

Verb Collocations and CEFR Criterial Features



Satoru UCHIDA
Kyushu University

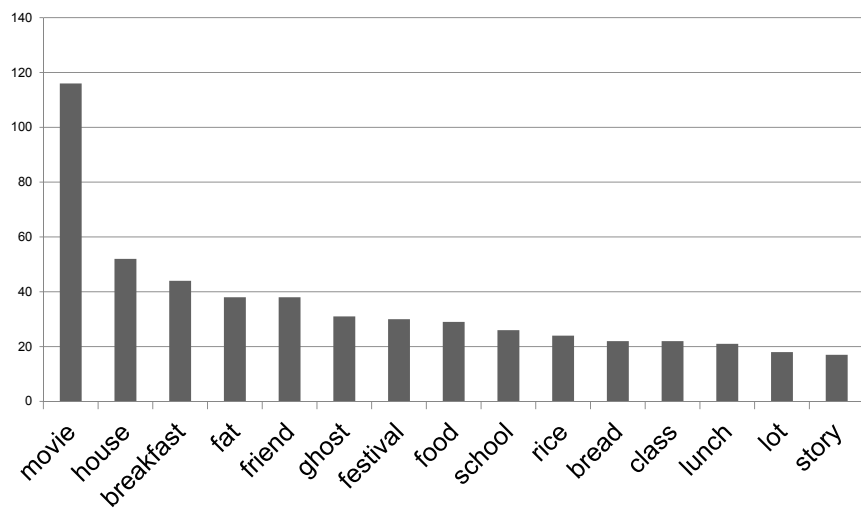
A small quiz...

Q. Can you list a few nouns that come after the verb “make”?

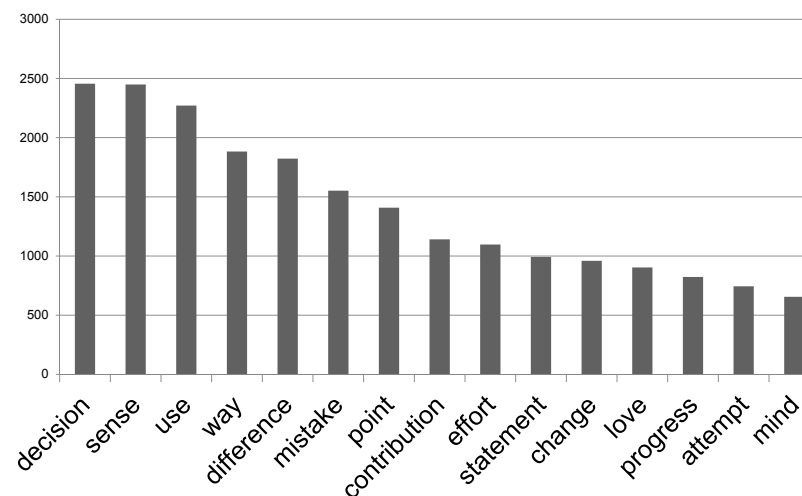
make + (a/an/the) **noun**

1. cake
2. breakfast / lunch / dinner
3. friends

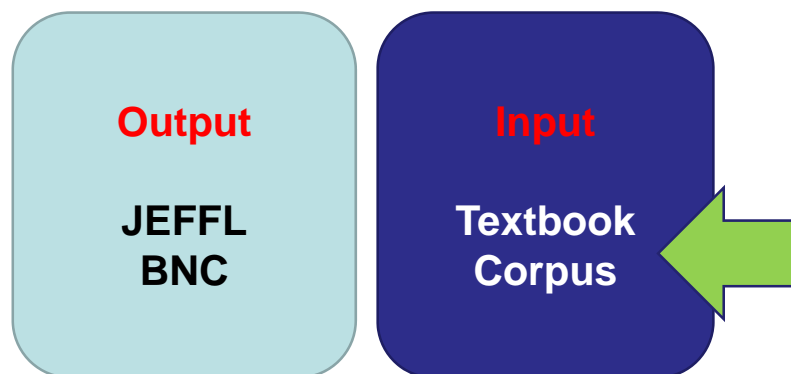
JEFFL corpus



BNC



What about learner's *input*?



4

Introduction

【Aims】

- (1) To clarify how verb collocations differ among the CEFR levels
- (2) To examine if verb collocations can be used as criterial features for the CEFR levels

5

Target of the study

- get, have, make

VERB + nouns/adjectives/verbs

[within 4 spans]

6

Overview

- Data: Textbook corpus
- Methodology: correspondence analysis
- Results of correspondence analysis: get, have, make
- Discussion: Level discrepancies
- Conclusion and Future Tasks

7

Data

- Textbook corpus

A1	A2	B1	B2	C	Total
164,585	278,750	486,787	582,763	302,149	1,815,034

- *C1 and C2 are treated as one category due to the limitation of C2 data.
- *Vocabulary sections in each textbook are not included in this study.

Target tags

```
<part="READING" page="XX">...
... <w c7="VVN" c5="VVN" hw="make"
    pos="VERB">made</w> ...
</part>
```

【node】

hw="get", hw="have", hw="make"

【collocators】

Pos="VERB", pos="SUBST" (noun), pos="ADJ"

Contingency table

- words * CEFR levels

words	A1	A2	B1	B2	C
ache(n)	0	5	0	0	0
all(a)	5	4	5	6	7
any(a)	41	25	10	8	7
bad(a)	4	4	6	2	3
be(v)	3	9	16	17	20
big(a)	6	6	6	2	3
book(n)	7	5	1	1	0
brother(n)	13	3	1	1	0
car(n)	5	9	5	4	2
card(n)	6	4	1	2	0

Methodology

- Correspondence analysis
- R ver. 3.2.2
- MASS library: corresp function

CEFR level

	A1	A2	B1	B2	C1	C2
Word1	36	37	4	3	2	4
Word2	5	7	25	34	19	10
Word3	34	44	6	22	23	41
Word4	1	23	23	6	4	17
Word5	8	7	5	2	36	33
Word6	40	16	1	2	5	4
Word7	6	7	11	23	22	43
Word8	6	17	56	42	11	14
Word9	43	35	9	19	16	17
Word10	40	43	46	22	38	9

Collocators

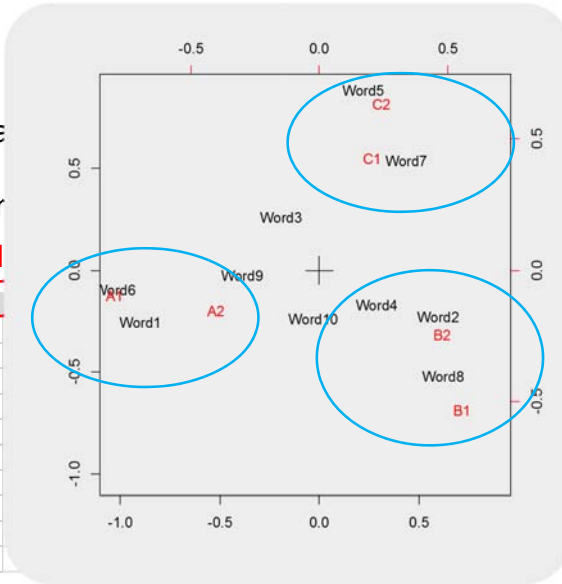
Methodology

- Correspondence analysis
- R ver. 3.2.2
- MASS library: correspondence analysis

CEFR I

	A1	A2
Word1	36	37
Word2	5	7
Word3	34	44
Word4	1	23
Word5	8	7
Word6	40	16
Word7	6	7
Word8	6	17
Word9	43	35
Word10	40	43

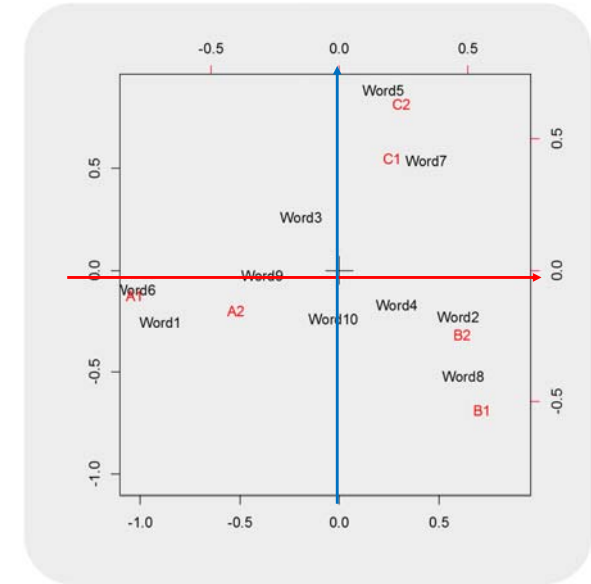
Collocators



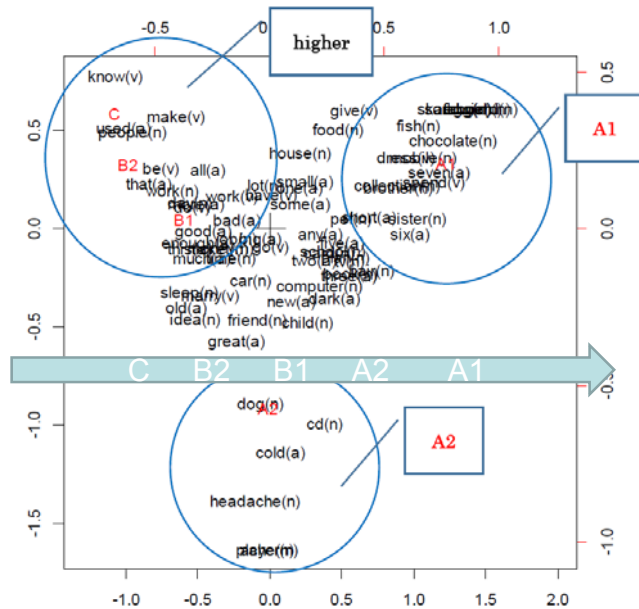
How to read the result

X axis:
First principle component

Y axis:
Second principle component



get



get

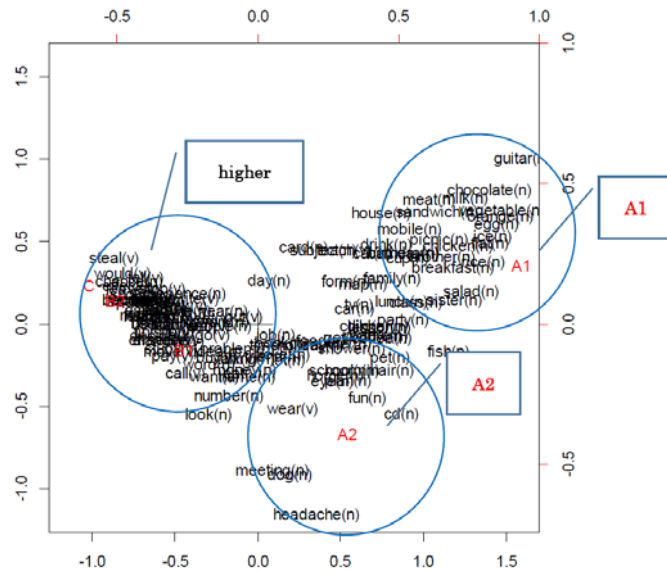
Scores of words in A1 and A2 clusters

A1	[.1]	[.2]	CEFR'
egg(n)	1.3704	0.6132	A1
flour(n)	1.3704	0.6132	A2
orange(n)	1.3704	0.6132	A1
sandwich(n)	1.3704	0.6132	A1
skateboard(n)	1.3704	0.6132	B2
chocolate(n)	1.2714	0.4529	A1
seven(a)	1.1725	0.2926	A1
spend(v)	1.1395	0.2392	A1
mobile(n)	1.0593	0.3627	A1
fish(n)	1.031	0.5279	A1
sister(n)	1.024	0.0521	A1

A2	[.1]	[.2]	CEFR
ache(n)	-0.0151	-1.6312	B1
player(n)	-0.0151	-1.6312	A1
headache(n)	-0.108	-1.3837	A1
cold(a)	0.07676	-1.1341	A1
cd(n)	0.38078	-0.9899	A1
dog(n)	-0.0654	-0.8892	A1
great(a)	-0.2335	-0.5672	A1
child(n)	0.25922	-0.4703	A1
friend(n)	-0.0877	-0.4599	A1
idea(n)	-0.52	-0.4543	A1
old(a)	-0.5805	-0.3965	A1

- Blue: sorted in descending order
- Red: sorted in ascending order

have



have

Scores of words in A1 and A2 clusters

A1	[.1]	[.2]	CEFR
guitar(n)	1.5898	1.0056	A1
vegetable(n)	1.4734	0.6934	A1
orange(n)	1.4588	0.6544	A1
egg(n)	1.4401	0.6043	A1
ice(n)	1.4151	0.5374	A1
flat(n)	1.3992	0.4948	A1
chocolate(n)	1.3951	0.8166	A1
rice(n)	1.3569	0.3813	A1
salad(n)	1.2903	0.2029	A1
milk(n)	1.2506	0.7703	A1
chicken(n)	1.2044	0.478	A1
sister(n)	1.1856	0.1468	A1

A2	[.1]	[.2]	CEFR
fish(n)	1.15305	-0.1649	A1
cd(n)	0.86213	-0.5402	A1
hair(n)	0.81203	-0.2748	A1
pet(n)	0.80028	-0.1982	A1
coffee(n)	0.74308	-0.0807	A1
lesson(n)	0.734	-0.019	A1
dinner(n)	0.7195	-0.0082	A1
baby(n)	0.71053	-0.0711	A1
picture(n)	0.71053	-0.0711	A1
fun(n)	0.66123	-0.4468	A1
child(n)	0.64405	-0.0061	A1
room(n)	0.56826	-0.2753	A1

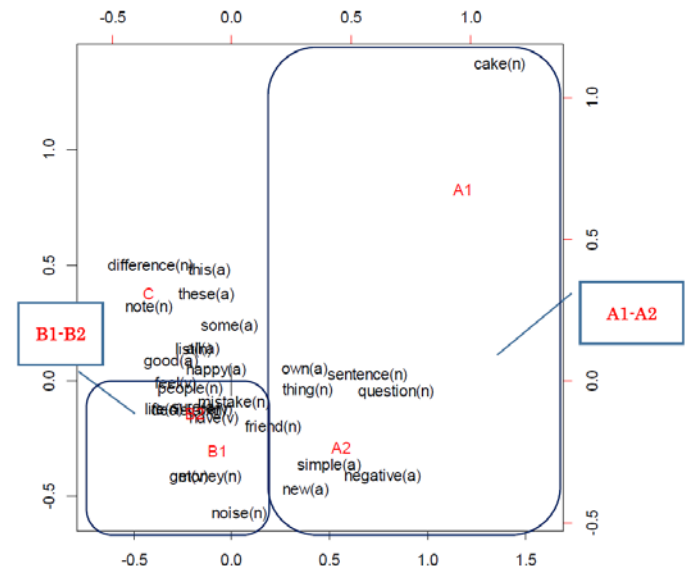
have

Scores of words in higher levels

higher	[.1]	[.2]	CEFR
steal(v)	-0.871	0.4031	A2
would(v)	-0.8253	0.30976	A1
chance(n)	-0.7826	0.2595	A2
effect(n)	-0.7727	0.23838	A2
leave(v)	-0.7566	0.21282	A1
show(v)	-0.7416	0.18359	A1
spend(v)	-0.7242	0.14033	A1
come(v)	-0.7236	0.15496	A1
happen(v)	-0.7208	0.14714	A1
become(v)	-0.7189	0.14912	A1
ask(v)	-0.6799	0.08836	A1
tell(v)	-0.6787	0.28465	A1

look(v)	-0.6688	0.06482	A1
make(v)	-0.6641	0.0597	A1
learn(v)	-0.6598	0.04904	A1
say(v)	-0.6461	0.15988	A1
try(v)	-0.6443	0.02849	A2
give(v)	-0.6388	0.1617	A1
use(v)	-0.6379	0.00513	A1
help(v)	-0.6333	0.00728	A1
take(v)	-0.6129	0.13842	A1
know(v)	-0.6095	0.13851	A1
be(v)	-0.6032	0.10487	A1
finish(v)	-0.5954	-0.0459	A1
arrive(v)	-0.5891	-0.0943	A1

make



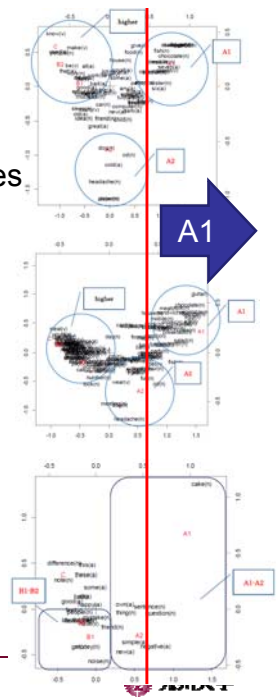
make

Scores of words in A1-A2 and B1-B2 clusters

A1-A2	[.1]	[.2]	CEFR	B1-B2	[.1]	[.2]	CEFR
cake(n)	1.3686	1.374	A1	get(v)	-0.225	-0.4113	A1
question(n)	0.8349	-0.044	A1	money(n)	-0.113	-0.41	A1
negative(a)	0.7675	-0.405	A2	have(v)	-0.0924	-0.1554	A1
sentence(n)	0.6888	0.0293	A1	decision(n)	-0.2207	-0.1236	B1
simple(a)	0.4935	-0.358	A2	be(v)	-0.0861	-0.1175	A1
thing(n)	0.3839	-0.034	A1	life(n)	-0.346	-0.1169	A1
new(a)	0.3758	-0.465	A1	sure(a)	-0.1893	-0.1104	A1
own(a)	0.3663	0.0557	A1	people(n)	-0.2142	-0.0306	A1
friend(n)	0.2068	-0.193	A1	feel(v)	-0.2876	-0.007	A1
noise(n)	0.0366	-0.568	A1				
mistake(n)	0.0062	-0.087	A2				

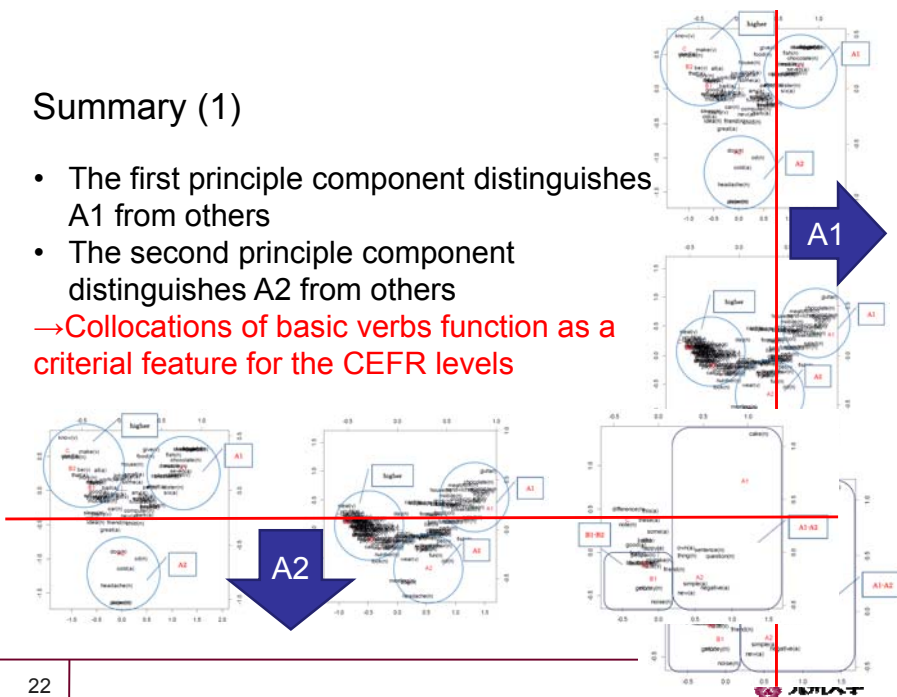
Summary (1)

- The first principle component distinguishes A1 from others
 - The second principle component distinguishes A2 from others
- Collocations of basic verbs function as a criterial feature for the CEFR levels



Summary (1)

- The first principle component distinguishes A1 from others
 - The second principle component distinguishes A2 from others
- Collocations of basic verbs function as a criterial feature for the CEFR levels



Summary (2)

get:

- A1: concrete objects (e.g. *egg, flour, orange, sandwich*)
- A2: becoming *get* (e.g. *get cold, get old, get dark*)
- Higher: some idiomatic expressions (e.g. *get to know, get used to*)

have:

- A1 and A2: concrete objects (e.g. [possession] *guitar, sister* [eating] *vegetable, orange*)
- Higher: perfective (e.g. *have stolen, have left, have shown*)

make:

- A1 and A2: creative *make* (e.g. *cake, sentence, question*)
- B1 and B2: earning *make* (*make money*), doing *make* (*make decision*), causative (*make X feel*)

Discussion: Level discrepancies

have (A1) + chance (A2) ... higher than B

have (A1) + effect (A2) ... higher than B

→implies a need for considering “combinations of words”

Scores of words in higher levels

higher	[1]	[2]	CEFR
steal(v)	-0.871	0.4031	A2
would(v)	-0.8253	0.30976	A1
chance(n)	-0.7826	0.2593	A2
effect(n)	-0.7727	0.23838	A2
leave(v)	-0.7566	0.21282	A1
show(v)	-0.7416	0.18359	A1
spend(v)	-0.7242	0.14033	A1
come(v)	-0.7236	0.15496	A1
happen(v)	-0.7208	0.14714	A1
become(v)	-0.7189	0.14912	A1
ask(v)	-0.6799	0.08836	A1
tell(v)	-0.6787	0.28465	A1

look(v)	-0.6688	0.06482	A1
make(v)	-0.6641	0.0597	A1
learn(v)	-0.6598	0.04904	A1
say(v)	-0.6461	0.15988	A1
try(v)	-0.6443	0.02849	A2
give(v)	-0.6388	0.1617	A1
use(v)	-0.6379	0.00513	A1
help(v)	-0.6333	0.00728	A1
take(v)	-0.6129	0.13842	A1
know(v)	-0.6095	0.13851	A1
be(v)	-0.6032	0.10487	A1
finish(v)	-0.5954	-0.0459	A1
arrive(v)	-0.5891	-0.0943	A1

Conclusion and Implication

- Verb collocations statistically distinguish CEFR-levels. Especially, A1 and A2 are distinguished statistically from higher levels
- The nouns that collocates with the target basic verbs denote concrete objects
- Collocations should be taken into account when discriminating CEFR levels

→”B level barrier” (80% of JE belong to A level)

→From Concrete (A level) to Abstract (B level)

Future tasks

A sample list of collocations for each CEFR level

make v. (A1)

+ NOUNS

A1: cake, dinner, breakfast ...

A2: question, noise, mistake ...

B1: decision

B2: ...

Thank you for your attention!

- Satoru UCHIDA