CLS | Centre for Language Studies
Radboud University

DONDERS INSTITUTE

CLST | Centre for Language and Speech Technology
Radboud University

Native and non-native speakers' idiom production : What can read speech tell us?

Jing Liu[1] & Helmer strik[1,2,3]

[1]Center for language studies, [2]Centre for Language and Speech Technology,
[3]Donders Institute for Brain, Cognition and Behaviour

Radboud University Nijmegen, the Netherlands

MWE 2022 Workshop
06/25/2022

Idiom production

- Holistic representation in mental lexicon

Idiom production

- **Holistic representation in mental lexicon**

- Comprehension

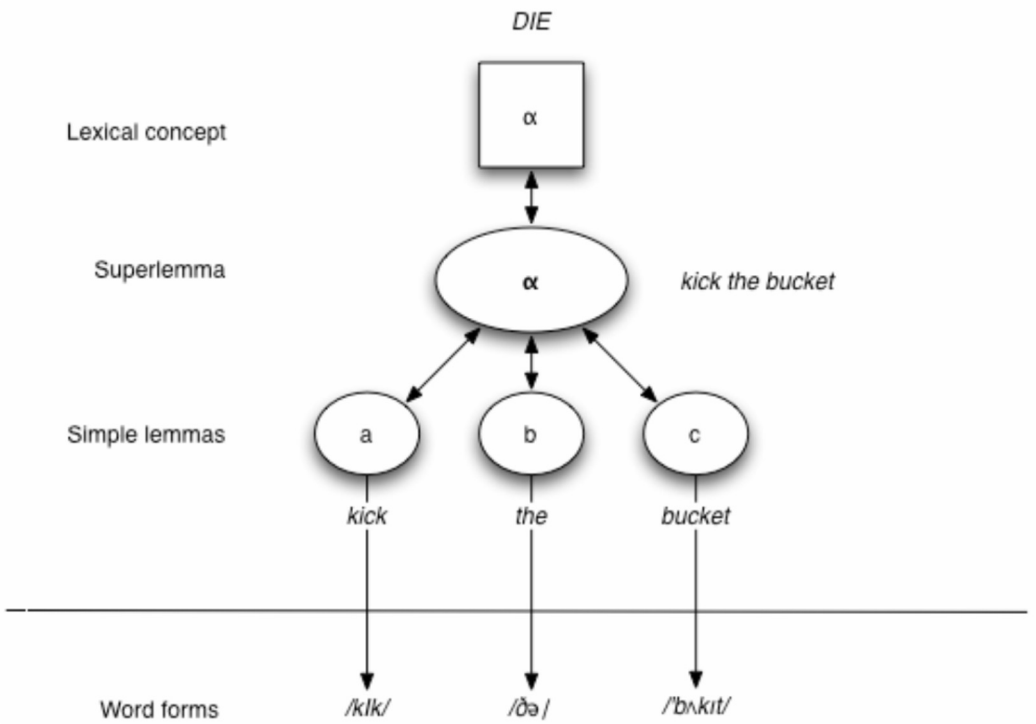
Eye-tracking (Jiang et al., 2020; Siyanova-Chanturia et al.,2011)

Self-paced reading paradigm (Conklin & Schmitt, 2008)

Idiom production

- **Holistic representation in mental lexicon**

- Comprehension: Shorter fixation/reading time (e.g. Jiang et al., 2020)
- Production: Superlemma Theory (Kuiper et al., 2007)

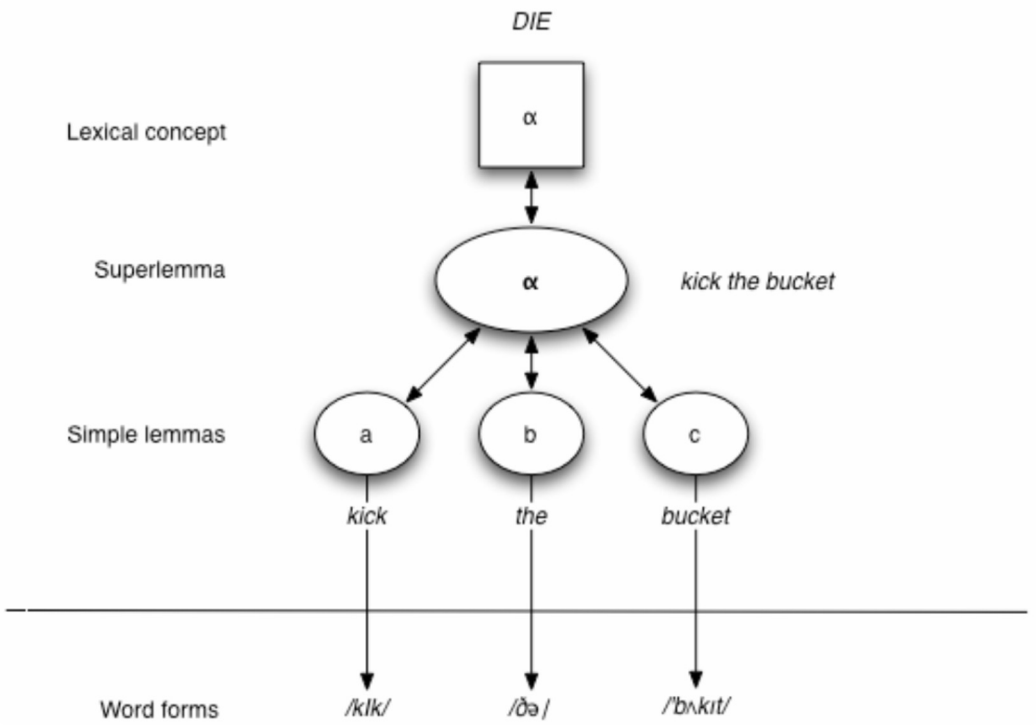


- Unitary idiomatic concept

Idiom production

- **Holistic representation in mental lexicon**

- Comprehension: Shorter fixation/reading time (Jiang et al., 2020)
- Production: Superlemma Theory (Kuiper et al., 2007)



- Unitary idiomatic concept

-> holistic phonological encoding

-> higher co-occurrences

Less hesitation, pauses and dysfluencies

Idiom production

- **Holistic representation in mental lexicon**

- Comprehension

Eye-tracking (Jiang et al., 2020; Siyanova-Chanturia et al.,2011)

Self-paced reading paradigm (Conklin & Schmitt, 2008)

- Production

Superlemma Theory (Kuiper et al., 2007)

Longer formulation time (Goldberg, 2017)

Q: articulation level?

Idiom production

- **Holistic representation in mental lexicon**

- Comprehension

Eye-tracking (Jiang et al., 2020; Siyanova-Chanturia et al.,2011)

Self-paced reading paradigm (Conklin & Schmitt, 2008)

- Production

Superlemma Theory (Kuiper et al., 2007)

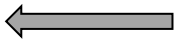
Longer formulation time (Goldberg, 2017)

- Influence of language proficiency (Zheng et al., 2021)

Q: L2 speakers?

Research design

- Focus on articulation level



Read speech

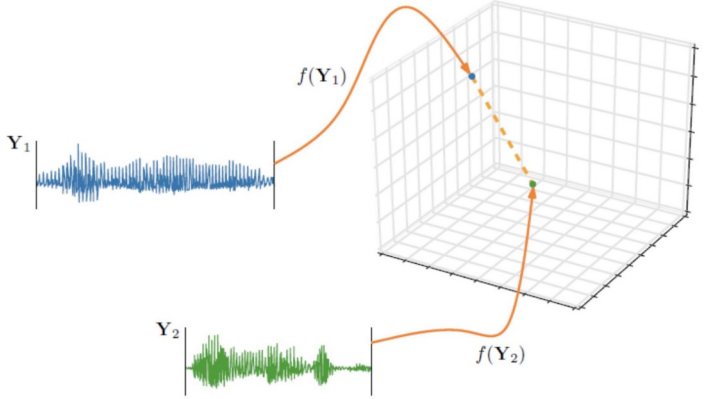
Research design

- **Focus on articulation level**

Read speech

- **Bottom-up investigation**

- Include all possible acoustic features



Multidimensional acoustic embedding (Beguš & Zhou, 2022)

Research questions

Q1: What are the differences of idiomatic and non-idiomatic speech produced by **native** speakers?

Q2: What are the differences of idiomatic and non-idiomatic speech produced by **L2** speakers?

Methodology

- **Idiom selection**

44 idioms were selected from 393 idioms based on:

Property	Explanation	Native	L2
Frequency	Relative degree to which participants have come across an idiom in speech or in print	0.94	0.88
Familiarity	How well speakers know the meaning of an idiom	0.94	0.93
Transparency	The extent to which the original metaphorical meaning can be deduced from its literal analysis	0.94	0.90
Imageability	The extent to which an idiom can evoke an image	0.91	0.89
Objective knowledge	Obtained from a multiple-choice test of meaning recognition	0.91	0.91
Usage	Frequency subjects indicate having used an idiom	0.91	0.86

Methodology

- **Data collection**

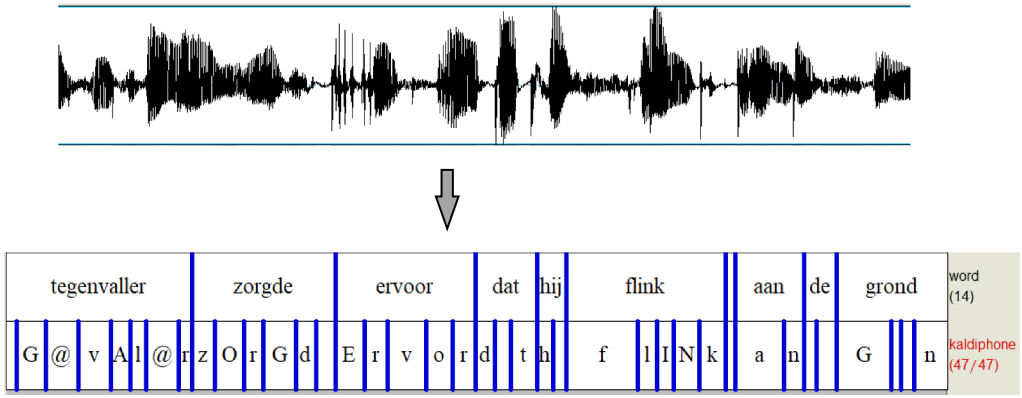
L2 speech data (ISLA corpus)

- Produced by 22 German L1 learners of Dutch L2 (mean age = 24.76; SD = 3.46)
- Each idiom is embedded in a sentence
- 1413 recordings altogether

Native speech material (Spoken Dutch Corpus: CGN)

- Corresponding 44 idioms produced by native Dutch speakers
- Each idiom is embedded in the sentence
- 175 recordings altogether

Data analysis



Step 1: Data pre-processing

- Forced alignment

-> Get temporal information for each word

Automatic: CLARIAH

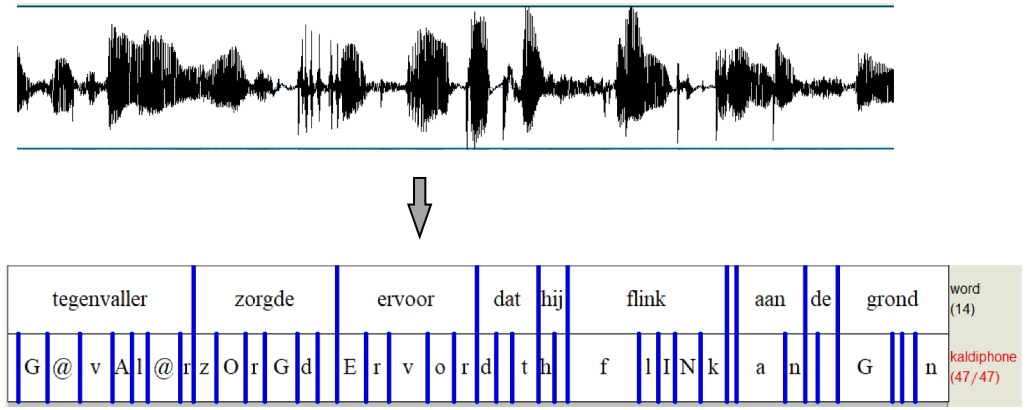
Manual checking

- Idiom extraction

Automatic: Lemma-based matching

Manual checking

Data analysis



	Feature 1	Feature 2	...	Feature n
Word 1	XXX	XXX		XXX
Word 2	XXX	XXX		XXX
...				
Word n	XXX	XXX		XXX

Step 2: Feature extraction on the word level

- Extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS: Eyben et al., 2016)

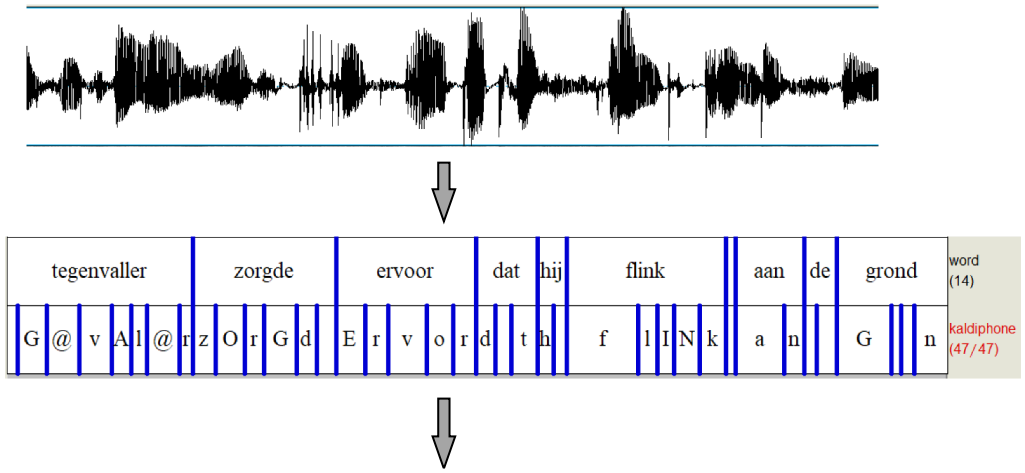
88 Frequency related, energy related, spectral and temporal parameters

- Praat features

17 speech rate, articulation rate, intensity, formants, center of gravity related features

-> 105 features altogether

Data analysis



	Feature 1	Feature 2	...	Feature n
Word 1	XXX	XXX		XXX
Word 2	XXX	XXX		XXX
...				
Word n	XXX	XXX		XXX

Step 2: Feature extraction on the word level

- Extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS: Eyben et al., 2016)

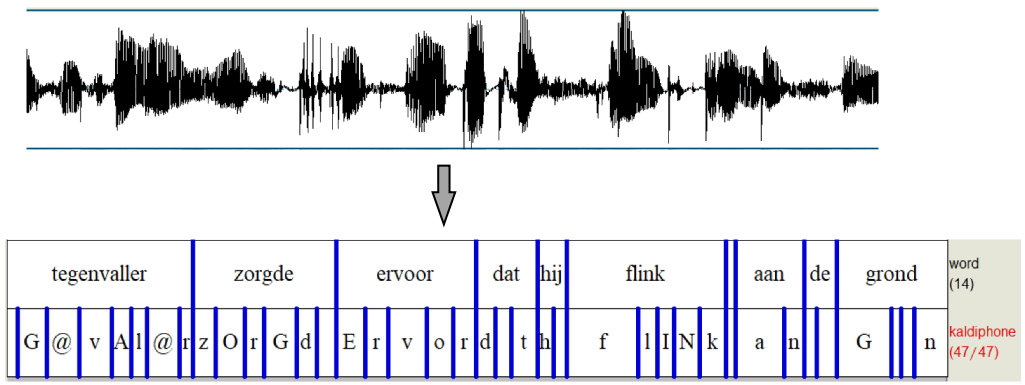
- Praat features

-> 105 features altogether

Characteristic for some types of atypical read speech

1) Read Dutch COPD speech (Van Bemmel et al., 2021)
2) Dysarthric speech (Wei et al., 2021)
3) Non-native speech (Bosland, 2022)

Data analysis



	Feature 1	Feature 2	...	Feature n
Word 1	XXX	XXX		XXX
Word 2	XXX	XXX		XXX
...				
Word n	XXX	XXX		XXX

	Feature # 1	Feature #2	...	Feature #n
Word 1	XXX	XXX		XXX
Word 2	XXX	XXX		XXX
...				
Word n	XXX	XXX		XXX

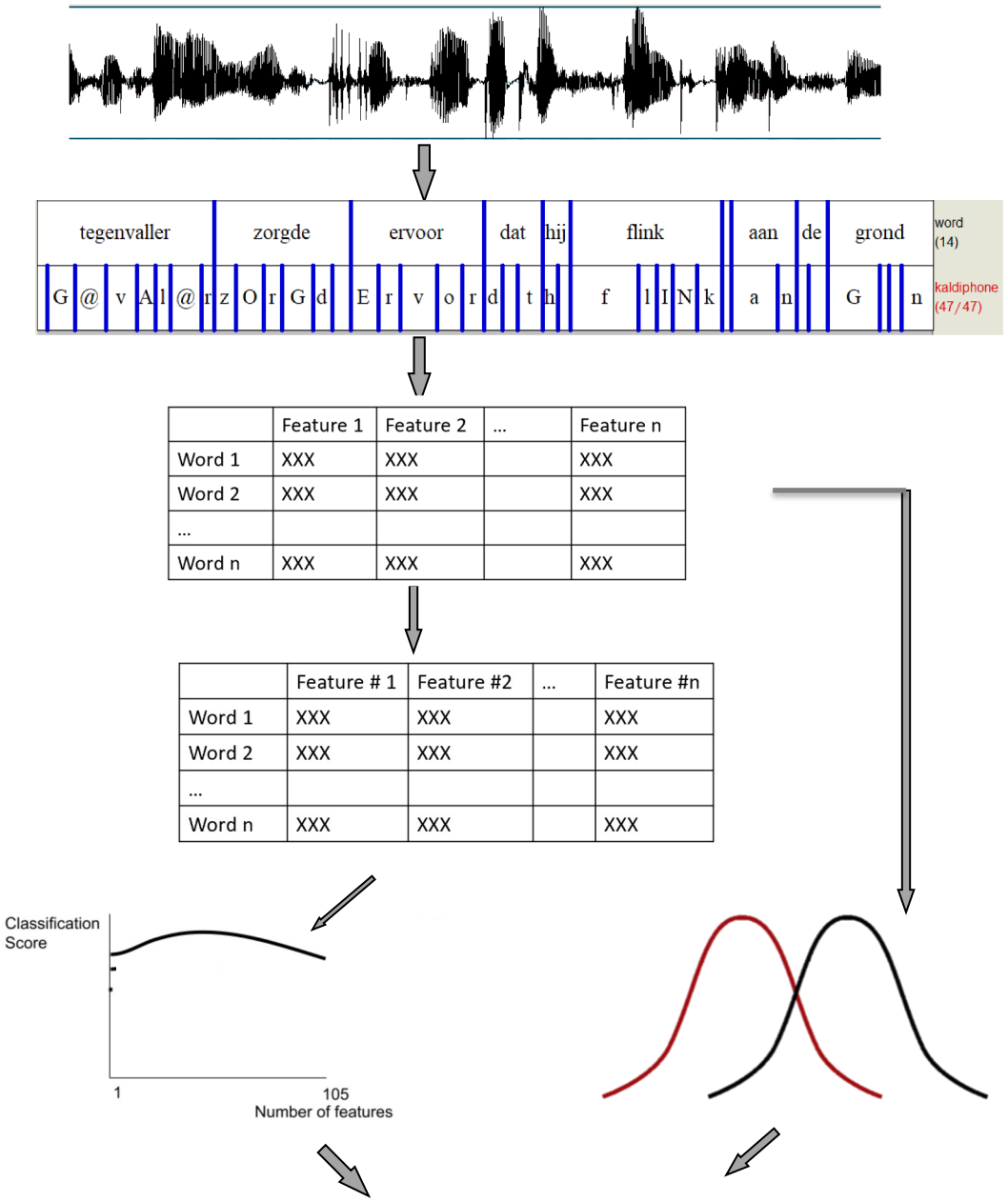
Step 3: Feature ranking

- Outlier detection

To exclude mispronounced words

- Recursive Feature Elimination (RFE: Guyon et al., 2002)

Data: $D = \{X, L\}$ // a training data set with n number of features
// where $X = \{f_1, f_2, f_3, ..., f_n\}$ and L are the labels
X' // predefined initial feature subset ($X' \subset X or X' = \{\}$)
θ // a stopping criterion
Result: X'_{opt} // an optimal subset
begin
  initialize:
  $X_{opt} = X'$
  $\phi_{opt} = E(X', A)$ // evaluate X' by using an mining algorithm A
  while (θ *not reached*) **do**
    $X_g = generate(X)$ //subset generation for evaluation
    $\phi = E(X_g, A)$ // X_g current subset evaluation by A
    if ($\phi > \phi_{opt}$) **then**
      $\phi_{opt} = \phi$
      $X'_{opt} = X_g$
  return X'_{opt}



Principal Component Analysis

Step 4: Statistical analysis

- T-tests

1.) Individual effects of features

2.) Rank all the features based on significance and effect size

- Support Vector Machine Model (SVM)

Select the most characteristic combination of features based on performance

- Principal Component Analysis (PCA)

Reduction of overlapping features

Results

Idioms v.s. Non-idioms produced by native speakers

T-tests: 45 significant features; SVM: 34 features

-> 13 overlapping features

Table 1

Feature categories after PCA

Category	Features	%variance
Loudness	Loudness percentile 20.0 - 50.0; intensity mean; loudness std	23.965
Spectral flux	Spectral Flux std; Spectral Flux normalized std	14.381
Speaking rate	Articulation rate(syllable/sec); Speech rate(syllable/sec)	10.956
Formant	F3; F2bandwidth normalized std; F2 mean amplitude	8.416
F0	F0 semitone percentile20.0; pitch std; F0semitone Falling Slope	7.348

Results

Idioms v.s. Non-idioms produced by L2 speakers

T-tests: 45 significant features; SVM: 26 features

-> 9 overlapping features

Table 2

Feature categories after PCA

Category	Features	%variance
Loudness and spectral flux	Loudness percentile20-80; intensity mean; intensity min equivalent Sound Level dB; spectral Flux mean	41.746
Formant	F3amplitude mean; F2amplitude mean; F1amplitude mean	22.108
F0	F0 semitone percentile20.0	10.258

Conclusion

- **Speaking rate**

Native speakers: **shorter** speech and articulation rate in idiomatic expressions

L2 speakers: no significant difference in speech and articulation rate

-> The superlemma representation may be influenced by language proficiency

Conclusion

- **Speaking rate**

Native speakers: **shorter** speech and articulation rate in idiomatic expressions

L2 speakers: no significant difference in speech and articulation rate

-> The superlemma representation may be influenced by language proficiency

- **Loudness**

Higher loudness in idiomatic expressions for both native and L2 speakers

-> Higher confidence in the more frequent fixed collocations (Cucchiarini,2000)

-> Syntactic structure of the idioms (most are verb phrases)

Limitations & future research

- **Data differ between idioms and non-idioms**

Formant as one of the characteristic features (different words)

Loudness as one of the characteristic features (different syntactic functions)

-> Next step: Collect similarly structured non-idiom materials

- **Limited (native) speech data may lead to overfitting**

-> Next step: Collect more (native) speech data

References

[1] Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. (2011). Seeing a phrase “time and again” matters: The role of phrasal frequency in the processing of multiword sequences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 37(3), 776.

[2] Goldberg, N. (2017). Production of idiomatic expressions: Formulation and articulation

[3] Zheng, H., Bowles, M. A., & Packard, J. L. (2021). NS and NNS processing of idioms and nonidiom formulaic sequences: What can reaction times and think-alouds tell us?. Applied Psycholinguistics, 1-26.

[4] Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners’ fluency by means of automatic speech recognition technology. The Journal of the Acoustical Society of America, 107(2), 989-999.

[5] Kuiper, K., Van Egmond, M. E., Kempen, G., & Sprenger, S. (2007). Slipping on superlemmas: Multi-word lexical items in speech production. The Mental Lexicon, 2(3), 313-357.

[6] Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., ... & Truong, K. P. (2015). The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE transactions on affective computing, 7(2), 190-202.

References

[7] Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? Applied Linguistics, 29(1), 72–89. https://doi.org/10.1093/applin/amm022.

[8] Jiang, S., Jiang, X., & Siyanova-Chanturia, A. (2020). The processing of multiword expressions in children and adults: An eye-tracking study of Chinese. Applied Psycholinguistics, 41(4), 1–31. https://doi.org/10.1017/s0142716420000296.

[9] Beguš, G., & Zhou, A. (2022, May). Interpreting intermediate convolutional layers in unsupervised acoustic word classification. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8207-8211). IEEE.

[10] Bemmel, L. V., Harmsen, W., Cucchiarini, C., & Strik, H. (2021, September). Automatic Selection of the Most Characterizing Features for Detecting COPD in Speech. In International Conference on Speech and Computer (pp. 737-748). Springer, Cham.

[11] Xue, W., van Hout, R., Boogmans, F., Ganzeboom, M., Cucchiarini, C., & Strik, H. (2021). Speech Intelligibility of Dysarthric Speech: Human Scores and Acoustic-Phonetic Features. In Interspeech (pp. 2911-2915).

Acknowledgements

- Special thanks to ISLA Team who inspired me to explore the magic of idioms and their support in formulating hypotheses.

Dr. Strik (Helmer)

Dr. Cucchiarini (Catia)

Dr. Hubers (Ferdy)

ISLA
Idiomatic Second Language Acquisition

NWO
Nederlandse Organisatie voor Wetenschappelijk Onderzoek

- Special thanks to AASPAA group (Loes van Bemmel, Xing Wei etc.) whose help was invaluable in developing the scripts.

Thanks for your attention!

For more information:


@JLiu_Compuling


@Jing-L97

Radboud Universiteit