Collocation: which words go together

VOCABULARY PRACTICE
ADVANCED | C1

Name: _____ Date: _____

1 Match each advanced term with its definition.

1 corpus linguistics ____	**a** the study of language through large, systematically compiled collections of authentic text
2 n-gram ____	**b** a contiguous sequence of a given number of words extracted from a text for statistical analysis
3 mutual information score ____	**c** a statistical measure of how strongly two words are associated compared with chance co-occurrence
4 collostructional analysis ____	**d** a corpus-based method examining the statistical association between a grammatical construction and specific words
5 formulaic language ____	**e** language produced from memorised multi-word chunks rather than constructed word by word
6 idiom principle ____	**f** the theory that much language is generated by choosing from a large stock of semi-fixed phrases
7 open-choice principle ____	**g** the theory that language is generated by freely selecting individual words according to grammatical rules
8 prefabricated pattern ____	**h** a ready-made multi-word sequence retrieved as a single unit during production
9 collocational priming ____	**i** the psycholinguistic process by which repeated exposure to word combinations shapes automatic expectations
10 native-like selection ____	**j** the ability to choose the word combination a native speaker would naturally use, from among grammatically equivalent options

2 Choose the most precise term for each context.

1 Studying language through large text collections is corpus linguistics / individual introspection.

2 A contiguous sequence of words for statistical analysis is an n-gram / a single word.

3 A statistical measure of word association strength is mutual information score / a subjective impression.

Collocation: which words go together

4 Examining statistical association between a construction and words is collostructional analysis / general grammar teaching.

5 Language produced from memorised multi-word chunks is formulaic language / word-by-word construction.

6 The theory that language draws on semi-fixed phrases is the idiom principle / the open-choice principle.

7 The theory that language is built freely word by word is the open-choice principle / the idiom principle.

8 Choosing the combination a native speaker would use is native-like selection / a grammatically valid but unnatural choice.

3 Match the verbs with the noun phrases to form academic collocations.

1 compile ____	**a** an n-gram from raw text
2 extract ____	**b** collocational strength
3 quantify ____	**c** a prefabricated pattern
4 identify ____	**d** lexical retrieval
5 prime ____	**e** formulaic production
6 automatise ____	**f** a representative corpus
7 model ____	**g** native-like selection
8 disambiguate ____	**h** a polysemous collocate

Collocation: which words go together

VOCABULARY PRACTICE
ADVANCED | C1

4 Complete the analysis with terms from the box.

collocational priming | collostructional analysis | corpus linguistics | formulaic language | idiom principle | mutual information score | n-gram | native-like selection | open-choice principle | prefabricated pattern

Advances in 1 _____ have transformed how researchers study collocation, allowing systematic extraction of every 2 _____ from vast digital text collections rather than relying on introspective judgement alone. A 3 _____ quantifies how strongly two words associate compared with chance, helping distinguish genuine collocation from coincidental co-occurrence. More sophisticated 4 _____ examines statistical association between specific grammatical constructions and the words that populate them, revealing patterns invisible to unaided intuition. This research broadly supports the 5 _____, the theory that much language production draws on stored, semi-fixed 6 _____ rather than the 7 _____, which assumes words are freely selected and combined according purely to grammatical rules. Psycholinguistic research on 8 _____ suggests that repeated exposure to particular word combinations creates automatic expectations, explaining why proficient speakers achieve 9 _____ almost effortlessly, while learners must painstakingly acquire the same patterns through extensive exposure to 10 _____.

5 Read the text and answer the questions briefly.

Rethinking how language is produced

The idiom principle, proposed as an alternative to the traditional open-choice principle, suggests that fluent language production relies far more heavily on retrieving prefabricated multi-word chunks from memory than on assembling sentences word by word according to grammatical rules alone. Corpus evidence increasingly supports this view, revealing that natural language is saturated with formulaic sequences far more extensively than earlier grammatical theories acknowledged, suggesting that fluent production depends substantially on a speaker's stock of memorised chunks rather than purely generative grammatical competence. This reframing has significant implications for language teaching, since it suggests that explicit attention to collocation and formulaic sequences deserves a more central role in instruction than traditional approaches focused primarily on individual vocabulary items and abstract grammatical rules have typically afforded it.

1 What does the idiom principle suggest about fluent language production? _____

2 What does corpus evidence increasingly support? _____

Collocation: which words go together

3 What does natural language turn out to be saturated with? _____

4 What does fluent production depend substantially on, according to this view? _____

5 What significant implications does this reframing have? _____

6 What have traditional approaches typically focused on instead? _____

6 Replace the words in bold with a precise term or collocation.

1 The researchers measured how strongly two words associate compared to chance. _____

2 The multi-word sequence was retrieved as a single memorised unit. _____

3 The theory holds that language draws on stored, semi-fixed phrases. _____

4 She chose the word combination a native speaker would naturally use. _____

OVER TO YOU

Write a 160-190 word critical commentary on how corpus linguistics has reshaped understanding of collocation and formulaic language, distinguishing observation from theoretical claim.

[ ] Use at least twelve target expressions. [ ] Include four academic collocations. [ ] Maintain a precise, balanced register.

Collocation: which words go together

VOCABULARY PRACTICE
ADVANCED | C1

ANSWER KEY

1 Advanced terms

1 a 2 b 3 c 4 d 5 e 6 f 7 g 8 h 9 i 10 j

2 Precise terms

1 corpus linguistics 2 an n-gram 3 mutual information score 4 collostructional analysis 5 formulaic language 6 the idiom principle 7 the open-choice principle 8 native-like selection

3 Academic collocations

1 f compile a representative corpus 2 a extract an n-gram from raw text 3 b quantify collocational strength 4 c identify a prefabricated pattern
5 d prime lexical retrieval 6 e automatise formulaic production 7 g model native-like selection 8 h disambiguate a polysemous collocate

4 Analysis

1 corpus linguistics 2 n-gram 3 mutual information score 4 collostructional analysis
5 idiom principle 6 prefabricated pattern 7 open-choice principle 8 collocational priming
9 native-like selection 10 formulaic language

5 Reading - suggested answers

1 That it relies far more heavily on retrieving prefabricated multi-word chunks from memory than on assembling sentences word by word.
2 This view (the idiom principle).
3 Formulaic sequences.
4 A speaker's stock of memorised chunks rather than purely generative grammatical competence.
5 Implications for language teaching, suggesting explicit attention to collocation deserves a more central role.
6 Individual vocabulary items and abstract grammatical rules.

6 Precise terms and collocations

1 a mutual information score 2 a prefabricated pattern 3 the idiom principle 4 native-like selection

Over to you

Answers will vary. Look for conceptual precision, evidence-aware reasoning and balanced evaluation.