Chapter 1
Intelligent CALL:
Using Pattern Matching to Learn English

John Blake
https://orcid.org/0000-0002-3150-4995
University of Aizu, Japan

ABSTRACT

This chapter shows readers the importance and application of pattern matching in learning languages; specifically, the application of natural language processing to address specific problems of Japanese learners of English at a public university. The chapter introduces the concepts of patterns, detection, and detection methods. The author turns to the pedagogic application of pattern matching, first discussing the relevant theory, then describing hacks developed by language teachers and learners. The final section describes and evaluates iCALL tools developed at the University of Aizu, including a mobile app and the Pronunciation Scaffolder, a real-time presentation script annotator.

INTRODUCTION

Patterns permeate every aspect of life (Carbone, Gronov, & Prusinkiewicz, 2004) from predicting weather based on cloud formations to distinguishing friends from foe based on behaviour patterns. Some of these patterns are learnt simply through exposure, but others may involve various combinations of intuition and experience. Pattern recognition is also a key tenet in language learning and has been the focus of scholarly articles for many years. In fact, almost a century ago, Sapir (1925) published an article focusing on sound patterns. Actively assisting language learners to notice and use patterns is thought to help learners master a new language more effectively and more efficiently (Ellis, 1994). More recently, Hunston and Francis (2000) coined the term pattern grammar to describe the contextual usage patterns associated with particular word senses.

Most children begin learning their first language through extensive listening (Asher, 1972). During the receptive phase when children listen but are unable to speak, language patterns in the various interrelated language systems (e.g., phonemes, morphemes, lexis and grammar) are thought to be acquired. Second language learning may differ from first language learning depending on various factors, such

DOI: 10.4018/978-1-7998-2591-3.ch001

as the type of tuition, purpose of learning, relationship between mother tongue and second language, and the age of the learner (Cook, 2016; Johnson, 2017). For example, adult learners already literate in their mother tongue may first focus on reading a second language and start identifying visual language patterns before audio language patterns. Learners of a second language with the same script as their first language are more able to draw on their existing knowledge of orthographic patterns. Lexis in languages that are linguistically close share more in common than lexis in language that are linguistically distant. The greater the linguistic distance, the greater the difficulty to learn the second or additional language (Piske, MacKay, & Flege, 2001).

This chapter aims to show readers the importance and application of pattern recognition in learning languages; specifically, it focuses on the application of natural language processing to assist Japanese learners of English at a public university. These learners studied English for at least six years before entering university and study for four more years during their university program. English language learning was a compulsory school subject from the first year of junior high school when students are approximately thirteen years old. The vast majority have extensive passive vocabularies, general knowledge of many grammatical forms, and detailed knowledge of some obscure grammatical oddities that are often tested on university entrance exams. Despite their extensive knowledge of rules and words, few Japanese university students can converse freely in English. King (2013), having observed thirty English language classes in multiple Japanese universities concluded that there was "a robust trend, with minimal variation, toward silence" (p. 337).

This chapter begins by reviewing the origins of computer assisted language learning (CALL) and the transition to intelligent CALL. This is followed by a discussion of patterns in language and pattern detection. The chapter then introduces the concept of discovery learning and describes the pedagogic application of pattern matching. Pattern matching and detection tools that were developed specifically for learning English are then introduced. Specifically, two iCALL tools, both of which were developed at the University of Aizu are discussed in depth. The first iCALL tool is a mobile app called WordBricks and the second, the Pronunciation Scaffolder, a real-time presentation script annotator. WordBricks (Purgina, Mozgovoy, & Blake, 2019) was designed to help language learners acquire knowledge of syntactical patterns through a gamified environment in which users attempt to piece together jigsaw-like texts. The second tool is the Pronunciation Scaffolder (Blake, 2019) that targets learners of English who can read and understand texts but have difficulty in reading aloud. It was designed to help Japanese learners more easily read prepared presentation scripts using colour, size and symbols to visualize various pronunciation features. The chapter concludes by discussing future research directions and noting the increasing importance of deep learning in iCALL.

The main objectives of this chapter are to:

- introduce the general field of computer-assisted language learning (CALL),
- situate intelligent computer-assisted language learning (iCALL) within the CALL paradigm,
- define patterns by distinguishing between randomness, repetition and patterns,
- describe, explain and exemplify patterns in language,
- show how iCALL tools enhance discovery learning through pedagogic pattern matching,
- describe and evaluate two iCALL tools.

BACKGROUND OF CALL

CALL originated in the 1960s and 1970s in the days of microcomputers. Early innovators had to overcome significant resistance from teachers at the chalkface. The resistance was based on various misconceptions and fears (Roberts, 1993), including the fear of being replaced by a computer. This fear of technology threating jobs is not new. English workers in the textile industry in the early 19th century known as the Luddites vehemently opposed the introduction of machines and fell victim to technological unemployment.

In the pre-CALL days, the main technology used in language learning was the audio cassette. Well-resourced schools often had a dedicated language lab where students could access audio materials. The popularity of the audio-lingual method in the 1960s and 1970s led to a rapid increase in the number of language labs. Language labs nowadays tend to be multimedia labs in which students can access audio-visual media locally and via the internet. In the 80s and 1990s, audio and video cassette players were the technology that most classroom teachers of English as a foreign language (EFL) tended to use on a daily basis. In one of the early CALL books, (Kenning & Kenning, 1983) described how to write simple programs in BASIC. The perceived necessity to learn a programming language was a barrier than many language teachers did not attempt to overcome.

In a practical resource book for EFL teachers, Dudeney (2000) describes how to use the internet as a language learning resource. With ubiquitous access to the internet in many developed countries, there is a rapid trend to using online resources. Nowadays, language learners in Japanese universities often consult dictionaries and other language resources on their own web-enabled devices instead of asking their teachers. Teachers who used to display materials to classes by writing on whiteboards may now present materials using slideshow software, such as PowerPoint. This move helped classroom teachers realize how technology could increase the efficiency of classroom teaching. As laptop computers became more affordable and language learning websites became more user-friendly, CALL moved into the mainstream.

The prevalence of smartphones among learners has caused a shift away from accessing the internet via a computer. Almost all university students in Japan possess a smartphone (Tateno et al., 2019). This transition to online learning using handheld devices enables learners to access the same kinds of functionalities and opportunities previously only available to computer users. CALL is, therefore, no longer technically accurate. A new term mobile assisted language learning (MALL) appeared to catch this new technological and pedagogic development (for more details on CALL and MALL see Figueroa, 2020). Yet, as Microsoft and others began to develop platforms that work on different platforms, the boundary between mobile and computer interfaces is less distinct. Windows 10, for example, is designed to work with touchscreen devices in any orientation (portrait or landscape) and on any viewport size (e.g., mobile, tablet, laptop or monitor). Technology-enhanced language learning is perhaps a suitable catch-all term that does not limit its usage to one type of device.

The affordability and ubiquitous nature of smartphones in many countries has resulted in schools adopting bring-your-own-device (BYOD) classrooms, in which students learn on their preferred web-enabled device. This reduces the reliance on institutions to provide devices, but raises the issue of ensuring compatibility across the various platforms. Song and Kong (2017) provide a succinct description of the affordances and constraints of BYOD on pedagogic practices. Technical troubleshooting tends to be left to students themselves who solve connection, access and software problems on their own devices.

Intelligent CALL

The harnessing of natural language processing (NLP) to help language learners has taken CALL to a new level, in which systems appear to make intelligent decisions. This use of NLP in CALL is termed intelligent CALL (iCALL) (Finkbeiner & Knierim, 2008). Arguments in support of the use of artificial intelligence in CALL have been made since the 1980s. Bailin first published on using artificial intelligence in CALL in 1988, and later created a bibliography of iCALL (Bailin et al., 1989). To help create a shared understanding of iCALL, Oxford (1993, as cited in Levy, 1997, p. 220) enumerated nine principles for an Intelligent CALL (iCALL) framework to ensure that the focus of developers and practitioners was firmly on language, learners and learning rather than the technology.

In recent years, natural language processing with the help of artificial intelligence has made considerable progress in being able to "understand" language. Axiomatically, computers do not actually understand language, but follow algorithms to perform tasks, such as extracting information from "unstructured" raw language and placing it into "structured" language. "Unstructured" language refers to natural language that is not compartmentalized into tables that could be housed in a spreadsheet or database. "Structured" language refers to language that is compartmentalized. Each of the compartments can be referred to using coordinates. This structured language (i.e., data in tables) can be easily accessed by computer software while natural language is less easily accessed. The late Adam Kilgarriff an influential corpus and computational linguist and co-founder of Sketch Engine (Kilgarriff et al., 2014), often argued that all language is inherently structured by language systems, such as morphology and syntax (Kilgarriff, 2005).

Intelligent CALL can use natural language processing to search raw "unstructured" language (e.g., articles, books, essays) and categorize items. This categorization may help learners learn a language. Computational linguists, generally speaking, are focused on publishing cutting-edge research, which tends to mean creating a system that is more accurate, more powerful or faster than existing systems. There are, however, some researchers with a strong interest in language acquisition who are working in iCALL. Elena Volodina and her colleagues in Scandinavia have also been working extensively in the iCALL field on Scandinavian languages rather than English. Volodina, Borin, Lofsson, Arnbjörnsdóttir and Leifsson (2012) describe a system architecture that they developed to house iCALL resources and applications designed for any Scandinavian language. Natalia Bogach and her team at Peter the Great St. Petersburg Polytechnic University has been working on a computer-assisted prosody teaching tool that uses pitch visualization technology in Android mobile application (Boitsova et al., 2018). In the United States, Evgeny Chukharev-Hudilainen harnesses natural language processing to design iCALL tools that provide feedback on writing (Chukharev-Hudilainen, 2019; Chukharev-Hudilainen & Saricaoglu, 2016) and pronunciation (Qian, Chukharev-Hudilainen, & Levis, 2018). These tools are adaptive, scalable and effective, but what is particularly notable about the tools is their basis on empirically-proven pedagogic principles. Headway is also being made in Japan. The theme of the 2019 Conference of Japanese Association of Language Teachers Computer Assisted Language Learning Special Interest Group was artificial intelligence and machine learning in language education. This no doubt will act as a catalyst to researchers with proficiency in linguistics, education and computer science. Some language researchers in Japan have already developed tools that can be classed as iCALL. Gary Ross and his team at Kanazawa University are working on web speech technology and are using AI to analyze data on spoken patterns collected from an innovative online conversation practice platform in which users act out one-side of a dialogue (Ross, 2018).

Patterns

An individual instance cannot be considered a pattern. When instances are repeated, the repetitions are either random or patterned. The concept of randomness simply means that no pattern is discernable. Imagine that each word in a dictionary is written on separate slips of paper and placed in an opaque box. The box is shaken, and then ten slips of paper are picked up one by one. The words are placed in order of selection face up on a desk to create a sentence. There is a slim chance that a grammatically correct sentence is formed, but there is a substantially higher probability that the sentence is non-sensical and ungrammatical. In this case the words were selected randomly with no bias. However, what if the paper sizes differed with long words written on longer slips and short words on shorter slips? There was no bias in frequency, but this bias in surface area increases the odds that the selector can pick up large or small slips at will. Assuming a hand is used, the hand size may dictate which size pieces of paper are easier to be picked up. Various factors can creep in and bias randomness. In this example, physical factors, such as thickness and curvature of the paper, can alter the outcome. Humans are adept at seeing patterns, but conversely are poor at creating randomness. When asked to choose random numbers, random animals or create multiple choice quizzes in which the answers are randomized, we perform less well than automated systems.

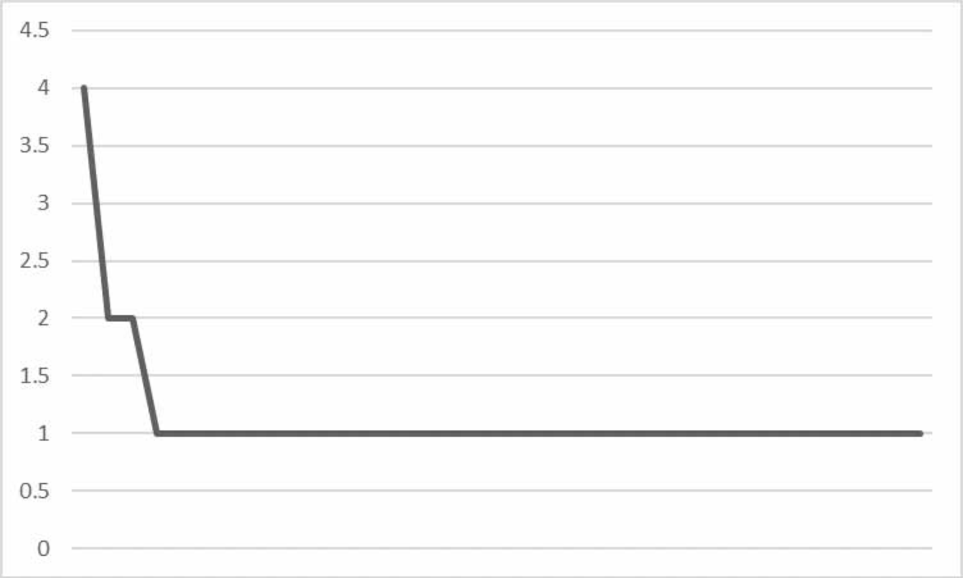
Repetition is a necessary precursor to a pattern, but repetition does not necessarily result in a pattern. The number of instances necessary to create a pattern that can be perceived varies depending on its mode, medium and complexity. For example, when learning a language, some words occur more frequently and some words occur less frequently. In English, the pattern of the most common word, the definite article *the*, is rather complex but despite the multiple reasons that can explain the usage of the definite article, the most noticeable pattern is that it is almost always followed by a noun, which may or may not be the subsequent word.

Patterns in language

Being able to recognize which words are most frequently used can help learners prioritize which words to learn first. For example, should a learner attempt to learn a language from reading a seminal book, such as learning Arabic from the Qur'an, Latin from Odes by Horace or Greek from the Odyssey by Homer. It would make little sense to focus on *hapax legomena,* that is words that only occur once in a text. However, in authorship analysis *hapax legomena* help to provide a fingerprint for the lexical repertoire of a writer (Savoy, 2012). Language courses and coursebooks are frequently designed based on the frequency with high frequency words and grammatical constructions being prioritised. This principle is found in both books aimed at children learning their mother tongues and learners of second or additional languages (Thornbury, 2002).

In an interesting twist, although a single instance of a particular word cannot be considered a pattern, there is actually a pattern formed by all the single instances of words. This pattern is so reliable that, to date, each time this pattern has been observed, the predicted result was correct. The pattern is one of the few scientific laws that relate to language, and is described by Zipf's law (Altmann & Gerlach, 2014), which is illustrated in Figure 1.

Figure 1. Plot of the frequency of words in frequency order for the example sentence



Patterns permeate language (Kilgarriff, 2005). Arguably, language is patterns. Patterns can be found at every level in the language system from phonemes and graphemes to morphemes to lexis to phases to clauses to sentences to paragraphs to texts. Table 1 provides example items and patterns from three different language systems.

Table 1. Examples of a few language systems

Language system	Example item	Example pattern
• Phonemic	• /k/	• Realized by letters c, ck, k, ch
• Morphemic	• -ment	• Verb to noun transitions (e.g., advertise, advertisement)
• Lexical	• Collocation	• Black and….. [white, blue]

Words that frequently co-occur are said to collocate. The association strength of collocations can be measured statistically and from this the probability of one word following another can be predicted (Evert, 2009). Collocation is the basis of predictive text that is a feature of not only short messaging services (e.g., texting) but is now rolled out as a feature in Google's email service, Gmail. Patterns manifest themselves in many forms. Patterns related to word length and word frequency in English are described to show how noticing patterns can help learners of English.

Word Length

Language learners pick up on prominent patterns. Learners differ in their abilities to perceive patterns with some language learners are more adept at aural, visual or spatial patterns. More visually inclined learners may notice that there is a pattern in English (and many other languages) that relates to word length. There is a length pattern in many languages with comparatively short words occurring between longer words. In English the shortest words consist of one or two letters. The grammatical classes for the majority of these words are personal pronouns (e.g., I, he, it), indefinite articles (i.e., a/an) and prepositions (e.g., in, at, on). Personal pronouns are frequently used as the first word in simple sentences, especially sentences that occur in coursebooks aimed at beginners. Indefinite articles follow a strict pattern, occurring before singular nouns. Prepositions also occur before nouns, but their syntax is not as strict as indefinite articles, and they may occur in other positions, such as when the preposition collocates with a verb.

The length pattern is based on the part of speech (POS) of words. This pattern is easy to notice and visualize. A simple script can be used that automatically annotates each word with its POS tag. Another script can be executed that colorizes the words associated with each POS tag. The result would be that words in the same POS class will appear the same colour. This can help reveal patterns that learners may not have discovered otherwise. One pattern that commonly occurs in scientific texts is the stacking of prepositional phrases (Benelhadj, 2019). Colorizing prepositions and their associated noun phrases may help show language learners how details are packed into scientific texts using multiple prepositional phrases.

The first sentence in the abstract of this chapter is reproduced below. Prepositions are shown in bold. Noun phrases following prepositions are underlined.

This chapter aims to show readers the importance and application **of** pattern matching **in** learning languages*; specifically, it will focus* **on** the application **of** natural language processing *to address specific problems* **of** Japanese learners **of** English **at** *a* public university.

The sentence with all prepositional phrases removed now reads:

This chapter aims to show readers the importance and application …; specifically, it will focus … to address specific problems ….

The general meaning of the sentence can still be inferred from the grammatical subjects, verbs and objects, but the specific details are lacking. The prepositional phrases extracted from this sentence are placed in alphabetical order here:

at a public university
in learning languages
of English
of Japanese learners
of natural language processing
of pattern matching
on the application

The pattern of a short word, prepositions in these cases, preceding noun phrases can be seen. The high frequency of prepositional phrases is typical of academic and scientific writing, and so the pattern of stacking prepositional phrases should be noticed by those reading and writing such texts.

Word Frequency

Word length is closely related to grammatical categories. Word frequency, however, trumps length in terms of importance. Most language coursebooks present vocabulary in frequency order introducing words that are more commonly used before those less commonly used words. Word frequency is a key indicator that can help learners assess whether candidate words are worth the effort of learning. Learners of English soon realize that function words are used far more frequently that content words, and so these grammatical words deserve more attention. Function words are the grammatical “glue” that hold a sentence together. These words are often guessable. The same sentence is reproduced below without the function words.

___chapter aims __ show readers ___ importance __ application __ pattern matching __ learning languages; specifically, it ___focus __ __ application __ natural language processing __ address specific problems __ Japanese learners __ English __ __ public university.

This time, the same sentence is reproduced below without the content words. These content words are unlikely to be guessed. This is because function words belong to a closed set of words while content words do not.

This ___ ____ to ___ ____ the ___ and ___ of ___ ____ in ___ ____; ___, ___ will ___ on the ___ of ___ ____ ___ to ___ ____ ___ of ___ ____ of ___ at a ___ ____.

Table 2. Table of function and content words

Function words		Content words			
• a	• at	• address	• aims	• application	• chapter
• and	• in	• English	• focus	• learners	• importance
• of	• on	• it	• Japanese	• language	• languages
• the	• this	• learners	• learning	• matching	• natural
• to	• will	• pattern	• problems	• processing	• public
		• readers	• show	• specific	• specifically
		• university			

Each word in the sentence is classified as a function or content word in Table 2.

It can be seen that the number of content words far outstrips the number of function words. In fact, three of the function words are repeated, *of* (4), *the* (2) and *to* (2), while none of the content words are repeated. Some content words are related semantically, but take different grammatical forms. Examples of semantically-related words are given here:

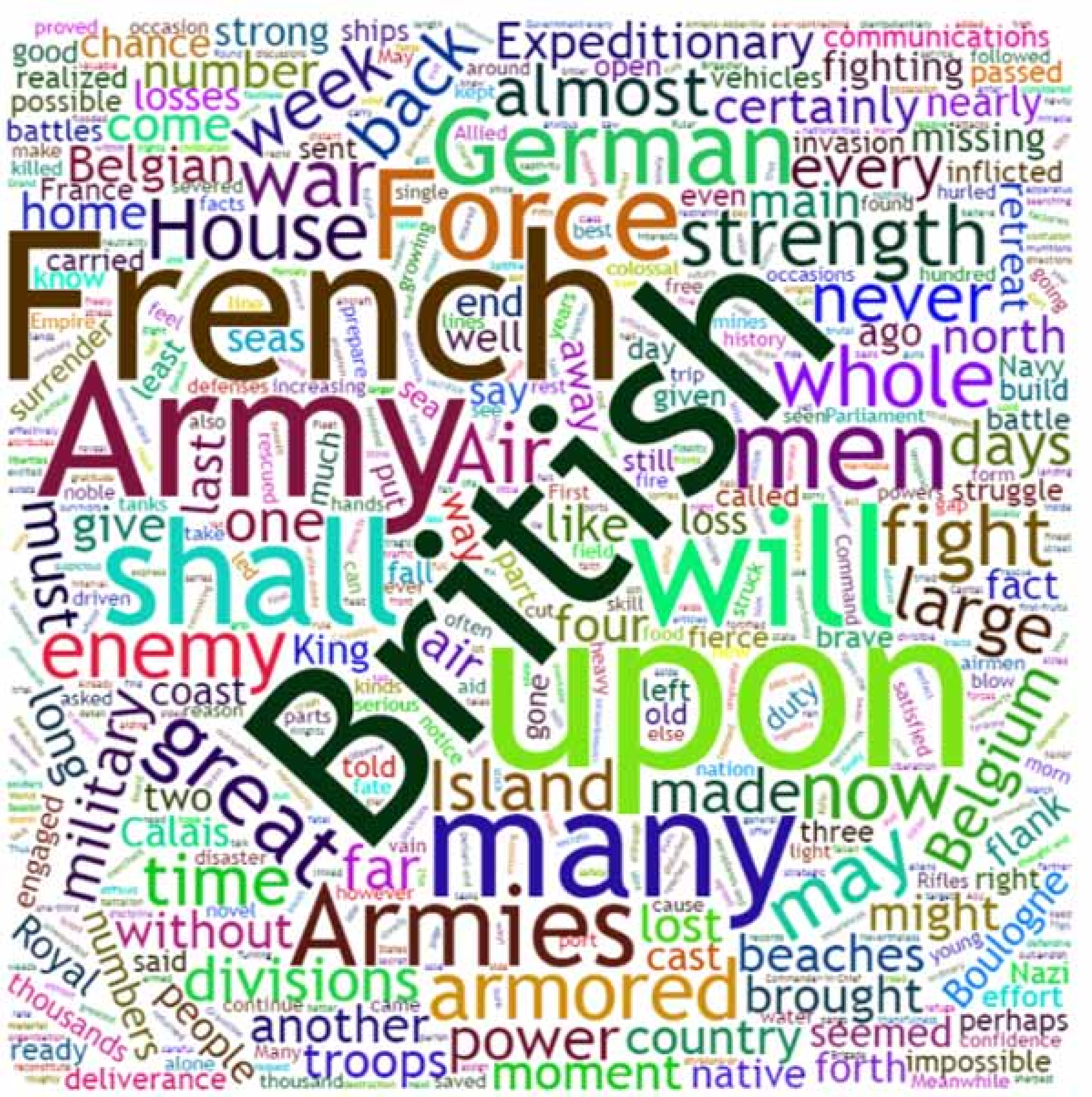
specific (adjective) cf. specifically (adverb)
language (singular noun) cf. languages (plural noun)

learning (-ing form, uncountable noun) cf. learners (plural noun)

When the words are listed in frequency order and plotted against frequency, the resultant graph is non-linear and shows an inverse relationship between frequency order and frequency (i.e., the higher the frequency, the lower the order). The graph shown in Figure 1 illustrates the non-linear pattern that is described by Zipf's law, which states that the frequency of a word is inversely proportional to its rank in frequency. This enables language users to draw upon a "miniature lexicon for efficient communication" (Kanwal, Smith, Culbertson, & Kirby, 2017). As predicted by Zipf's law, the tail of the graph contains a large proportion of *hapax legomena*, words occurring once, which are usually expected to comprise between 40 to 60% of any corpus.

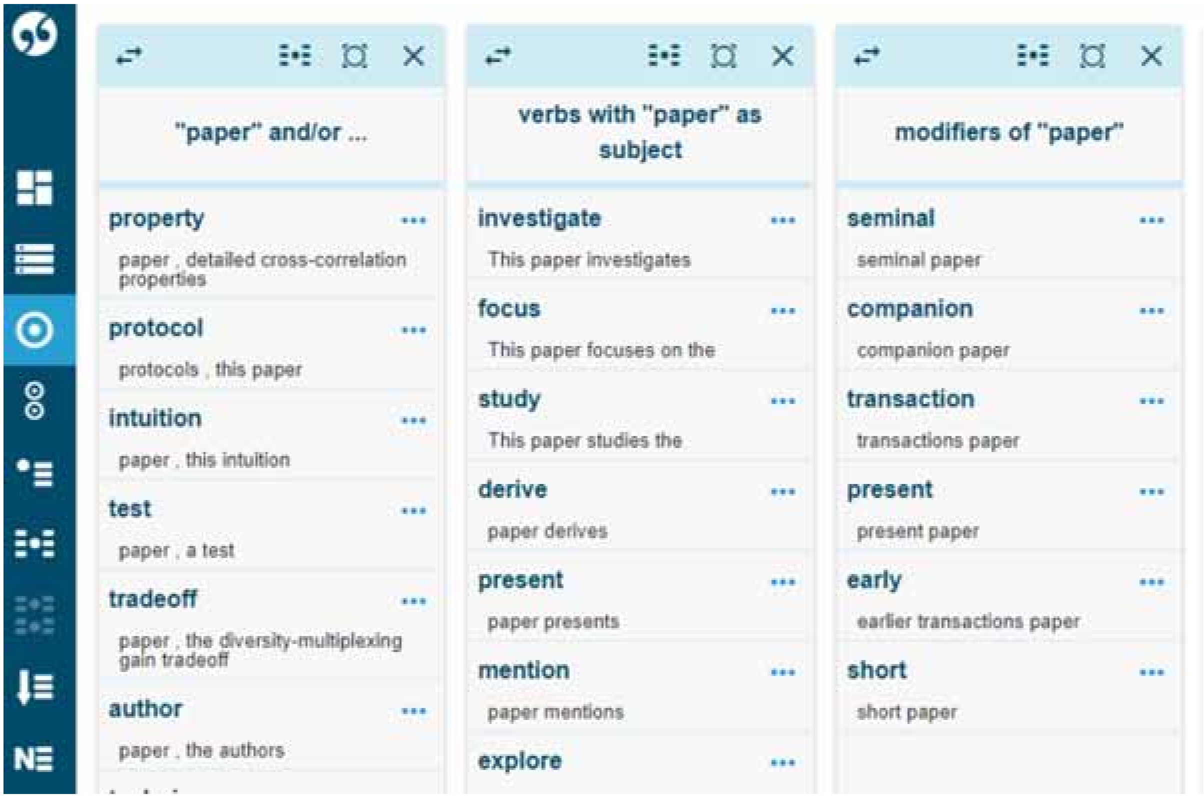
Word frequency lists can be created by using a script to count the instances of each word in a text. There are a number of online text profilers providing easy-to-use interfaces that can be used to identify not only a frequency list, but a breakdown of the words in a text according to published frequency lists, including academic word lists. Relative word frequency can be visualized using word clouds. Word clouds show the proportional frequency of content words in any text by using larger fonts for more frequent words and smaller fonts for less frequent words. To avoid function words dominating the cloud, stop lists of function words are usually utilized. Figure 2 shows a word cloud of a draft of this chapter. Word clouds can be used to provide a lexical summary of a text or corpus, priming readers with not only the specific words that appear in a text, but engaging readers to work out what the text will be about (by spotting a pattern in the words). The word cloud in Figure 2 is of a famous speech by Sir Winston Churchill, a British politician. Those with a knowledge of British history will be able to immediately recognize the speech from the word cloud.

Figure 2. Word cloud of speech by Winston Churchill (1940)



Word frequency can also be used to work out which words tend to co-occur. Sketch engine, a tool designed by Kilgarriff et al. (2014) provides word sketches using measures of association. Figure 3 shows an extract of a word sketch for the word *paper* in a tailormade corpus of scientific research abstracts in the field of information theory.

Figure 3. Word sketch for the word paper in information theory abstracts



"paper" and/or ...

property
paper , detailed cross-correlation properties

protocol
protocols , this paper

intuition
paper , this intuition

test
paper , a test

tradeoff
paper , the diversity-multiplexing gain tradeoff

author
paper , the authors

verbs with "paper" as subject

investigate
This paper investigates

focus
This paper focuses on the

study
This paper studies the

derive
paper derives

present
paper presents

mention
paper mentions

explore

modifiers of "paper"

seminal
seminal paper

companion
companion paper

transaction
transactions paper

present
present paper

early
earlier transactions paper

short
short paper

Pattern Detection

The innate ability to discover patterns helps humans learn how languages are “encoded, processed and modified” for communicative purposes (Mattson, 2014, p.7). Good language learners are thought to be good pattern matchers, who can notice systematic variations in language features, such as syntax, verb inflections and collocations (Sykes, 2015). It is only possible to learn from a pattern if the pattern is noticed. The noticing hypothesis (Schmidt, 2010) is not a new concept as it has been discussed for over thirty years. Teachers of English as an additional language tend to foreground patterns making it easier for language learners to focus on the target language. Substitution drills can be used for controlled practice to stimulate learners to recall the target pattern. Teachers may use generative situation to maximize learners’ opportunity to practice particular structures.

For example, a teacher of intermediate English learners may get students to work in pairs and underline every instance of passive voice in a newspaper article. On completion, student groups could discuss the reason for the use of passive voice for each instance. The teacher can then explain that passive voice tends to be used to focus on the action or process rather than the actor or the doer. This can be done to hide the doer because *inter alia* they are unimportant or obvious. Following this focus on receptive skills, the class could then move onto productive skills, possibly rewriting a narrative that is written in active voice. This type of learning closely matches Nation’s four stands model (Nation, 2001; Nation

& Newton, 2009) in which classroom activities are divided into meaningful input, meaningful output, language-focused instruction and fluency practice.

Various simple patterns can be discovered using a corpus and a concordancer using the key word in context (KWIC) function, which aligns instances of the same word to the center of the viewport (Scott & Tribble, 2006). The KWIC function is a simple feature commonly built into corpus software and helps users identify collocations and colligations, that is the grammatical patterns associated with a lexical item. The processing power of computers is so vast that they can discern patterns that humans are unable to detect. Subtle patterns, however, may necessitate the use of more sophisticated statistical models, such as principal component analysis and factor analysis.

DISCOVERY LEARNING

Discovery learning is premised on the noticing hypothesis. Many language learners have experience of noticing a language feature and from that moment have a sense that they learnt it. This leads many to accept the hypothesis despite the lack of experimental and empirical evidence. The hypothesis makes sense and continues to underpin many language teaching approaches and methods. Skehan (2016) notes that noticing needs to be followed by pattern detection to enable learners to generalize from the language noticed. Noticing involves the learner searching for particular language features. On discovering the language features, learners have the opportunity to learn. Discovery learning is a type of inductive learning, in which learners generate their own rules to explain discovered phenomena. Bruner (1961) advocated inquiry-based instruction for language learning. Such learning is said to be engaging, motivating and promotes autonomy. Mayer (2004) later argues convincingly that a guided learning approach is more effective than the pure discovery learning approach. DDL tends to make use of collections of texts or corpora.

Data driven learning (Johns, 2002) puts students at the centre of the learning process. Feedback from students using data driven learning approaches is not always positive. Common complaints are the difficulty of trying to understand ungraded language. As the proportion of unknown words increases in the text or corpus that they are examining, the difficulty of the task increases. Cotos, Link and Huffman (2017) found that graduate students who undertook guided DDL using a specialist corpus were able to improve the "rhetorical, formal, and procedural aspects of genre knowledge" (p. 104). Mizumoto and Chujo (2015) provide a meta-analysis of DDL approaches in EFL in Japan, and showed that in primary education focusing on vocabulary acquisition DDL was effective.

The use of corpora in lessons places the focus on authentic language and tends to adopt a frequentist perspective (Boulton, 2017; Boulton & Cobb, 2017). Words can be ranked according to raw frequency, disproportionate frequency as measured by keywords and simple KWIC concordance line results can be used. Students shift through language data for a particular purpose. Lee and Swales (2006) describe an approach that raises rhetorical consciousness through technology-enhanced interactions using specialist and self-compiled corpora. Purposes vary, but searching for particular patterns of language, such as collocation and colligation. Baker (2006) claimed that students were able to make generalizations about linguistic patterns using authentic corpus examples. To discover collocations, students can load a specialist corpus into a concordancer, such as AntConc (Anthony, 2019). Colligation tasks include searching for the typical tense or voice that as associated with particular verbs. The verb *bear*, for example, tends to be used in past simple tense and passive voice, as in "He was born yesterday." To reduce the cognitive overload, a number of researchers have tried to filter corpus results to manageable proportions (Frankenberg-Garcia, 2014; Huang, L.-S, 2008; Mizumoto & Chujo, 2015; Smart, 2014).

Pedagogic Pattern Matching

Another discovery learning approach, however, is to harness an iCALL tool that uses automatic pattern matching and visualization to focus learners on the most important patterns. Pattern matching and visualization can be used to help language learners notice and learn language features in context. The use of pattern matching tools in language learning classes enables teachers to adopt learning approaches that are data-driven or data-informed (Godwin-Jones, 2017). The pedagogic power of pattern matching is predicated on the premise that to learn a particular language feature it is necessary to notice the language feature. The advantage of using iCALL tools to learners is that the software development team has already selected the patterns worthy of attention, and so learners spend less time filtering out unnecessary language, and can focus on the target language more easily than in typical DDL.

Intelligent CALL tools can involve both top-down move-analysis approaches and bottom-up linguistic feature approaches (Cotos et al., 2017). By combining the power of natural language processing with the behaviour and interaction properties of websites and applications, learners can be automatically guided and assisted to discover patterns that were either preselected or can be discovered using an algorithm. The automated mechanism to help learners see new patterns comprises two steps: matching and manipulating. The matching, viz. the identification step, can be achieved via pattern matching or recognition. Regular expressions (Regex) can be created to search texts for specific combinations or permutations of letters or characters. Regex are, in the words of Christiansen and Torkington, (2003) akin to “mutant wildcards on steroids” (p. 180). Regex are powerful search tools that can be used to identify predetermined patterns. Once a particular pattern is matched, the discovered language features can be highlighted. This could be achieved by using JavaScript to control the behaviour of elements in a webpage to emphasize the matched language feature by altering its colour or size. For example, software designed to focus on verb phrases that users input could, on matching a verb phase execute a function that colorizes the matched string of words and provides a label stating the voice and tense of the verb phase.

As an example, consider the lexical set of quantifiers can be matched using simple regular expressions and classified into four distinct categories based on the grammatical category of nouns that they can precede. The four categories (Q1 to Q4) are listed here:

Q1: quantifiers for singular nouns, e.g.{each, every, one}
Q2: quantifiers for plural nouns, e.g. {both, many, a number of, several}
Q3: quantifiers for uncountable nouns, e.g. {much, little}
Q4: quantifiers for uncountable and plural nouns, e.g. {all, a lot of, lots, most, plenty of, some}

For each of these categories a separate regular expression can be created. When the regular expression matches any of the elements, the matched word can be highlighted in bold. Each of the four categories can be assigned a different background colour. When any element is matched, the assigned background colour can be displayed behind the matched string. The result will be an online text in which quantifiers are colour-coded to highlight the grammatical categories associated with each quantifier. This manipulated text can be used by learners from absolute beginners through to upper intermediate. The colorized texts helps learners differentiate between the four grammatical categories of quantifiers.

ICALL TOOLS FOR LANGUAGE LEARNING

There are numerous iCALL resources that use pattern matching and detection. These include online concordancing tools, word sketches and text profilers. In a Japanese university dedicated to computer science, a number of online pattern-matching tools have been developed for language learning purposes. WordBricks (Purgina et al., 2019) is a gamified mobile app designed to help learners acquire syntax by experimenting with jigsaw-like word bricks to create syntactically accurate sentences. At the same university, the Pronunciation Scaffolder was created to help undergraduate Japanese students deliver scripted presentations. This iCALL tool uses natural language processing to automatically visualize various pronunciation features (e.g., intonation, word stress, sounds) of texts submitted using colour, size and symbols. The following subsections describe and evaluate each of these tools in turn.

WordBricks

WordBricks is designed to be able to work for any language, but was developed for Japanese learners of English. Although this language learning game uses pattern matching, the motivation for the tool was to provide a gamified language learning experience by representing words or phrases in sentences as jigsaw pieces and then allow players to fit the jigsaw pieces together in order. WordBricks allows sentences that are syntactically accurate regardless of the semantics. This focus on form allows learners the freedom to play with words in the same way as children can enjoy nursery rhymes with lines like “The cow jumped over the moon …. And the dish ran away with the spoon.” Children acquiring language may replace the nouns and create their own semantically-incorrect sentences, such as “The monkey jumped over the sun.”

Languages comprise syntactic elements that combine in specific orders. Some languages, such as English, are fairly rigid in the order of words. Deviations from the expected order may be grammatically possible and treated as marked (i.e., less frequent) forms. Other deviations are treated as grammatically incorrect. WordBricks assigns syntactic elements to a particular class. Each class is attributed with a set of grammatical attributes that determine the elements that co-occur with it. For example, the verb “love” is assigned as a transitive verb. The attributes assigned could include that the actor occurs to the left of the verb in the subject position, the receiver to the right in the object position, any adverb of manner is placed after the receiver while adverbs of frequency can occur between the subject and the verb. WordBricks provides visual clues using shape and colour to show users how each syntactic element works.

This tool relies on pattern-matching. The possible permutations are derived on each placement of a brick. A script is executed that checks whether the adjoining bricks allow the placement of a brick. This can be visually deduced, since if the available position in a clause is rectangular, then only a brick that is rectangular will be permitted. The user can match the pattern visually or linguistically. The idea being that as the user gains familiarity with the bricks, mastery over the allowable syntactic combinations will increase.

WordBricks was positively received by both learners and teachers. When used for teaching grammar, the simplistic rules described in typical textbook grammars mask the true complexity of natural language. This simplification of grammar may be pedagogically justified, but does not sit well with computer scientists seeking to create expert systems that can deal with all possible use cases. Some grammatical rules are difficult to implement in WordBricks because of vaguely defined categories, semantic concepts, and the necessity for world knowledge to disambiguate different items.

WordBricks is gamified, and so this can help to overcome the lack of inertia that demotivated learners of English may display. Teenage learners of English in Japan have spent many hours completing grammar drills and rote learning vocabulary to pass university English entrance examinations. This grammar game provides a novel way to help learners review or recall grammatical structures and hone their knowledge about language, particularly the part-of-speech of words. Learners are encouraged to experiment and construct syntactically accurate sentences. At the same time, they discover how word choice affects brick type. Brick type is primarily determined by part-of-speech (POS), but the POS in turn affects the compatibility of connecting bricks. For example, adjective bricks may follow copula verbs or precede nouns, but most adjective bricks are unable to precede copula verbs or follow nouns. Descriptive grammars state various grammatical rules, but with WordBricks users experiment in a virtual lab and try various positions for words, but only positions that are grammatically possible will be allowed.

WordBricks was tested with pre-intermediate Japanese learners who were studying grammar. At the University of Aizu computer science majors enrolled in an EFL course were divided into a control group ($N = 11$) and an experimental group ($N = 10$). All students were given pre-tests and post-tests on two units both of which focus on uncountable and countable nouns, which are challenging for Japanese learners of English. The post-test for countability involved identifying grammatical errors in sentences while in the post-test for articles learners were asked to complete sentences with an appropriate noun form. During the study all participants studied with the same teacher for four months. The control group were taught about countable and uncountable nouns using a teacher-centered, grammar-focused approach that students were familiar with. The experimental group were provided with tablet PCs loaded with WordBricks with predesigned exercises, based on the same course syllabus.

The results of this preliminary study (Park, Purgina, & Mozgovoy, 2016) were positive with the experimental group scoring relatively higher than the control group in both post-tests as shown in Table 3.

Table 3. Pronunciation features

Unit	**Test**	**Group**	**Number of students**	**Mean score** *(M)*	**Standard deviation** *(SD)*
• First	• Pre-test	• Experimental	• 10	• 15.90	• 4.43
		• Control	• 11	• 15.18	• 5.04
	• Post-test	• Experimental	• 10	• 24.20	• 4.02
		• Control	• 11	• 21.00	• 5.80
• Second	• Pre-test	• Experimental	• 10	• 4.20	• 2.57
		• Control	• 11	• 6.00	• 2.72
	• Post-test	• Experimental	• 10	• 11.60	• 2.84
		• Control	• 11	• 9.18	• 4.17

Descriptive statistics of the pre- and post-tests from the first unit show that the pre-test mean scores in the experimental group ($M = 15.90$, $SD = 4.43$) and the control group ($M = 15.18$, $SD = 5.04$) were similar. In the post-test, the experimental group ($M = 24.20$, $SD = 4.02$) performed slightly better than the control group ($M = 21.00$, $SD = 5.80$). The pre-test scores differed more in the second unit with the control group having a higher mean score (*M* = 6.00, *SD* = 2.72), which increased in the post-test (*M* = 9.18 *SD* = 4.17), yet the post-test mean score in the experimental group almost tripled to 11.60 (*SD* = 2.84). These results show that the experimental group outperformed the control group on the post-test. Although there is a strong likelihood that students were engaged more deeply with WordBricks and that was the cause of the increase, sceptics could argue that other causal factors were at work.

Pronunciation Scaffolder

The Pronunciation Scaffolder (Blake, 2019) is designed to help learners of English who need to read presentation scripts aloud. The impetus for the creation of this tool was to help Japanese undergraduate students at the University of Aizu who need to deliver presentations in English to fulfil their graduation requirements. However, for students who lack confidence and practice in speaking, this is particularly difficult. Students tend to write the full script of their presentation to compensate for their lack of ability (or confidence) to produce grammatically accurate speech when reading notes. Despite having the exact wording prepared in advance, students still struggle to read the words. The Pronunciation Scaffolder does not address specific issues of public speaking, but focuses on basic reading aloud skills, such as where to pause, how long to pause for, which syllables to stress and which words carry less stress.

The Pronunciation Scaffolder provides users with the choice of which features they want visualized. There are three categories, six features and nine function buttons. Pronunciation features are divided into three categories: core, optional and advanced. The core features are recommended for lower-level learners and include pausing, intonation and content words. The optional features are word stress, sounds of the letter "s" and two graphemes, "-ed" and "th." These features are designed to provide additional assistance to Japanese learners with specific difficulties. The advanced feature is linking, which is subdivided into linking from consonant-ending words and vowel-ending words. The annotations are created using rule-based parsing to identify patterns. Once a pattern is identified, the pattern is annotated and visualized. The visualization is realized by using JavaScript to colorize the text, alter its size or insert symbols, such as arrow heads. Table 4 shows the pronunciation features that can be visualized.

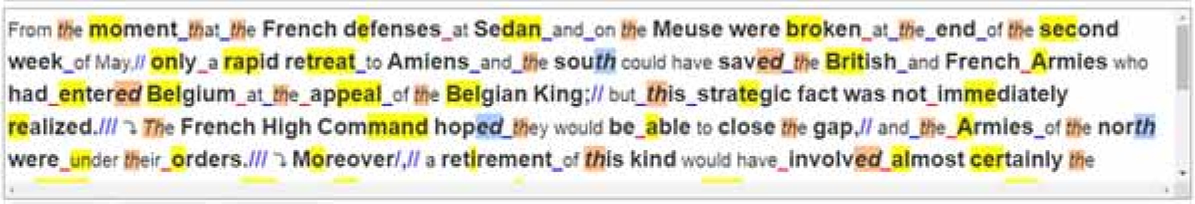
Table 4. Pronunciation features

#	Feature	Purpose of annotation
• 1	• Pausing	• show short, medium and long pauses
• 2	• Intonation	• show falling and rising intonation
• 3	• Rhythm	• stress content words more than function words
• 4	• Word stress	• locate syllable carrying primary stress
• 5	• Sounds	• disambiguate between letters realized by different sounds
• 6	• Linking	• shows elision, intrusion and linking

Figure 4 shows the political speech delivered by Sir Winston Churchill in 1940, which incidentally is the same speech used for the word cloud shown in Figure 3. Visualization is shown for pausing, intonation, content words, word stress, -ed endings, th sounds and linking from consonant-ending words.

Figure 5 shows the key to the annotations that uses colour and symbols. For example, word stress is shown by bold letters on a bright yellow background.

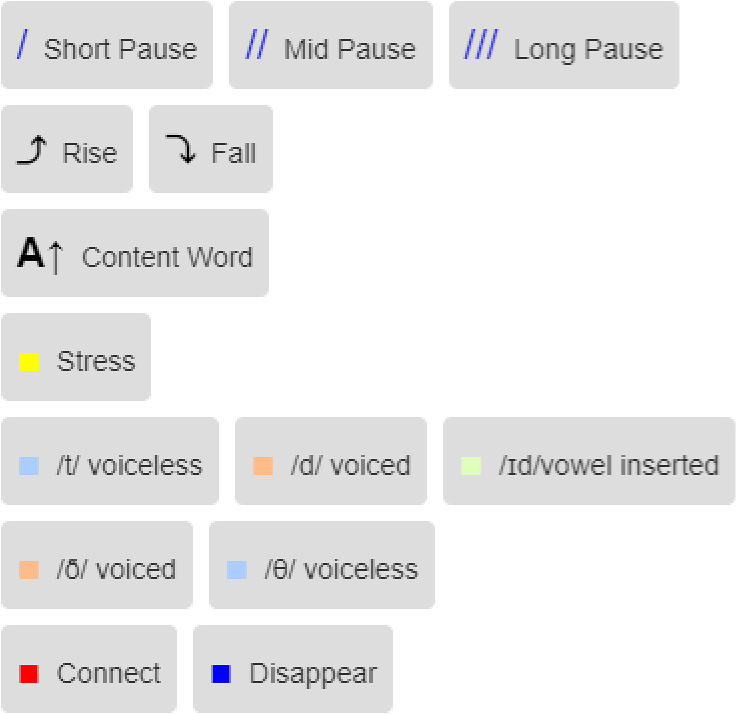
Figure 4. Speech annotated by the Pronunciation Scaffolder



From the moment_that_the French defenses_at Sedan_and_on the Meuse were broken_at_the_end_of the second
week_of May,// only_a rapid retreat_to Amiens_and_the south could have saved_the British_and French_Armies who
had_entered Belgium_at_the_appeal_of the Belgian King;// but_this_strategic fact was not_immediately
realized./// ⤵ The French High Command hoped_they would be_able to close the gap,// and_the_Armies_of the north
were_under their_orders./// ⤵ Moreover/,// a retirement_of this kind would have_involved_almost certainly the

Regular expressions parse for predetermined patterns to identify each pronunciation feature. One straightforward pattern is to identify words ending in -ed. Once the word has been identified, the -ed ending needs to be classified into one of the three possible pronunciations, namely voiced /d/, voiceless /t/ or vowel-added //ɪd/. Regular expressions were created for each of the three categories. This tool is currently unable to discriminate between words ending in -ed that have two possible pronunciations. For example, the word *learned* as an adjective is pronounced differently from the word *learned* as a verb. Should part-of-speech (POS) tagging be used prior to the rule-based parsing, this ambiguity could be resolved by matching both the POS tag and the raw text.

Figure 5. Key to the annotations used



A preliminary study of effectiveness was undertaken with twenty university students enrolled in a compulsory English course at a private university in Thailand. To test the effectiveness of version 2 of the Pronunciation Scaffolder students listened to a text-to-speech engine reading a presentation script, and then practised reading aloud. Students selected the text, web-enabled device and recording software themselves. They audio-recorded themselves reading both a raw script and an annotated script. Assessors ($N = 5$) who worked at the same university, but who were unaware of context or research aim were asked to judge the quality of the delivery of both presentations. Assessors were asked "which recording (if any) sounds better?" The rather vague terms "better" and "sounds" were chosen to avoid a framing bias in the evaluation. Recordings were played in random order. For the higher proficiency students ($N = 4$) assessors rated both recordings as being the same, but for the less proficient students ($N = 16$) assessors rated the reading of the annotated script as being better. The Pronunciation Scaffolder improved the quality of the pronunciation of the presentation for all lower-level students in this study. Although this is a small-scale study, this adds weight to anecdotal evidence that lower-level students benefit from using this tool.

FUTURE RESEARCH DIRECTIONS

This chapter has shown the ubiquity of patterns, and how harnessing pattern matching and detection can enhance data-driven language learning. The integration of NLP has brought traditional CALL into a new era of intelligent CALL. The number of language learning platforms that integrate iCALL technology is increasing rapidly. The increased access to Wi-Fi combined with year-on improvements in software and hardware provide a firm foundation for iCALL to take a more prominent role in language learning. NLP can be used to assist language learners analyze both native and learner language. This can be achieved through sophisticated NLP pipelines or relatively simple rule-based algorithms. Advances in artificial intelligence, such as neural networks and deep learning have already brought about radical breakthroughs in machine translation, and appear set to radically alter the world of technology-enhanced language learning.

Future research will focus on the development of iCALL tools, measuring their accuracy, speed and usability. These aspects are likely to be those foremost in the minds of computer scientists and computational linguists. However, the pedagogical side of iCALL tools is the primary consideration and so the effectiveness and the efficiency of the tools needs to be evaluated through both empirical and experimental studies. With any experiment study involving humans, particularly those that focus on learning processes, a particular challenge is to create a study from which statistically valid generalisations can be made. The solution to this problem is beyond the remit of this chapter, but one which needs to be addressed in order to evaluate the efficacy of iCALL.

REFERENCES

Altmann, E. G., & Gerlach, M. (2014). Statistical laws in linguistics. *Proceedings of the Flow Machines Workshop: Creativity and Universality in Language*. Academic Press. Paris, France. June 18-20, 2014.

Anthony, L. (2019). *AntConc* [computer software]. Tokyo, Japan: Waseda University. Available at http://www.antlab.sci.waseda.ac.jp/

Asher, J. J. (1972). Children's first language as a model for second language learning. *Modern Language Journal*, *56*(3), 133–139.

Bailin, A. (1988). Artificial intelligence and computer-assisted language instruction: A perspective. *CALICO Journal*, *5*(3), 25–50.

Bailin, A., Chappelle, C., Levin, L., Mulford, G., Neuwirth, C., Sanders, A., ... Underwood, J. (1989). A bibliography of intelligent computer-assisted language instruction. *Computers and the Humanities*, *23*(1), 85–90. doi:10.1007/BF00058771

Baker, P. (2006). *Using corpora in discourse analysis*. New York, NY: Continuum.

Benelhadj, F. (2019). Discipline and genre in academic discourse: Prepositional phrases as a focus. *Journal of Pragmatics*, *139*, 190–199. doi:10.1016/j.pragma.2018.07.010

Blake, J. (2019). Pronunciation Scaffolder ver. 3.0 [online tool]. Retrieved from https://jb11.org/pron-scaff.html

Boitsova, E., Pyshkin, E., Takako, Y., Bogach, N., Lezhenin, I., Lamtev, A., & Diachkov, V. (2018). StudyIntonation courseware kit for EFL prosody teaching. *Proceedings of 9th International Conference on Speech Prosody 2018* (pp. 413-417). Poznań, Poland: International Speech Communication Association. 10.21437/SpeechProsody.2018-84

Boulton, A. (2017). Research timeline: Corpora in language teaching and learning. *Language Teaching*, *50*(4), 483–506. doi:10.1017/S0261444817000167

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, *67*(2), 348–393. doi:10.1111/lang.12224

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, *31*, 21–32.

Carbone, A., Gronov, M., & Prusinkiewicz, P. (2004). *Pattern formation in biology, vision, and dynamics*. Singapore: World Scientific.

Christiansen, T., & Torkington, N. (2003). *Perl cookbook: Solutions & examples for Perl programmers*. O'Reilly Media.

Chukharev-Hudilainen, E. (2019). Empowering automated writing evaluation with keystroke logging. In E. Lindgren, & K. P. H. Sullivan (Eds.), *Observing writing: Insights from keystroke logging and handwriting* (pp. 125–142). Leiden, The Netherlands: Brill Publishing. doi:10.1163/9789004392526_007

Chukharev-Hudilainen, E., & Saricaoglu, A. (2016). Causal discourse analyzer: Improving automated feedback on academic ESL writing. *Computer Assisted Language Learning*, *29*(3), 494–516. doi:10.1080/09588221.2014.991795

Churchill, W. (1940). We shall fight on the beaches [political speech]. House of Common, UK. June 4, 1940.

Cook, V. (2016). *Second language learning and language teaching* (5th ed.). London, UK: Routledge. doi:10.4324/9781315883113

Cotos, E., Link, S., & Huffman, S. R. (2017). Effects of DDL technology on genre learning. *Language Learning & Technology*, *21*(3), 104–130.

Dudeney, G. (2000). *The Internet and the language classroom: A practical guide for teachers*. Cambridge, UK: Cambridge University Press.

Ellis, N. C. (1994). Implicit and explicit language learning. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 79–114). Amsterdam, The Netherlands: John Benjamins.

Evert, S. (2009). Corpora and collocations. In A. Lüdeling, & M. Kytö (Eds.), *Corpus linguistics: An international handbook*, 2, 1212–1248. Berlin, Germany: De Gruyter Mouton.

Figueroa, J. F. (2020). Bridging the language gap with emergent technologies. In C. Huertas-Abril, & M. Gómez-Parra (Eds.), *International approaches to bridging the language gap* (pp. 83–101). Hershey, PA: IGI Global. doi:10.4018/978-1-7998-1219-7.ch006

Finkbeiner, C., & Knierim, M. (2008). Developing L2 strategic competence online. In F. Zhang, & B. Barber (Eds.), *Handbook of research on computer-enhanced language acquisition and learning* (pp. 377–402). Hershey, PA: IGI Global; doi:10.4018/978-1-59904-895-6.ch022

Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL*, *26*(2), 128–146. doi:10.1017/S0958344014000093

Godwin-Jones, R. (2017). Data-informed language learning. *Language Learning & Technology*, *21*(3), 9–27.

Huang, L.-S. (2008). Using guided, corpus-aided discovery to generate active learning. *English Teaching Forum, 46*(4), 20-27.

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English* (Vol. 4). Amsterdam, The Netherlands: John Benjamins. doi:10.1075cl.4

Johns, T. (2002). Data-driven learning: The perpetual challenge. *Language and Computers*, *42*(1), 107–117.

Johnson, K. (2017). *An introduction to foreign language learning and teaching* (3rd ed.). London, UK: Routledge.

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, *165*, 45–52. doi:10.1016/j.cognition.2017.05.001 PMID:28494263

Kenning, M. J., & Kenning, M.-M. (1983). *An introduction to computer assisted language teaching*. Oxford, UK: Oxford University Press.

Kilgarriff, A. (2005). Language is never, ever, ever random. *Corpus Linguistics and Linguistic Theory*, *1*(2), 263–276. doi:10.1515/cllt.2005.1.2.263

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., ... Suchomel, V. (2014). The sketch engine: Ten years on. *Lexicography*, *1*(1), 7–36. doi:10.100740607-014-0009-9

King, J. (2013). Silence in the second language classrooms of Japanese universities. *Applied Linguistics*, *34*(3), 325–343. doi:10.1093/applin/ams043

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, *25*(1), 56–75. doi:10.1016/j.esp.2005.02.010

Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford, UK: Oxford University Press.

Mattson, M. P. (2014). Superior pattern processing is the essence of the evolved human brain. *Frontiers in Neuroscience*, *8*, 1–12. doi:10.3389/fnins.2014.00265 PMID:25202234

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *The American Psychologist*, *59*(1), 14–19. doi:10.1037/0003-066X.59.1.14 PMID:14736316

Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies*, *22*, 1–18.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524759

Nation, I. S. P., & Newton, J. (2009). *Speaking*. New York: Routledge.

Park, M., Purgina, M., & Mozgovoy, M. (2016). Learning English grammar with WordBricks: Classroom experience. *Proceedings of 2016 IEEE International Conference on Teaching and Learning in Education* (pp. 220-223). Washington, DC: IEEE.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215. doi:10.1006/jpho.2001.0134

Purgina, M., Mozgovoy, M., & Blake, J. (2019). WordBricks: Mobile technology and visual grammar formalism for gamification of natural language grammar acquisition. *Journal of Educational Computing Research*.

Qian, M., Chukharev-Hudilainen, E., & Levis, J. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *Language Learning & Technology*, *22*(1), 69–96.

Roberts, P. (1993). *Computer assisted language learning: RSA/UCLES diploma TEFLA distance training programme*. London, UK: International House.

Ross, G. (2018). The development of a learning management system utilizing modern mobile and modern web technologies. *Proceedings of the 12th International Technology, Education and Development Conference 2018* (pp. 9440-9446). Valencia, Spain: IATED. 10.21125/inted.2018.2336

Sapir, E. (1925). Sound patterns in language. *Language*, *1*(2), 37–51. doi:10.2307/409004

Savoy, J. (2012). Authorship attribution based on specific vocabulary. *ACM Transactions on Information Systems*, *30*(2), 1–30. doi:10.1145/2180868.2180874

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010* (pp. 721-737). Singapore: National University of Singapore, Centre for Language Studies.

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Amsterdam, The Netherlands: John Benjamins. doi:10.1075cl.22

Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 17–40). Amsterdam, The Netherlands: John Benjamins. doi:10.1075/bpa.3.02ske

Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL*, *26*(2), 184–201. doi:10.1017/S0958344014000081

Song, Y., & Kong, S. C. (2017). Affordances and constraints of BYOD (Bring Your Own Device) for learning and teaching in higher education: Teachers' perspectives. *The Internet and Higher Education*, *32*(1), 39–46. doi:10.1016/j.iheduc.2016.08.004

Sykes, A. H. (2015). The good language learner revisited: A case study. *Journal of Language Teaching and Research*, *6*(4), 713–720. doi:10.17507/jltr.0604.02

Tateno, M., Kim, D., Teo, A., Skokauskas, N., Guerrero, A., & Kato, T. (2019). Internet addiction, smartphone addiction, and Hikikomori trait in Japanese young adult: Social isolation and social network. *Psychiatry Investigation*, *16*(2), 115–120. doi:10.30773/pi.2018.12.25.2 PMID:30808117

Thornbury, S. (2002). *How to teach vocabulary*. Harlow, UK: Longman.

Volodina, E., Borin, L., Lofsson, H., Arnbjörnsdóttir, B., & Leifsson, G. Ö. (2012). Waste not; want not: Towards a system architecture for ICALL based on NLP component re-use. *Proceedings of the SLTC 2012 workshop on NLP for CALL* (pp. 47-58). Linköping University Electronic Press.

ADDITIONAL READING

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. Sebastopol, CA: O'Reilly Media.

Clark, A., Fox, C., & Lappin, S. (Eds.). (2013). *The handbook of computational linguistics and natural language processing*. New York, NY: John Wiley & Sons.

Friedl, J. E. (2006). *Mastering regular expressions*. Sebastopol, CA: O'Reilly Media.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

SaiKrishna, V., Rasool, A., & Khare, N. (2012). String matching and its applications in diversified fields. *International Journal of Computer Science Issues*, *8*(1), 219–226.

Watt, A. (2005). *Beginning regular expressions*. Indianapolis, IN: Wiley Publishing.

KEY TERMS AND DEFINITIONS

Algorithm: A set of rules to be followed.

Discovery learning: A theory in which learners reflect on their experiences to discover new ideas.

Hapax legomenon: Words that occur only once within a particular text or corpus of texts.

iCALL: iCALL stands for intelligent computer-assisted language learning, which refers to the use of natural language processing for language learning purposes.

Machine learning: When computers use artificial intelligence to learn from data or experience.

NLP: NLP stands for natural language processing, which uses computational methods to analyze natural language.

Pattern: Pattern describes items that are organized regularly not randomly.

Regex: Regex and Regexp stand for regular expressions, which are powerful search expressions that can match characters, words and/or strings.

Rule-based parsing: A process that use rules related to syntactic structure to divide written texts into components.

Zipf's law: This law states that the frequency of a word is inversely proportional to the rank in frequency of the word.