**On-line Supplementary Material**

Table S1: Characteristics of the NSs and NNSs (standard deviation in brackets)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Group |  | Gender |  |  |  |  | Exposure (weekly hours)  |  | Self-reported Proficiency |
|  |  | Age now | Age when startedlearning English |  | Reading | Watching | Listening | Communication | Average |  | Speaking | Reading | Writing | Listening | Average |
|  |  | M | F |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| NS(*n* = 27) |  | 6 | 21 |  | 43.52 | 0.00 |  | 18.67 | 13.41 | 3.81 | 16.37 | 13.06 |  | 5.00 | 5.00 | 5.00 | 5.00 | 5.00 |
|  |  |  |  |  | (13.45) | (0.00) |  | (13.96) | (10.78) | (5.45) | (18.25) | (7.98) |  | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| NNS(*n* = 55) |  | 15 | 40 |  | 20.77 | 7.87 |  | 5.40 | 11.91 | 9.59 | 10.62 | 9.38 |  | 4.13 | 4.31 | 3.91 | 4.27 | 4.15 |
|  |  |  |  |  | (1.69) | (4.57) |  | (6.65) | (13.16) | (12.03) | (10.8) | (7.6) |  | (0.75) | (0.66) | (0.84) | (0.8) | (0.57) |

Table S2. Stratified sampling at the three frequency bands (percentage of total number for each part of speech is presented in parentheses)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Total number of lexical lemmas a | Adjective  | Adverb  | Noun  | Verb |
| 1K | 834.00 | 110.00 | 96.00 | 410.00 | 218.00 |
|  |  | (13.2) | (11.5) | (49.2) | (26.1) |
| 2K | 965.00 | 173.00 | 70.00 | 522.00 | 200.00 |
|  |  | (17.9) | (7.6) | (54.1) | (20.7) |
| 3K | 970.00 | 184.00 | 61.00 | 527.00 | 198.00 |
|  |  | (19.0) | (6.3) | (54.3) | (20.4) |
| Total  | 2769.00 | 467.00 | 227.00 | 1459.00 | 616.000 |
|  |  | (16.9) | (8.2) | (52.7) | (22.3) |
| Proposed frequency inthe CPWT (out of 20 for each level)  |  | 3.37 | 1.64 | 10.54 | 4.45 |

a As the focus of the present study is on lexical lemmas, the total count here excludes grammatical lemmas.

Table S3. Frequencies and percentages of correct/incorrect answers in the *CPWT*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | NSs (*n =* 27) |  | NNSs (*n =* 55) |
|  |  | Correct | % | Incorrect | % |  | Correct | % | Incorrect | % |
| 1K Level (*k =* 16) |  | 410.00 | 95.0 | 22.00 | 5.1 |  | 585.00 | 66.5 | 295.00 | 33.5 |
| 2K Level (*k =* 16) |  | 406.00 | 94.0 | 26.00 | 6.0 |  | 471.00 | 53.5 | 409.00 | 46.5 |
| 3K Level (*k =* 15) |  | 376.00 | 92.8 | 29.00 | 7.2 |  | 330.00 | 40.0 | 495.00 | 60.0 |

Table S4. Summary of continuous variables

|  |  |  |
| --- | --- | --- |
| Variable | Range (adjusted range) | Mdn |
| Average weekly exposure to English  | 0.50 - 36.25 (-1.52 - 3.19) hours | 0.03 |
| Age when started to learn English | 0.00 - 19.00 (-0.72 - 3.03) years | -0.72 |
| Node lemma length  | 4.00 - 14.00 (-1.31- 2.75) characters | -0.09 |
| Collocate lemma length  | 2.00 - 15.00 (-1.85 - 3.78) characters | -0.12 |
| Collocate lemma frequency (log)  | 291 - 12314090 (-2.81 - 2.36 log) points | 0.13 |
| Total CPWT score | 0.00 - 47.00 (-3.07 - 0.85) points | 0.42 |

Note. The second column shows the range of the variables. The adjusted range after transformation and/or centering, is presented in parentheses. Medians refer to the predictor values in the models. All variables are centred, and their means are zero.

Table S5. The best-fit model for variables that predict appropriacy in the *CPCT* (*N* = 12491, *log* *likelihood* = -7699.6)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Fixed effects | *β* | *SE* | *z value* | *p* | *Exp (β)* | ***≈ d*** |
| (Intercept) | 0.39 | 0.14 | 2.72 | .007\*\* | 1.47 | 0.21 |
| CPWT score | 0.44 | 0.05 | 8.18 | < .001\*\*\* | 1.55 | 0.24 |
| Frequency Level: 2K | -0.27 | 0.19 | -1.44 | 0.15 | 0.77 | -0.15 |
| Frequency Level: 3K | -0.75 | 0.19 | -3.98 | < .001\*\*\* | 0.47 | -0.41 |
| Random effects | Variance | SD |  |  |  |  |  |
| Subject | 0.16 | 0.40 |  |  |  |  |  |
| Subject| Frequency Level: 2K | 0.05 | 0.23 |  |  |  |  |  |
| Subject| Frequency Level: 3K | 0.09 | 0.30 |  |  |  |  |  |
| Item | 0.28 | 0.52 |  |  |  |  |  |
| Item|Group: Nonnative | 0.23 | 0.48 |  |  |  |  |  |

Table S6. The best-fit model for variables that predict COCA frequency of appropriate responses in the *CPCT* *(N =6156, R2marginal = .38,R2conditional = .52)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Fixed effects |  |  |  |  |  |  |
|  | *Estimate* | *SE* | *t value* | *Pr(>|t|)* | *≈ d* |
| (Intercept) | 5.88 | 0.11 | 55.25 | < .001\*\*\* | 5.74 |
| Node lemma length | -0.18 | 0.06 | -2.77 | .008\*\* | -0.17 |
| Collocate lemma length | -0.10 | 0.02 | -6.76 | < .001\*\*\* | -0.10 |
| Log collocate lemma frequency | 0.86 | 0.02 | 46.72 | < .001\*\*\* | 0.84 |
| CPWT score | 0.10 | 0.02 | 4.61 | < .001\*\*\* | 0.10 |
| Frequency Level: 2K | -0.44 | 0.15 | -3.00 | .004\*\* | -0.42 |
| Frequency Level: 3K | -0.65 | 0.15 | -4.20 | < .001\*\*\* | -0.63 |
| Random effects | *Variance* | *SD* |  |  |  |  |
| Residual | 0.81 | 0.90 |  |  |  |  |
| Subject | 0.02 | 0.13 |  |  |  |  |
| Item | 0.18 | 0.43 |  |  |  |  |
| Item|Group: Nonnative | 0.05 | 0.21 |  |  |  |  |

Table S7. The best-fit model for variables that predict strength of the association (MI) of appropriate responses in the *CPCT (N = 6156, R2marginal = .50,R2conditional = .62)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Estimate* | *SE* | *t value* | *Pr(>|t|)* | ***≈ d*** |
| (Intercept) | 4.65 | 0.18 | 25.21 | < .001\*\*\* | 3.06 |
| Log collocate lemma frequency | -1.73 | 0.03 | -65.11 | < .001\*\*\* | -1.14 |
| Configuration: VN | -0.67 | 0.18 | -3.63 | < .001\*\*\* | -0.44 |
| Collocate lemma length | -0.15 | 0.02 | -6.82 | < .001\*\*\* | -0.10 |
| CPWT score | 0.14 | 0.03 | 4.56 | < .001\*\*\* | 0.09 |
| Frequency Level: 2K | 0.25 | 0.22 | 1.13 | 0.26 | 0.17 |
| Frequency Level: 3K | 0.87 | 0.22 | 3.85 | < .001\*\*\* | 0.57 |
| Random effects: | *Variance* | *SD* |  |  |  |  |
| Residual | 0.04 | 0.19 |  |  |  |  |
| Subject | 0.48 | 0.70 |  |  |  |  |
| Item | 0.10 | 0.31 |  |  |  |  |
| Item|Group: Nonnative | 1.70 | 1.30 |  |  |  |  |

**Appendix S1: Target items for the CPWT**

| **Frequency** **Level**  | **Lemma**  | **PoS** | **Length** |
| --- | --- | --- | --- |
| 1K | nice | adjective | 4 |
| 1K | social | adjective | 6 |
| 1K | important | adjective | 9 |
| 1K | recently | adverb | 8 |
| 1K | truth | noun | 5 |
| 1K | steps | noun | 5 |
| 1K | day | noun | 3 |
| 1K | memory | noun | 6 |
| 1K | evidence | noun | 8 |
| 1K | interest | noun | 8 |
| 1K | pressure | noun | 8 |
| 1K | interview | noun | 9 |
| 1K | series | noun | 6 |
| 1K | position | noun | 8 |
| 1K | chairs | noun | 6 |
| 1K | forget | verb | 6 |
| 1K | wait | verb | 4 |
| 1K | walk | verb | 4 |
| 1K | play | verb | 4 |
| 1K | realized | verb | 9 |
|  |  |  |  |
| 2K  | tired | adjective | 5 |
| 2K | tough | adjective | 5 |
| 2K | expensive | adjective | 9 |
| 2K | obviously | adverb | 9 |
| 2K | customer | noun | 8 |
| 2K | era | noun | 3 |
| 2K | violence | noun | 8 |
| 2K | influence | noun | 9 |
| 2K | perception | noun | 10 |
| 2K | content | noun | 7 |
| 2K | regulations | noun | 10 |
| 2K | task | noun | 4 |
| 2K | bridge | noun | 6 |
| 2K | advice | noun | 6 |
| 2K | conflict | noun | 8 |
| 2K | participate | verb | 11 |
| 2K | commit | verb | 6 |
| 2K | expanding | verb | 8 |
| 2K | blame | verb | 5 |
| 2K | achieve | verb | 7 |
|  |  |  |  |
| 3K | immediate | adjective | 9 |
| 3K | innocent | adjective | 8 |
| 3K | practical | adjective | 8 |
| 3K | rapidly | adverb | 6 |
| 3K | emotions | noun | 8 |
| 3K | violation | noun | 9 |
| 3K | bench | noun | 5 |
| 3K | talent | noun | 6 |
| 3K | mixture | noun | 7 |
| 3K | volunteer | noun | 9 |
| 3K | stability | noun | 9 |
| 3K | advocate | noun | 8 |
| 3K | electricity | noun | 11 |
| 3K | assistance  | noun | 10 |
| 3K | consultant | noun | 10 |
| 3K | urge | verb | 4 |
| 3K | negotiate | verb | 9 |
| 3K | detect | verb | 8 |
| 3K | justify | verb | 7 |
| 3K | impose | verb | 6 |

**Appendix S2: The CPWT with expected answers**

**Instructions**

This test is designed to measure your productive knowledge of the most 3000 frequent words in English. The first letter of each target word and the number of missing letters are provided to help you identify the appropriate response. Please stick to the first letter and the number of missing letters, as indicated by the spaces provided.

**Here is an example:**

You need to finish this q \_ \_ \_ \_ \_ \_. We don’t have much time left.

*The target word is* ***quickly****.*

Please write the missing **target word** in each sentence in the assigned column.

**Section I:** 1,000 Word Frequency Level

|  |  |  |
| --- | --- | --- |
|  | **Context** | **Target word** |
| 1 | The weather is n\_ \_ \_ this week. We may go out. | nice |
| 2 | Please don't f\_ \_ \_ \_ \_ our meeting this afternoon. | forget |
| 3 | She found out the t\_ \_ \_ \_ about him. | truth |
| 4 | She felt afraid to enter, so she took a few s\_ \_ \_ \_ back. | steps |
| 5 | The nurse asked me to w\_ \_ \_ a few minutes before I could see the doctor. | wait |
| 6 | There have been many changes r\_ \_ \_ \_ \_ \_ \_. | recently |
| 7 | The two brothers w\_ \_ \_ to school every morning. | walk |
| 8 | It's a beautiful d\_ \_. We might take the kids out. | day |
| 9 | Being too busy, I have almost no s\_ \_ \_ \_ \_ life. | social |
| 10 | He has a great m\_ \_ \_ \_ \_ . He almost remembered every word the teacher said. | memory |
| 11 | The children p\_ \_ \_ football in the field near their houses. | play |
| 12 | The judge asked for further e\_ \_ \_ \_ \_ \_ \_ before sending the man to prison. | evidence |
| 13 | It is i\_ \_ \_ \_ \_ \_ \_ \_ to prepare for your test to get a good mark. | important |
| 14 | They showed an i\_ \_ \_ \_ \_ \_ \_ in the art subject. | interest |
| 15 | He can work under p\_ \_ \_ \_ \_ \_ \_. | pressure |
| 16 | I am pleased to tell you that you have been chosen for the i\_ \_ \_ \_ \_ \_ \_ \_. | interview |
| 17 | The kids loved the ‘Harry Potter’ book s\_ \_ \_ \_ \_. | series |
| 18 | I r\_ \_ \_ \_ \_ \_ \_ that something was wrong. | realised |
| 19 | I will try to get this p\_ \_ \_ \_ \_ \_ \_. I actually need a job.  | position |
| 20 | We need more c\_ \_ \_ \_ \_ in the classroom. | chairs |

**Section II:** 2,000 Word Frequency Level

|  |  |  |
| --- | --- | --- |
|  | **Context** | **Target word** |
| 1 | He is a good c\_ \_ \_ \_ \_ \_ \_. He always buys from our shop. | customer |
| 2 | I want to p\_ \_ \_ \_ \_ \_ \_ \_ \_ \_ in this training. It will be useful for my job. | participate |
| 3 | I wish I was born in another e\_ \_. Living in this age does not suit me. | era |
| 4 | I cannot join the afternoon's class since I'm too t\_ \_ \_ \_. | tired |
| 5 | I cannot c\_ \_ \_ \_ \_ to the job yet. I need to know how much you will pay me first. | commit |
| 6 | V\_ \_ \_ \_ \_ \_ \_ is everywhere in this city. There are so many fights and killings in the streets. | violence |
| 7 | TV programs can have a bad i\_ \_ \_ \_ \_ \_ \_ \_ on children. Parents need to be careful about that. | influence |
| 8 | The workers are e\_ \_ \_ \_ \_ \_ \_ \_ the road. It is good to have wider roads. | expanding |
| 9 | There is a general p\_ \_ \_ \_ \_ \_ \_ \_ \_ that the company is losing power in the market | perception |
| 10 | The book I read was very enjoyable. I liked its c**\_ \_ \_ \_ \_ \_**. | content |
| 11 | He is o\_ \_ \_ \_ \_ \_ \_ \_ a rich man. His taste in clothes and shoes clearly show that. | obviously |
| 12 | We should follow the school r\_ \_ \_ \_ \_ \_ \_ \_ \_ \_ or we can get in trouble. | regulations |
| 13 | Fighting fire is a very hard t\_ \_ \_. To do this job well, you need a lot of training. | task |
| 14 | I cannot b\_ \_ \_ \_ her for breaking the glass. She did not mean to break it. | blame |
| 15 | They liked the idea of building a new b\_ \_ \_ \_ \_ across the river. | bridge |
| 16 | My father gave me some good a\_ \_ \_ \_ \_ on how to study for tests. | advice |
| 17 | He had a t\_ \_ \_ \_ life that was full of problems and hardships. | tough |
| 18 | He works so hard to a\_ \_ \_ \_ \_ \_ \_ his dream. | achieve |
| 19 | We need to end the c\_ \_ \_ \_ \_ \_ \_ between these countries. It can become another war.  | conflict |
| 20 | This shirt is too e\_ \_ \_ \_ \_ \_ \_ \_. I don't have enough money to buy it. | expensive |

**Section III:** 3,000 Word Frequency Level

|  |  |  |
| --- | --- | --- |
|  | **Context** | **Target word** |
| 1 | This is a burning issue which needs an i\_ \_ \_ \_ \_ \_ \_ \_ action.  | immediate |
| 2 | I u\_ \_ \_ you to study harder. Your test marks have not been very good. | urge |
| 3 | Sandra couldn’t control her e\_ \_ \_ \_ \_ \_ \_ when she knew that she lost the game. | emotions |
| 4 | We must report what he has done to the police. This is a v\_ \_ \_ \_ \_ \_ \_ \_ of the law. | violation |
| 5 | Things are changing very r\_ \_ \_ \_ \_ \_. This is too fast for me. | rapidly |
| 6 | Let's sit on this b\_ \_ \_ \_. It looks comfortable. | bench |
| 7 | She is a great singer with so much t\_ \_ \_ \_ \_. | talent |
| 8 | She has a natural ability to n\_ \_ \_ \_ \_ \_ \_ \_ prices with sellers.  | negotiate |
| 9 | The city is a m\_ \_ \_ \_ \_ \_ of old and new buildings. I like how they stand side by side. | mixture |
| 10 | The students were asked to leave the classrooms after smoke was d\_ \_ \_ \_ \_ \_ \_ in the school.  | detected |
| 11 | She is a v\_ \_ \_ \_ \_ \_ \_ \_ in this hospital. So, she won't get paid for her help. | volunteer |
| 12 | After many wars, the country needs some s\_ \_ \_ \_ \_ \_ \_. We cannot have more changes at this time. | stability |
| 13 | I actually cannot understand why you did this. Can you please j\_ \_ \_ \_ \_ \_ your action?  | justify |
| 14 | He is an a\_ \_ \_ \_ \_ \_ \_ of women education. He always supports this cause. | advocate |
| 15 | This small town has no e\_ \_ \_ \_ \_ \_ \_ \_ \_ \_. People depend only on sunlight. | electricity |
| 16 | She always tries to i\_ \_ \_ \_ \_ her own views on the rest of the team in a way that is not acceptable.  | impose |
| 17 | Nurses offer great a\_ \_ \_ \_ \_ \_ \_ \_ \_ to doctors. Everyone knows how helpful nurses are. | assistance  |
| 18 | She is such an i\_ \_ \_ \_ \_ \_ \_ girl. I couldn’t believe she was involved in the crime. | innocent |
| 19 | He was happy to work as a c\_ \_ \_ \_ \_ \_ \_ \_ \_ to the company, giving his views on serious issues. | consultant |
| 20 | She has lots of lovely ideas, but they're not very p\_ \_ \_ \_ \_ \_ \_ \_ for our business. | practical |

***End of the test!***

*Thank you****!***

**Appendix S3: Target Items for the CPCT**

| **Frequency** **Level**  | **Target Noun\*** | **Length** |
| --- | --- | --- |
| 1K | image | 5 |
| 1K | tree | 4 |
| 1K | chance | 6 |
| 1K | benefit | 7 |
| 1K | strategy | 8 |
| 1K | name | 4 |
| 1K | treatment | 9 |
| 1K | economy | 7 |
| 1K | baby | 4 |
| 1K | skills | 6 |
|  |  |  |
| 2K | threat | 6 |
| 2K | strength | 8 |
| 2K | discussions | 10 |
| 2K | opinion | 7 |
| 2K | difficulties | 10 |
| 2K | conversation | 11 |
| 2K | notion | 6 |
| 2K | title | 5 |
| 2K | mistake | 7 |
| 2K | smile | 5 |
|  |  |  |
| 3K | knife | 5 |
| 3K | payment | 7 |
| 3K | habit | 5 |
| 3K | implication | 11 |
| 3K | criteria | 8 |
| 3K | alternative | 11 |
| 3K | salary | 6 |
| 3K | recommendation | 14 |
| 3K | reputation | 10 |
| 3K | request | 8 |

*\* Each target noun was presented twice to elicit two collocation types: MN/ VN*

**Appendix S4: The CPCT with suggested answers**

**Productive Collocation Test**

**Instructions:**

This test aims at assessing your productive knowledge of collocations (pairs of words that usually go together, e.g., ‘heavy rain’, ‘make a mistake’). You will be presented with a sentence with a gap. Your task is to read the sentence carefully and to provide as many words as possible (adjectives in the first section and SINGLE-WORD verbs in the second section) that often go with the underlined noun. Please separate your responses with a comma.

**Here are two examples:**

*Example 1*: Adjective + noun collocations:

He was the \_\_\_\_\_\_\_\_\_\_\_\_\_ **manager** in the company.

**Possible answers**: *top, general, assistant, senior, regional*

*Example 2*: Verb + noun collocations

He \_\_\_\_\_\_\_\_\_ a **meeting** last week.

**Possible** **answers**: *held, attended, arranged, scheduled, organized*

**Section 1:** Modifier + Noun Collocations

1,000 Frequency Level

|  |  |  |
| --- | --- | --- |
| # | **Sentence**  | **Possible modifiers**  |
|  | Her \_\_\_\_\_\_\_\_\_\_\_\_ **image** has recently changed.  | positive, popular, public |
|  | There was a/an \_\_\_\_\_\_\_\_\_\_\_\_ **tree** next to the house. | big, tall, small, dead, large, green |
|  | This is a/an \_\_\_\_\_\_\_\_\_\_\_\_ **chance**. | good, real, great, fair, excellent |
|  | This is the \_\_\_\_\_\_\_\_\_\_\_\_ **benefit** of the new system. | great, added, additional,  |
|  | They have a/an \_\_\_\_\_\_\_\_\_\_\_\_ **strategy**. | different, grand, new |
|  | I don't know her \_\_\_\_\_\_\_\_\_\_\_\_ **name**. | real, big, full, common |
|  | The doctor gave her \_\_\_\_\_\_\_\_\_\_\_\_ **treatment**. | special, standard |
|  | The \_\_\_\_\_\_\_\_\_\_\_\_ **economy** needs a push. | local, national, weak |
|  | Look at that \_\_\_\_\_\_\_\_\_\_\_\_ **baby**. | little, beautiful, poor, dead, tiny |
|  | \_\_\_\_\_\_\_\_\_\_\_\_ **skills** are very important. | social, basic,  |

|  |  |  |
| --- | --- | --- |
| # | **Sentence**  | **Possible modifiers**  |
|  | This was a/an \_\_\_\_\_\_\_\_\_\_\_\_ **threat** to the business. | serious, real, biggest |
|  | They used their \_\_\_\_\_\_\_\_\_\_\_\_ **strength**. | military, economic, full, great, |
|  | There has been a/an\_\_\_\_\_\_\_\_\_\_\_\_ **discussion** of the issue. | public, serious, online, open, detailed, brief, general, further |
|  | She has a/an \_\_\_\_\_\_\_\_\_\_\_\_ **opinion** about this situation. | strong, different, personal,  |
|  | She may have \_\_\_\_\_\_\_\_\_\_\_\_ **difficulties**. | serious, great, major, practical |
|  | They are having a/an \_\_\_\_\_\_\_\_\_\_\_\_ **conversation** on the matter. | private, casual, brief, serious, |
|  | The writer took this \_\_\_\_\_\_\_\_\_\_\_\_ **notion** to a new level. | traditional, general,  |
|  | What is the \_\_\_\_\_\_\_\_\_\_\_\_ **title**? | official, original, working, full, |
|  | This was a/an \_\_\_\_\_\_\_\_\_\_\_\_ **mistake**. | big, terrible, huge, serious, major |
|  | She has such a/an \_\_\_\_\_\_\_\_\_\_\_\_ **smile**. | big, broad, wide, warm, bright |

2,000 Frequency Level

3,000 Frequency Level

|  |  |  |
| --- | --- | --- |
| # | **Sentence**  | **Possible modifiers**  |
|  | There was a/an \_\_\_\_\_\_\_\_\_\_\_\_ **knife** in the kitchen. | sharp, dull, long, large, bloody, hot |
|  | This is the \_\_\_\_\_\_\_\_\_\_\_\_ **payment**. | direct, monthly, late, final, |
|  | They have a/an \_\_\_\_\_\_\_\_\_\_\_\_ **habit**. | bad, old, healthy, nasty, personal, annoying, |
|  | This is the most \_\_\_\_\_\_\_\_\_\_\_\_ **implication** of the findings. | important, practical,  |
|  | The manager had \_\_\_\_\_\_\_\_\_\_\_\_ **criteria** in mind. | specific, certain, important, different, |
|  | I think this is a/an \_\_\_\_\_\_\_\_\_\_\_\_ **alternative**. | good, possible, attractive, real,  |
|  | The \_\_\_\_\_\_\_\_\_\_\_\_ **salary** is not bad. | average, monthly,  |
|  | This is a/an \_\_\_\_\_\_\_\_\_\_\_\_ **recommendation**. | strong, specific, general,  |
|  | She is a scientist with a/an \_\_\_\_\_\_\_\_\_\_\_\_ **reputation**. | good, bad, national, great, professional, solid |
|  | This is a/an \_\_\_\_\_\_\_\_\_\_\_\_ **request**. | special, written, specific, reasonable, ill, simple, unusual |

**Section 2:** Verb + Noun Collocations

1,000 Frequency Level

|  |  |  |
| --- | --- | --- |
| # | **Sentence**  | **Possible verbs**  |
|  | They worked hard to \_\_\_\_\_\_\_\_\_\_\_\_ the **image**. | create, improve, produce |
|  | Farmers \_\_\_\_\_\_\_\_\_\_\_\_ several **trees**. | planted, grew, cut, climbed |
|  | They \_\_\_\_\_\_\_\_\_\_\_\_ a **chance** to win. | had, got, took, stood |
|  | They started to\_\_\_\_\_\_\_\_\_\_\_\_ the **benefits** of the system. | receive, enjoy  |
|  | The manager \_\_\_\_\_\_\_\_\_\_\_\_ several **strategies**. | applied, developed, used, adopted |
|  | She could not \_\_\_\_\_\_\_\_\_\_\_\_ his **name**. | change, remember, write, call, mention |
|  | Ali \_\_\_\_\_\_\_\_\_\_\_\_ **treatment** at several hospitals. | received, sought, |
|  | The new programs \_\_\_\_\_\_\_\_\_\_\_\_ the **economy**. | grew, hurt, improved, fixed, strengthened |
|  | The mother is \_\_\_\_\_\_\_\_\_\_\_\_ a **baby**. | having, holding, delivering, feeding, expecting,  |
|  | She needs to \_\_\_\_\_\_\_\_\_\_\_\_ several **skills**. | develop, learn, improve,  |

|  |  |  |
| --- | --- | --- |
| # | **Sentence**  | **Possible verbs**  |
|  | They have \_\_\_\_\_\_\_\_\_\_\_\_ several **threats** recently. | received, addressed, reduced, faced, |
|  | They worked hard to \_\_\_\_\_\_\_\_\_\_\_\_ their **strength**. | build, gain, increase |
|  | Today's show will \_\_\_\_\_\_\_\_\_\_\_\_ the **discussion**. | continue, join, encourage,  |
|  | They have \_\_\_\_\_\_\_\_\_\_\_\_ their **opinion** recently. | written, expressed, voiced, changed, shared, offered |
|  | They haven't \_\_\_\_\_\_\_\_\_\_\_\_ any **difficulty**. | experienced, faced, had, reported, |
|  | She has \_\_\_\_\_\_\_\_\_\_\_ the **conversation** with them. | joined, continued, started, recorded, |
|  | They \_\_\_\_\_\_\_\_\_\_\_\_ this **notion**. | supported, accepted, challenged,  |
|  | He has \_\_\_\_\_\_\_\_\_\_\_\_ the **title** of the youngest artist. | won, held, earned, |
|  | He \_\_\_\_\_\_\_\_\_\_\_\_ his **mistake**. | admitted, corrected, made, realized, repeated |
|  | She \_\_\_\_\_\_\_\_\_\_\_\_ her **smile**. | managed, hid, forced, flashed, |

2,000 Frequency Level

3,000 Frequency Level

|  |  |  |
| --- | --- | --- |
| # | **Sentence**  | **Possible verbs**  |
|  | Mom \_\_\_\_\_\_\_\_\_\_\_\_ a **knife** to cut the chicken. | held, used, grabbed, pulled, carried, drew, threw |
|  | The bank \_\_\_\_\_\_\_\_\_\_\_\_ several **payments** this month. | made, received, accepted, missed, reduced |
|  | How can I \_\_\_\_\_\_\_\_\_\_\_\_ that **habit**? | change, break, kick, develop,  |
|  | We need to \_\_\_\_\_\_\_\_\_\_\_\_ the **implications** of this finding. | understand, discuss, consider |
|  | They \_\_\_\_\_\_\_\_\_\_\_\_ several **criteria**. | met, developed, fit, satisfied, established, used |
|  | The manager asked us to \_\_\_\_\_\_\_\_\_\_\_\_ **alternatives**. | find, offer, provide, consider, develop |
|  | The company did not \_\_\_\_\_\_\_\_\_\_\_\_ his **salary** for months. | pay, raise, cut |
|  | The doctor \_\_\_\_\_\_\_\_\_\_\_\_ several **recommendations**. | made, provided, offered,  |
|  | They \_\_\_\_\_\_\_\_\_\_\_\_ their **reputation** over the years. | earned, built, developed, gained, established, ruined, damaged, protected |
|  | They \_\_\_\_\_\_\_\_\_\_\_\_ a **request** for more information. | received, refused, denied, considered, sent |

***End of test***

*Thank you very much for your help!*

**Appendix S5: The three steps for scoring the CPCT**

*Step 1 (Initial screening):* This step was only intended for initial screening of the responses based on word-level properties (i.e., accurate PoS and recognizable spelling). The same research assistant who scored the CPWT coded the provided responses as 'accurate' or 'inaccurate'. Similar to the CPWT, minor mistakes in grammar and spelling were ignored. A response that was recognized as a modifier/verb form (depending on the test section) was coded as 'accurate'. A response of the wrong PoS and/or one that is unrecognizable was scored as 'inaccurate'. Additionally, all gaps that were not filled were coded as 'inaccurate'. Possible responses in the MN section included adjectives, nouns, or determiners. For example, *amazing* (adjective), *first (determiner), dream* (noun adjunct) were all coded as 'accurate' responses for the noun node 'chance' because they were all modifiers. Additionally, phrasal verbs and prepositional verb constructions in the VN section (*jump at* – *chance*) of the test were coded as 'accurate' as several native speakers provided them as responses.

An 'inaccurate' code was given to adjectival responses like *spical, unexpective* as the spelling was not recognizable. On the other hand, responses like *marvelous, fabulous, and excellent* (MN) and *take, achieve, and illustrate* (VN) were all coded as 'accurate' as these are recognizable modifies/verbs, respectively (despite wrong spelling for words like *marvelous*). The same coding procedures was applied by the second research assistant (same as the one who scored the CPWT) for 30% of the data points. Inter-rater reliability was high (ICC =1.00,95%, (*CI*) [.99, 1.00]), so only codes provided by the first rater were considered in the analysis. It should be noted that these accuracy codes for the CPCT data were only meant for initial screening to specify which responses are to be evaluated for Step 2 and Step 3. Thus, these scores were not subject to any further statistical analysis.

*Step 2 (Data recording):* Only responses that were coded as 'accurate' in Step 1 were admitted to the second step. The research assistant, who was trained on using the COCA (Davies, 2008-) interface, looked up each provided pair individually (e.g., MN: *marvellous - chance, fabulous - chance, excellent - chance*; VN: *take - chance, achieve - chance, illustrate - chance*) and recorded: (1) the frequency of the collocating lemma, (2) the frequency of the noun lemma, and (3) the frequency of the pair (in a window of -1 for MN collocations and -2 for VN collocations). MI values were calculated based on these values. For this search procedure, we did not set up any specified thresholds for collocation frequency or MI. Thus, all responses that passed the initial screening (Step 1) had their data entered into the specified data points. When a search resulted in no matches (e.g., *fabulous chance*), the frequency of the pair was recorded as (0) with a missing MI value.

*Step 3 (*Appropriacy *classification):*In the final step, each provided 'accurate' response (see Step 1 above) was classified dichotomously as 'appropriate' (1) or 'inappropriate' (0) based on specified thresholds for COCA frequency and MI. An MI threshold of 1 was set for an 'appropriate' collocation based on Evert's (2008) ranking approach to operationalizing collocations. As for frequency, we opted for a more lenient threshold than that employed by Nguyen and Webb (2017, 50 occurrences) to allow a range of values for the frequency analysis (Research Question 2). Thus, we employed a collocation frequency threshold of 30. It is worthy of notice that our approach is purely corpus-based and does not involve any phraseological judgement (Howarth, 1996).

Based on Step 3, a pair like *take a chance* was coded as an 'appropriate' collocation with a frequency of 7144 and an MI score of 3.55, but a pair like *achieve a chance* was deemed 'inappropriate' with a frequency of 3 and an MI score of -3.11. It should be noted that all responses that were initially coded as 'inaccurate' in Step 1 due to unrecognizable spelling, wrong PoS, or missing answer were automatically coded as 'inappropriate' in Step 3.

**References**

Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (eds.), *Corpus linguistics. An international handbook* (extended manuscript of Chapter 58, <https://lexically.net/downloads/corpus_linguistics/Evert2008.pdf>). Mouton de Gruyter.

Howarth, P. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making*. Max Niemeyer.

Nguyen, T.M.H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research, 21*(3), 298–320.