

Vocabulary acquisition:

Word structure, collocation, word-class, and meaning

Nick C. Ellis

School of Psychology,
University of Wales Bangor,
Bangor,
Gwynedd,
Wales.
LL57 2DG

e-mail: N.Ellis@bangor.ac.uk

phone: (+44) 1248 382207

fax: (+44) 1248 382599

Ellis, N. C. (1997). Vocabulary acquisition: Word structure, collocation, grammar, and meaning. In M. McCarthy & N. Schmidt (Eds.) (1997), *Vocabulary: description, acquisition and pedagogy* (pp. 122-139). Cambridge: Cambridge University Press.

Introduction

Language is sequential. Speech is a sequence of sounds. Writing is a sequence of symbols. Learning to understand a language involves parsing the speech stream into chunks which reliably mark meaning. The learner doesn't care about linguists' analyses of language. They don't care about grammar or whether words or morphemes are the atomic units of language. From a functional perspective, the role of language is to communicate meanings, and the learner wants to acquire the label-meaning relations.

This task is made more manageable by the patterns of language. Learners' attention to the evidence to which they are exposed soon demonstrates that there are recurring chunks of language. There are limited sets of sounds and of written alphabet. These units occur in more or less predictable sequences (to use written examples, in English 'e' follows 'th' more often than 'x' does, 'the' is a common sequence, 'the [space]' is frequent, 'dog' follows 'the [space]' more often than it does 'book [space]', 'how [space] do [space] you [space] do?' occurs quite often, etc.). A key task for the learner is to discover these patterns within the sequence of language. At some level of analysis, the patterns refer to meaning. It doesn't happen at the lower levels: 't' doesn't mean anything, nor does 'th', but 'the' does, and 'the dog' does better, and 'how do you do?' does very well, thank you. In these cases the learner's goal is satisfied, and the fact that this chunk activates some meaning representations makes this sequence itself more salient in the input stream.

The learner is searching for sequential patterns with reliable reference. The structure of language is such that the easiest of these to identify are words (or, more properly, morphemes) since they occur most frequently, their bounds are often marked by pauses, and their short length makes the sequence more easily learned. It is this fact which permits the present analysis of language acquisition in a collection on second language *vocabulary*. However, it should be clear from this introduction that it is only the frequency and brevity of words which gives them this privileged status, and that the learner's search for meaningful chunks is even better satisfied by larger sequences of collocations or lexical phrases.

What is it to learn a new word? Minimally we must recognise it as a word and enter it into our mental lexicon. But there are several lexicons specialised for different channels of Input/Output (I/O). To understand speech the auditory input lexicon must categorise a novel sound pattern (which will be variable across speakers, dialects, etc.); to read the word the visual input lexicon must learn to recognise a new orthographic pattern (or, in an alphabetic language, learn to exploit grapheme-phoneme correspondences in order to access the phonology and hence match the word in the auditory input lexicon); to say the word the speech output lexicon must tune a motor programme for its pronunciation; to write it the spelling output lexicon must have a specification for its orthographic sequence. We must learn its syntactic properties. We must learn its place in lexical structure: its relations with other words. We must learn its semantic properties, its referential properties, and its roles in determining entailments (for example, the word 'give' is only properly understood when we know that it relates a giver, a gift, and a recipient). We must learn the conceptual underpinnings that determine its place in our entire conceptual system. Finally we must learn the mapping of these I/O specifications to the semantic and conceptual meanings. There is no single process of learning a word. Rather these processes are logically, psychologically, and pedagogically separable.

This chapter will argue that these several different aspects of vocabulary acquisition are subserved by two separable types of learning mechanism: (i) the acquisition of a word's form, its I/O lexical specifications, its collocations, and its grammatical class information all result from predominantly unconscious (or implicit) processes of analysis of sequence information; (ii) acquisition of a word's semantic and conceptual properties, and the mapping of word form labels onto meaning representations, result from conscious (or explicit) learning processes. I will deal with each aspect in turn.

Implicit Learning of Language Form: Sequencing and chunking the speech stream

In 1971 the Advanced Research Projects Agency of the US Department of Defence (ARPA) funded several research projects to advance the art of speech recognition by computers to handle connected speech. The models which resulted five years later were compared and evaluated. Several of them (e.g. Hearsay-II, HWIM) were developed using state-of-the-art artificial intelligence (AI) inferencing techniques simulating native intelligence and native-speaker-like linguistic competence. One system, HARPY, contrasted with its rivals in rejecting logic-based techniques and instead modelled language as a statistical process - it made simple use of simple transition probabilities between linguistic units (phonemes within words, and words within sentences) (Reddy, 1990). HARPY was the only one of the systems to meet the set of performance specifications set by ARPA and it clearly outperformed the other systems. Its success has led to the now-accepted wisdom among researchers on spoken-language analysis that earlier enthusiasm for 'intelligent', top-down approaches was misplaced, and that systems which are capable of delivering sufficient results in language analysis are likely to be based on simple (e.g. Markovian) models of language structure which lend themselves to empirical statistics-gathering and sequential frequency calculation (Sampson, 1987), even though they are known to be too crude to be fully adequate as a representation of native-speaker linguistic ability (Miller & Chomsky, 1963). Indeed researchers like Newell (1980, 1990) found these results so impressive that they suggested that in addition to engineering systems for speech analysis, our psychological models of human language processing should become more statistical and less logic-based. Hence the subsequent rise of connectionist models of language processing (Rumelhart & McClelland, 1986).

These types of statistical approaches assume that our phonological memory systems automatically, and often unconsciously, abstract regularities or patterned 'chunks' from the collective evidence of the stream of speech to which we are exposed. Patterns which recur frequently in the speech stream are automatically 'chunked' because our pattern recognition systems become preferentially tuned to perceive them in future. The term 'chunking' was

coined by George Miller in his classical review of short-term memory (Miller, 1956). It is the development of permanent sets of associative connections in long-term memory and is the process which underlies the attainment of automaticity and fluency in language. Newell (1990) argues that it is the main principle of human cognition: "A chunk is a unit of memory organisation, formed by bringing together a set of already formed chunks in memory and welding them together into a larger unit. Chunking implies the ability to build up such structures recursively, thus leading to a hierarchical organisation of memory. Chunking appears to be a ubiquitous feature of human memory. Conceivably, it could form the basis for an equally ubiquitous law of practice." (Newell, 1990, p. 7). Its role in language acquisition is reviewed in McLaughlin (1987), Schmidt (1992) and Ellis (1996).

Melton (1963) showed that, when learning letter or digit sequences, the more stimuli are repeated in short-term memory (STM), the greater the long-term memory (LTM) for these items, and in turn, the easier they are to repeat as sequences in STM. This process occurs for all phonological material. Repetition of sequences in phonological STM allows their consolidation in phonological LTM. The same cognitive system which supports LTM for phonological sequences supports the perception of phonological sequences. Thus the tuning of phonological LTM to regular sequences allows more ready perception of input which contains regular sequences. Regular sequences are thus perceived as chunks and, as a result, L2-experienced individuals' phonological STM for regular L2 sequences is greater than for irregular ones (e.g. it is easier for you to perceive and repeat back the number sequence 2345666 than it is 8395327).

Such influences of LTM on STM make the relationship between these systems truly reciprocal and underlie the development of automaticity (LaBerge & Samuels, 1974; McLaughlin, 1987). Examples of these interactions in the domain of language include the effects of: long-term lexical knowledge on STM for words (Brown & Hulme, 1992), long-term phonological knowledge on STM for non- and foreign-language-words (Treiman & Danis, 1988; Gathercole & Baddeley, 1993; Ellis & Beaton, 1993b), long-term grammatical knowledge on STM for phrases (Epstein, 1967), and long-term semantic knowledge on STM for word strings (Cook, 1979).

If we are concerned with the acquisition of language form, either as perceptual units or as motor programs for output, then the ubiquitous quantitative law, *the power law of practice* applies (Anderson, 1982). The critical feature in this relationship is not just that performance, typically time, improves with practice, but that the relationship involves the power law in which the amount of improvement *decreases* as a function of increasing practice or frequency. Anderson (1982) showed that this function applies to a variety of tasks, including for example, cigar rolling, syllogistic reasoning, book writing, industrial production, reading inverted text, and lexical decision. For the case of language acquisition, Kirsner (1994) has shown that lexical recognition processes (both for speech perception and reading) and lexical production processes (articulation and writing) are equally governed by a power law relation between access time and number of exposures. Newell (1990; Newell & Rosenbloom, 1981) formally demonstrate that the following three assumptions of chunking as a learning mechanism could lead to the power law of practice. (1) People chunk at a constant rate: every time they get more experience, they build additional chunks. (2) Performance on the task is faster, the more chunks that have been built that are relevant to the task. (3) The structure of the environment implies that higher-level chunks recur more rarely. Chunks represent environmental situations. The higher the chunk in the hierarchy, the more subpatterns it has; and the more subpatterns, the less chance there is of it being true of the current situation. For example (i) at a sublexical level, if one chunk is the trigram ‘the’ and another the bigram ‘ir’ then one will see each of these situations more frequently than the higher level chunk ‘their’. For example (ii) at a supralexic level, if one chunk is the collocation ‘words in’ and another ‘their best order’, then one will see each of these situations more frequently than the higher level collocation ‘words in their best order’ (“Prose = words in their best order;—poetry = the best words in the best order.” Coleridge, *Table Talk*, 12 July 1827). That example (ii) nests example (i) within it also demonstrates this principle. These three assumptions interact as follows: the constant chunking rate and the assumption about speedup with chunking yields exponential learning. But as higher level chunks build up, they become less and less useful, because the situations in which they would help do not recur. Thus the learning slows down, being drawn out from an exponential towards a power law.

If we apply these general principles to language, then the general process of acquisition of L2 structure is as follows: Learning vocabulary involves sequencing the phonological properties of the language: the categorical units, syllable structure, and phonotactic sequences. Learning discourse involves sequencing the lexical units of the language: phrases and collocations. Learning grammar involves abstracting regularities from the stock of known lexical sequences.

Chunking and Lexical form

For the case of vocabulary acquisition, Gathercole et al. (1991, pp. 364-365) take a similar position to Melton: “Nonword repetition ability and vocabulary knowledge develop in a highly interactive manner. Intrinsic phonological memory skills may influence the learning of new words by constraining the retention of unfamiliar phonological sequences, but in addition, extent of vocabulary will affect the ease of generating appropriate phonological frames to support the phonological representations”. This is as true for second and foreign language as for native language. The novice L2 learner comes to the task with a capacity for repeating native words. This is determined by (i) constitutional factors (e.g., a good brain, especially the relevant biological substrates in the left temporal, parietal, and frontal areas), (ii) metacognitive factors (e.g. knowing that repetitive rehearsal is a useful strategy in STM tasks), (iii) cognitive factors (e.g., phonological segmentation, blending, articulatory assembly abilities). These latter language processing skills occur at an implicit level (Ellis, 1994a,b) in input and output modules which are cognitively impenetrable (Fodor, 1983) but whose functions are very much affected by experience (hence, for example, frequency and regularity effects in reading (Paap, McDonald, Schvaneveldt & Noel, 1987), spelling (Barry & Seymour, 1988), and spoken word recognition (Marslen-Wilson, 1987; Kirsner, 1994)). The degree to which such skills and knowledge (pattern recognition systems for speech sounds, motor systems for speech production) are transferable and efficient for L2 word repetition is dependent on the degree to which the allowable arrangement of phonemes in the L2

approximate to those of the native language (Ellis & Beaton, 1993a,b; Odlin, 1989). Phonotactic regularity of a novel word allows its pronunciation to better match the learner's settings of excitatory and inhibitory links between sequential phonological elements (Estes, 1972) for input processes such as phonological segmentation or for output as articulatory assembly (Snowling, Chiat & Hulme, 1991), either per se or as expectations of phonological sequences as influenced by regularities in the learner's lexicons (Gathercole et al., 1991). Either way, this long-term knowledge affects phonological STM. The reverse is also true - repetition of foreign language forms promotes long-term retention (Ellis & Beaton, 1993a; Ellis & Sinclair, in press). As learners' L2 vocabulary extends, as they practise hearing and producing L2 words, so they automatically and implicitly acquire knowledge of the statistical frequencies and sequential probabilities of the phonotactics of the L2. Their input and output modules for L2 processing begin to abstract knowledge of L2 regularities, thus to become more proficient at short-term repetition of novel L2 words. And so L2 vocabulary learning lifts itself up by its bootstraps.

Although learners need not be aware of the *processes* of such pattern extraction, they will later be aware of the *product* of these processes since on the next time they experience that pattern it is the patterned chunk that they will be aware of in working memory, not the individual components (for illustration, while young children learn about clocks they devote considerable attention to the position of hands on an analogue face in relation to the pattern of numerals; when experienced adults consult their watch they are aware of *the time*, and have no immediate access to lower-level perceptual information concerning the design of the hands or numerals; Morton, 1967).

Chunking and Collocations, Idioms, and Lexicalised Phrases

It is becoming increasingly clear that fluent language is not so completely open-class as followers of Chomsky would have us believe. Just what are the meaningful units of language acquisition (Peters, 1983)?

Sinclair (1991), as a result of his experience directing the Cobuild project, the largest lexicographic analysis of the English language to date, proposed *the principle of idiom* - “a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. To some extent this may reflect the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort; or it may be motivated in part by the exigencies of real-time conversation. However it arises, it has been relegated to an inferior position in most current linguistics, because it does not fit the open-choice model” (Sinclair, 1991, p. 110). Rather than its being a rather minor feature, compared with grammar, Sinclair suggests that for normal texts, the first mode of analysis to be applied is the idiom principle, since most of text is interpretable by this principle. Comparisons of written and spoken corpora demonstrate that collocations are even more frequent in spoken language (Butler, 1995).

Lexical phrases [or, depending on the author: holophrases (Corder, 1973), prefabricated routines and patterns (Hakuta, 1974), formulaic speech (Wong-Fillmore, 1976), memorised sentences and lexicalized stems (Pawley & Syder, 1983), or formulas (R. Ellis, 1994)] are as basic to SLA as they are to L1 (Nattinger & DeCarrico, 1989; Kjellmer, 1991; Renouf & Sinclair, 1991) and so instruction relies as much on teaching useful stock phrases as it does on teaching vocabulary and grammar. The EFL learner is introduced to phrases such as ‘Excuse me’, ‘How do you say ____ in English?’, ‘I have a headache’, etc. Phrase books provide collections of such useful utterances for purposes of everyday survival in a foreign country; ‘Naturalistic’ methods condone their acquisition because they allow the learner to enter into further conversation; ‘Audiolingual’ methods promote practice of structured collections of such patterns so that the learner might abstract structural regularities from them. Whatever the motivation, most methods encourage learners to pick up such phrases: “for a great deal of the time anyway, language production consists of piecing together the ready-made units appropriate for a particular situation and ... comprehension relies on knowing which of these patterns to predict in these situations. Our teaching therefore would

centre on these patterns and the ways they can be pieced together, along with the ways they vary and the situations in which they occur” (Nattinger, 1980, p. 341).

While language snobs may deride formulas as choreographed sequences in comparison with the creative dance of open language use, Pawley and Syder (1983) give good reason to believe that much of language is in fact closed-class. They provide two sources of evidence: native-like selection and native-like fluency. Native speakers do *not* exercise the creative potential of syntactic rules of a generative grammar (Chomsky, 1965) to anything like their full extent. Indeed if they did, they would not be accepted as exhibiting native-like control of the language. While such expressions as (1) ‘I wish to be wedded to you’, (2) ‘Your marrying me is desired by me’, and (3) ‘My becoming your spouse is what I want’, demonstrate impeccable grammatical skill, they are unidiomatic, odd, foreignisms when compared with the more ordinary and familiar (4) ‘I want to marry you’. Thus native-like selection is not a matter of syntactic rule alone. Speaking natively is speaking idiomatically using frequent and familiar collocations, and the job of the language learner is to learn these familiar word sequences. That native speakers have done this is demonstrated not only by the frequency of these collocations in the language, but also by the fact that conversational speech is broken into ‘fluent units’ of complete grammatical clauses of four to ten words, uttered at or faster than normal rates of articulation. A high proportion of these clauses, particularly of the longer ones, are entirely familiar memorised clauses and clause sequences which are the normal building-blocks of fluent spoken discourse (and at the same time provide models for the creation of (partly) new sequences which are memorable and in their turn enter the stock of familiar usages - for example ‘I’m sorry to keep you waiting’, ‘Mr. Brown is so sorry to have kept you waiting’, etc. can allow the creation of a lexicalised sentence stem ‘*NP be-tense sorry to keep-tense you waiting*’). “In the store of familiar collocations there are expressions for a wide range of familiar concepts and speech acts, and the speaker is able to retrieve these as wholes or as automatic chains from the long-term memory; by doing this he minimises the amount of clause-internal encoding work to be done and frees himself to attend to other tasks in talk-exchange, including the planning of larger units of discourse.” (Pawley & Syder, 1983, p. 192).

An important index of nativelike competence is that the learner uses idioms fluently. So language learning involves learning sequences of words (frequent collocations, phrases, and idioms) as well as sequences within words. For present purposes such collocations can simply be viewed as big words - the role of chunking in phonological memory in learning such structures is the same as for words. It is a somewhat more difficult task to the degree that these utterances are longer than words and so involve more phonological units to be sequenced. It is a somewhat less difficult task to the degree that the component parts cluster into larger chunks of frequently-encountered (at least for learners with more language experience) sequences comprising morphemes, words, or shorter collocations themselves. Despite these qualifications the principle remains the same - just as repetition aids the consolidation of vocabulary, so it does the long-term acquisition of phrases (Ellis & Sinclair, in press).

Sequence Analysis of Grammatical Word Class

But word sequences have characteristic structures all of their own, and the abstraction of these regularities is the acquisition of grammar. There are good reasons to consider that sequence information is central to the acquisition of word grammatical class. Slobin (1973) proposed that 'paying attention to the order of words and morphemes' is one of the most general of children's 'operating principles' when dealing with L1, and word order is similarly one of the four cues to part of speech in the Bates and MacWhinney (1981) Competition Model of L2 processing. More recently, Tomasello (1992) has proposed that young children's early verbs and relational terms are *individual* islands of organisation in an otherwise unorganised grammatical system - in the early stages the child learns about arguments and syntactic markings on a *verb-by-verb* basis, and ordering patterns and morphological markers learned for one verb do not immediately generalise to other verbs. Positional analysis of each verb island requires long term representations of that verb's collocations, and thus these accounts of grammar acquisition posit vast amounts of long term knowledge of word sequences. Only later are syntagmatic categories formed from abstracting regularities from

this large dataset in conjunction with morphological marker cues (at least in case-marking languages).

What might these processes of positional analysis of grammatical word class entail? I will next describe a detailed computational example for English in order to show the power of these mechanisms.

Computational models of English word-class acquisition

Kiss (1973) provided the first computational model of the acquisition of grammatical word class from accumulating evidence of word distributions. An associative learning program was exposed to an input corpus of 15,000 words gathered from tape recordings of seven Scottish middle class mothers talking to their children who were between one and three years of age. The program read the corpus and established associative links between the words and their contexts (here defined as their *immediate successor*). Thus, for example, the program counted that *the* was followed by *house* 4.1% of the time, by *horse* 3.4%, by *same* 1%, by *put* never, etc., that *a* was connected to *horse* 4.2%, to *house* 2.9%, to *put* never, etc., etc. For computational reasons (this work was done in the days of punched cards) such ‘right-context’ distributional vectors were only computed for 31 frequent words of the corpus. These vectors constituted a level of associative representation which was a network of transitions. Next a classification learning program analysed this information to produce connections between word representations which had strengths determined by the degree of similarity between the words in terms of the degree to which they tended to occur together after a common predecessor (i.e. the degree of similarity based on their ‘left-contexts’). This information formed a level of representation which was a network of word similarities. Finally the classification program analysed this similarity information to produce a third network which clustered them into groups of similar words. The clusters that arose were as follows: (*hen sheep pig farmer cow house horse*) (*can are do think see*) (*little big nice*) (*this he that it*) (*a the*) (*you I*). It seemed that these processes discovered word classes which were nounlike, verblike, adjectivelike, articlelike, pronounlike, etc. Thus the third level of representation,

which arises from simple analysis of word distributional properties, can be said to be that of word-class.

Kiss' work shows that a simple statistical system analysing sequential word probabilities can be remarkably successful in acquiring grammatical word-class information for a natural language like English. Other demonstrations include Sampson (1987), Charniak (1993), and Finch and Chater (1994).

Conclusions Concerning the Acquisition of Language Form

This section has shown that general learning mechanisms of chunking and sequence analysis, operating in the particular domain of phonological memory, allow the acquisition of (i) the phonotactic patterns of a language, (ii) word form, (iii) formulas, phrases and idioms, (iv) word collocation information, (v) grammatical word class information. As long as the speech stream is attended, then a sufficient mass of exposure will guarantee the automatic analysis of this information. However, it is important to note that acquisition can also be speeded by making the underlying patterns more salient as a result of explicit instruction or consciousness-raising (see, Ellis & Laporte, in press; Ellis, 1995; Ellis, 1996).

Such an account entails that individuals who are deficient at phonological chunking and analysis should have difficulties in the acquisition of these various aspects of lexis. There is too much evidence in support of this claim to properly review here (see Ellis, 1996; Ellis & Sinclair, in press). However, the summary facts are as follows: (i) STM (measured as the longest sequence of digits that an individual can immediately repeat in correct order) is a reliable predictor of long-term acquisition of L1 vocabulary and syntax (Ellis, 1996; Blake, Austin, Cannon, Lisus & Vaughan, 1994; Adams and Gathercole, in press; Daneman and Case 1981); (ii) phonological STM (measured as the longest sequence of nonwords that can be repeated in order) is a reliable predictor of later vocabulary acquisition in both L1 (Gathercole & Baddeley, 1990, 1993) and L2 (Service, 1992); (iii) the verbal STM deficiency of developmental dyslexic children (Ellis & Miles, 1981; Ellis, 1994c) results in poor

syntactic development both in L1 (Scarborough, 1991) and L2 (Sparks, Ganschow, Javorsky, Pohlman & Patton, 1992). To put it bluntly, learners' ability to repeat simple strings of numbers or even nonword gobbledygook is a remarkably good predictor of their ability to acquire sophisticated language skills both in L1 and L2.

Some people have difficulty acquiring lexis because of their problems in sequencing and chunking in phonological memory.

Explicit Learning of Lexical Meaning

Acquiring Word Meanings

Unlike young children learning their native language, older L2 and FL learners have already developed rich conceptual and semantic systems which are already linked to L1. In the first instance at least, the acquisition of L2 words usually involves a mapping of the new word form onto pre-existing conceptual meanings or onto L1 translation equivalents as approximations. Ijaz (1986) demonstrated that even advanced adult ESL learners are heavily influenced by native language transfer: "the second language learners essentially relied on a *semantic equivalence hypothesis*. This hypothesis facilitates the acquisition of lexical meanings in the L2 in that it reduces it to the relabelling of concepts already learned in the L1. It confounds and complicates vocabulary acquisition in the L2 by ignoring crosslingual differences in conceptual classification and differences in the semantic boundaries of seemingly corresponding words in the L1 and L2." (Ijaz, 1986, p. 443).

Even so, the L2 learner still has to determine the reference of a new label in the context of its first-noticed occurrence. There are good reasons to believe that a rich source of L2 vocabulary is the context provided during reading (Sternberg, 1987): (i) People who read more know more vocabulary. This relationship between print exposure and vocabulary appears to be causal in that it holds even when intelligence is controlled (Stanovich & Cunningham, 1992). (ii) Moderate-to-low-frequency words - precisely those words that differentiate between individuals of high and low vocabulary size - appear much more often in common reading matter than they do in common speech. (iii) There is opportunity for the reader to study the

context, to form hypotheses at leisure and cross validate them, to have time to infer meanings. The word is frozen in time on the page, whereas in speech it passes ephemerally.

But word meanings do not come from mere exposure during reading, rather, as Sternberg (1985, p. 307) argues, “simply reading a lot does not guarantee a high vocabulary. What seems to be critical is not sheer amount of experience but rather what one has been able to learn from and do with that experience. According to this view, then, individual differences in knowledge acquisition has priority over individual differences in actual knowledge.” Jensen (1980, pp. 146-147) argues this position even more strongly: “Children of high intelligence acquire vocabulary at a faster rate than children of low intelligence, and as adults they have a much larger vocabulary, not primarily because they have spent more time in study or have been more exposed to words, but because they are capable of educing more meaning from single encounters with words... The crucial variable in vocabulary size is not exposure per se, but conceptual need and inference of meaning from context, which are forms of education.” Learners can be profitably trained in strategies of education. Sternberg (1987) identified three basic subprocesses: *selective encoding* (separating relevant from irrelevant information for the purposes of formulating a definition), *selective combination* (combining relevant cues into a workable definition), and *selective comparison* (relating new information to old information already stored in memory). He categorised the types of available cue and moderating variables such as (i) the number of occurrences of the unknown word, (ii) the variability of contexts in which multiple occurrences of the unknown word appear, (iii) the importance of the unknown word to understanding the context in which it is embedded, (iv) the helpfulness of the surrounding context in understanding the meaning of the unknown word, and (v) the density of unknown words (too high a proportion of unknown words will thwart attempts to infer meaning). Subjects trained in use of these moderating variables or given practice in the processes of inferencing from context showed marked gains over control subjects in vocabulary acquisition from texts in a pretest-posttest design (see Nagy, this volume, for more on inferencing from context).

Not only does such training promote inferencing from context, but also this active derivation of meaning makes the vocabulary more memorable. Thus Hulstijn (1992) showed

that inferred word meanings were retained better than those given to the reader through the use of marginal glosses.

Unlike the processes of sequence learning involved in the acquisition of word-form, inferring the meaning of new words is neither an automatic or implicit process. It involves conscious application of strategies for searching for information, hypothesis formation and testing.

Some people have difficulty in acquiring L2 lexis because they fail properly to infer the meanings of new lexis.

Linking Meaning and Form

Whether they access the meaning by inference from context, by asking someone, or by looking the word up in a dictionary, learners must consolidate the memory of this label-meaning pair if it is not to be an ephemeral knowing. As with the analysis of meaning, there are conscious, strategic processes which can facilitate this. Repetition of label-meaning pairs gets the learner some way, but, as Bower and Winzenz (1970) showed, mnemonic strategies can take them much further. Bower and Winzenz had subjects do an vocabulary learning task which involved learning to associate 15 arbitrary pairs of words (e.g. *horse-cello*) under one of four conditions: (i) Repetition: they were asked to verbally rehearse each pair, (ii) Sentence Reading: subjects saw each pair of words in a simple sentence, and were told to read it and use it to associate the two critical words, (iii) Sentence Generation: subjects were shown each pair of words and asked to construct and say aloud a meaningful sentence relating the two words, (iv) Imagery: subjects were asked to visualise a mental picture or image in which the two referents were in some kind of vivid interaction. The mean recall results in each condition were strikingly different: Repetition 5.6, Sentence Reading 8.2, Sentence Generation 11.5, Imagery 13.1.

Imagery and semantic mnemonic strategies are thus highly effective in long-term L1 paired-associate learning. They are equally useful in L2 vocabulary learning

Imagery Mediation using Keyword Methods

Atkinson and Raugh (1975) compared learning of FL vocabulary by means of keyword mnemonics with a control condition in which subjects used their own strategies. Keyword condition subjects were presented with a Russian word and its English translation together with a word or phrase in English that sounded like the Russian word. For example, the Russian word for *battleship* is *linkór*. American subjects were asked to use the word *Lincoln*, called the keyword, to help them remember this. Subjects who used the keyword method learned substantially more translations than a control group and this advantage was maintained up to six weeks later.

In this method the first stage of recalling the meaning of a foreign word involves the subject remembering the native keyword which sounds like the foreign word. The second stage involves accessing an interactive image containing the referent of the keyword and ‘seeing’ the object with which it is associated (this is the equivalent of the Imagery mediation condition of Bower & Winzenz, 1970). By naming this object the learner accesses the native translation. This two-stage route serves as a crutch in early acquisition; with enough use the link between FL word and native translation becomes direct.

Although it is a highly effective technique (see Levin & Presley, 1985 for review), it does have some limitations: (i) it is of little use with abstract vocabulary and keywords of low imageability, (ii) it is much less effective in productive vocabulary learning than in learning to comprehend the L2 because imagery association in the keyword technique allows retrieval of a keyword which is merely an approximation to the L2 form (Ellis & Beaton, 1993a, b). The keyword technique does not have any in-built ‘mnemonic tricks’ to help spelling or pronunciation. In sum, although imagery mediation does not contribute to the lexical productive aspects of L2, it does help forge L1-L2 linkages.

Semantic Mediation

Sometimes FL words just remind us of the native word, a factor which usually stems from languages’ common origins or from language borrowing. Thus the German *Hund* (dog) may be more easily retained than the French *chien* because of its etymological and sound

similarity with the English *hound*. Such reminding, whether based on orthography, phonology, etymology or borrowing (e.g. 'le hot-dog') typically facilitates the learning of that FL word. If the reminding is not naturally there, one can create it using keywords and semantic rather than imagery mediation. By simply remembering the keyword and the native word in a mediating sentence it is possible to derive the translation (the equivalent of the Sentence Generation condition of Bower & Winzenz, 1970).

Beck, McKeown, and Omanson (1987) proposed that learners should focus on the meaning of new words and integrate them into pre-existing semantic systems. Crow and Quigley (1985) evaluated the effectiveness for ESL students of several such semantic processing strategies (such as the 'semantic field' approach where subjects manipulated synonyms along with the target words in meaningful sentences) and found them to be superior to 'traditional methods' over long time periods.

It can be advantageous to combine keyword reminders and elaborative processing. Brown and Perry (1991) contrasted three methods of instruction for Arabic students' learning of English vocabulary. The keyword condition involved presenting the new word, its definition, and a keyword, and learners were given practice in making interactive images; the semantic condition presented the new word, its definition, two examples of the word's use in sentences, and a question which they were required to answer using the new word; the keyword-semantic condition involved all of these aspects. A delayed testing over a week later demonstrated that the combined keyword-semantic strategy increased retention above the other conditions.

Stahl and Fairbanks (1986) performed a meta-analysis of nearly one hundred independent studies comparing the effectiveness of vocabulary instruction methods. This showed that the body of evaluative research to that data demonstrated (1) that vocabulary instruction is a useful adjunct to natural learning from context; (2) that the methods which produced highest effects on comprehension and vocabulary measures were those involving *both* definitional and contextual information about each to-be-learned word; (3) that several exposures were more beneficial for drill-and-practice methods; (4) that keyword methods produced consistently strong effects; and (5) that methods which provided a variety of knowledge

about each to-be-learned word from multiple contexts had a particularly good effect on later understanding of texts incorporating these words (rather than on tests which merely demanded accurate echoing of learned vocabulary definitions).

Taking their results together with the more recent ones reviewed here, it is clear that it truly matters what learners do in order to acquire the meaning of a new word. Successful learners use sophisticated metacognitive knowledge to choose cognitive learning strategies appropriate to the task of vocabulary acquisition. These include: inferring word meanings from context, semantic or imagery mediation between the FL word (or a keyword approximation) and the L1 translation, and deep processing for elaboration of the new word with existing knowledge.

Some people have difficulty acquiring lexis because they fail to use appropriate strategies for learning label-meaning associations.

Conclusions

This chapter has argued that much of language learning is the acquisition of memorised sequences of language (for vocabulary, the phonological units of language and their phonotactic sequences; for discourse, the lexical units of language and their sequences in clauses and collocations) and has demonstrated the interactions of short-term and long-term phonological memory in this learning process. Short-term representation and rehearsal allows the eventual establishment of long-term sequence information for language. In turn there are reciprocal interactions between long-term sequence representations and short-term storage whereby long-term sequence information allows the chunking of working memory contents which accord with these consolidated patterns, thus extending the span of short-term storage for chunkable materials. The better the long-term storage of frequent language sequences, the more easily they can serve as labels for meaning reference. The more automatic their access, so the more fluent is the resultant language use, concomitantly freeing attentional resources for analysis of the meaning of the message, either for comprehension or for production

planning. Finally, it is this long-term knowledge base of word sequences which serves as the database for the acquisition of language grammar.

However, the function of words is meaning and reference. And the mapping of I/O lexical form to semantic and conceptual representations is a cognitive mediation dependent upon conscious explicit learning processes. It is strongly affected by the degree to which learners engage, integrate and elaborate their semantic and conceptual knowledge. Metacognitively sophisticated language learners excel because they have cognitive strategies for inferring the meanings of words, for enmeshing them in the meaning networks of other words and concepts and imagery representations, and mapping the surface forms to these rich meaning representations. To the extent that vocabulary acquisition is about meaning, it is an explicit learning process.

References

- Adams, A-M., & Gathercole, (in press). Phonological working memory and speech production in pre-school children. *Journal of Speech and Hearing Research*.
- Anderson, J.R. (1982). Acquisition of Cognitive Skill. *Psychological Review*, **89**, 369-406.
- Atkinson, R.C. & Raugh, M.R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, **104**, 126-133.
- Barry, C., & Seymour, P.H.K. (1988). Lexical priming and sound-to-spelling contingency effects in nonword spelling. *Quarterly Journal of Experimental Psychology*, **40(A)**, 5-40.
- Bates, E., & MacWhinney, B. (1981). Second language acquisition from a functionalist perspective. In H. Winitz (Ed.), *Native Language and Foreign Language Acquisition, Annals of the New York Academy of Sciences*, **379**, 190-214.
- Beck, I.L., McKeown, M.G., & Omanson, R.C. (1987). The effects and use of diverse vocabulary instruction techniques. In M.G. McKeown & M.E. Curtis (Eds.), *The Nature of Vocabulary Acquisition* (pp. 147-163). Hillsdale, N.J.: Lawrence Erlbaum.

- Blake, J., Austin, W., Cannon, M., Lisus, A., & Vaughan, A. (1994). The relationship between memory span and measures of imitative and spontaneous language complexity in pre-school children. *International Journal of Behavioural Development*, **17**, 91-107.
- Bower, G.H. & Winzenz, D. (1970). Comparison of associative learning strategies. *Psychonomic Science*, **20**, 119-120.
- Brown, G.D.A., & Hulme, C. (1992). Cognitive psychology and second-language processing: the role of short-term memory. In R.J. Harris (Ed.) *Cognitive Processing in Bilinguals*. North Holland: Elsevier.
- Brown, T.S., & Perry, F.L. Jr. (1991). A comparison of three learning strategies for ESL vocabulary acquisition. *TESOL Quarterly*, **25**, 17-32.
- Butler, C. (1995). Between lexis and grammar: Repeated word sequences and collocational frameworks in Spanish. Paper presented to the 5th Dyffryn Conference on Vocabulary and Lexis, Cardiff, March 31-April 2, 1995.
- Charniak, E. (1993). *Statistical Language Learning*. Cambridge, Mass.: MIT Press.
- Chomsky, N. (1965). *Aspects of a Theory of Syntax*. Cambridge, Mass.: MIT Press.
- Cook, V. (1979). Aspects of memory in secondary school language learners. *Interlanguage Studies Bulletin - Utrecht*, **4**, 161-172.
- Corder, S.P. (1973). *Introducing Applied Linguistics*. Harmondsworth: Penguin.
- Crow, J.T., & Quigley, J.R. (1985). A semantic field approach to passive vocabulary acquisition for reading comprehension. *TESOL Quarterly*, **19**, 497-513.
- Daneman, M., & Case, R. (1981). Syntactic form, semantic complexity, and short-term memory: Influences on children's acquisition of new linguistic structures. *Developmental Psychology*, **17**, 367-378.
- Ellis, N.C. (Ed.) (1994a). *Implicit and Explicit Learning of Languages*. London: Academic Press.
- Ellis, N.C. (1994b). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp. 211-282). London: Academic Press.
- Ellis, N.C. (1994c). The cognitive psychology of developmental dyslexia. In G. Hales (Ed.) *Dyslexia Matters: A Celebratory Contributed Volume to Honour T.R. Miles*. London: Whurr Publishers Ltd.
- Ellis, N.C. (1995). Consciousness in second language acquisition: A review of recent field studies and laboratory experiments. *Language Awareness*, **4**, 123-146.

- Ellis, N.C. (1996). Sequencing in SLA: Phonological Memory, Chunking, and Points of Order. *Studies in Second Language Acquisition*, **18**, 91-126..
- Ellis, N.C., & Beaton, A. (1993a). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology*. **46A**, 533-558.
- Ellis, N.C., & Beaton, A. (1993b). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, **43**, 559-617.
- Ellis, N.C., & Laporte, N. (in press). Contexts of acquisition: Effects of formal instruction and naturalistic exposure on SLA. To appear in A. de Groot & J. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Ellis, N.C., & Miles, T.R. (1981). A lexical encoding deficiency I: Experimental evidence. In G. Th. Pavlidis and T.R. Miles (Eds.) *Dyslexia Research and Its Applications to Education*. Chichester: Wiley.
- Ellis, N.C., & Sinclair, S. (in press). Working Memory in the Acquisition of Vocabulary and Syntax: Putting Language in Good Order. *Quarterly Journal of Experimental Psychology. Special Issue on Working Memory*.
- Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Epstein, W. (1967). The influence of syntactical structure on learning. In N.J. Slamecka (Ed.), *Human Learning and Memory: Selected Readings* (pp. 391-395). New York: Oxford University Press.
- Estes, W.K. (1972). An associative basis for coding and organisation in memory. In A.W. Melton & E. Martin (Eds.), *Coding Processes in Human Memory*. Washington, D.C.: Winston.
- Finch, S., & Chater, N. (1994). Learning syntactic categories: A statistical approach. In M. Oaksford & G.D.A. Brown (Eds.) *Neurodynamics and Psychology*. London: Academic.
- Fodor, J.A. (1983). *The Modularity of Mind*. Cambridge, Mass.: MIT Press.
- Gathercole, S.E., & Baddeley, A.D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, **81**, 439-454.
- Gathercole, S.E., & Baddeley, A.D. (1993). *Working Memory and Language*. Hove, U.K.: Lawrence Erlbaum Associates.

- Gathercole, S.E., Willis, C., Emslie, H., & Baddeley, A.D. (1991). The influence of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, **12**, 349-367.
- Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning*, **24**, 287-298.
- Hulstijn, J. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. Arnaud & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics*. (pp. 113-125). London: Macmillan.
- Ijaz, I.H. (1986) Linguistic and cognitive determinants of lexical acquisition in a second language. *Language Learning*, **36**, 401-451.
- Jensen, A. (1980). *Bias in Mental Testing*. New York: Free Press.
- Kirsner, K. (1994). Implicit processes in second language learning. In N. Ellis (Ed.), *Implicit and Explicit Learning of Languages*. London: Academic Press.
- Kiss, G.R. (1973). Grammatical word classes: A learning process and its simulation. In G.H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*. Vol. 7. New York: Academic Press.
- Kjellmer, G. (1991). A mint of phrases. In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman.
- LaBerge, D., & Samuels, S.J. (1974). Towards a theory of automatic information processing in reading. *Cognitive Psychology*, **6**, 292-323.
- Levin, J.R., & Pressley, M. (1985). Mnemonic vocabulary acquisition: What's fact, what's fiction? In R.F. Dillon (Ed.), *Individual Differences in Cognition* (Vol. 2, pp. 145-172). Orlando, FL: Academic Press.
- Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word-recognition. In U.H. Frauenfelder & L.K. Tyler (Eds.), *Spoken Word Recognition*. Cambridge, Mass.: MIT Press.
- McLaughlin, B (1987). *Theories of Second Language Acquisition*. London: Edward Arnold.
- Melton, A.W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behaviour*, **2**, 1-21.
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, **63**, 81-97.

- Miller, G.A., & Chomsky, N. (1963). Finitary models of language users. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (pp. 419-491). New York: John Wiley and Sons.
- Morton, J. (1967). A singular lack of incidental learning. *Nature*, **215**, 203-204.
- Nattinger, J.R. (1980). A lexical phrase grammar for ESL. *TESOL Quarterly*, **14**, 337-344.
- Nattinger, J.R., & DeCarrico, J. (1989). Lexical phrases, speech acts and teaching conversation. In P. Nation & N. Carter (Eds.), *Vocabulary Acquisition, AILA Review*, **6**. Amsterdam: Free University Press.
- Newell, A. (1980). Harpy, production systems, and human cognition. In R.A. Cole (Ed.), *Perception and Production of Fluent Speech* (pp. 289-380). Hillsdale, N.J.: Erlbaum.
- Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, Mass.: Harvard University Press.
- Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), *Cognitive Skills and their Acquisition*. Hillsdale, N.J.: Erlbaum.
- Odlin, T. (1989). *Language Transfer*. Cambridge: Cambridge University Press.
- Paap, K.R., McDonald, J.E., Schvaneveldt, R.W., & Noel, R.W. (1987). Frequency and pronunciability in visually presented naming and lexical decision tasks. In M. Coltheart (Ed.) *Attention and Performance*, **XII**, London: Lawrence Erlbaum Associates.
- Pawley, A., & Syder, F.H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J.C. Richards & R.W. Schmidt (Eds.), *Language and Communication*. London: Longman.
- Peters, A.M. (1983). *The Units of Language Acquisition*. Cambridge: Cambridge University Press.
- Reddy, D.R. (1990). Machine models of speech perception. In R.A. Cole (Ed.), *Perception and Production of Fluent Speech* (pp. 215-242). Hillsdale, N.J.: Erlbaum.
- Renouf, A., & Sinclair, J.McH. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman.
- Rumelhart, D., & McClelland, J. (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. **2: Psychological and biological models** (pp. 272-326). Cambridge, Mass.: MIT Press.

- Sampson, G. (1987). Probabilistic models of analysis. In R. Garside, G. Leech & G. Sampson (Eds.), *The Computational Analysis of English*. Harlow, Essex: Longman.
- Scarborough, H.S. (1991). Early syntactic development of dyslexic children. *Annals of Dyslexia*, **41**, 207-221.
- Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, **14**, 357-385.
- Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology*, **45A**, 21-50.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C.A. Ferguson & D.I. Slobin (Eds.), *Studies of Child Language Development* (pp. 175-208). New York: Holt Rinehart Winston.
- Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, & Baddeley. *Applied Psycholinguistics*, **12**, 369-373.
- Sparks, R.L., Ganschow, L., Javorsky, J., Pohlman, J., & Patton, J. (1992). Test comparisons among students identified as high-risk, low-risk, and learning disabled in high-school foreign language courses. *Modern Language Journal*, **76**, 142-159.
- Stahl, S.A., & Fairbanks, M.M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, **56**, 72-110.
- Stanovich, K.E., & Cunningham, A.E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory and Cognition*, **20**, 51-68.
- Sternberg, R.J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence*. Cambridge: Cambridge University Press.
- Sternberg, R.J. (1987). Most vocabulary is learned from context. In M.G. McKeown & M.E. Curtis (Eds.), *The Nature of Vocabulary Acquisition* (pp. 89-105). Hillsdale, N.J.: Lawrence Erlbaum.
- Tomasello, M. (1992). *First Verbs: A Case Study of Early Grammatical Development*. Cambridge: Cambridge University Press.
- Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **14**, 145-152.

Wong-Fillmore, L. (1976). *The Second Time Around*. Unpublished doctoral dissertation, Stanford University.