

Degree of Divergence between Spoken and Written Vocabularies as a Means for Classification of L2 Learners

Shin ISHIKAWA

School of Languages & Communication

Kobe University

1 Introduction

SV vs WV

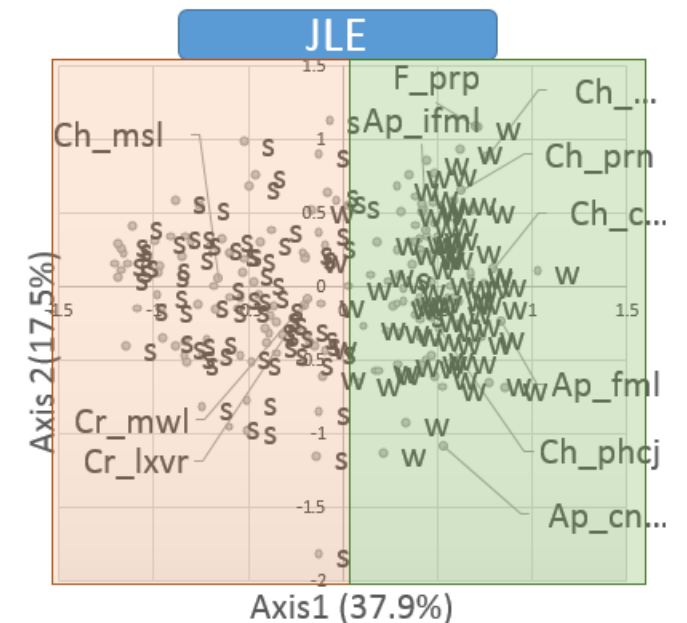
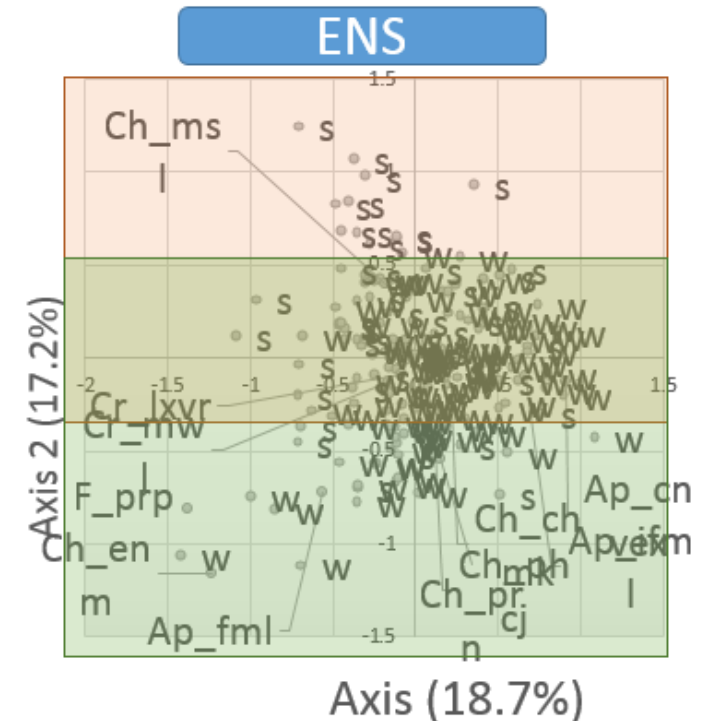
- Montgomery (2007)
- SV is relatively limited (5,000 to 10,000 words)

- Milton (1998: 25)
- Good coverage can be achieved with fewer words in speech than in writing.
- Several very useful written voc size tests may work less well if retransferred directly to an aural form

- Turk (1985)
- “S language is not W language.”
- 1) Repetition (S > W)
- 2) Choice of voc (S = familiar/ everyday W = formal and explicit)

Really different?

- Are SV and WV really different even when s/o talks or writes about the same topic in the same setting?
- Ishikawa (2015) compared the vocs in speeches and essays produced about the same topic by ENS and JLE, and it showed that SV/WV dif is observed clearly for JLE, but not for ENS.
- If so, Degree of Divergence in Spoken and Written Vocabulary (DDSWV) may function as a means to classify learners' L2 prof levs and/or varied int'l learners.



2 Research Design

Aims and RQs

- To clarify whether DDSWV can be utilized as a means to classify learners' L2 proficiency levels and/or varied int'l learners at the same proficiency levels.
- *RQ1 Does DDSWV classify L2 proficiency levels of EFL learners appropriately? (A2, B11, B12, B2+ on the CEFR-band)*
- *RQ2 Does DDSWV classify int'l learners at the same proficiency levels and native speakers appropriately? (EFL learners, ESL learners, and ENS)*

Four Kinds of Measurements of DDSWV

Lexical
Content
(VO)

Freq
Stability
(CV)

Dependency
on Freq
Words (FC)

Lexical
Variety
(HC)

Four Criteria to Measure DDSWV

- **1. Lexical Content (VO: Vocabulary Overlap)**

Examine the # of common words in the top 100 SV and the top 100 WV

- **2. Freq Stability (CV: Coefficients of Variation)**

Compare Standard Deviation/Mean in the top 100 SV in the top 100 WV

- **3. Dependency on High-freq Words (FC: Frequent Words Coverage)**

Compare the ratio of the sum of the top 100 words to the whole SV/WV

- **4. Lexical Variety (HC: Herdan's C)**

Compare LN (type)/ LN (token) in the whole SV and WV

Four Kinds of Measurements of DDSWV

Number of Common
Words
(+) Similar/ (-) Dif

S/W Ratio
Close to 1.0, Similar/ >1.0 or <1.0 Different

Lexical
Content
(VO)

Freq
Stability
(CV)

Dependency
on Freq
Words (FC)

Lexical
Variety
(HC)

Data

- In order to examine the DDSWV in a reliable way, we need the SW data highly homogeneous.
- The Spoken and Written modules of The ICNALE (International Corpus Network of Asian Learners of English) (Ishikawa, 2012-2016)

The ICNALE

- The world largest corpus of Asian L2 English learners
- Collection of speeches and essays produced by college students in ten countries and areas in Asia as well as ENS.
- Includes both of EFL learners (CHN, JPN, IDN, KOR, THA, TWN) and ESL learners (HKG, PHK, PAK, SIN)
- Designed as a database for sophisticated contrastive interlanguage analysis
- The number of the topics is 2: (Part-time job for college students; Non-smoking at the restaurants)
- Same time, same length, same settings
- Collection of learner proficiency info (A2, B11, B12, B2+)

ICNALE: The International Corpus Network of Asian Learners of English

Last Updated: 19-Dec-2014

KWIC

Collocation

Wordlist

Keywords

Setting

Sample **1-20** / Total **404**

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) ... [Next Page](#) [Last Page](#)

Results

Sorting: 1st key 2nd key 3rd key

the same boat and feel the same way that I	do	. We have a tremendous amount of reading to do	ENS1
do . We have a tremendous amount of reading to	do	just to try to keep up with our workload For	ENS1
perform at a high level . On that basis I	do	n't believe that it is important for students to have	ENS1
work part time jobs and those who are cashed up	do	n't That 's it in a nutshell really and end	ENS1
a part time job out of necessity because my parents	do	not have enough money to pay for all of my	ENS1
money situation has improved no surprises there but I also	do	n't stress out as much as some of my mates	ENS1

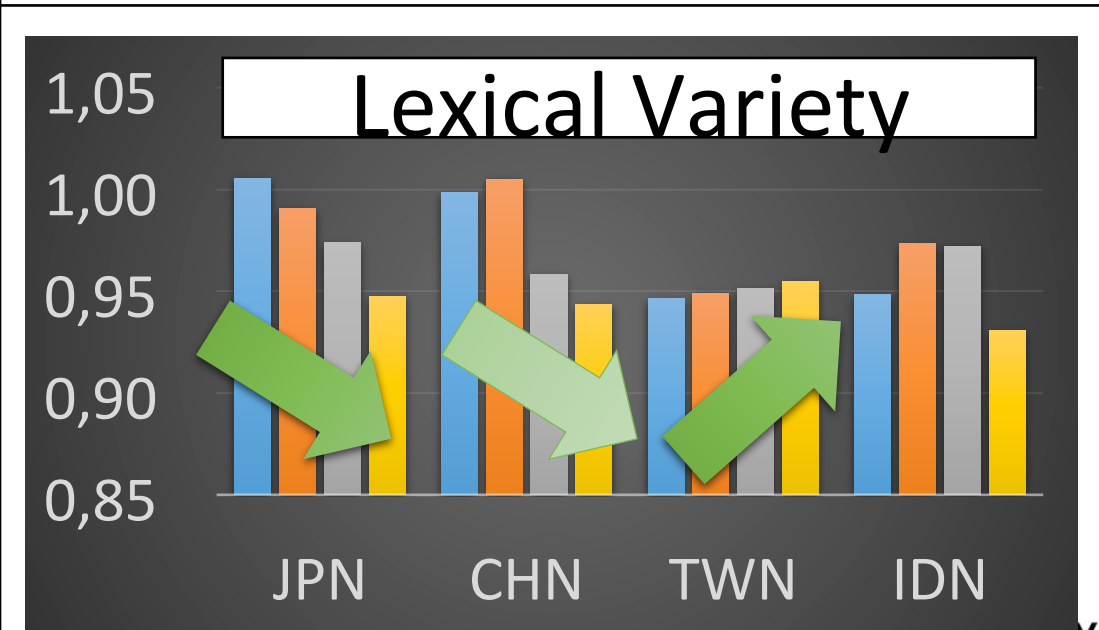
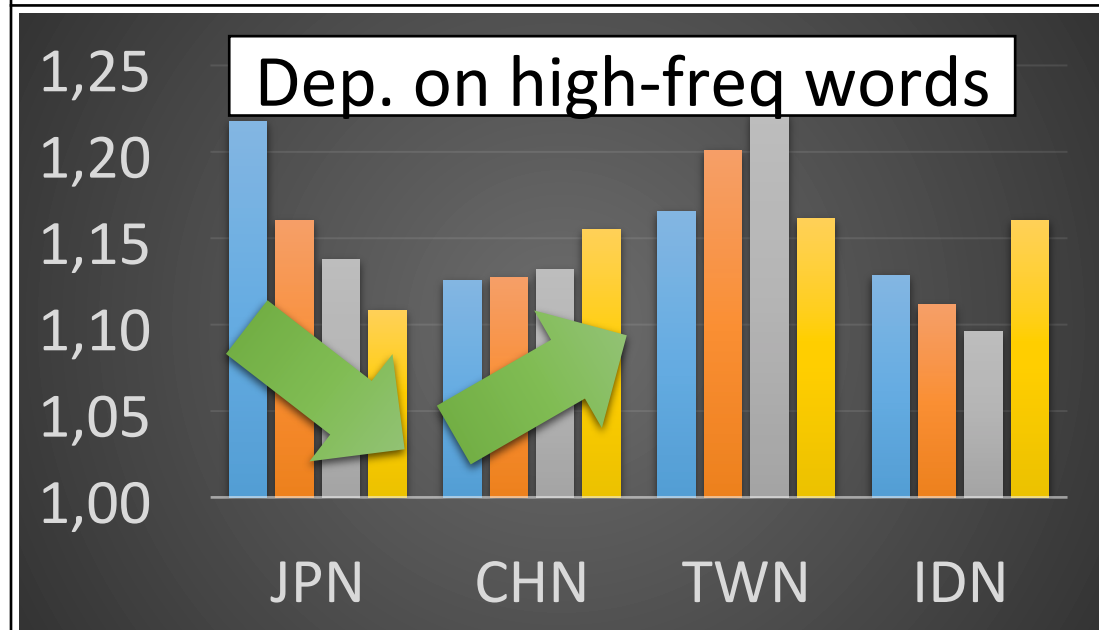
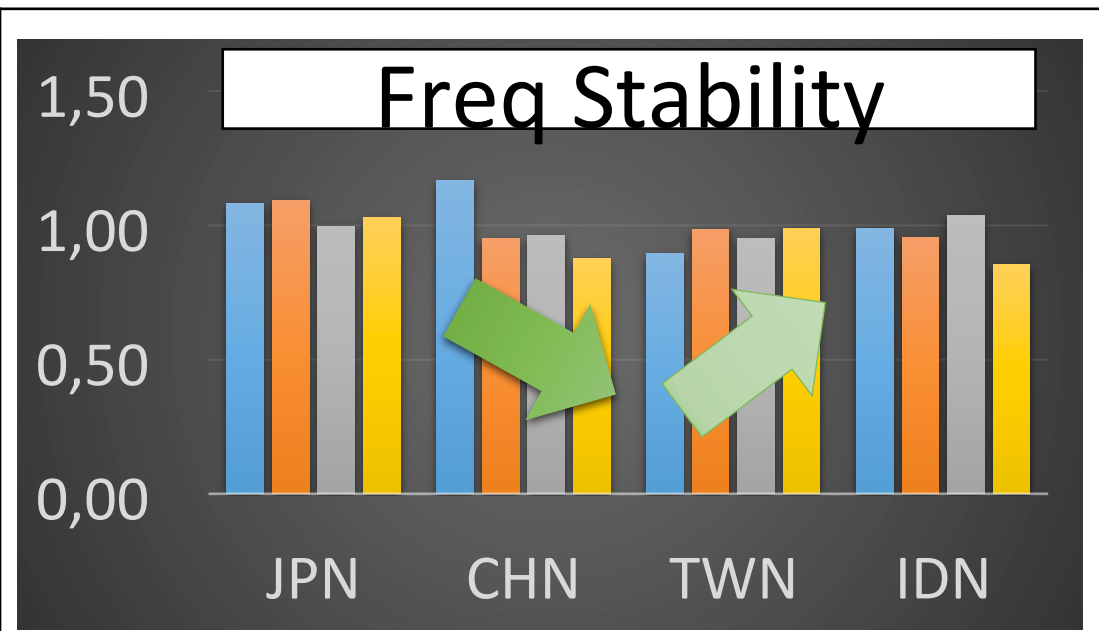
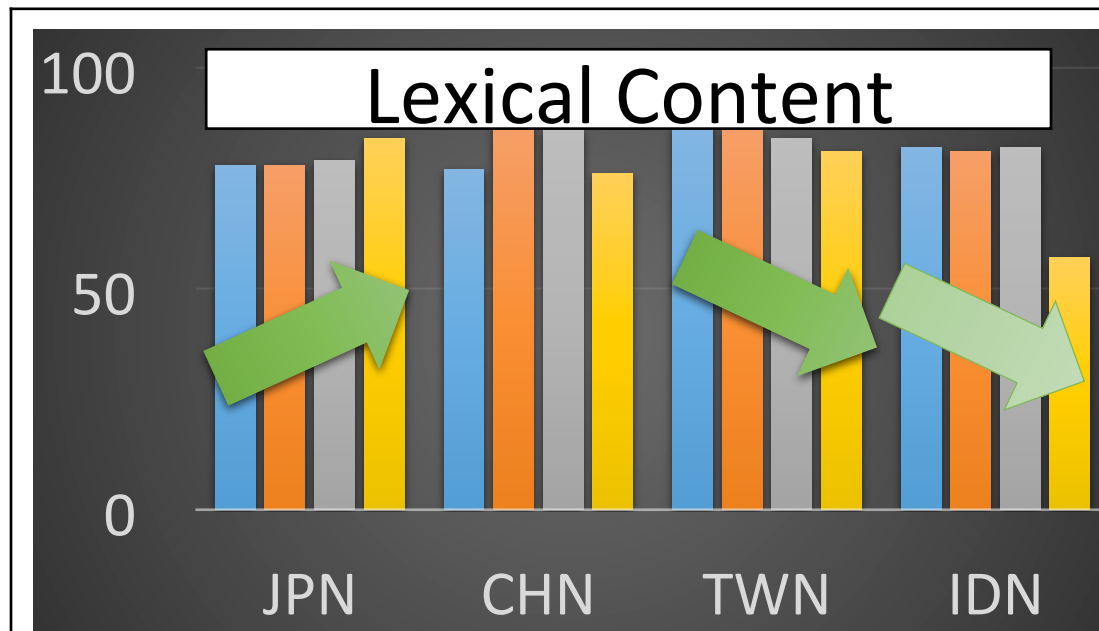


The samples analyzed in the current study

- RQ1: Learners in JPN, CHN, TWN, IDN at A2, B11, B12, B2 levls respectively
- RQ2: Learners in JPN, CHN, TWN, IDN, PHL, SIN all at B12 lev and ENS
- Analyzed 2,748 speeches (60 sec) + 3,420 essays (200-300 words) in total

3 Results & Discussions

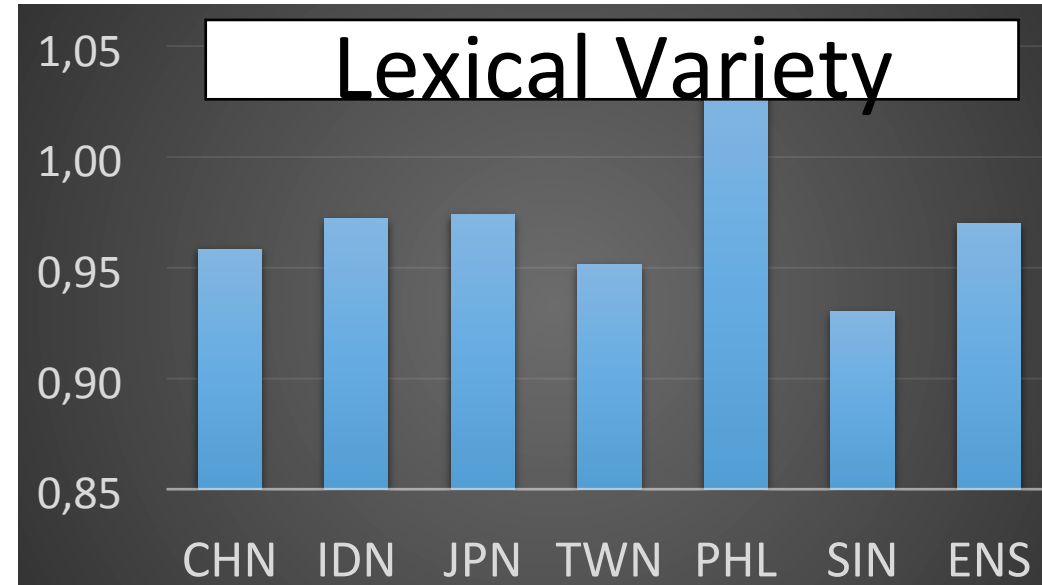
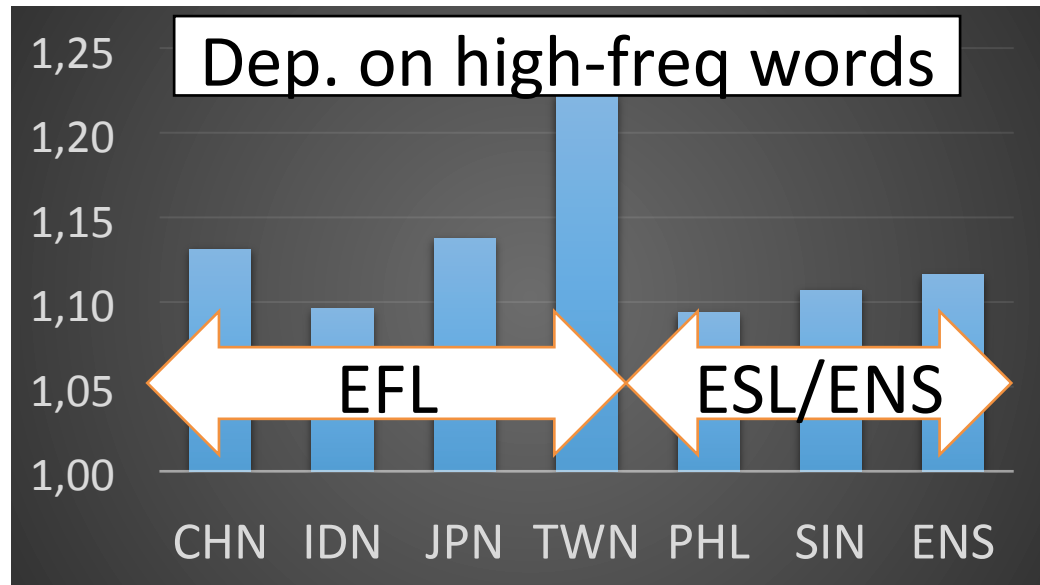
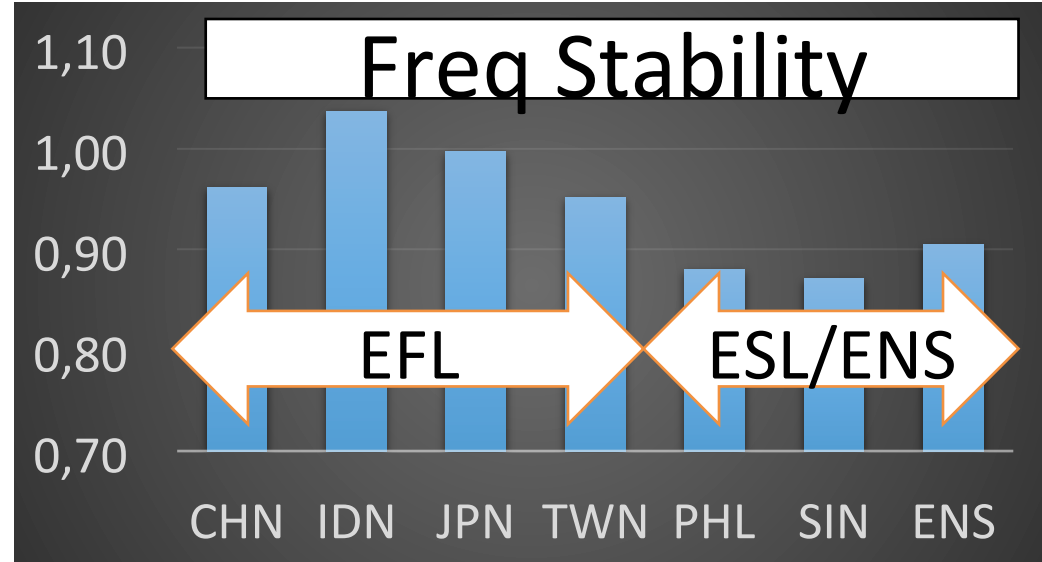
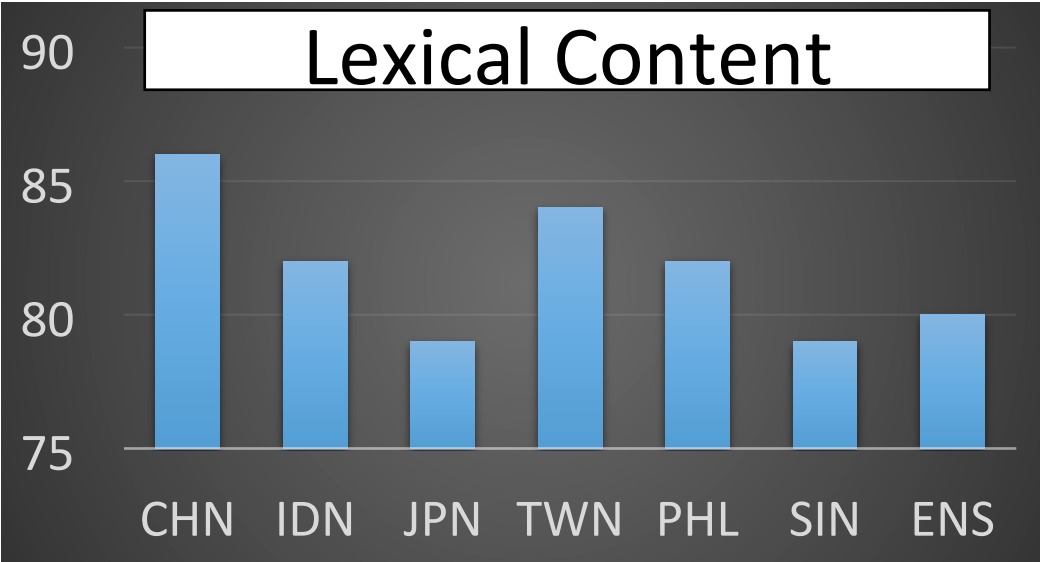
RQ1: Classification of L2 Prof. Levs



Findings

- Four types of DDSWV can function as a means to classify L2 learners' proficiency levels, but their patterns of change are not universal.
- Among four indices, Lexical Variety (3+) and Lexical Content (3) are relatively better markers.
- Different learners show different patterns of change.
- For JLE, DDSWV increases in terms of lexical content (LO), but decreases in terms of dependency on high freq words (FC) and lexical variety (HC) as the proficiency level goes up.
- For CLE, DDSWV increases in terms of dependency on high freq words (FC), but decreases in terms of freq stability (CV) and lexical variety (HC) as the proficiency level goes up.

RQ2: Classification of Int'l learners



Findings

- Some types of DDSWV, dependency on high-freq words (FC) and freq stability (CV), can function as a means to classify int'l EFL/ESL learners at the same prof levs.
- However, ESL learners and ENS cannot be classified appropriately.
- For EFL learners, DDSWV tends to be larger in terms of dependency on high freq words (FC) and freq stability (CV) than for ESL learners and ENS.

4. Summary and Implication

RQs

- *RQ1 Does DDSWV classify L2 proficiency levels of EFL learners appropriately? (A2, B11, B12, B2+)*
- Partly yes. Lexical Variety(HC) and Lexical Content (LO) are relatively better markers.
- However, these indices cannot be used universally.

- *RQ2 Does DDSWV classify int'l learners at the same prof levls and English native speakers appropriately? (EFL learners, ESL learners, and ENS)*
- Partly yes. Freq Stability (CV) and Dependency on high freq words (FC) distinguish EFL learners from ESL learners. However, ESL learners and ENS cannot be classified appropriately.

Different aims and different types of DDSWV

For classification of L2 proficiency levels

- Lexical Variety (HC)
- Lexical Content (LO)

For classification of int'l learners at the same proficiency levels

- Freq Stability (CV)
- Dependency on high freq words (FC)

Pedagogical implications

- Our corpus-based study has shown how learners' development in L2 proficiency levels influence the degree of divergence in spoken and written vocabularies. Also, it has revealed that the divergence appears differently for different learners in Asia.
- Generally speaking, learners are not necessarily conscious of the “mode” of their L2 productions. Teachers need to help learners to naturally “notice” the similarities and differences observed in speeches and writings.

Thank you.