

The Language of Academic Prose

First draft of Chapter 2

About Vocabulary

James Thomas

March 2013

1. Introduction	50
How many words do native speakers know?	50
How many words are needed for academic prose?	51
2. Meaning	52
Labelling	52
Synonyms	55
Polysemy and Meaning Potential	56
3. Words in context	58
Key words	58
Topic Trails	60
4. Groups of words	61
Lexical sets	62
Academic Word List	62
Phrasal Expressions List	64
Academic Formulas List	65
5. The structure of vocabulary: morphology	66
Conversion and Morphology = Word Families	67
Conversion and pronunciation	70
Singular and Plural	71
6. The grammar of vocabulary: colligation	73
Introduction to Structures, Frames and Patterns	73
Frames and patterns	74
Collocation	76
Cognitive Profiles	80
Colligation	81
Collocation within colligation	86
7. Vocabulary study	87
Learning a word	88
Mindmap	91
8. Pronunciation	92

1. Introduction

This chapter is the first of three dealing specifically with academic vocabulary in the Message, and the systematic use of that vocabulary. General vocabulary concepts are outlined as are matters of studying *per se*. We delve further into some of the lists of academic vocabulary which have been produced by various researchers for students, teachers and authors.

As we saw in the Introduction, there are many words and phrases that are used in academic prose and many more that are not. Those that are, are often used in standard English, but sometimes in different ways, e.g. *hold, obtain*. Once this chapter has covered a range of general concepts concerning vocabulary, the following chapters build on this presenting academic vocabulary and its use in context. Chapter 3 deals with nouns and things that do “nouny things”. Chapter 4 does likewise with verbs. And in Chapter 5 we study other elements that typify the Message units in academic prose. Then Chapter 6 moves on to focus on the words and phrases that express the Organisational language, (O) such as discourse markers. From then on the book moves on to other aspects of writing.

The major aim of this book is to improve the accuracy, idiomaticity and sophistication of your academic prose. As we know from the Introduction, being a professional in anything requires a deep understanding of the tools of the trade. Our main tool is language. If you sometimes feel that there is too great an emphasis on linguistics or metalinguistic knowledge, do understand that the associated illustrative texts, sentences and lists contain and exemplify the vocabulary that will help you to move beyond being an apprentice writer.

How many words do native speakers know?

There are two constructs to define in this question. What is a word? What does it mean to know a word?

The issues surrounding the definition of a word complicate estimating both the number of words in a language and in an individual’s vocabulary size. Aitchison reports: An educated adult speaker of English can understand, and potentially use, at least 50,000 words, with *word* provisionally defined as “dictionary entry” ... This guesstimate is based on informal tests with British English university students. (2012: 7-8).

Another research project into vocabulary size, *Test Your Vocab*, presents somewhat different and valuable findings. Hundreds of thousands of people have self-assessed their vocabulary size using an online Yes-No test format: native-speakers come in at much less than 50,000. Follow the link below which show and discuss the data related to both native and non-native speakers of English. It is not uninteresting.

Task 1: Take the test at <http://testyourvocab.com/>. What is your score and how does it compare with other people in a similar situation to you?

ELF: Test Your Vocab provides data of native and non-native speaker results which are linked in the Forum, You and Test your vocab.

Task 2: Does having a foreign language vocabulary larger than most non-native speakers, but still much smaller than native speakers, confer any great advantage?

English has a huge vocabulary but we do not use very much of it. The authors of the online study found that the British dictionary they use contains around 70,000 headwords (and many more derived forms), of which only about 45,000 occur in the 100-million-word BNC.

The most frequent objects of confer: benefit, rights, right, degree, authority, status, advantage, title.

Task 3: Why are there 25,000 words in that dictionary which are not in the BNC? What does this tell us?

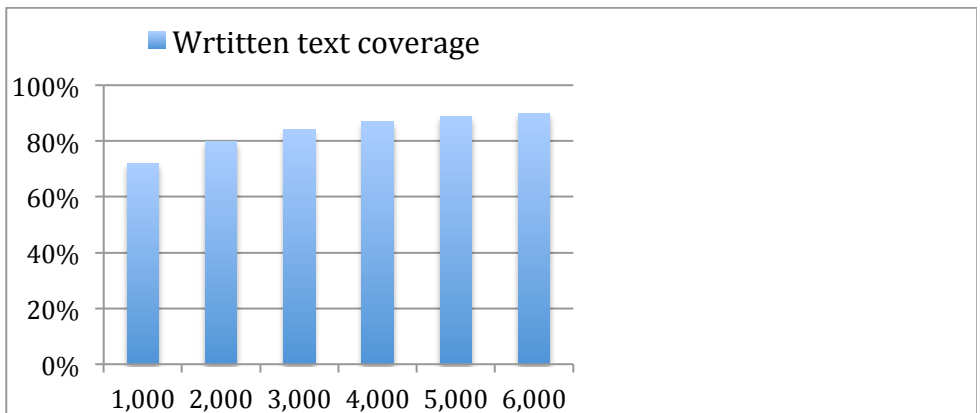
ELF: You can discuss this in the online forum called *Too many words in the dictionary*.

How many words are needed for academic prose?

In Ch.1 we saw that 27.2% of words in the Cherry Picking article and 29.3% in the Hedging abstract were beyond the 2,000 word threshold. As the graph below shows, one needs on average a vocabulary of about 6,000 words to expect to know 90% of words in text generally. And even this means that one in ten words could be unknown. For one thing, this complicates inferring the meaning of an unknown word from context.

Task 4: What factors influence an individual's likelihood of knowing many of a text's off-list words?

Task 5: What does this graph tell a learner of English about the value of learning another 1,000 words?



Task 6: How many words do you need?

The question of how many words you need for academic prose has been implied a number of times. We know about a text that:

- about 75% consists of the most frequent _____ lemmas
- 10 to 15% consists of the AWL which has _____ word families
- about 15% are the words of your field _____ lexemes

This may appear naïve and/or simplistic, but it does show that the vocabulary needed for writing academic prose in no way approaches the above estimates of general vocabulary size. It also shows that you can target your vocabulary study.

2. Meaning

It is quite likely that you consider *meaning* as one of the most important features of a word that you need to know for both receptive (listening and reading) and productive (writing and speaking) purposes.

Meaning is a multi-faceted construct and a deep awareness of its many faces is beneficial to advanced learners. Words mean different things to different people, and words mean different things in different contexts. As in all fields, terminology helps us think about similarities, differences and groupings.

Labelling

Picture a car, a skyscraper, a computer. When searched for on the internet, the thousands of images that appear for each of these rather general words all have sets of features in common. When we see one of these things, it has a direct relationship with several things that have hitherto been called a car, skyscraper, computer. (Wittgenstein P.I. § 67)

But what about a three-wheeled solar-powered, computer-controlled vehicle with room for one person only? What about a vast, residential building stretching high into the sky and wider than it is tall with hydroponic gardens that provide the inhabitants with all their fruit and vegetable needs? What about e-book readers, tablets, smart phones, a 1960s computer lab?

Some of these things lie outside our prototypical view of cars, skyscrapers and computers. But when our language does not have a specific enough word, we use a more general word, and then specify our meaning using other words. Sometimes, there is a term e.g. *interlocutor*, that is known to specialists. In our normal use of language, we express this concept using the language that we have in common with our reader or the *person we are speaking with*, i.e. our interlocutor.

These are some of the important considerations we bear in mind in the process of choosing vocabulary when writing academic prose.

Task 7: How close is the definition of a word to the meaning of a word?

Before you look up these seven words from the AWL, make some notes about what each of them mean to you.

react

scheme

valid

persist

subordinate

trend

whereas

What do you learn from comparing your understanding with the definitions provided in dictionaries?

Task 8: How close are foreign language equivalents in meaning?

In this table, write your L1 equivalent(s) of the given word and make some notes beside it about any differences between the range of meaning that the two words have.

challenging _____

accumulate _____

inevitable _____

intrinsic _____

sequence _____

channel (vb) _____

this _____

primitive (n) _____

know _____

Can “closeness” be measured? We have to consider the possibility that the distance between a word and its foreign language equivalent means that NSs' and NNs' perceptions differ.

One approach to thinking about the meaning of a word or lexeme is digging into the elements it consists of. For example, how do you know when you are looking at a *graph*? In other words, what are some features that all graphs share, and what are some features that are optional or possible?

Task 9: Indicate which of these features are *obligatory* (✓) and which are *optional* (?) in your conception of a *graph*.

two dimensional	have X and Y axes	labels
represent data	absolute data	relative data
bars	lines	title
square	circular	colour
Other features:		

This is looking inside a word, but we can look outside a word too. Our understanding of words comes partly from knowing where they fit *in the scheme of things*. We will now consider some of the classic semantic relationships, not only for the sake of vocabulary study, but also in preparation for their use in writing full text.

Task 10: Use arrows to match these terms to their definitions. Then copy these examples beside the appropriate term.

small, little; correspondence, email; kill, stab; hand, finger; high, low; key.

ELF: This can be done online: [Definitions and examples of semantic terms.](#)

Brief definition	Term	Examples
words that have multiple meanings	antonym	
words that have opposite meanings	hypernym	
a more specific way of doing a verb	meronym	
a word that is more general	polysemy	
a part of something	synonym	
a word that means “the same”	troponym	

Task 11: Now apply these terms in the following ways:

POLYSEMY: how many different meanings for the noun *code* can you think of? And for the verb *to know*?

What are the ANTONYMS of these words?
biased visible rise above lend show

Which of these words are TROPONYMS of *show*? *demonstrate, prove, establish, present, bias, depict, lend, contradict, validate, picture*

What is the HYPERNYM of these words?
message, letter, essay, article, column, paper

What are some SYNONYMS of *method*?

What MERONYM relationships can be described in your field?

ELF: This can be done online: Examples of semantic notions.

Task 12: Read the following text and say which of the above terms this text is referring to.

One of the most influential of all processing approaches to meaning is based on the idea that the meaning of a word is given by how it is embedded within networks of other meanings. Thus words obtain their meanings by their place in a network of associations. The meaning of dog might involve an association with barks, four legs, furry.

In a network, nodes are connected by links that specify the relation between the link nodes; the most common link is an "ISA" link which means that the lower level node "is a" type of the higher-level node. For example, a house ISA dwelling and a car ISA vehicle. A camper van is both. A Ferrari is a type of car, but to some people it is a status symbol, and to others, a product. This one-to-many and many-to-one is a short-coming of this system. So is the fact that many real world entities do not fall logically into hierarchies; abstractions even less so.

Task 13: Do you have any concepts in your field of study that can be organised in a multi-level *is a type of* structure?

For example: a Ferrari is a sports car is a car is a vehicle is a means of transport. A piccolo is a type of flute is a woodwind instrument is a musical instrument. Biologists have different types of cells.

Synonyms

Here are two learner-dictionary definitions of SYNONYM.

- a word that has the same meaning as another word. For example 'scared' is a synonym for 'afraid'. (Macmillan)

About the suffix **nym**: -onyma is a dialectical form of onoma, the Greek for **name**.

- a word or phrase which has the same or nearly the same meaning as another word or phrase in the same language The words 'small' and 'little' are synonyms. (Cambridge)

These definitions do not imply that the words can be used interchangeably, as we shall see on page 80.

Take the word *graph* again, and consider if *chart* and *diagram* are its synonyms. One way of investigating how closely they are related is to compare their primitives. i.e. their most basic or essential features.

Task 14: Complete the table by adding more PRIMITIVES as column headings and by making notes in the rest of the table.

This process is called COMPONENTIAL ANALYSIS.

word	shape	purpose			
graph					
chart					
diagram					

Task 15: Draw your own tables to repeat this procedure with some of these two sets of related words.

1. picture graph diagram illustration chart grid table plot
2. construct primitive data element concept

You can now depict the extent of their similarity by creating a tri-partite VENN DIAGRAM for graph, chart, diagram. For a simple example, see the famous *Nerd Geek Dork* Venn diagram linked in the e-learning course.

Polysemy and Meaning Potential

Many words are POLYSEMOUS, that is, they have more than one meaning. Because context selects the intended meaning, we can also say that words have MEANING POTENTIAL. The intended meaning of the polysemous words in these phrases emerges from just a little bit of context:

poly is the Greek for many and sem is the Greek for sign.

- a sentence of ten years, a sentence of ten words
- have a lot of experience, have a lot of experiences
- the panel will decide, a wooden panel

to have designs on someone, the art director came up with some new designs

I wonder what the future holds, English has no future
someone feels tense, English has no future tense
who is responsible for this child, responsible for this mess?
to lecture on microbiology, to lecture my teenage son
my son is wasting his potential, POLYSEMOUS words have meaning potential

The high proportion of polysemic English owes much to conversion. Each meaning of a word is its potential, and as these invented examples show, very little context is needed to DISAMBIGUATE the words.

Task 16: Have some fun writing a sentence or two that follow logically from each of these four sentences.

We booked a table for our son's birthday.

The secretary booked a table for the pianist who was very busy.

The policeman booked a waiter for spitting while driving through a red light.

The new restaurant in the red light district booked a topless pianist for our son's birthday.

Task 17: Answer the following questions about these sentences above. Explain the evidence for your answers.

What is the difference between "booking a table *for*" in the first two sentences?

Who was very busy, the secretary or the pianist?

What gender(s) are the secretary, the boss, the pianist and the policeman?

Who was driving through a red light?

What is the difference between a red light that you drive through and a red light district, that you might also drive through?

Thus, not very much context is required to understand the intended meaning of a POLYSEMOUS word. As your answers show, there are many levels of context ranging from neighbouring words to an awareness of the culture. When we write, it is our responsibility to provide readers with the necessary contexts so that the meaning they receive is the one that we send.

Task 18: Write something similar to the *book* sentences using one of the following words: science, power, force, energy.

Experiment with its lemmas, word forms, word family, multi-word units. Stringnet navigator will offer some quick and easy assistance.

This brings to a close our study of words that are related through meaning relationships.

3. Words in context

When reading, the meaning of a text emerges from various aspects of its composition, the most obvious of which is the use of vocabulary. Grammatical and syntactic features play their roles as do the contextual domains field, tenor and mode. When writing, we view this process from the other end which means that our micro choices are constrained by macro contexts. We must know why we choose the words we choose, and we need to know how to combine them.

The following formulation might be a slight exaggeration, but it makes a useful point: it is not the words which tell you the meaning of the phrase, but the phrase which tells you the meaning of the individual words in it. (Stubbs 2001:18)

Our words have meanings and functions, sometimes thought of as bricks and mortar respectively. Function words, i.e. prepositions, articles/determiners, conjunctions, pronouns and “particles”, are among the most frequent 2,000 lemmas and often manifest the additive principle. In this frequency

range we also find most of the words that express common notions such as time, place, people, cause and effect, priority, society, and the relationships between things.

The bricks and mortar co-operate to create the meaning of a text. The words carrying most of the propositional meaning are lexical words and are key words in a given text. But we do not understand a text fully by just understanding the key words.

Key words

Key words are nouns, verbs, adjectives and adverbs in that order: texts have approximately twice as many nouns as verbs, and verbs tend to be twice as frequent as adjectives. It is the nouns that tell most of the story of our work. They are a text’s KEY WORDS.

An interesting and novel way of learning from the vocabulary of a text is to work with a WORD CLOUD which shows the *keyness* of words by relative sizes. A good read of a word cloud allows us to predict what a text is about, but *caveat emptor* ☺, not all predictions come true. The word cloud below was created by pasting in the whole text of the *Happiness* research paper we mentioned in Ch. 1. Note that word clouds typically employ lists of STOP WORDS to keep function words out of the picture, literally, in this case.

Task 20: Does this process enable you to predict the content of this text?

Topic Trails

These are strings of words that represent various topics that run through a text forming trails. For example, in the introduction to a chapter entitled, *The Meaning of Things in Time and Space* (Kral 2012: 209), the following topic trails can be observed.

ELF: The full text (225 words) can be read in the online course.

Meaning	Time	Space	Attitude	People
systems of meaning	history	environment		
relatedness	being	socialisation		
disguised	evolution	transmission		
identities	development	dwelling		
cultural practices	become	spaces		
	generations	domestic		

Different readers perceive text differently, which can be shown in the Topic Trails they identify and even in the words and phrases that they consider represent them. Differing trails is an indicator of how readers respond differently.

What we learn from TOPIC TRAILS as authors of academic prose is how our words relate to each other, not in close proximity but across larger stretches of text as we tell our story. The process of identifying topic trails helps us to think about words in new ways.

Task 21: Read Kral's paragraph in the e-learning course and complete the remaining two columns above.

Task 22: Find another short text, perhaps an abstract. Make a word cloud of a text and find trails in the output.

Beyond the text

Studying a specialist corpus reveals the key words of its field and the words they operate with. This is a way of generalising the vocabulary of a field, not just a text. For example, the lists in the following table contain the most frequent key words in their respective corpora. It is worth noting that general academic words occur in several columns, whereas the more field-specific words do not.

Task 23: Fill the empty columns with key words from other corpora.

Open the ARC corpus, click Word list, click All lemmas. Read down the list and copy the lexical words into the column above. If you have a specialist corpus, repeat the process and enter the words into the fourth column.

Impact	IRC	ARC	
system	use		
water	word		
change	system		
increase	sentence		
model	model		
rate	set		
concentration	example		
year	language		
starting	result		
time	rule		
country	information		
area	feature		
result	structure		

4. Groups of words

Task 24: How do you think tens of thousands of words are most likely stored in our brain?

- **alphabetically** e.g. hospice, hospitable, hospital, hostage
- **by part of speech**, e.g. verbs are stored together, nouns are stored together, etc.
- **as families of words**, e.g. mode, modal, modality
- **as sets of topic words** e.g. book, library, reader, read, publish
- **as sets of synonyms** e.g. mad, crazy, insane, stupid, ridiculous
- **by word association** e.g. comment => say, no, opinion, speak, remark
- **in chunks** i.e. words are not stored individually, rather as members of phrases of various types

Dictionaries organise words alphabetically. Within a dictionary entry we find a variety of word forms and combinations. But our brain does not

store words alphabetically. It is an extraordinary thing that the educated native speaker with over 30,000 words crammed into less than 1.5 kgs of grey matter can almost always retrieve the appropriate word instantaneously.

The groupings of words in our brains are more associative. And some of these networks and associations are of considerable interest to the study and teaching of vocabulary.

In this text book, the target words have been selected primarily from the AWL but grouped according to patterns and frames, which is not a grouping offered in the previous task.

Lexical sets

This is a very common way of presenting vocabulary in language text books: words about holidays, pollution, cooking, etc. In academia, there are also many such field-related sets of terms. There are some which belong to science in general. For example, *research, result, experiment, findings, biased, hypothesis, conclusion, evidence, process, system, method, factor*. Learning lists is of limited value as all you learn is a list. Studying the relationships between the words and the company they keep is far more valuable for both receptive and productive language use. And this is precisely what this book does.

Academic Word List

Task 25: Pre questions

Can the words of the AWL typically be found in the indexes of academic books?

Given that there are 570 word families, how many words would you expect the whole AWL to contain?

If you studied one WF a week, how long would it take study them all?

What would you study about each item in the WF?

The Academic Word list (AWL) was mentioned in the previous chapter. We now look specifically at how students of academic prose can benefit from it. Created in 2000 by Averil Coxhead, it consists of 570 word families, many of which are common in educated native speaker speech and writing. The word families of the AWL occur to a small extent in newspaper writing (c. 4.5%) and to a very small extent in fiction, (c.1.4%), but in academic prose they constitute a whopping 10% approximately. This means that one in ten words is likely to be from the AWL, the other 90% being (a) general English words and (b) the technical language of the field, as we saw on page 52.

Note that the AWL contains single words only, i.e. no multi-word units or lexemes, no chunks, no phrases, no discourse markers.

Since it would take about ten years to get through the whole list if you studied one word family a week, it is necessary to exploit this list in pro-

ductive and meaningful ways. Almost all of the examples we studied in the Introduction use AWL words. And as we study many other language features, you are constantly exposed to these salient words. Growing your vocabulary is not just a matter of learning new words/lexemes, phrases and chunks, but observing ones that you already know in certain contexts. Add them to your dossiers.

A lecture on comparative psychology, whose text has 2,502 words (tokens), contains 98 word families from the AWL, almost 10%. Twelve of them occur from between 4 and 20 times each. The top ten of these word families are *classic, task, respond, adapt, complex, reinforce, environment, aspect, similar, neutral*.

Conversely, a student's work which did not have a strong academic feel was processed by the Vocabulary Profiler and found to contain 7% of AWL families.

ELF: [Link to the AWL homepage provided.](#)

Task 26: Take a piece of your academic writing, paste it into the Vocabulary profiler.

See if it has approximately 75% general words (blue and green), 10% AWL (yellow), and the rest the words of your article (red).

Repeat the process with other texts you have written.

Here are five sentences from the above-mentioned lecture notes, using some of these top ten word families.

Task 27: Circle the target words in each sentence and respond to the notes under each one.

Target words: *classic, task, respond, adapt, complex, reinforce, environment, aspect, similar, neutral*.

Corp ex. 1. *In particular it has been shown that, given similar species specific constraints to those I discussed earlier, that new-born humans classically condition well.*

Read this sentence aloud, pausing at the end of each information unit. Then underline each of them.

Corp ex. 2. *Learning is just one way for an individual to adapt to its environment.*

Comment on the two uses of *to*.

Corp ex. 3. *A number of investigators have, however, found effective probability (matching), reward-shift and reversal learning in fish, amphibia and reptiles provided [that] the stimuli, reinforcers and task setting are tailored to suit the species under investigations.*

This sentence contains three sets of noun groups, each with three items. In the first set, underlined, each item consists of two words. Underline the other two sets. The skeleton of this

sentence is **someone found something in creatures under specific research conditions.**

Corp ex. 4. *Given this, however, we still need to discover whether the principles of animal learning apply to those aspects of human behaviour which are not linguistically mediated.*

This sentence, and the one below, open with a good demonstration of how we tell our stories. Note the use of *discover whether*, and the simple present use of *apply to*. Note also the skeleton: **principles apply to aspects of behaviour which ...**

Corp ex. 5. *Similarly, heart-rate reduction might decrease the unpleasant effect of shock and so be regarded as an operant.*

Does this sentence appear to be extracted from a hypothesis? Why not? Comment on the use of *be* – shouldn't it be *is*?

Task 28: Using your own corpora, search for these words and observe them in context. Whatever you learn from your observations, add to your dossier.

classic, task, respond, adapt, complex, reinforce, environment, aspect, similar, neutral.

Up to this point we have been dealing with sets of isolated words. The next sets are groups of words that are believed to be stored in the brain as single units.

Phrasal Expressions List

The Phrasal Expression List was compiled from frequency data in a corpus of general English. Despite not being academically oriented, the following four statements from the article which launched it, account for our interest in chunks of language. Martinez and Schmitt (2012) comment:

Formulaic language is ubiquitous in language use.

Meanings and functions are often realised by formulaic language.

Formulaic language has processing advantages.

Formulaic language can improve the overall impression of L2 learners' language production.

Task 29: Beside each of the above, make a few notes about its relevance to you as a language learner.

Academic Formulas List

A decade after the AWL was created, Simpson-Vlach and Ellis developed the Academic Formulas List (AFL) with the principles similar to the Phrasal Expressions List, but basing their research on academic corpora. By *formula* they mean short phrases that occur significantly in their corpora, and that have some function in the prose.

Here are some of their formulae. They appear in a great variety of academic contexts.

from the point of view of	BE the case
in other words	BE more likely to
the extent to which	it/ there may be
at the same time	based on the
the relationship between	FOCUS on the
exactly the same	ASSUME that the

Task 30: The left column are all _____ while the right column all contain _____.

Don't just read over them. Think about how, when and why each one is used. Say them aloud. Look for them in your academic reading and add them to your dossiers.

The AFL formulae are very well represented in the psychology lecture notes mentioned above. Here are just ten of them with their occurrences. The most frequent structure is THE NP OF THE NP e.g. *the basis of NP, the value of NP*. We study this important structure in Ch.3.

Corp ex. 6. ... *electricity might form the basis of Descartes' elusive 'animal spirits'*

Corp ex. 7. *If the value of the first reinforcer is now reduced, ...*

Task 31: Do these formulae look like organisational (O) or message (M) language? Why?

<i>the number of</i>	6
<i>the nature of the</i>	2
<i>this is not</i>	5
<i>the presence of</i>	4
<i>the ability to</i>	3
<i>in order to</i>	5
<i>different from the</i>	2
<i>the value of</i>	2
<i>part of the</i>	7
<i>a series of</i>	3

Why is it inevitable that this language would emerge from searching a general academic corpus?

ELF: Some of them have been prepared in the e-learning course. Open them. Click the Sketch Engine's Sample button and choose a relatively low number, say 250 lines. Sort them *left* and see if anything recurs significantly. Ditto *right*. Whatever you learn from your observations, add to your dossier.

The following extract comes from the academic article which Nick Ellis and Rita Simpson-Vlach, the linguists who developed the AFL, wrote when they launched it in 2010.

Task 32: Read it and indicate what is said about the following pieces of information.

the AFL concerns academic prose only

the AFL represents general academic language, i.e. it is not field specific

the lists are structured with a pedagogical purpose in mind

formulaic sequences contain more than one word

The AFL includes formulaic sequences, identifiable as frequent recurrent patterns in written and spoken corpora that are significantly more common in academic discourse than in non-academic discourse and which occupy a range of academic genres. It separately lists formulas that occur frequently in both academic spoken and academic written language, as well as those that are more common in either written or spoken genres. A major novel development this research brings to the arena is a ranking of the formulas in these lists according to an empirically derived psychologically valid measure of utility, called 'formula teaching worth'(FTW). Finally, the AFL presents a classification of these formulas by pragma-linguistic function, with the aim of facilitating their inclusion in EAP curricula.

This section has mainly introduced the AWL and AFL, both of which will be frequently cited in the remainder of this book.

5. The structure of vocabulary: morphology

Under the heading of MORPHOLOGY in Ch 1, we saw how the stem of a word undergoes inflection and word formation processes. Taken from the point of view of our morphology hierarchy, a WORD FAMILY is all the words created by these processes.

Here we study WORD FORMATION and other processes that form WORD FAMILIES. Language acquisition researchers have shown that it is naïve to assume that learners know all the forms of the words that they know. Let us see if we can challenge these findings.

Task 33: Here are some examples. Complete the table, using your intuition first. Then use other resources.

	work	indicate	context	abstract
Conjugation	<i>works worked working</i>			
Declension	<i>work, (the) works</i>			
Derivation	<i>workable, unworkable, worker</i>	<i>indication indicator</i>		
Compounding	<i>word day, work sched- ule</i>			
Combined processes	<i>unworkable</i>		<i>decontextual- ized</i>	

Conversion + Morphology = Word Families

In the section on polysemy (p.56), we observed that many words have different meanings. In contrast, we now consider the same word forms used as different parts of speech: it is a standard feature of English that some verbs can be used as intransitive and transitive verbs (*to hold*); many past participles are used as a verb and as an adjective (*personalised, desired*); present participles are used as verbs and adjectives (*training, winning*), and as nouns (gerunds); many nouns can be both countable and uncountable (*experience, scholarship*); the word *fast* is most commonly used as an adjective, but it can also be an adverb, noun and a verb. The term for this is CONVERSION. Here are some AWL examples.

Task 34: Underline any of these alternations that you observe in the following sentences. The sentence are taken from the Happiness article unless indicated otherwise.

Experiment

Corp ex. 8. *This is a subset of the corpus used in the experiments reported in (Mishne 2005).*

Corp ex. 9. *Mishne (2005) experimented with affect.*

Contrast

Corp ex. 10. *We evaluate our corpus-based approach in a classification task and contrast our wordlist with emotionally-annotated wordlists produced by experimental focus groups.*

Corp ex. 11. The table visualizing the essentials of the contrast between the modal and main verb uses is reproduced below. (BNC Sci)

List

Corp ex. 12. First, we determined the time distributions associated with the concepts in our list of salient words.

Corp ex. 13. The happiness load for a 24-hour day listed in the ANEW list of words.

Abstract

Corp ex. 14. A conception of law is a general, abstract interpretation of legal practice as a whole. (BNC Sci)

Corp ex. 15. While it is possible, in the abstract, to treat policies in isolation from other policies, in practice any new policy will be adopted ... (BNC Sci)

Corp ex. 16. The book was an abstract of a work which had not appeared, ... (BNC Sci)

Corp ex. 17. If we abstract those features which result from social context the exercise becomes ... (BNC Sci)

Approach

Corp ex. 18. We evaluate our corpus-based approach in a classification task and contrast our wordlist with emotionally-annotated wordlists produced by experimental focus groups.

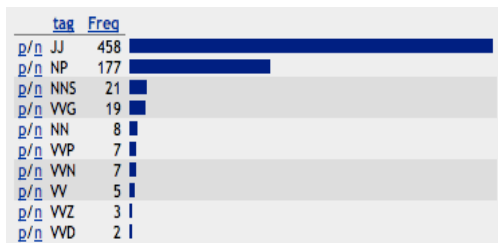
Corp ex. 19. These examples should give you some idea of how to approach your own list of foods. (BNC Sci)

Factor

Corp ex. 20. Anticipation seems factored into each day's happiness, ...

Corp ex. 21. Instead, the word love in our list is neutral, with a happiness factor of only 48.7.

To find a word form's parts of speech using the Sketch Engine, start with a Word forms query. Clicking Node Tags produces a graph like this. Tags starting with N are nouns, V are verbs, J for adjective.



Click on the p (in p/n) to see sentences containing the word.

Task 35: Which of the above corpus illustrated words might this graph represent?

Task 36: Try this with five different words. Is it usual for one or two tags to be significantly more frequent than the others, as is the case above?

Task 37: On page 56 we saw the following nouns. Are the verb forms the same as these noun forms, or do they differ?

picture graph diagram illustration chart grid table plot

Use the Sketch Engine procedure above to verify your answers.

This feature of English which sees the same word form function as different parts of speech can hardly be a feature of languages whose endings indicate part of speech. For example, most German verbs end in -en, Czech verbs in -at/-et/-nout, Italian verbs in -are/-ere/-ire. In such languages parts of speech are regularly distinguished by endings.

Task 38: In other languages you know, can the word form of an adjective be the same word form for a different part of speech?

To summarise, conversion is the use of the same word form in different parts of speech, regardless of meaning.

To continue, it is also necessary to study the quite different forms of words as they change their parts of speech, e.g. *paradigm, paradigmatic*, a pair which also has significant differences in pronunciation. Other examples: *sign, signify, significant; differ, difference, different, differentiate; pronounce, pronunciation*.

Task 39: Copy these words into this table. Underline the stressed syllable in each word.

ELF: The online version of this activity is called Conversion.

Noun	Adjective	Verb

Task 40: The Happiness article contains these verbs whose nouns are quite different word forms. Can you provide them?

live	_____	derive	_____
compare	_____	produce	_____
evaluate	_____	conclude	_____

Task 41: Is this true of the adjectives in the task above?

These verbs also have adjectival forms. Find some corpus examples of the adjectival form. For example, the word family of *alter* includes *altered* as a past tense which is also used as an adjective.

If you use the Sketch Engine, search for each word from and select adjective as its part of speech.

alter	_____	an altered state, altered circumstances
live	_____	
compare	_____	
evaluate	_____	
derive	_____	
produce	_____	
conclude	_____	

Another example of word formation we saw in Ch 1 (p.21) was the making of verbs and adverbs with prefixes and suffixes.

Task 42: Can you recall some of these?

Conversion and pronunciation

Here are some more academic words which function as more than one part of speech; in doing so, their pronunciation changes.

Task 43: Look in a dictionary to find the meanings and pronunciations of these words. Add them to your dossier.

ELF: This can be done online at Conversion and pronunciation.

To find your own examples, follow the steps on page 68.

appropriate	(adj)	_____
	(vb)	_____

arithmetic	(adj)
	(n)
elaborate	(adj)
	(vb)
minute	(adj)
	(n)
segment	(vb)
	(n)
alternate	(adj)
	(vb)
extract	(vb)
	(n)
project	(vb)
	(n)
contract	(vb)
	(n)

Singular and Plural

It would be reasonable to believe that the singular and plural forms of a word are nothing more than that. However, this is not always the case. Here, for example, we see that *factor* and *factors* are clearly used in quite different ways. This data comes from the Impact corpus.

a	factor	of
important	factor	in
a	factor	in
major	factor	in
significant	factor	in
a	factor	which
crucial	factor	in
risk	factor	for
a	factor	,
platelet	factor	4
a	factor	.
key	factor	in
critical	factor	in
important	factor	.
a	factor	that

the	factors	which
the	factors	that
other	factors	,
such	factors	as
these	factors	,
risk	factors	for
of	factors	.
other	factors	which
other	factors	.
of	factors	,
these	factors	are
of	factors	which
social	factors	,
two	factors	:
these	factors	.

Task 44: You can discover these differences with other words for yourself.

Search for a noun in the **Word Form** field. Then click on Frequency, and fill in the first three levels: 1L, Node, 1R as shown here. Lists like those for *factor* is generated when you click on Make Frequency List.

The screenshot shows a web interface for creating a frequency list. It is divided into three columns: 'first level', 'second level', and 'third level'. Each column has a radio button at the top, a dropdown menu for 'Attribute' (set to 'word'), a checked checkbox for 'Ignore case', and a list of options for 'Position'. The 'third level' is selected. The 'Position' lists for each level are: 6L, 5L, 4L, 3L, 2L, 1L, Node, 1R, 2R. A 'Make Frequency List' button is at the bottom.

Note that you need to repeat this for both word forms.

This frequency list tool is very useful for creating such lists. Stringnet navigator provides something similar and is easier to use, but you cannot control the parameters and its only source is the full BNC.

As mentioned above, pedagogical research has shown that we cannot assume that knowing some words in a word family guarantees that we know all of them. The researchers found that “reading facilitates derivational knowledge in particular because derivational suffixes are more common in the written mode than in the oral mode and are particularly associated with formal and academic discourse.” (Schmitt and Zimmerman 2002:149).

Task 45: Here are some erroneous word form examples from my learner corpus. Can you correct them?

ILC 1: *combine the visual and auditive style*

ILC 2: *declination is another type*

ILC 3: *In general, he had troubles with word stresses*

ILC 4: *... however, the student's works belonged to the most elaborated ones.*

ILC 5: *This mis-order also appeared in her first written homework*

ILC 6: *the only deficiency that re-occurred also in the student's assignments was*

Schmitt and Zimmerman also point out that speakers of Romance languages have an advantage because many academic words have Romance language roots.

Task 46: Which of the following languages are the Romance languages that have contributed to the vocabulary of English academic prose?

Japanese Albanian Czech French Farsi Portuguese Latin English Dutch Greek

In conducting their research, stylistic, collocational appropriateness colligational accuracy were not tested. For your active use of any vocabulary, however, such facets are crucial. And they are the focus of Chapters 3 and 4.

6. The grammar of vocabulary: colligation

Introduction to Structures, Frames and Patterns

We are now going to formalise our understanding of the term STRUCTURE. It has been used in this book already, not always as a technical term.

A structure is a string of parts of speech containing no specific words unless they are part of the structure itself, e.g. *if, when, way, be, have*. Some structures have names such as passive, conditional, perfect, transitive, causative. It is our highest level of abstraction in clause structure.

We know each other by sight, but we have not been formally introduced.

Structures also involve PREPOSITIONS, typically bound. The prepositions in structures are free when they are part of an obligatory adverbial. Remember that this way of thinking about prepositions is a useful learning strategy.

Task 47: The left column below contains some structures that are typical of academic prose.

On the right are some partial examples. Add missing words. For example, if the structure contains a Noun Phrase (NP) but the example doesn't, find something typical in your corpora and write it in.

ELF: There is a similar activity in the online course using these five structures.

NP verb prep NP	_____	accounts for	_____
NP verb <i>way</i> prep	_____	clear the way for	_____
NP verb NP prep NP	_____	provides an opportunity for	_____
NP <i>about</i> NP	_____	information about	_____
NP <i>among</i> Pl-NP	_____	disagreement among	_____

Corpus Query Language

We use Corpus Query Language (CQL) to search for structures.

For example:

Structure	Components	CQL
passive	BE + past participle	[lemma = "be"][tag = "V.N"]
perfect	HAVE + past participle	[lemma = "be"][tag = "V.N"]
with preposition	NP verb preposition adjective	[tag = "NN"][tag = "VVZ"][tag = "IN"][tag = "JJ"] (this structure has 52.4 HPM in ARC)

This is possible because our corpora use part of speech tags and lemmatize words: every word is stored as a so-called TRIPLE. These hidden codes can be "unhidden" in the SkE by selecting Tag and Lemma in VIEW OPTIONS. Here is an extract from one of the CQLs in the above table.

```

fields /NNS/field varies /VVZ/vary in /IN/in different /JJ/different
semantics /NNS/semantics consists /VVZ/consist of /IN/of simple /JJ/simple
speakers /NNS/speaker depends /VVZ/depend on /IN/on many /JJ/many

```

Task 48: Which of the three structures in the above table do these corpus extracts exemplify? Underline the components.

Corp ex. 22. *The system consists of multiple blackboards each of which store ...*

Corp ex. 23. *The algorithm for this choice depends on many phenomena...*

Corp ex. 24. *Our approach relies on distributional and frequency statistics ...*

Frames and patterns

Frames and patterns are structures that contain lexical words. I use the term FRAME for the collocations and colligations of nouns, as we are about to see, and the term PATTERN for the collocations and colligations of verbs. This is the basis of the rest of this chapter and much of Chapters 3 and 4.

Task 49: Pre questions

What is the difference between collocation and colligation?

What pairings count as collocation and colligation?

Why study this?

Is a multi-word unit (lexeme) a collocation?

You have already met the terms COLLOCATION and COLLIGATION a number of times in this book. Here are two of these references from Chapter 1:

- When authors express themselves using pre-existing chunks of language, their readers find the text easier to process if the lexical words combine with each other in familiar ways (collocation), and words are used in their typical grammar structures (colligation).

- The program, Stringnet Navigator clearly shows that every word exists in patterns of associated words (collocation) and associated grammar structures (colligation).

Our purpose here is to develop an understanding of how you best combine words while writing academic prose. The distinction between collocation and colligation is that the former consists of a pair of LEXICAL WORDS, whereas the latter contains a lexical word and its grammatical company. The parts of speech regarded as lexical are noun, verb (not modal or auxiliaries), adjective and adverb. Grammatical company includes such things as an obligatory preposition, a clause starting with *wh* or *that*, infinitive, -ing forms, used in passive.

You know a word by the company it keeps.

(Firth 1957)

Both of these terms entered our lexicon at the same time, thanks largely to the work of the British linguist, J.R. Firth (1890 – 1960). Interestingly, in



my CorpCorp, the word collocation occurs 1,235 pm whereas colligation occurs 49.1 pm. Given that the study of collocation became a 'growth industry' in the 1990s because of the ready availability of empirical data, it is not surprising to see this in a corpus of texts dedicated to using corpora in language teaching. However, empirical data is no less valuable to the study of colligation, but the use of this term is diluted firstly, by having competitors: valency, pattern, structure, complementation, transitivity – none of which mean quite the same, but in essence, the study of colligation was already well catered for. Secondly, the study of colligation by students does not lead to a more idiomatic use of English, which was the promise that made collocation the Next New Thing in language teaching. From

a learner's perspective, colligation is about accuracy, and this goal of language teaching was anything but new.

The collocations and colligations of nouns and verbs play different roles in creating your message. This is why I use the term PATTERN when working with verbs and FRAMES when working with nouns: two terms for two notions.

The upshot is that if you want your writing to be accurate, start by paying attention to colligation. If you want your writing to be idiomatic, pay attention to collocation. If you want both, study both! As we will soon see, we now have the mechanisms for studying both at the same time.

Task 50: **Indicate if the following are collocation, colligation, neither or both.**

go on + -ing	as ADJ a NOUN as possible
can't help +ing	for the sake of
flout maxim	empirical evidence
future holds	adjective + enough + to-ing
go on + to-inf	X is the new Y
knock-on effect	lose sth vs. lose s.o.
radically alter	rocket science

Collocation

Task 51: Pre-question: read this rich sentence (aloud) and answer the questions below.

Following the phraseological approach, collocations are defined [in this paper] [as a fixed expression] [characterized by] [relative transparency in meaning] and [a restricted binary co-occurrence] of [lexical units] between which [a syntactic relation holds] (Mel'čuk 1998; Nesselhauf, 2004) [RN 1904]

- Where are the collocations defined? _____
- How are they characterised? _____
- How open is the co-occurrence? _____
- How are the items related? _____
- How are they defined? _____
- How many items in a collocation? _____
- What co-occurs? _____
- What *holds*? _____

Transparency in this context means that the meaning of the collocation can be understood from knowing the meanings of the collocates. This differs from the meanings of many lexemes (e.g. glass ceiling, white elephant), phrasal verbs (e.g. shrug off, wipe out) and idioms (e.g. pigs might fly, water off a duck's back).

The authors' use of *lexical units* is their solution to the slippery concept of *word*, as we discussed in the previous chapter. The important point here is that collocations are not restricted to two words (binary), but to two semantic units, either or both consisting of one or more words (lexemes).

The two units in a collocation are the NODE and the COLLOCATE. They do not have to be adjacent and they can appear in both orders: N => C and C

=> N. In the following two examples, *recover motion* is a collocation, the first in the active, the second in the passive VOICE.

Corp ex. 25. *However, it is possible to recover both motions precisely if we start with even a very crude estimate of either.* (IRC)

Corp ex. 26. *In an analogous fashion, the motion p can be recovered when q is known.* (IRC)

In the next examples, we see *recover information*, C + N then N + C.

Corp ex. 27. *the classification systems of the large corpora are insufficiently delicate to recover the information required.* (CC Johns)

Corp ex. 28. *... the speed of analysis and the amount of information recovered.* (BNC)

Likewise in these two examples of *recover fully*:

Corp ex. 29. *The body of the adult human, however, can often withstand this chemical onslaught and ultimately recover fully.* (BNC)

Corp ex. 30. *... university schools of English have been dominated by "historical critics" attempting the "impossible" task of fully recovering the meaning imputed to a poem ...* (BNC)

Task 52: In the above examples, circle the node and box the collocate.

State the parts of speech that are collocating in each pair.

How did you decide which was the node, and which was the collocate?

And in the next example, we also see the collocation *recover meaning*, beloved of literary scholars and linguists. It is not unusual for another member of the word family to collocate similarly. The next two examples show *information* and *meaning* collocating with *recoverable*.

Corp ex. 31. *... they may well contain information which is ultimately recoverable.* (BNC)

Corp ex. 32. *... the author has provided a text from which some meaning is recoverable.* (BNC)

The next two examples come from my learner corpus – each example is the work of a duo or trio – I rarely ask students to submit long pieces of writing composed by individuals. The first two examples illustrate the distance that often exists between a node and a collocate.

ILC 7: *[M...] is the visual type of student, not the auditory and this could explain why her progress with the listening skills is not so fast.*

ILC 8: *As the difference between her current level of listening abilities and the level of the recordings used in the lesson was not that large.*

Task 53: What are the adjectives collocating here with *progress*, and with *difference*? Look in a corpus to see if they are typical – this is the Hoey Procedure at work.

Task 54: Create several sentences that contains *data* and *questionnaire*. with at least 11 words per sentence.

Here are some more extracts from my LC. Do you see the problems?

ILC 9: *A part of the study is the data concluded from a questionnaire ...*

ILC 10: *... some hypotheses suggest its importance for....*

Do you *conclude* data from something? Do hypotheses *suggest*? Ask your corpus.

The collocate of a node can also be the node of another collocate. For example, in a sentence we studied in Chapter 1, we saw the following collocation CHAIN: large number, number of studies, studies measure, measure discharge. This has been more poetically referred to as a ‘collocational cascade’.

When corpus software generates lists of collocates, your choice of parameters influence the content of the list. In the Sketch Engine’s Collocation

Attribute: lemma In the range from: -5 to: 5
Minimum frequency in corpus: 3
Minimum frequency in given range: 2
T-score MI MI3 log likelihood min. sensitivity logDice
T-score MI MI3 log likelihood min. sensitivity logDice
Show functions: logDice Sort by: logDice
Make Candidate List Save Options

secret-j
LFSR-n
decrypt-v
reader-n
ks3-n
ks2-n
failure-n
compromise-v
trajectory-n
precision-n
faithfully-r
Mills-n
Rj-n
multiply-v
motion-n
alignment-n
shift-n
temporal-j
impossible-j
trust-v
key-n
transformation-n
RFID-j
tag-n
ambiguity-n

Candidates form (pictured), all of the parameters can be changed. Another, albeit obvious factor, that influences the content of the collocate list, is the corpus itself.

The screen shot on the right was generated by the Sketch Engine using the settings shown above but with the Attribute *lempos*. The list here contains lemmas with their part of speech (POS), hence “lempos”. The node is *recover* as it appears in the IRC, showing how very differently this rather standard word is used in Informatics.

In addition, a polysemous word, e.g. *table* will return different collocates when it refers to tabling a mo-

tion, tables with columns and rows, tables in relational databases, a water table, and a piece of furniture. In a general corpus, highly polysemous words return incoherent lists. When writing academic prose, a corpus of your field is therefore very valuable.

Task 55: In the table below, there are three lists of collocates of one noun from three different corpora.

Read through the lists and try to work out which one common science noun they are the collocates of. Secondly, determine which corpus has generated which list: (a) science writing in the BNC, (b) Corpus Corpus (c) IRC. Thirdly, notice that any collocates occurring in all three columns will imply their general use in science: the remainder are likely to be field-specific.

empirical	empirical	suggest	
provide	accumulate	adduce	
there	anecdotal	supporting	
There	ISO	witness	
support	reuse	empirical	
little	suggest	admissible	
transfer	provide	oral	
suggest	symmetry	taking	
body	Stahls	criminal	
gather	experimental	obtain	
qualitative	strong	sufficient	
anecdotal	eSCM-SP	support	
effectiveness	substantial	prosecution	
substantial	binding-as-initial-filter	medical	
further	against	expert	
sufficient	little	documentary	
gain	collect	hearsay	
substantiate	piece	historical	
statistical		trial	

Task 56: The fourth column is for you to add the collocates of this word from your field's corpus. If you use the SkE, set the parameters as in the screen shot above. Why?

Looking in and looking around

In a previous section we mentioned synonyms, and used componential analysis and Venn diagrams to look inside *chart*, *diagram*, etc. As we look around words at the collocates of synonyms and see the company they keep, we find that they are used differently. In the following table, we compare two adjectives, *small* and *little* and see that the nouns they typically modify are different and so are the adverbs that modify them.

	adverb	adjective	noun
	<i>mostly* pitifully* sufficiently* extremely* rather*</i>	<i>small</i>	<i>proportion* minority* quantity* scale* amount* business</i>
	<i>precious* strikingly* surprisingly remarkably</i>	<i>little</i>	<i>while* bit girl thing</i>
	<i>comparatively relatively</i>	<i>small/little</i>	<i>boy smile village boat</i>

* The asterisk indicates that the “other” adjective is *not* used with it in my data.

ELF: This data is prepared in our online course for you to investigate.

Task 57: Try this with any pair of similar adjectives, especially any that you feel are interchangeable, e.g. *smart intelligent clever wise; viable, workable, functional; graphic, visual*.

These differences do not only appear between adjectives. Take any pair of words which are considered synonyms and observe how they are used. Some verbs: *force, compel; watch. observe; believe, trust*.

Cognitive Profiles

A cognitive profile is a set of sentences that revolves around a noun. Each sentence illustrates the noun’s relationship with a collocate. It was developed by the lexicographer, Patrick Hanks, who points out that collocates of nouns do not have the relationships that verbs have with their collocates. This is a reason I prefer to use the term FRAME for nouns, and leave PATTERN for verbs. Hanks writes:

Cognitive profiles consist of phraseologically well-formed, idiomatic statements for a noun. The goal of a cognitive profile is to organise as many as possible of the salient collocates of the target word into meaningful, informative, and idiomatic statements. A good cognitive profile uses all the salient collocates of the target word and so provides excellent guidance on its idiomatic use.

(Slightly adapted from Hanks (2012: 66-8).

For language learning, cognitive profiles give students the opportunity to express the relationship between a node and its collocates. Here is a cognitive profile of the AWL noun, *role*, based on the collocates from the BNC Sci.

Roles in theatre and cinema are performed by actors, and there are roles in a more abstract sense that can be played by inanimate things such as data, health and stress, as well as by people at work. Both senses exhibit either a temporary or a non-central function.

- People play and perform roles, and they can fulfil roles, such as teacher, guardian, mentor.
- Actors play leading and supporting roles on the stage and in film.
- The role of the press is seen as judge and jury.
- Roles are said to be crucial, important, vital, key and central.
- Roles are said to be active, passive, dominant.
- There are different types of roles, such as gender roles, traditional roles, managerial roles, leading roles, minor roles, and advisory roles.
- People are assigned roles by others, and assume roles themselves.
- Teachers are required to play a wide variety of roles.

The sentences reveal aspects of the word's meaning and use, and they are created freely. Corpus data allows the author of cognitive profile sentences to include other typical words and structures that are used with the collocates.

When the meanings of polysemic nouns are quite distinct, each meaning has its own cognitive profile. *Hearing* in speech therapy is different from that in law. A *cell* in biology is different from that in prisons. A *key* to a quiz is different from the *key* of a piece of music. These words would not conflate their cognitive profiles.

Task 58: Choose some nouns from your academic field, create a list of collocates and then compose a set of sentences.

Alternatively, you could start with your "happiness" spray diagrams on page 59.

Colligation

The colligation of a particular word is its relationship with "grammatical markers and grammatical categories" (McEnery 2012:130). Stringnet Navigator efficiently and convincingly demonstrates the patterned nature of language. This makes it a very useful tool for writers, but as mentioned, its only corpus is the BNC.

Given that there are many fewer grammar structures and grammar words (CLOSED CLASS) than there are lexical words (OPEN CLASS), the options are relatively limited. However, given that a structure realises the intended meaning potential of a word, knowing how our key words work is important to using them. Some modern dictionaries contain this information, but not for every instance of every word, especially not for multi-word units. This is why corpora are valuable when writing about specialised topics.

Patterns and frames exist within STRUCTURES. In a sentence, frames and patterns are the extended collocations that overlap, or cascade.

A. One structure only: verbs

Many words have one grammar structure only, for example, the following verbs, according to Cobuild Grammar Patterns I. (Francis, Hunston, Manning, 1996)

We have one soup on the menu and you will eat it before going on to the main course.

Task 59: What types of subjects precede the verb? A general someone or something or a limited one?

_____	deter	someone from doing something
_____	envy	someone
_____	familiarise	someone with something
_____	motivate	someone to do something
_____	rid	someone/something of something
_____	stem	from something
_____	supplant	something with something
_____	program	something
_____	exclude	someone from something
_____	distract	someone from something

Some examples of *familiarise/ize*.

Corp ex. 33. *Their scent markings help to familiarise them with their territory.* (BNC)

Corp ex. 34. *Such words do occur, however, in historical writings and you must familiarise yourself with them.* (BNC)

Corp ex. 35. *The rationale for it was to give the treatment group an opportunity to familiarize themselves with concordance strategies and to achieve optimal results* (Flowerdew, 1996, p. 112) *when using the concordance program.* (CC)

Task 60: Which of those verbs above are most likely to be significant in academic prose?

The two monosemic AWL verbs that are in Hanks' online *Pattern Dictionary of English Verbs* (PDEV) are *amend* and *analyse*. In the extract here, the first line tells us what types of things can be the subjects and objects of

100% [[Human | Computer | Process]] analyse [[Information | Anything]]
[[Human | Computer | Process]] examines [[Information | Anything]]
methodically and in detail, in order to understand and explain it

analyse. The second line and beyond paraphrases the information.

ELF: See the online course for more information about PDEV.

Task 61: It is claimed that words with one pattern have only one meaning – they are MONOSEMIC.

Investigate this claim on some of the words above, i.e. do the above words have one meaning only?


B: One structure only: nouns

The following task presents some structures that are used with nouns.

Task 62: Match these nouns with their patterns. There should be one each, but perhaps you will find others.

ELF: See the online task, One structure only: nouns.

projection	between pl-NP
resources	of NP
overlap	in NP
exploitation	to-inf
expertise	that



Task 63: Here are three sentences from three corpora for each of two AWL nouns that have one pattern only.

Corp ex. 36. ... it is difficult to predict the duration of the file transfers (IRC)

Corp ex. 37. The difference in duration of ventilation was not significant (BNC Sci)

Corp ex. 38. ... such as the location or duration of an event.

Corp ex. 39. ...and the flexibility to fill in the design details as choices are made (IRC)

Corp ex. 40. ... a multimodal interface which provides users with the flexibility to provide input using speech (ARC)

Corp ex. 41. *Using natural gas has given us the flexibility to run domestic heating from the same boilers (BNC Sci)*

What can you observe about the structures of *duration* and *flexibility*? Do you know of any other structures for these two nouns?

C. Multiple structures: verbs

Most words however, function in a variety of structures, not that this gives you any real choice: only one of the options will express your intended meaning. Language, like society, is full of conventions that dictate our behaviour, and constrains our freedom to choose.

We can offer you red and white wine, but if you're having beef, convention dictates that you will choose a red.

According to PDEV, the following AWL verbs have five or more structures:

deny, file, abandon, devote, submit, accompany, acknowledge, acquire, maintain

Task 64: You might find it interesting to notice Zipf's law at work.

Open each of these verbs in PDEV and see if the most frequent pattern/meaning is approximately twice as frequent as the second most, and if the second is twice as frequent as the third etc, as alluded to on p.58.

PDEV is here: <http://deb.fi.muni.cz/pdev>.

Task 65: Given that these are AWL verbs, which of them occur in your field, and with which of PDEV's patterns?

Add these, with sentence examples, to your dossiers.

The following verbs have various colligations, one which they share is **wh + to-inf**:

assess, debate, establish, illustrate, instruct, investigate, reveal, indicate, explain

Task 66: Search your corpus for useful examples of these words in this pattern. Add them to your dossier.

The CQL is [lemma="assess"][tag="WRB"][tag="TO"]
Copy this, changing the lemma as necessary.

ELF: In the online course, this search is prepared.

Here are two examples sentences that this CQL found.

Corp ex. 42. *It would leave too many alternatives and would not indicate which to choose . (ARC)*

Corp ex. 43. *Students worked from textbooks which explain how to conduct the experiment (BNC Sci)*

Task 67: Write five interview questions using these verbs in this collocation. Ask your questions.

E. Multiple structures: nouns

Learning nouns in the structures is a useful strategy for learning not only new words, but new ways of thinking about how the words are used. My database has 51 AWL nouns that are followed by a *that clause*, e.g.

coincidence, concept, equation, implication, symbol, tradition

and 41 AWL nouns that are followed by *between* and a *plural noun*, e.g.

bond, inconsistency, overlap, transition

and 29 AWL nouns that are preceded by *on*, e.g.

behalf, commission, display, target.

Many nouns work in a number of structures, which, as we have seen, affects their meaning and how we use them. For example, the *concept that*, *concept behind*, *concept of ...* express different semantics of *concept*. Similarly, the noun *role*: someone's *role as NP*, *role in NP*, and in the fuller structure:

it BE [possessive] *role* to-inf, e.g.

Corp ex. 44. *My role is to oversee the day-to-day research on the island (ukWaC)*

Corp ex. 45. *ASIO's role is to advise government on security threats (ukWaC)*

F. Preference for the passive

It has been found that some verbs are more typically used in the passive voice. For example, *establish* when used in academic prose in the passive:

Corp ex. 46. *Once the major premises have been established the principles fall into place. (BNC Sci)*

Corp ex. 47. *a new equilibrium is established between the condensate, ... (BNC Sci)*

G: Preference for the negative

There are other colligations that include a word's preference for the negative e.g. *afford, bother*; with modals e.g. *mind*. Specific words are used in negative environments. For example, the subjects of *ensue* tend to be negative as do the objects of *cause, harbour* and *exacerbate*. Adjectives such as *utterly* and *abject* intensify negative things. Even the preposition *amid* is often found amid negative contexts.

Task 68: What are the differences between:

to ask someone vs. to question someone

to give a lecture vs. to lecture someone

to label someone/ thing vs. to put a label on someone/ thing

H. Obligatory items

Some words keep company with others not by choice. We have seen that *put* must have an adverbial. So does the verb *place*, when used with that meaning: *put/place something somewhere*. Many VERBS OF MOTION have an obligatory destination or a path, and many VERBS OF EXCHANGE require two objects. Without these, their meaning would either be incomplete or different.

Another level of obligation is more pragmatic than semantic as it relates to the MAXIM OF QUANTITY mentioned in Chapter 1. For example, it might be necessary to provide all the elements in this structure:

[someone] *awards a scholarship* [to someone] [to study] [something]
[somewhere]

but only if the reader had no knowledge of them, and if it is relevant to your story telling.

Collocation within colligation

Chapter 1 contains screen shots of two web-based tools that display collocation data sorted within colligation categories. The first (p.xx) is the SkE which generates so-called Word Sketches (WS) of words in virtually any uploaded corpora, including your own. The other is *Just the Word* using the BNC only. These categorised lists provide much more focussed data than collocation lists and are very useful when writing.

The pair of WS extracts below come from the IRC and depict only the subject and object colligations, i.e. the verbs that collocate when the noun (a) “performs” the verb (subject) and (b) when it is affected by the verb (object). In the Sketch Engine, the columns of collocates can be sorted by frequency of co-occurrence, or by significance. In the Word Sketch, click on the Sorting button to toggle between these two sorts.



Task 69: Firstly, can you work out which very common science noun these are the collocates of?

Secondly, which one is sorted by frequency and which by significance? Write these as headings beside A and B here.

A:

object of	1186	2.0	subject of	763	2.4
perform	<u>120</u>	8.09	show	<u>88</u>	4.88
conduct	<u>84</u>	7.94	use	<u>72</u>	3.81
carry	<u>82</u>	5.84	involve	<u>48</u>	5.26
design	<u>49</u>	5.98	suggest	<u>19</u>	3.96
describe	<u>46</u>	5.28	go	<u>19</u>	1.73
repeat	<u>41</u>	7.06	take	<u>19</u>	1.15
make	<u>35</u>	1.64	begin	<u>17</u>	3.28
try	<u>33</u>	4.14	lead	<u>16</u>	3.27
report	<u>30</u>	4.92	demonstrate	<u>15</u>	5.41
begin	<u>17</u>	3.27	prove	<u>15</u>	4.47
use	<u>17</u>	1.72	illustrate	<u>11</u>	5.13
show	<u>14</u>	2.23	produce	<u>11</u>	2.77
devise	<u>13</u>	6.32	support	<u>10</u>	3.44
undertake	<u>13</u>	5.15	continue	<u>10</u>	3.11
discuss	<u>13</u>	4.2	last	<u>9</u>	5.71

B:

object of	1186	2.0	subject of	763	2.4
perform	<u>120</u>	8.09	verify	<u>4</u>	6.39
conduct	<u>84</u>	7.94	last	<u>9</u>	5.71
footprint	<u>8</u>	7.7	demonstrate	<u>15</u>	5.41
repeat	<u>41</u>	7.06	involve	<u>48</u>	5.26
devise	<u>13</u>	6.32	illustrate	<u>11</u>	5.13
replicate	<u>4</u>	6.27	investigate	<u>9</u>	4.99
design	<u>49</u>	5.98	show	<u>88</u>	4.88
carry	<u>82</u>	5.84	test	<u>8</u>	4.66
terminate	<u>5</u>	5.82	consist	<u>7</u>	4.58
cite	<u>7</u>	5.44	prove	<u>15</u>	4.47
describe	<u>46</u>	5.28	confirm	<u>9</u>	4.35
undertake	<u>13</u>	5.15	reveal	<u>9</u>	4.0
interpret	<u>7</u>	4.95	suggest	<u>19</u>	3.96
report	<u>30</u>	4.92	use	<u>72</u>	3.81
review	<u>5</u>	4.51	indicate	<u>8</u>	3.69

Task 70: As you compare these two lists, does one of them seem more valuable to you than the other?

FREQUENCY is simply the number of times a collocate co-occurs with a node. But SIGNIFICANCE is a statistical measure that takes into account how often the collocate occurs in the corpus generally. For example, despite the fact that *verify* is not a particularly frequent verb in the BNC (only 5.1 pm), it is more closely bound to *experiment* than other verbs it collocates with: *show* (524 pm), *use* (1059.4 pm), *involve* (202.2 pm), because these more general purpose verbs keep company with many nouns.

Significant collocates tend to be semantically richer than frequent collocates. When authors use semantically richer ingredients, their text is semantically richer. It's like baking a cake.

Task 71: Here is a hypothesis for you to test:

Significant collocates are more field related than high frequency collocates.

What value are the findings here?

7. Vocabulary study

In this book we have so far studied vocabulary in a number of ways. And there are more to come. Much of the work has involved observing how words behave in text, and how they co-operate with other words, both functional and lexical words. In the closing section of this chapter, we shall consider a range of issues in studying vocabulary.

Task 72: Try this little self assessment task:

ELF: This can be done online as an anonymous questionnaire.

Tick the columns accordingly.

0. I don't believe that this is an English word.

1. I've never seen this word in my life

2. I've seen it but I don't know what it means

3. I understand it in context

4. I know it and use it

	0	1	2	3	4
alternate					
compound					
discrete					
discretionary					
evolving					
expansive					
formate					
enabled					
stylised					
transitory					

We will revisit these words shortly.

Learning a word

You probably have your own list of a word's features that you focus on when you are learning a new word.

Task 73: What features are on your list?

Task 74: In the above self-assessment task, mark the parts of speech of the words you know.

A word's part of speech determines the structures and other combinations in which it functions, thereby influencing the meaning of the word, its use and often its pronunciation.

Task 75: Add some words beside them that demonstrate the POS, e.g. *discretionary* (Adj) *power*. Check your corpus as necessary.

Also note pronunciation as necessary, e.g. discretionary. The online dictionaries linked in the e-course play pronunciation.

When discussing vocabulary size at the beginning of this chapter, we saw Aitchison refer to British university students being able to *understand, and potentially use 50,000 words*, which raises the issue of ACTIVE and PASSIVE vocabulary, in layman's terms. We prefer the terms PRODUCTIVE, i.e. *use* in speaking and writing, and RECEPTIVE, i.e. *understand* when listening and reading.

Task 76: Do you feel that you have both productive and receptive use of all the words that you ticked in the self assessment task on page 88?

The difference between 3 (I understand it in context) and 4 (I know it and use it) is ultimately a statement of what you know about a word.

Task 77: Indicate the features which you need for receptive and productive vocabulary use. Add your own features at the bottom if you wish. Use this Likert scale:

(1) not necessary (2) helpful (3) essential.

There is room to comment if you wish.

ELF: This can be done as an anonymous questionnaire called Receptive and Productive Vocabulary.

Word Features	Receptive	Productive	Comment
antonyms/opposite			
colligation			
collocation			
connotation			
context			
derivation		3	
frequency			
meaning	1	1	
multiple meanings			
pronunciation			
register (formality)			
synonyms			
word family			
word forms			

As a language user aiming to improve, your responses in the above table leave no doubt as to what has to be included in your dossier when growing your vocabulary. For example, we saw on page 51, the typical objects of *confer* taken from the BNC.

Task 78: Can you remember some of these typical objects of *confer*?

There is plenty of semantic similarity between these objects. Now let us look at three dictionary definitions of *confer*:

ELF: There are screen shots in the e-learning course.

[TRANSITIVE] FORMAL

to give something such as authority, a legal right, or an honour to someone (Macmillan)

confer a title / degree / honour etc., to officially give someone a title etc, especially as a reward for something they have achieved (Longman)

2 [TRANSITIVE] **confer something (on/upon somebody)** to give somebody an award, a university degree or a particular honour or right (Oxford)

Task 79: How many of your essential word features are provided by the dictionary entries?

It is important to have sources for the *features of words* that you consider important. As we have seen, dictionaries provide different types and different amounts of information.

ELF: See the blocks in the right margin of the e-learning course for links to other sites.

Task 80: Look up *confer* at other sites and see what else you learn about *confer*.

Is *confer* among the most frequent 2,000 words in English? Is it an academic word?

Confer first came up as a collocate of *advantage*. Have a look at a list of collocates of *advantage* and consider alternative verbs. Do we *grant, award, put, give* an advantage?

Do you consider it worth adding to your productive vocabulary?

If you add it to your dossier what information will you provide there?

ELF: At this stage, you might like to join the discussion, How important is vocabulary? which is in Section 2 of the online course.

Task 81: Look up some of the self assessment task words in various dictionaries and add what you find useful to your dossiers.

Also give some consideration to what tools and processes you find most helpful.

Mindmap

A true mindmap is a visual feast, as mentioned briefly in the Preface. It typically contains a central notion surrounded by related ideas whose relationships are shown by lines and arrows. The ideas themselves are often shown pictorially rather than in words. And this is possibly how thoughts are stored in our minds. In any case, a mindmap is a useful and pleasing way of representing the company that words keep, on various levels of their relationships.

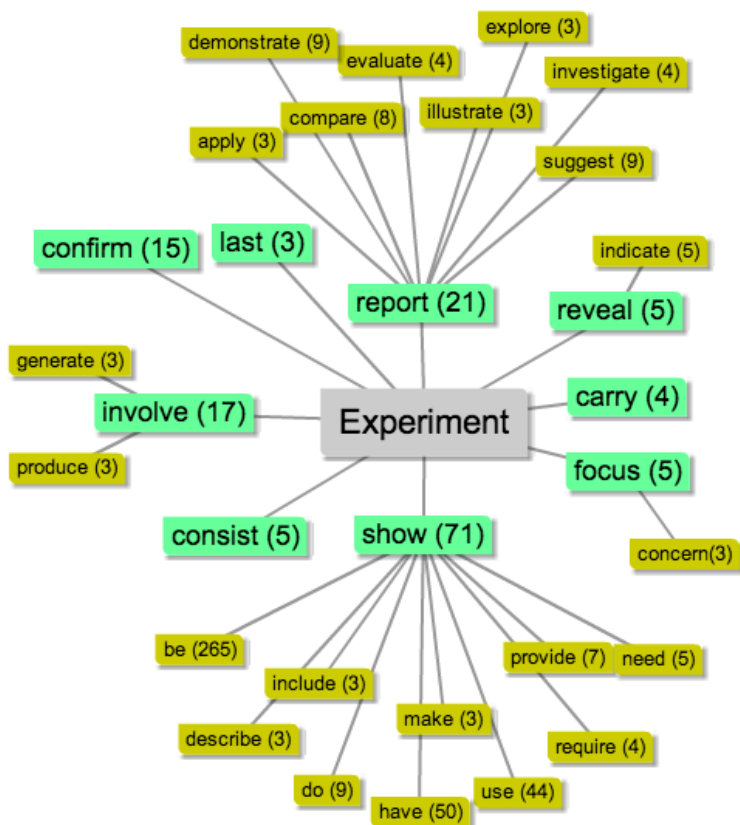
To create a mindmap, the essential ingredients are:

- its aim. This will influence its structure,
- the content, which comes partly from your head(s) and partly from other dictionaries, corpora, etc.
- the tools, e.g. Text2mindmap.com produces simple, word-based mindmaps, as you can see below.
- This mindmap diagrams the verbs which collocate with *experiment*, the noun. The first layer out from the node contains the heads word for its TROPONYMS (p.54). These are created by an algorithm in *Just the Word* and the *Sketch Engine*.
- The Clustering button can be seen on page 86. A cluster looks like this:

report 21	64	8.68	show 71	464	8.36
apply 3	compare 8		be 265	describe 3	do 9
demonstrate 9	evaluate 4		have 50	include 3	make 3
explore 3	illustrate 3		need 5	provide 7	require 4
investigate 4	suggest 9		use 44		

Task 82: Try text2mindmap. See what you learn about the word *innovation* in the process of creating a mindmap.

Once you have experimented with this, consider what you learn from this process, and if you would use it, adapt and develop it, use different software or none at all.



8. Pronunciation

Although pronunciation is not germane to academic prose *per se*, this book encourages you to learn and relearn many words, phrases, structures, frames and patterns. Your acquisition will be considerably enhanced by knowing the pronunciation of individual words as well as the stress, intonation and linking aspects of all types of word groups.

Task 1: Why do we need to improve our pronunciation? Once you have five reasons, put them in order of importance to you.

Learning the pronunciation of words and phrases is important so that you use the words with confidence. If you are about to use a word whose pronunciation you are not sure of, your fluency will be interrupted. If you mispronounce the word, your listener(s) will probably understand you, but listening is harder work for them and you do not make such a good impression. These factors are important for your professional standing.

It is also important to learn the correct pronunciation when you first learn a word so that when you are using it in various exercises, you can practise the correct pronunciation at the same time as other aspects of the word's usage.

Say words and phrases aloud. Turn back a few pages to: *Match the groups of subjects that are used with these verbs*. Say the collocations aloud. Several times. The book contains hundreds of corpus examples – read them aloud as well.

The last point here concerns your need to be an independent learner. Use internet pronunciation resources.

ELF: [Link to Howjsay provided.](#)