



# Investigating vocabulary in academic speech of hard and soft sciences through corpora

**Yen Dang**

[T.N.Y.Dang@leeds.ac.uk](mailto:T.N.Y.Dang@leeds.ac.uk)

School of Education, University of Leeds

BALEAP conference

*Leeds, 12-14 April, 2019*

# Why investigate vocabulary in academic spoken English?



UNIVERSITY OF LEEDS

## Academic written word lists

1. Champion & Elley's (1971)
2. Praninskas's (1972)
3. Lynn's (1973)
4. Xue & Nation's (1984)
5. Ghadessy's (1979)
6. Coxhead 's(2000)
7. Gardner & Davies's (2014)
8. Browne, Culligan, & Phillip's (n.d)

## Academic spoken word lists

1. Nesi's (2002)
2. Dang, Coxhead, & Webb's (2017)

# Why investigate vocabulary in academic spoken English?

---



UNIVERSITY OF LEEDS

1. The lack of a large and representative academic spoken corpus
2. The assumption that academic written word lists are sufficient for learners to comprehend academic speech.

## **Coxhead's (2000) Academic Word List covered**

- **4.41%** of **academic spoken** English (Dang & Webb, 2014)
- around **10%** of **academic written** English (e.g., Coxhead, 2000; Hyland & Tse, 2007)

# Academic Spoken Word List (Dang, Coxhead, Webb, 2017)

---



UNIVERSITY OF LEEDS

- 1,741 words with high frequency and wide range in a 13 million word academic spoken corpus
- May help EAP learners to recognize 92–96% of words in academic speech

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, 67(4), 959–997.

- Available on Yen Dang's website:  
<https://essl.leeds.ac.uk/education/staff/689/yen-dang>

# Research questions

---



UNIVERSITY OF LEEDS

1. What are the most frequent words in academic speech of hard sciences?
2. What are the most frequent words in academic speech of soft sciences?
3. To what extent do these words overlap with the items in the Academic Spoken Word List?

# Hard science spoken corpus (6.5 million words)



UNIVERSITY OF LEEDS

Hard-pure		Hard-applied	
<i>Subject</i>	<i>Running words</i>	<i>Subject</i>	<i>Running words</i>
Astronomy	593,062	Chemical Engineering	563,938
Biology	552,452	Computer Sciences	555,175
Chemistry	556,138	Cybernetics	555,401
Ecology & Geology	555,312	Electrical Engineering	550,181
Mathematics	450,481	Health & Medical Sciences	470,795
Physics	554,178	Mechanical Engineering	558,604
<i>Total</i>	<i>3,261,623</i>	<i>Total</i>	<i>3,254,094</i>

# Soft science spoken corpus (6.5 million words)



UNIVERSITY OF LEEDS

Soft-pure		Soft-applied	
<i>Subjects</i>	<i>Words</i>	<i>Subjects</i>	<i>Words</i>
Art	553,160	Business	513,133
Cultural Studies	498,393	Economics	610,998
History	554,214	Education	571,023
Philosophy	549,577	Law	616,398
Political Studies	545,059	Management	461,093
Psychology	555,880	Public Policy	485,016
<i>Total</i>	<i>3,256,283</i>	<i>Total</i>	<i>3,257,661</i>

## Selection criteria



UNIVERSITY OF LEEDS

---

Same statistical measures were used to select items for the Hard Science Spoken Word List and the Soft Science Spoken Word List .

**Dang, T. N. Y.** (2018a). The nature of vocabulary in academic speech of hard and soft-sciences. *English for Specific Purposes*, 51, 69-83.

**Dang, T. N. Y.** (2018b). The hard science spoken word list. *ITL – International Journal of Applied Linguistics*, 169(1), 44–71.

✓ 1,595 words = Hard Science Spoken Word List (HSWL)

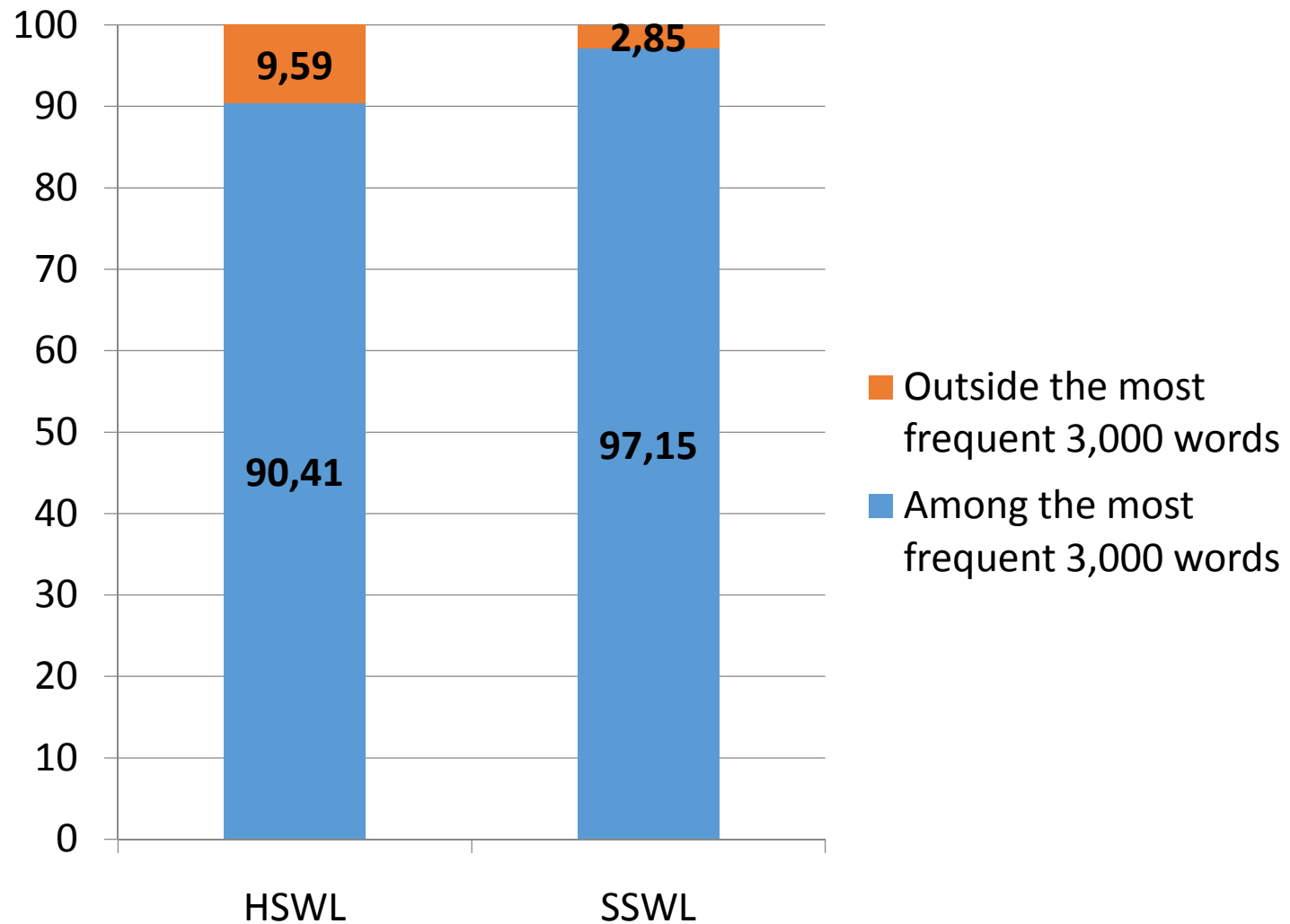
✓ 1,964 words = Soft Science Spoken Word List (SSWL)



# Percentage of different kinds of words in the HSWL and SSWL (%)



UNIVERSITY OF LEEDS



# Why so?



1. One distinctive feature of academic speech is 'on-line informational elaboration' (Csomay, 2006).
2. One important function of academic speech is classroom management (Biber, 2006)

*E.g., I'm planning on posting the exam at seven p.m. tonight*

*I hope you can play an active role in learning rather than sit down and wait for the material I give you you have to go to the library and search for the reference books*

3. Multi-words are more common in academic speech than academic writing (Biber et al., 2004; Simpson-Vlach & Ellis, 2010).

# Most frequent two-words terms in hard and soft science speeches



UNIVERSITY OF LEEDS

## Hard sciences

solar system  
pituitary gland  
kinetic energy  
northern hemisphere  
differential equation  
angular momentum  
lattice point  
binary search  
kinetic energy  
voltage across  
renal failure  
triple bottom

## Soft sciences

restored view  
world war  
personal identity  
social contract  
public health  
parental investment  
demand curve  
language teaching  
death penalty  
population size  
global warming

# Most frequent two-words terms in hard and soft science speeches



UNIVERSITY OF LEEDS

## Hard sciences

solar **system**  
pituitary gland  
kinetic **energy**  
**northern** hemisphere  
differential **equation**  
angular momentum  
lattice **point**  
binary **search**  
kinetic **energy**  
voltage **across**  
renal **failure**  
triple **bottom**

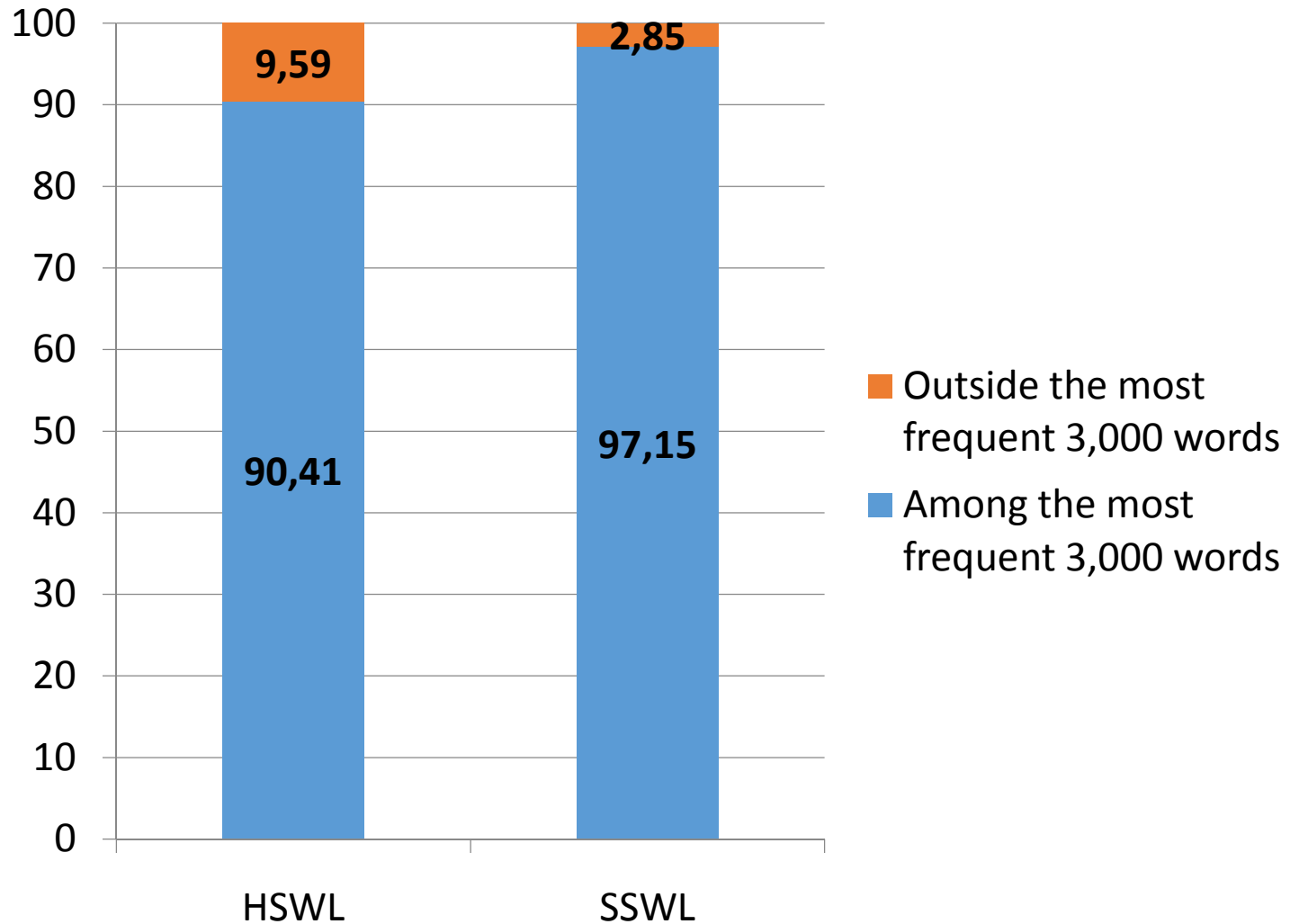
## Soft sciences

restored **view**  
**world war**  
**personal identity**  
**social contract**  
**public health**  
**parental investment**  
**demand curve**  
**language teaching**  
**death penalty**  
**population size**  
**global warming**

# Percentage of different kinds of words in the HSWL and SSWL (%)



UNIVERSITY OF LEEDS



# Nature of vocabulary in academic speech of hard and soft sciences

---



UNIVERSITY OF LEEDS

- 93.55% of the words unique to the SSWL are among the most frequent 3,000 words
- 54.58% of the words unique to the HSWL are outside the most frequent 3,000 words.

# Most frequent two-words terms in hard and soft science speeches



UNIVERSITY OF LEEDS

## Hard sciences

solar **system**  
pituitary gland  
kinetic **energy**  
**northern** hemisphere  
differential **equation**  
angular momentum  
lattice **point**  
binary **search**  
kinetic **energy**  
voltage **across**  
renal **failure**  
triple **bottom**

## Soft sciences

restored **view**  
**world war**  
**personal identity**  
**social contract**  
**public health**  
**parental investment**  
**demand curve**  
**language teaching**  
**death penalty**  
**population size**  
**global warming**

# HSWL and SSWL vs. ASWL

---



UNIVERSITY OF LEEDS

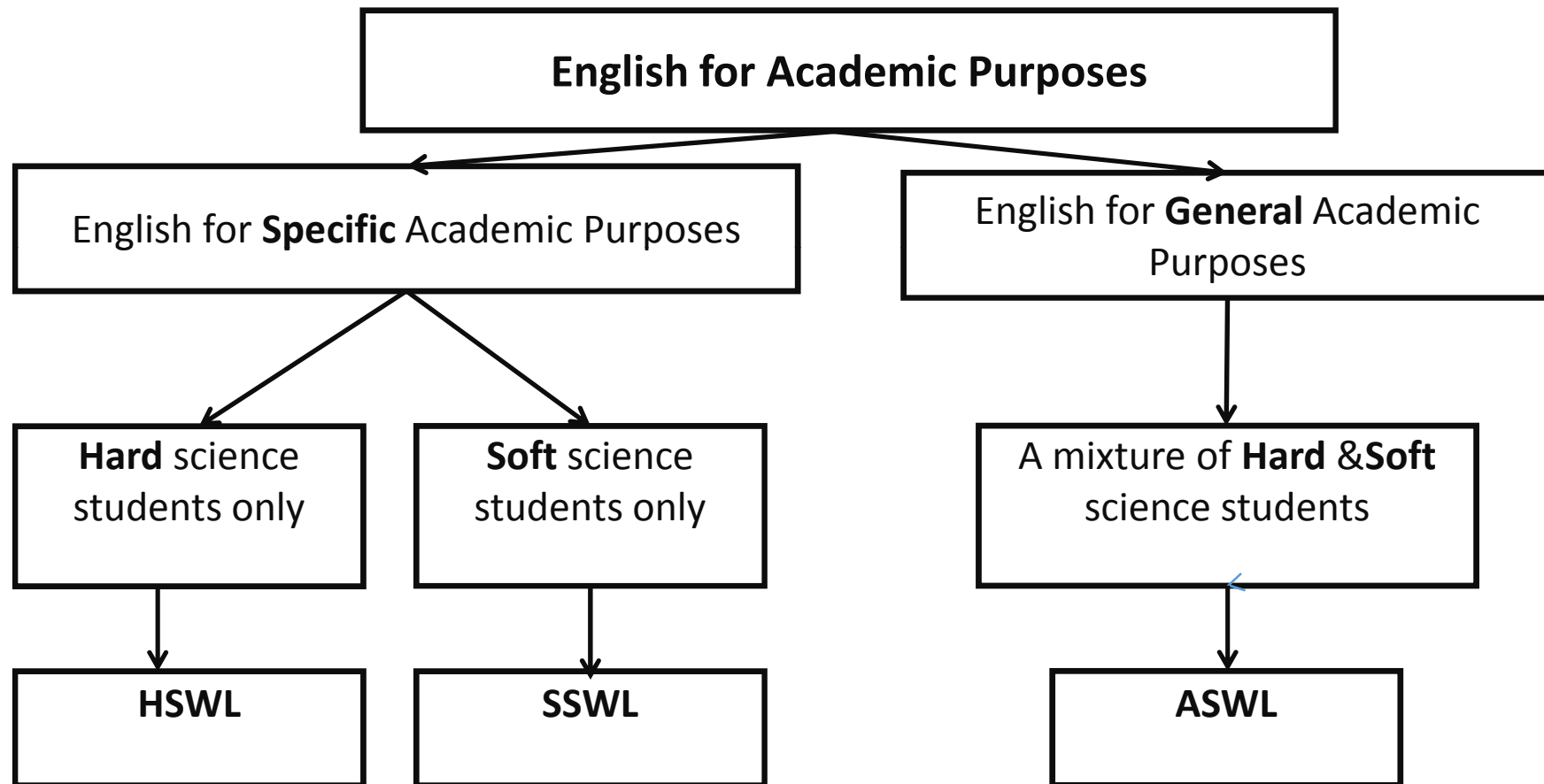
- 90.16% of the HSWL words appeared in the ASWL.
- 82.28% of the SSWL words appeared in the ASWL.



# Implications



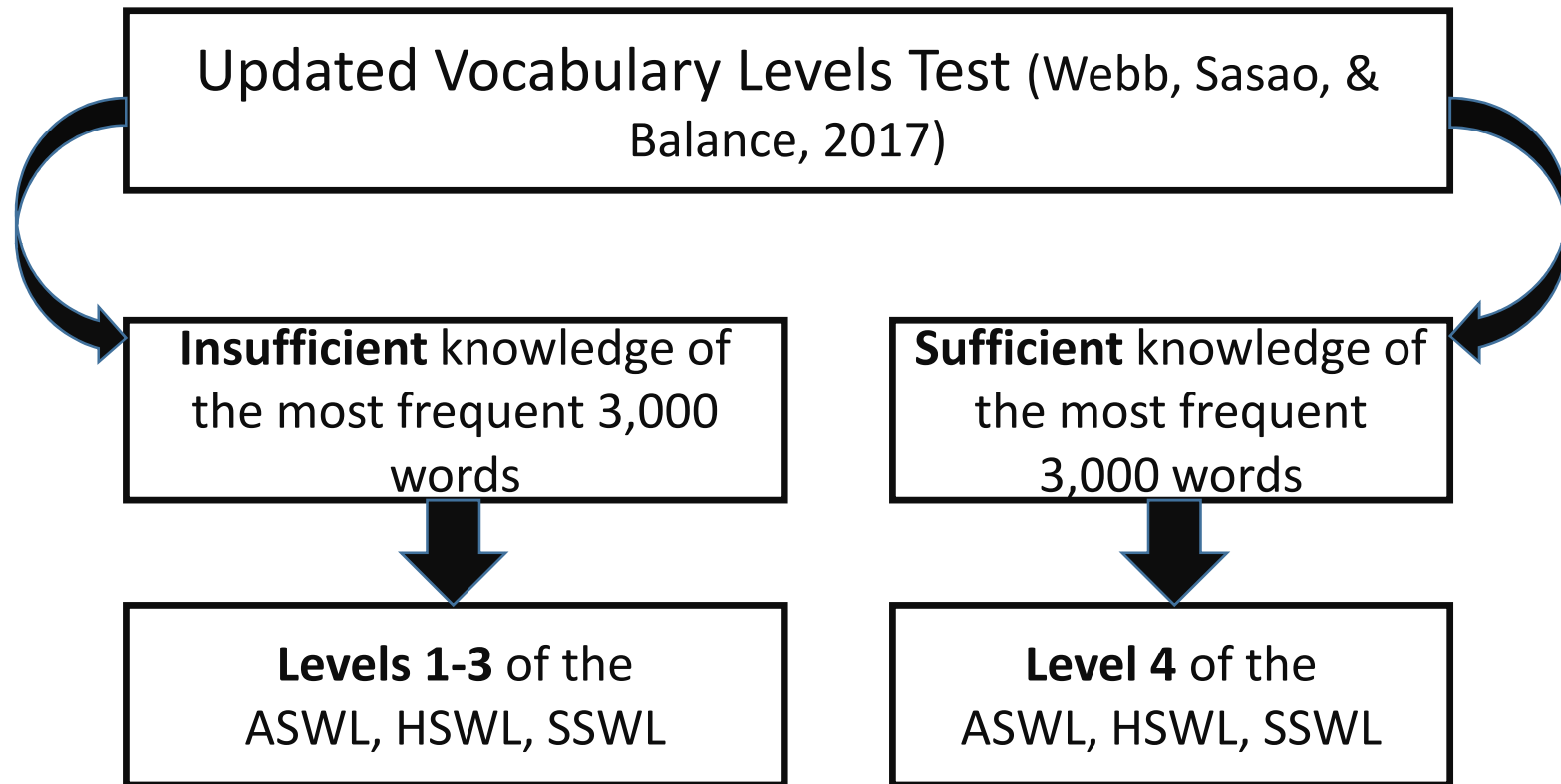
Depending on the learning and teaching contexts, teachers can select the relevant word list for their learners.



# Implications



Teachers should ensure that their learners have a good knowledge of the most frequent 3,000 words.



# Implications

---



UNIVERSITY OF LEEDS

Teachers should help learners to

- expand knowledge of these words
- reach the fluency level when using these words both receptively and productively.

# Future research



UNIVERSITY OF LEEDS

---

- Further examining multi-word units in hard- and soft-science speech
- Examining vocabulary of academic speech in a specific discipline
- Developing and trialling materials for EAP courses and tracking learning/effectiveness

# Acknowledgements



UNIVERSITY OF LEEDS

---

My sincere thanks to Cambridge University Press, Pearson, Dr. Lynn Grant, the lecturers at Victoria University of Wellington, the researchers in the International Corpus of English project, the British Academic Spoken English corpus project, the British Academic Written English corpus project, the Yale University Open courseware project, Massachusetts Institute of Technology Open courseware project, the University of California, Berkeley Open courseware project, and the Stanford University Open courseware project.



**Thank you!**