrieval should improve both for trained items and for semantically-related but untreated items, the result of improvements to the semantic system.

NEXT

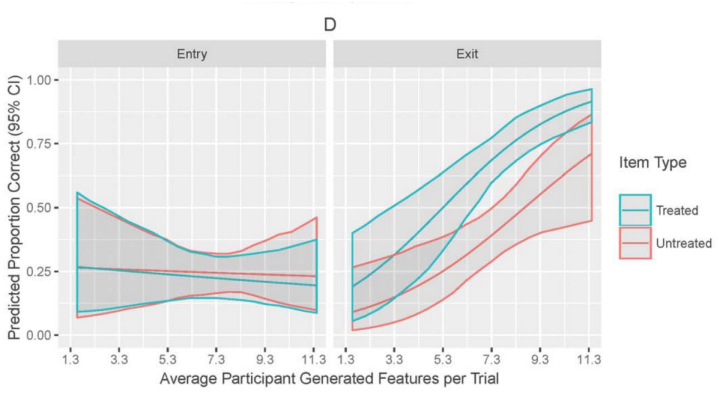
This predicted pattern of treatment outcomes has been shown relatively consistently across SFA studies (with some fair critiques), as summarized in the meta-analysis led by Yina Quique a few years ago.

These results have been interpreted as evidence for this hypothesized spreading activation account of SFA.

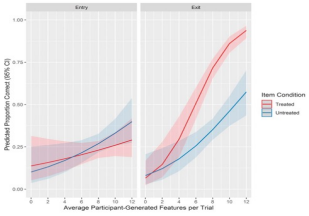
However, this evidence only provides a link between this behavioral intervention as a whole, and a pattern of outcomes which is consistent with theory. SFA includes a number of different components. If we want to understand more about our treatment components, our argument is that we also need to look closer at what those ingredients are doing within the treatment itself.

We need more evidence to establish that “feature generation” is in fact an active ingredient?

## Practice-related predictors of treatment response in SFA




Gravier, et al., (2018)



Evans et al., (2020)

University of Pittsburgh

VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

An initial study by Michelle Gravier in 2018 and follow-up by Will Evans in 2020 found that the number of semantic features generated per trial was predictive of naming outcomes for treated words, with some evidence that the relationship also holds for semantically-related, untreated words. These figures show this result – each additional feature generated per trial (the x-axis) was associated with greater odds of improvement from pre-treatment to post-treatment. We have interpreted this finding as evidence that the act of feature generation might be a key active ingredient in SFA.

However, this finding could just as well support the claim that that people who are able to generate more features are also those who respond to treatment – but not necessarily that the feature generation is the causal agent.

One way of testing these competing hypothesis is through a comparative effectiveness trial – testing the outcomes of SFA while experimentally manipulating the number of features individuals with aphasia generate in each trial. This is the focus of ongoing work at VA Pittsburgh.

Overarching Question:

What is the within-trial evidence for semantic feature generation as an active ingredient in SFA?

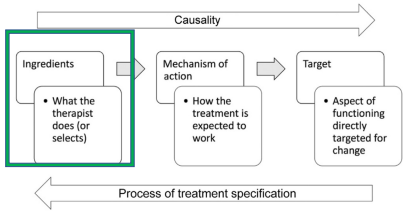



**Fig 1** Tripartite structure of treatment theory.

**Claim:**

**IF** feature generation is a key ingredient in SFA → **THEN** it will facilitate successful naming at the trial level during treatment.

The study I'm presenting today took a complementary approach to examining whether feature generation might be an active ingredient in SFA – by zooming in and looking closely at the within-trial effects of the feature generation rather than at treatment outcomes.

Our over-arching research question for this study was "What is the within-trial evidence for semantic feature generation as an active ingredient in SFA?"

Our claim is that, IF feature generation is a key active ingredient

**THEN** it will facilitate successful naming at the trial level during treatment.

## Study Overview

- Retrospective analysis of trial-level treatment data
- N = 44 (unilateral LCVA, 6mo+)
- Intensive Semantic Feature Analysis Treatment (~51 hours/3 weeks)
- Moderate-Severe to Mild Aphasia; CAT mean: 52.3 (4.7)
- 15 words treated per participant (5/list) for most participants
- 26k total trials (288k semantic feature attempts)
- Analysis: Generalized linear mixed-effects models in R

University of Pittsburgh

VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

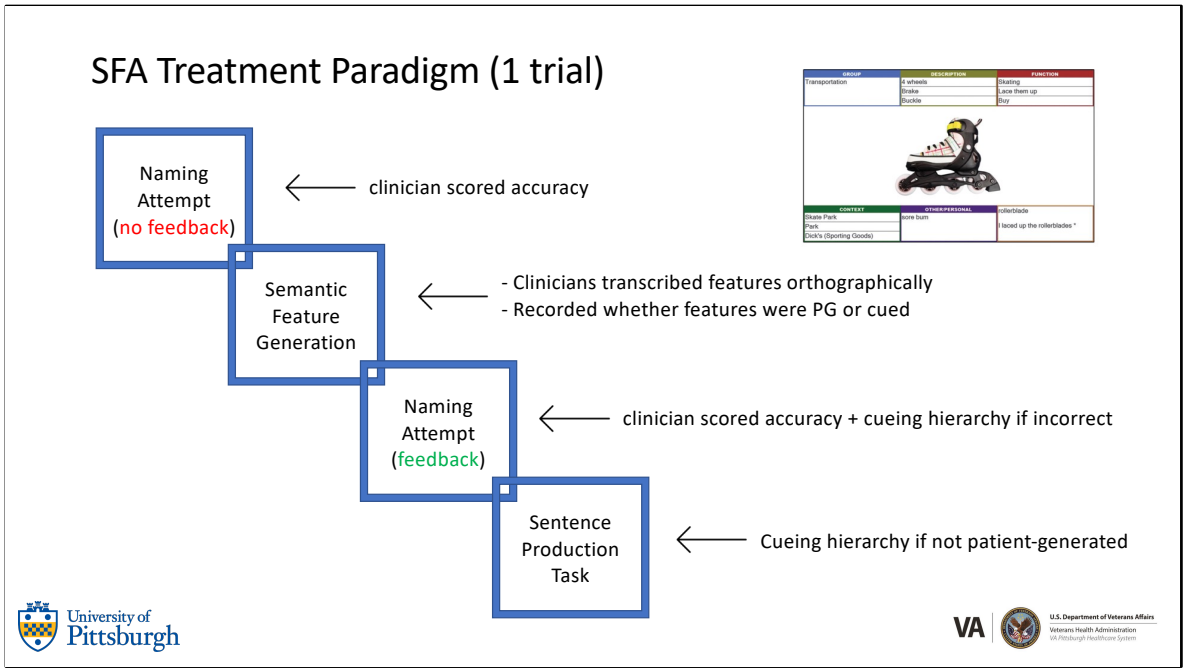
To answer this question, we went back and looked at trial-level treatment data from the clinical SFA trial with 44 total participants reported by Evans et al., in 2020.

Participants received 51 hours of SFA over 3-4 weeks on average.

Most participants received treatment on 15 words sequentially across 3 lists.

There were more than 26 thousand total trials across participants – nearly two hundred and ninety thousand feature generation attempts.

We used generalized linear mixed-effects models to analyze this data




Before I outline our specific research questions, I want to take a minute to review the components of a single trial in this SFA study.

First, participants were asked to name the target word, but they were not provided with feedback on their response. The clinician scored accuracy online.

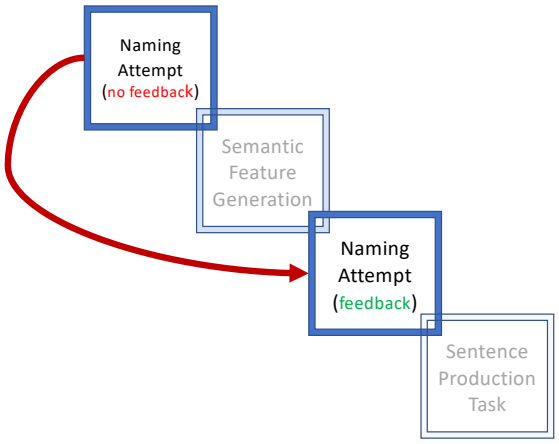
Second, they completed a typical feature generation task. They were asked to generate 11 total semantic features across function, context, description, category, and personal association feature types. During feature generation, clinicians transcribed each feature orthographically and recorded whether features were "patient-generated"– that is if they were generated by participants with minimal cueing.

Third, participants were again asked to name the target item. The clinician again scored accuracy online and participants received feedback on their response and a cueing hierarchy for the correct production if indicated.

Finally, participants were asked to use the target in a sentence with semantically-rich content from any of the features they generated.

# RQ 1

1. Does generating features during a treatment trial facilitate subsequent naming within the trial?




University of Pittsburgh

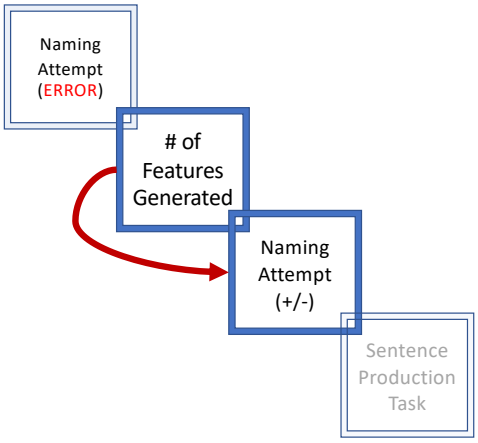
VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

Given this treatment paradigm, and the extensive data collected by the treating clinicians, we asked the following research questions.

First, Does generating features during a treatment trial facilitate subsequent naming within the trial? That is – Does naming accuracy improve from the initial to the final within-trial naming attempt, after generating semantic features.

# RQ 2

2. Does generating more features increase the odds of successful retrieval for trials *with initial naming difficulty?*




University of Pittsburgh

VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

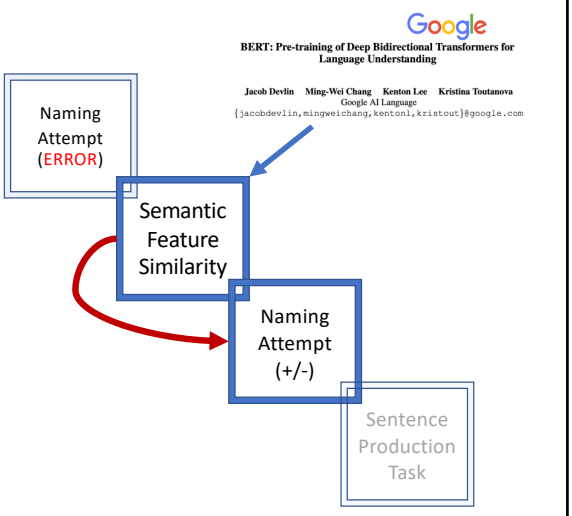
Second,

when a participant is unable to name the item correctly on the first naming attempt,

does generating more features within that trial increase the odds of a successful response on the second naming attempt?

# RQ 3

3. Does generating features more semantically similar to the target increase the odds of successful retrieval for trials *with initial naming difficulty*?




University of Pittsburgh

VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

Third and finally,

when a participant is unable to name the item correctly on the first naming attempt,

does generating features that are more semantically similar to the target item increase the odds of a correct response on the second naming attempt?

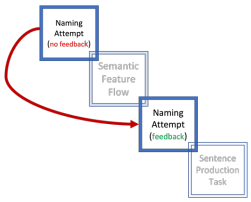
Our theory is that the spreading activation account of SFA suggests that generating features with close semantic similarity to the target would induce greater activation of the surrounding semantic network, and therefore increasing the odds of naming the item correctly on the second attempt.

Taken together,

If the answer is 'yes' to each of these research questions, we suggest that this would support the claim that feature generation is a *within-trial* active ingredient in SFA,

consistent with at least some formulations of the spreading activation mechanism.

Results RQ1: Does generating features during a treatment trial facilitate subsequent naming within the trial?




DV: Within-trial naming accuracy (0, 1)

IVs:

- ***Timepoint (initial vs. final naming attempt)***
- Within-session trial order
- Treatment session
- Two-way interaction between timepoint and session
- Two-way interaction between timepoint and trial

University of Pittsburgh

VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

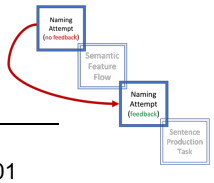
Now I'll walk through the statistical model and results for each question in turn.

Within-trial naming accuracy was predicted by timepoint (the key parameter of interest)

, within-session trial number, the treatment session number for that item, the two-way interaction between timepoint and session, and the two-way interaction between timepoint and trial.

The latter two parameters were included based on theoretically possible interactions that we wanted to account for.

## Question 1. Does generating features during a treatment trial facilitate subsequent naming within the trial?




*Question 1*

| Parameter | Log-Odds | SE | 95% CI | z | p |
|---|---|---|---|---|---|
| (Intercept) | 2.07 | 0.29 | (1.49, 2.64) | 7.04 | < .001 |
| timepoint | 1.00 | 0.10 | (0.81, 1.18) | 10.48 | < .001 |
| session.z | 0.61 | 0.08 | (0.45, 0.77) | 7.44 | < .001 |
| trial.z | 0.37 | 0.04 | (0.29, 0.45) | 9.05 | < .001 |
| timepoint * session.z | 0.07 | 0.03 | (0.01, 0.12) | 2.4 | 0.016 |
| timepoint * trial.z | 0.09 | 0.04 | (0.01, 0.16) | 2.27 | 0.023 |

*Random intercepts (slopes): participant (timepoint, session, trial); item (1)*

*Note: Inclusion of theoretically possible interactions between timepoint and model covariates (session, trial) significantly improved model fit but did not substantially change timepoint estimate. Timepoint was sum coded (-0.5, 0.5).*

University of Pittsburgh

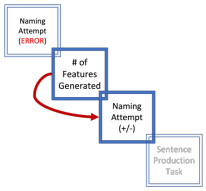
VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

*Here are the results for question 1, which show a robust and large effect of timepoint.*

*Participants were ~2.7 times more likely to name the word correctly after completing the feature-generation paradigm, regardless of how they did generating features. This effect accounts for the trial and session for each item.*

Question 2. Does generating more features increase the odds of successful retrieval for trials with initial naming difficulty?

DV: Within-trial naming accuracy (0, 1) on the final trial




IVs:

- ***The number of within-trial patient-generated features***
- Trial Number
- Session Number

*Note 1 : On trials where the first naming response was incorrect*
*Note 2: Interactions between features generated and trial, session did not improve model fit*

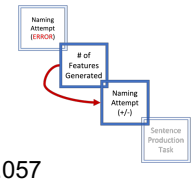
University of Pittsburgh

VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

For question 2, within-trial naming accuracy was predicted the number of within-trial patient generated features

as well as within-session trial number and the treatment session number for that item.

Question 2. Does generating features during a treatment trial facilitate subsequent naming within the trial?




| Parameter | Log-Odds | SE | 95% CI | z | p |
|---|---|---|---|---|---|
| (Intercept) | 0.34 | 0.18 | (-0.01, 0.69) | 1.9 | 0.057 |
| features-gen.c | 0.14 | 0.03 | (0.08, 0.19) | 4.85 | < .001 |
| trial.z | 0.26 | 0.05 | (0.17, 0.35) | 5.54 | < .001 |
| session c | 0.43 | 0.08 | (0.27, 0.58) | 5.41 | < .001 |

*Random intercepts (slopes): participant (features-gen.c, session.z, trial.z); item (features-gen.c)*

University of Pittsburgh

VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

*The results for question 2 indicate that there was a small, but reliable effect of feature generation. That is, when participants were unable to produce the target word initially, generating more features during the trial increased the odds of correctly producing the target word on the second naming attempt.*

Question 3. Does generating features more semantically similar to the target increase the odds of successful retrieval for trials with initial naming difficulty?


DV: Within-trial naming accuracy (0, 1) on the final trial

IVs:

- ***Cumulative within-trial semantic similarity score***
- Trial number
- Session number

*Note 1: On trials where the first naming response was incorrect*
*Note 2: Interactions between features generated and trial, session did not improve fit*

University of Pittsburgh

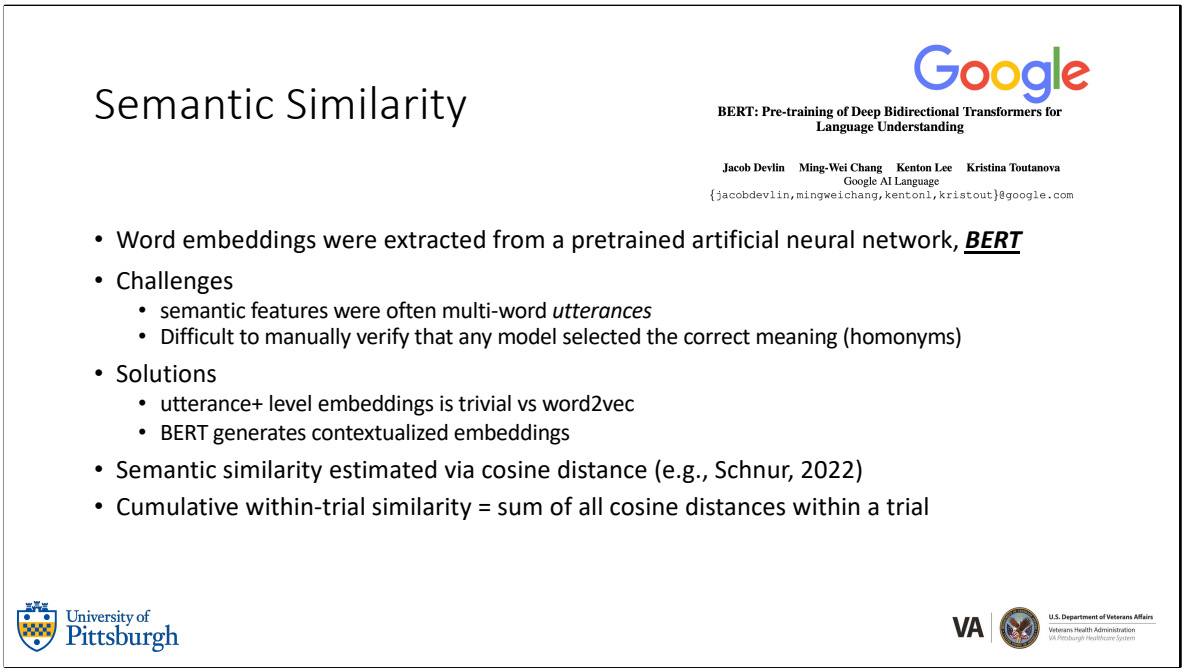
VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

Finally, for question 3, within-trial naming accuracy was predicted by the cumulative within-trial semantic similarity between patient generated features and the target word.

We chose this metric to represent semantic similarity a-priori as we felt is best aligned with the concept of total semantic activation from a trial

Again, we also included within-session trial number and the treatment session number for that item.

## Semantic Similarity




To estimate semantic similarity between patient-generated features and the target items, we extracted word embeddings from a pretrained semantic model BERT, which stands for Bidirectional Encoder Representations from Transformers.

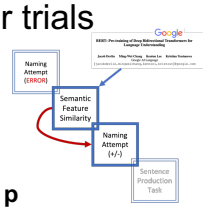
We elected to use BERT for two reasons.

- First - BERT generates contextualized embeddings – this helps to identify the correct meaning, especially for homonyms with two distinct meanings (for example a financial bank versus a river bank).

- Second, extraction of utterance level embeddings is trivial vs word2vec, which can only provide embeddings for single tokens in the corpus.

To compare embedding vectors, estimated the cosine similarity between them, which is an established approach that has been used previously in aphasia with models like word2vec.

Question 3. Does generating features more semantically similar to the target increase the odds of successful retrieval for trials with initial naming difficulty?




| Parameter | Log-Odds | SE | 95% CI | z | p |
|---|---|---|---|---|---|
| (Intercept) | 0.69 | 0.16 | (0.37, 1.01) | 4.21 | < .001 |
| similiarity.z | 0.37 | 0.08 | (0.21, 0.53) | 4.47 | < .001 |
| trial.z | 0.27 | 0.05 | (0.18, 0.37) | 5.61 | < .001 |
| session.z | 0.47 | 0.08 | (0.31, 0.64) | 5.56 | < .001 |

*Random intercepts (slopes): participant (similarity.z, session.z, trial.z); item (similarity.z)*

*Inclusion of theoretically possible interactions between features generated and model covariates (session, trial) did not significantly improved model fit*

University of Pittsburgh

VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

*So when we use our cumulative within-trial semantic similarity metric as a predictor, what do we find?*

*The results for question 3 indicate that there was a small, but reliable effect of similarity, suggesting that when participants were unable to produce the target word initially,*

*generating features that were, together, more semantically similar to the target word, increased the odds of correctly producing the target word on the final trial.*

Feature generation is likely a within-trial active ingredient in SFA

✅ The odds of a correct response improve within-trial after generating features
✅ Generation of more features increases the odds of a correct response (trials w/ anomia)
✅ Generation of semantically similar features increases the odds of a correct response (trials w/ anomia)

Consistent w/ hypothesized SFA mechanisms
- Restorative spreading activation account
- Also, compensatory/strategic circumlocution account (e.g., Bolowosky, 2022)

Demonstrates utility of analyzing treatment data to contribute to theory

Limitations:
- No method of examining effects on response generalization
- More research need to validate BERT similarity score
- Non-trivial correlations between features generated and semantic similarity

University of Pittsburgh

VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

In summary we found that

✅ The odds of a correct response improve within-trial after generating features
✅ Generation of more features within-trial increases the odds of a correct response on trials that begin with an instance of anomia
✅ Generation of semantically similar features within-trial increases the odds of a correct response on trials that begin with an instance of anomia

Together, this provides evidence that feature generation may be a *within-trial* active ingredient, consistent with the spreading activation account of SFA.

However, these findings are similarly consistent with alternative accounts of SFA – such as a strategic circumlocution mechanism discussed by Victoria Bolowsky on Sunday.

NEXT

Much of our recent work at the University of Pittsburgh has focused on treatment

specification and specifically asking empirical question about how our treatments might be working based on treatment data.

I think its worth emphasizing that these types of analyses can make meaningful contributions both to clinical practice recommendations and aphasia theory. Collecting treatment-level data is quite feasible with widely available software. And as shown by the ongoing comparative effectiveness trial, it can also provide critical preliminary support programmatic research.

NEXT

There are some important limitations to this work.

- Because this analysis focused on treatment data, we were also not able to examine whether within-trial feature generation was associated response generalization to related untreated words – a key hypothesized outcome of the spreading activation account of SFA.
- It would be beneficial to spend more time validating the semantic similarity scores estimated from the BERT model
- There were also non-trivial correlations between the number of features generated and the cumulative semantic similarity score for these features. One question we need to examine is how distinct these two predictors are.

## Future Directions

- Are there differences in the facilitative effects of feature generation if the features generated within a trial are more *taxonomic vs. thematic* in their relationship to the target?
- Do individual differences in semantic or phonological abilities moderate the facilitative effect of feature generation?
- Does successful feature generation predict successful sentence generation in the final treatment component?

We do have a few additional goals for this work.

By some accounts, the feature-target relationship is not unidimensional – it may be that the types of relationships between features generated and the target moderate the facilitative effects of feature generation. We are currently working on methods to automate categorizing features as taxonomically or thematically related to the target.

Second, we would like examine whether there might be individual differences in the facilitative effect of feature generation (e.g., phonological or semantic abilities) which might provide additional insight into candidacy for SFA.

Finally, we'd also like to consider how we might be able to better understand the role that the sentence generation component has on SFA outcomes. Specifically, Does it facilitate stimulus generalization to connected speech as intended?

# Acknowledgements & References


- NIH-NIDCD NRSA (F31DC019853-01)
- VA RR&D Award I01RX000832 awarded to Michael W. Dickey and Patrick J. Doyle
- Additional thanks to Yina Quique, Michelle Gravier, Alyssa Autenwreith, Emily Boss, Mary Sears, Jen Golovin, Miranda Babiak, and Jen Golovin



- Van Stan et al., 2018. The Rehabilitation Treatment Specification System: Implications for Improvements in Research Design, Reporting, Replication, and Synthesis
- Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of memory and language*, *43*(2), 182-216.
- Boyle, M., & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology*, *4*(4), 94-98.
- Quique, Y. M., Evans, W. S., & Dickey, M. W. (2019). Acquisition and generalization responses in aphasia naming treatment: A meta-analysis of semantic feature analysis outcomes. *American journal of speech-language pathology*, *28*(1S), 230-246
- Gravier, M. L., Dickey, M. W., Hula, W. D., Evans, W. S., Owens, R. L., Winans-Mitrik, R. L., & Doyle, P. J. (2018). What matters in semantic feature analysis: Practice-related predictors of treatment response in aphasia. *American Journal of Speech-Language Pathology*, *27*(1S), 438-453
- Evans, W. S., Cavanaugh, R., Gravier, M. L., Autenreith, A. M., Doyle, P. J., Hula, W. D., & Dickey, M. W. (2021). Effects of semantic feature type, diversity, and quantity on semantic feature analysis treatment outcomes in aphasia. *American journal of speech-language pathology*, *30*(1S), 344-358.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01
- Schnur, T. T., & Lei, C. M. (2022). Assessing naming errors using an automated machine learning approach. Neuropsychology.


University of Pittsburgh

VA U.S. Department of Veterans Affairs Veterans Health Administration VA Pittsburgh Healthcare System

I'd like to thank our funding sources which supported this work and thank you for your attention. I look forward to your questions.