CHAPTER 4

Salience in Language Usage, Learning and Change

Nick C. Ellis

4.1 Salience in Psychology, Learning Theory and Psycholinguistics

Psychological research uses the term 'salience' to refer to the property in a stimulus of standing out from the rest. Salient items or features are attended, are more likely to be perceived and are more likely than others to enter into subsequent cognitive processing and learning. Salience can be independently determined by physics and the environment and by our knowledge of the world. It is useful to think of three aspects of salience, one relating to psychophysics and the other two to what we have learned:

1 The physical world, our embodiment, and our sensory systems come together to cause certain sensations to be more intense (louder, brighter, heavier, etc.) than others. These phenomena are the subject of research in psychophysics (Gescheider 2013).
2 As we experience the world, we learn from it, and our resultant knowledge values some associations more heavily than others. We know that some stimulus cues are associated with outcomes or possibilities that are important to us, while others are negligible (James 1890a chapter 11; Gibson 1977).
3 We also have expectations about what is going to happen next in known contexts, we are surprised when our expectations are violated, and we pay more attention as a result. These phenomena are the subject of research in associative learning and cognition (Anderson 2009; Shanks 1995).

4.1.1 Three Aspects of Salience

4.1.1.1 Psychophysical Salience

Loud noises, bright lights and moving stimuli capture our attention. Salience arises in sensory data from contrasts between items and their context. These stimuli deliver intense signals in the psychophysics of our data-driven perception. Stimuli with unique features compared to those of their neighbors (Os in a field of Ts, a red poppy in a field of yellow), 'pop out' from the scene, whereas they will not in a context of shared features (Os among Qs) (Treisman and Gelade 1980). These are aspects of bottom-up processing (Shiffrin and Schneider 1977).

4.1.1.2 Salient Associations

Attention can also be driven by top-down, memory-dependent, expectation-driven processing. Emotional, cognitive and motivational factors affect the salience of stimuli. These associations make a stimulus cue 'dear'. A loved one stands out from the crowd, as does a stimulus with weighty associations ($500,000.00 vs. $0.000005, however similar the amount of pixels, characters or ink in their sensation), or one which matches a motivational state (a meal when hungry, but not when full). The units of perception are influenced by prior association: 'The chief cerebral conditions of perception are the paths of association irradiating from the sense-impression, which may have been already formed' (James 1890b: 82). Psychological salience is hugely experience-dependent: *hotdog, sushi* and 寿司 mean different things to people of different cultural and linguistic experience. This is why, *contra* sensation, the units of perception cannot be measured in physical terms. They are subjective. Hence George Miller's definition of the units of short-term memory as 'chunks': 'We are dealing here with a process of organizing or grouping the input into familiar units or chunks, and a great deal of learning has gone into the formation of these familiar units' (Miller 1956: 91; see also Chapter 6 of this volume).

4.1.1.3 Context and Surprisal

The evolutionary role of cognition is to predict what is going to happen next. Anticipation affords survival value. The Rational Analysis of Cognition (Anderson 1990, 1991) is guided by the principle that human psychology can be understood in terms of the operation of a mechanism that is 'optimally adapted' to its environment in the sense that the behavior of the mechanism is as efficient as it conceivably could be, given the structure of the problem space and the input-output mappings it must solve. We find structure in time (Elman 1990). The brain is a prediction machine (Clark 2013). One consequence is that it is surprisal, when prediction goes wrong, that maximally drives learning from a single trial. Otherwise, the regularities of the usual course of our experiences sum little by little, trial after trial, to drive our expectations. Cognition is probabilistic, its expectations a conspiracy tuned from statistical learning over our experiences (see this volume, Chapters 2 and 3 on frequency and Chapter 6 on chunking).

4.1.2 Salience and Learning

Rescorla and Wagner (1972) presented a formal model of conditioning which expresses the capacity of any cue (conditioned stimulus (CS); for example, a bell in Pavlovian conditioning) to become associated with an outcome (unconditioned stimulus (US); for example, food in Pavlovian conditioning) on any given experience of their pairing. This formula unified over the course of eighty years' research in associative learning, and it elegantly encapsulates the three factors of psychophysical salience, psychological importance and surprisal. The role of US surprise and of CS and US salience in the process of conditioning can be summarized as follows:

$$dV = ab(L - V).$$

The associative strength of the US to the CS is referred to by the letter V, and the change in this strength which occurs on each trial of conditioning is called dV. On the right-hand side, a is the salience of the US, b is the salience of the CS, and L is the amount of processing given to a completely unpredicted US. So the salience of the cue and the psychological importance of the outcome are essential factors in any associative learning. As for $(L - V)$, the more a CS is associated with a US, the less additional association the US can induce: 'But habit is a great deadener' (Beckett 1954). Alternatively, with novel associations where V is close to zero, there is much surprisal, and consequently much learning.

This is arguably the most influential formula in the history of learning theory. Physical salience, psychological salience and expectation/surprisal all affect what we learn from our experiences of the world. These factors affect what we learn about language from our language usage too, because linguistic constructions as symbolic form-function pairings are cue-outcome associations (de Saussure 1916; Bates and MacWhinney 1987a; Ellis 2006a; Robinson and Ellis 2008). Linguistic constructions as symbolic units relate the defining properties of their morphological, syntactic and lexical form with particular semantic, pragmatic and discourse functions (Goldberg 1995; Tomasello 2003; Trousdale and Hoffmann 2013).

4.1.3 Measurement and Methodology

4.1.3.1 Psychophysical Salience

Measuring the sensory/psychophysical salience of the linguistic form involves the physical factors that determine how easy it is to hear a particular structure. In his landmark study of first language acquisition, Brown breaks down the measurement of perceptual salience, or 'clarity of acoustical marking' (1973: 343), into 'such variables as amount of phonetic substance, stress level, usual serial position in a sentence, and so on' (1973: 463). Slobin (1973) proposed operating principles for child L1 acquisition suggesting that word-final morphemes are particularly salient. This resonates with the general serial-position curve relating memory to serial position, where items at the beginning and end of a list are preferentially recalled (Greene 1986).

Many grammatical form-function relationships in English, like those associated with grammatical particles and inflections such as the third-person singular *-s*, are of low salience in the language stream. This is illustrated in my companion Chapter 6 (in this volume) on chunking, in Figure 6.2, panel 2. The reason for this is the well-documented effect of frequency and entrenchment in the evolution of language: grammaticalized morphemes tend to become more phonologically fused with surrounding material because their frequent production leads to lenition processes, resulting in the loss and erosion of gestures (Zuraw 2003; Jurafsky et al. 2001; Bybee 2003, 2010; Bybee and Moder, Chapter 7 of this volume). As Slobin (1992: 191) put it, 'Somehow it's hard to keep languages from getting blurry: speakers seem to "smudge" phonology wherever possible, to delete and contract surface forms, and so forth.'

The basic principles of automatization that apply to all kinds of motor activities and skills (like playing a musical instrument, playing a sport, or cooking) are that, through repetition, sequences of units that were previously independent come to be processed as a single unit or chunk (Ellis 1996, Chapter 6 of this volume). The more frequently speakers use a form, the more they abbreviate it: this is a law-like relationship across languages (Zipf 1935). Zipf (1949) summarized this in the *principle of least effort* – speakers want to minimize articulatory effort and hence encourage brevity and phonological reduction. They tend to choose the most frequent words, and the more they use them, the more automatization of production causes their shortening. Frequently used words become shorter with use.

Grammatical functors are the most frequently used words of a language. In informal and rapid speech, this tendency to give short shrift to function words and bound morphemes, exploiting their frequency and predictability, deforms their phonetic structure and blurs the boundaries between these morphemes and the words that surround them. Of the strong syllables in a corpus examined by Cutler and Carter (1987), 86 percent occurred in open-class words and only 14 percent in closed-class words. The pattern was reversed for weak syllables, with 72 percent in closed-class words and 28 percent in open-class words.

Clitics, accent-less words or particles that depend accentually on an adjacent accented word and form a prosodic unit together with it, are the extreme examples of this: the /s/ of 'he's', /l/ of 'I'll' and /v/ of 'I've' can never be pronounced in isolation. Thus, grammatical function words and bound inflections tend to be short and low in stress, even in speech that is produced slowly and deliberately (Bates and Goodman 1997) and in speech directed to children (Goodman et al. 1990), with the result that these cues are difficult to perceive. When grammatical function words (*by, for, no, you*, etc.) are clipped out of connected speech and presented in isolation at levels where their open-class equivalents (*buy, four, know, ewe*, etc.) are perceived 90 to 100 percent correctly, adult native speakers can recognize them only 40 to 50 percent of the time (Herron and Bates 1997).

Thus, grammatical functors are extremely difficult to perceive from bottom-up evidence alone. Fluent language processors can perceive these elements in continuous speech because their language knowledge provides top-down support. But this is exactly the knowledge that learners lack. It is not surprising, therefore, that in L1 acquisition young children are unable to acquire grammatical forms until they have a critical mass of content words, providing enough top-down structure to permit perception and learning of those closed-class items that occur to the right or left of 'real words' (Bates and Goodman 1997: 51–52). As we will discuss in Section 4.2.3, the low salience of these forms is even more telling for second language learners.

4.1.3.2 Salient Associations

Measuring the psychological salience of linguistic forms is a difficult matter. It is subjective. However, we might start from William James's observations that

> Every nameable thing, act, or relation has numerous properties, qualities, or aspects. In our minds the properties of each thing, together with its name, form an associated group. If different parts of the brain are severally concerned with the several properties and a farther part with the hearing, and still another with the uttering, of the name, there must inevitably be brought about (through the law of association which we shall later study) such

> a dynamic connection among all these brain-parts that the activity of any one of them will be likely to awaken the activity of all the rest (James 1890a: 55)

and then try and measure the richness and reliability of association.

Early psychological efforts to measure meaning in these terms (e.g., Osgood 1957) can seem naïve in their focus upon individual lexical items presented out of context. They lack any linguistic analysis of phrasal meaning, sentence-level meaning, pragmatics or usage (Levinson 1995; Lyons 1995; Sinclair 1996). Nevertheless, they provide a starting point. A pioneer was Paivio (1971, 1986), who gathered ratings for words on dimensions of meaningfulness, imageability and concreteness, and then investigated the degree to which these factors affected the learning and memory for words. His ratings were guided by the operationalization that

> The psychological meaning of a stimulus pattern is defined by the total set of reactions typically evoked by it. The reactions may be verbal or nonverbal, so that the potential meaning reactions to a word would include word associations, referent images, non-verbal motor reactions, and affective reactions. (Paivio 1986: 120–121)

His measurement of imagery instructed participants that

> Nouns differ in their capacity to arouse mental images of things or events. Some words arouse a sensory experience, such as a mental picture or sound very quickly and easily . . . Any word which, in your estimation, arouses a mental image (i.e. a mental picture or sound, or other sensory experience) very quickly and easily should be given a high imagery rating. (Paivio et al. 1968: 4)

He showed that when people are asked to learn lists of words, the greater the imageability of a word, i.e. the degree to which it arouses a mental image, the more likely it is to be learned and recalled.

Another way of operationalizing the definition of meaning is to measure the 'ease of predication' of the word, i.e. the ease with which what the word refers to 'can be described by simple factual statements' (e.g. a *dog* is a type of animal, a *dog* barks when angry, a *dog* has four legs, a *dog* wags its tail when pleased, a *dog* often lives in a kennel, etc., *vs*. an *idea* . . ., Jones 1985). When Jones (1985) had subjects rate 125 nouns for ease of predication, there was a high correlation (r = 0.88) between this measure and imageability as measured by Paivio. Predication time (the mean number of seconds taken to produce two predicates for each word) correlated r = −0.72 with imageability (Jones 1988).

Paivio (1971, 1986) developed 'Dual Coding Theory' as an explanation of such findings. In this model, abstract words (words of low imageability and concreteness) have only verbal semantic representations in memory, and only these representations and those for concepts associated in meaning are accessed and activated following an episode of exposure to the word. In contrast, words with high imageability are represented not only in this semantic system but also in an imagery code, as 'sensory images awakened' (James 1890a: 265). 'Concrete terms such as *house* readily evoke both images and words as associative (meaning) reactions, whereas abstract words such as *truth* more readily arouse only verbal associations. The meaning of the latter is primarily intraverbal' (Paivio 1971: 85). His model and his method were quite remarkable, given their roots in behaviorist times. These ideas live on in more modern theories of perceptual symbol systems and grounded cognition (Barsalou 1999, 2008; Bergen and Chang 2013; Lakoff 1987). The idea is illustrated in my Chapter 6 on chunking, in Figure 6.2, panel 4. Cognitive neuropsychological studies confirm the dissociation of imagery and verbal representational systems in cases of agnosia (Warrington 1975). Brain imaging studies confirm how words evoke different modalities of imagery in different regions of the brain (Mitchell et al. 2008; Pulvermüller 1999).

Whatever you might want to add to an operationalization of salience, in terms of individual psychology regarding emotional content, physiological response, subjective report or whatever, or in terms of linguistic theory relating to phrasal meaning, sentence-level meaning, discourse meaning, pragmatics or usage, or whatever, is almost certainly sensible. But these face-valid and fairly simple methods at least allow us a start at measuring the psychological salience of linguistic items in terms of concreteness, imageability and reliability of interpretation.

4.1.3.3 Context and Surprisal

We automatically acquire knowledge of common sequences of events through implicit learning. There is considerable psychological research demonstrating that humans have dissociable, complementary systems for implicit and explicit learning and memory. Explicit learning involves attention and awareness and the generation and testing of hypotheses in a search for understanding. It is the conscious learning that we associate with education and the schoolroom. Implicit learning, in contrast, is the acquisition of knowledge about the underlying structure of a complex stimulus environment by a process which takes place naturally, automatically and without conscious operations.

It happens throughout our waking life. Simple attention to the stimulus suffices for implicit learning mechanisms to induce statistical or systematic regularities in the input environment (Ellis 1994b; Reber 1993; Stadler and Frensch 1998; Rebuschat 2015). Much of the research concentrates upon the implicit learning of sequences of behaviors (Cleeremans and McClelland 1991; Reber and Squire 1998), tones (Aslin and Newport 2012; Saffran et al. 1999) or artificial grammars (Reber et al. 1980), and these experiments show that from repeated experience of sequential behavior, learners automatically acquire knowledge of the underlying patterns of sequential dependencies. From infancy onwards, our unconscious learning systems come to predict what is likely to happen next.

Psycholinguistic research demonstrates that language users have tremendous knowledge of the sequential patterns of language at all levels (Ellis 1996, Chapter 6 this volume), despite their never having consciously counted any of these statistics in their language usage. This is true of child language learners, too (Ambridge et al. 2015). The frequency tuning under consideration here is 'computed' automatically by the learner's system during language usage. The statistics are implicitly learned and implicitly stored (Ellis 2002); learners do not have conscious access to them. Nevertheless, every moment of language cognition is informed by these data, as language learners use their model of usage to understand the actual usage of the moment as well as to update their model and predict where it is going next.

Much of the time, language processing, like walking, operates successfully using automatized, implicit processes. We only think about walking when it goes wrong, when we stumble, and conscious processes are called in to deal with the unexpected. From that episode we might learn where the uneven patch of sidewalk is, so that we will not fall again. Similarly, when language processing falters and we do not understand, we call the multi-modal resources of consciousness to help deal with the novelty. Processing becomes deliberate and slow as we 'think things through.' This one-off act of conscious processing, too, can seed the acquisition of novel explicit form-meaning associations (Ellis 2005). It allows us to consolidate new constructions as episodic 'fast-mapped' cross-modal associations (Carey and Bartlett 1978). These representations are then also available as units of implicit learning in subsequent processing. Broadly, it is not until a representation has been noticed and consolidated that the strength of that representation can thereafter be tuned implicitly during subsequent processing (Ellis 2006a, 2006b); thus, the role of noticing and consciousness in language learning (Ellis 1994b; Schmidt 1994).

Surprisal is inversely related to probability. Research operationalizations of surprisal in language involve computing norms in corpora of usage and then looking for violations of those norms. The simplest possible case is the unconditional probability (i.e. relative frequency) of, say, a word in a corpus. '*The* ... ' is less surprising than is '*Discombobulate* ... '. A slightly more complex example is a simple forward transitional probability such as the probability of the word *y* directly following the word *x* (compare *'strong tea', 'strong computers', 'powerful tea', 'powerful computers'*) or a conditional probability such as the probability of a particular verb, given a construction (*'give'* is much more likely in a ditransitive than is *'kick'*) (Gries and Stefanowitsch, 2004). Gries and Ellis (2015) review such measures. More complex applications include the conditional probability of a word, given several previous words in the same sentence or, to include a syntactic example, the conditional probability of a particular parse tree, given all previous words in a sentence (Hale 2011; Demberg and Keller 2008; Hale 2004; see also Chapter 3 in this volume on the different types of frequency effect).

Boo! However one measures the phenomenon, surprisal calls attention, and attention triggers learning. Surprise is consciousness kicking in. Consciousness is the interface (Ellis 2005). 'Paying attention – becoming conscious of some material – seems to be the sovereign remedy for learning anything, applicable to many very different kinds of information. It is the universal solvent of the mind' (Baars 1997, Section 5).

4.1.4 Acquisition and Processing

4.1.4.1 Psychophysical Salience

For first language acquisition, Brown concluded that 'some role for salience is guaranteed; the child will not learn what he cannot hear' (1973: 463), and that as a determinant of learning, salience is more important than frequency of experience. Pye (1980) found that perceptual salience, 'defined in terms of susceptibility to word and sentence stress and lack of disjuncture caused by a syllable boundary' (1980: 58), was the best predictor of children's order of acquisition of person markers in Quiche Mayan.

Goldschneider and DeKeyser (2001) performed a detailed meta-analysis of the 'morpheme order studies' that, in the 25 years following Brown's (1973) descriptions of first language acquisition, investigated the order of second language (L2) acquisition of the grammatical functors, progressive *-ing*, plural *-s*, possessive *-s*, articles *a, an, the*, third-person singular present *-s* and regular past *–ed*. These studies show remarkable commonality in the orders of acquisition of these functors across a wide range of learners of English as a second language (ESL). The meta-analysis investigated whether a combination of five determinants (perceptual salience, semantic complexity, morphophonological regularity, syntactic category and frequency) could account for the acquisition order. Scores for perceptual salience were composed of three subfactors: the number of phones in the functor (phonetic substance), the presence/absence of a vowel in the surface form (syllabicity) and the total relative sonority of the functor. The major determinants that significantly correlated with acquisition order were perceptual salience $r = 0.63$, frequency $r = 0.44$, morphophonological regularity $r = 0.41$. When these three factors were combined with semantic complexity and syntactic category in a multiple regression analysis, this combination of five predictors jointly explained 71 percent of the variance in acquisition order, with salience having the highest predictive power.

Field (2008) had second-language learners of English listen to authentic stretches of spoken English and, when pauses occurred at random intervals, they had to transcribe the last few words. The recognition of grammatical functors fell significantly behind that of lexical words, a finding that was robust across first languages and across levels of proficiency.

It is clear, therefore, that linguistic forms of low psychophysical salience are both more difficult to perceive and to learn.

4.1.4.2 Salient Associations

Child and second language acquisition (SLA) research shows that salience of association, as defined above in terms of concreteness and imageability, is a potent influence upon acquisition and learning. Brown (1973) discussed the salience of a morpheme's associations in terms of its consistency of function and semantic complexity:

> The primary determinant of the order in which mature forms are acquired is semantic complexity, forms making 'concrete' reference less complex than forms making abstract reference. Most difficult of all are the forms lacking any consistent semantic correlate. (1973: 343)

> In semantic terms, the grammatical morphemes appear to modulate the meanings of naming words, like nouns and verbs, and of the relations expressed by combining and ordering naming words. The grammatical morphemes add number, tense, aspect, specificity or nonspecificity, containment or support. These modulations are inconceivable without the major meanings they modify and for this reason alone grammatical morphemes could not be acquired before content words and rules of combination and order. (1973: 454)

Gilhooly and Logie (1980) reported norms for the age of acquisition, concreteness, and imageability of 1,944 nouns. Age-of-acquisition and concreteness of a word correlated −0.50, age of acquisition and imageability and words correlated −0.72. This is supported by SLA research: in the learning of foreign language vocabulary, imageable words are learned more easily than abstract words (0.37 > *r* > 0.53) (Ellis and Beaton 1993).

Language processing research also demonstrates robust effects of semantic richness: words associated with relatively more semantic information are recognized faster and more accurately, due to their possessing richer, better-specified semantic representations (Pexman et al. 2013). For a wide range of current evidence using methodologies including ERP, fMRI, TMS and behavioral approaches in both intact and patient populations, see Pexman et al. (2014).

The European Science Foundation (ESF) crosslinguistic and longitudinal research project (Perdue 1993) examined how 40 adult learners picked up the language of their social environment (Dutch, English, French, German and Swedish) by everyday communication. Analysis of the interlanguage of these L2 learners resulted in its being described as the 'Basic Variety'. All learners, independent of source language and target language, developed and used it, with about one-third of them fossilizing at this level, in that although they learned more words, they did not further complexify their utterances in the respects of morphology or syntax. In this Basic Variety, most lexical items stem from the target language, but they are uninflected.

> There is no functional morphology. By far most lexical items correspond to nouns, verbs and adverbs; closed-class items, in particular determiners, subordinating elements, and prepositions, are rare, if present at all . . .. Note that there is no functional inflection whatsoever: no tense, no aspect, no mood, no agreement, no casemarking, no gender assignment. (Klein 1998: 544–545)

More than half of English spontaneous speech consists of functors such as *the, of, and, a, in, to, it, is, to, was, I, for, that, you, he, be, with, on, by* and *at* (Leech et al. 2001). Their abstractness and semantic lightness makes them more difficult to acquire than concrete, imageable words. They also suffer, as Brown noted, as a result of their lack of consistency of function. Contingency of form-function mapping, as I explain in Chapter 6, Section 6.2, is a powerful determinant of learnability.

4.1.4.3 Context and Surprisal

Contemporary learning theory holds that learning is driven by prediction errors: that we learn more from the surprise that comes when our predictions are incorrect than when our predictions are confirmed (Rescorla and Wagner 1972; Wills 2009; Rumelhart et al. 1986; Clark 2013), and there is increasing evidence for surprisal-driven language acquisition and processing (Smith and Levy 2013; Demberg and Keller 2008; Dell and Chang 2014; Pickering and Garrod 2013; Jaeger and Snider 2013).

In first language acquisition, Brown (1973) considered how predictability worked against the acquisition of grammatical morphemes:

> In a face-to-face conversation between well-acquainted persons the meanings signaled by grammatical morphemes are largely redundant, they are largely guessable from linguistic and non-linguistic context. And so they are dispensable in child speech and in nonliterate adult dialects in a way that content words and word order are not. (Brown 1973: 452)

The same applies in second language acquisition. Grammatical morphemes are often redundant and overshadowed by more salient lexical cues to tense or number (e.g. *Tomorrow*, I*'ll* do the shopping; *Yesterday* I walk*ed; Seven* boy*s*) (Terrell 1991; Pica 1983). If a learner knows these lexical cues and has processed them, then subsequent processing of the morphological cues in these contexts affords no further information.

In usage-based linguistics, surprisal has been studied in particular in studies of structural priming and of online processing. Surprising structures – e.g. when a verb that is strongly attracted to the ditransitive is used in the prepositional dative – prime more strongly than non-surprising structures (Jaeger and Snider 2013; see also Chapter 8 in this volume, on priming). Analysis of a large corpus of eye-movements recorded while people read text demonstrate that measures of surprisal account for the costs in reading time that result when the current word is not predicted by the preceding context (Demberg and Keller 2008).

Contemporary corpus pattern analysis also focuses upon the tension between predictability in context and surprisal. Hanks (2011: 2) talks of norms and exploitations as the *Linguistic Double Helix*:

> Much of both the power and the flexibility of natural language is derived from the interaction between two systems of rules for using words: a primary system that governs normal, conventional usage and a secondary system that governs the exploitation of normal usage.

The *Theory of Norms and Exploitations* (Hanks 2013) is a lexically based, corpus-driven theoretical approach to how words go together in collocational patterns and constructions to make meanings. Hanks emphasizes that the approach rests on the availability of new forms of evidence (corpora, the internet) and the development of new methods of statistical analysis and inferencing. Partington (2011), in his analysis of the role of surprisal in irony, demonstrates that the reversal of customary collocational patterns (e.g. *tidings of great joy, overwhelmed*) drives phrasal irony (*tidings of great horror, underwhelmed*). Similarly, humor and jokes are based on surprisal that is pleasurable: we enjoy being led down the garden path of a predictable parse path, and then have it violated by the joke-teller.

These various approaches thus converge upon the conclusion that in language processing, surprisal aids acquisition, while redundancy hinders it.

4.2 Salience in Language Change: the Linguistic Cycle, Erosion and Grammaticalization

According to each of the three different aspects of salience considered here – psychophysical salience, salience of associations, and predictability/surprisal – grammatical morphology and grammatical functors score low. Learning theory therefore predicts that these constructions should be more difficult to acquire. These factors can be expected, in turn, to play out in language change.

From patterns of language usage, processing and acquisition, dynamic processes over diachronic timescales and synchronic states, there emerge what de Saussure (1916: 135) termed *Panchronic* principles, generalizations of language that exist independently of time, of a given language or of any concrete linguistic facts. One of these is the 'Linguistic Cycle' (Hodge 1970; Givón 1971; van Gelderen 2011), which describes paths of grammaticalization from lexical to functional category, followed by renewal. Givón (1979: 209) schematized the process as

> 'Discourse > syntax > morphology > morphophonemics > zero' and, more memorably, as 'Yesterday's syntax is today's morphology'.

Hopper and Traugott (2003, chapter 5) focus upon morphologicalization as 'Lexical item in specific syntactic context > clitic > affix', which leads in turn to 'the end of grammaticalization: loss' (Hopper and Traugott 2003: 140). Sometimes the form alone is lost; more usually, a dying form is replaced by a newer, usually periphrastic form with a similar meaning (Hopper and Traugott 2003: 172). The periphrastic replacement is salient, both psychophysically (it is several lexical items long) and, as an innovation, insofar as it is less predictable and automatic.

Some well-known examples involve negatives, where full negative phrases are reanalyzed as words and affixes and are then renewed by full phrases again. In French, negative statements were originally formed by the use of *ne* before the verb. For emphasis, *ne* often came to be reinforced by particles which once had been independent nouns (e.g. *pas* (step): *Je ne vais pas* (I'm not going), *Il ne marche pas* (he's not walking); *goute* (drop): *Je n'ai goute d'argent* (I have no money), etc.). These particles underwent grammaticalization, with *pas* assuming special status as the default neutral obligatory negative adverb (e.g. *Je ne pense pas* (I don't think so)), though before the twelfth century it was used with verbs of motion where its semantic connection is clear. In modern French, the *ne* is, as often as not, omitted entirely (e.g. *Je suis pas allée* (I didn't go)); its use depends on sociolinguistic factors such as age, gender, style of speech, phonology and clause type (Dewaele 2004), the general pattern emergent from subpatterns of regularities of usage.

Erosion has a particularly dramatic effect in sounds such as suffixes or prefixes that perform important grammatical functions. In this way, while Latin had different forms for all six combinations of person and number in the present tense, French has just three different forms for the present tense of *–er* verbs (four for *–ir*, *-re* and *–oir* type verbs), and modern English has just two.

The psychological and associative learning processes in usage reviewed in Section 4.1 affect both language learning and language change. In what follows, I summarize a psycholinguistic analysis of these processes, particularly through the lens of second language learners, whose lack of facility accelerates the process.

4.2.1 *Usage Leads to Change: Lower Salience and Homophony*

4.2.1.1 *Usage Leads to Erosion*

Frequently used words become shorter with use. As summarized in Section 4.1.3.1, considerable practice with a particular token results in automaticity of its production and sound reduction, assimilation and lenition – the loss and overlap of spoken gestures (Bybee 2003, 2006; see also Chapter 7 of this volume). Zipf's law describes the law relating frequency and length that occurs in all languages (Zipf 1935; also, see Ellis, this volume, Section 6.7). Salience eventually influences the form of language as a whole, causing some grammatical markers to 'wear away' entirely (McWhorter 2002).

4.2.1.2 Erosion Leads to Homophony

As different words shorten, they meet together into a limited number of monosyllables, so the most frequent words of the language tend also to be the most ambiguous ones (Köhler 1986; Polikarpov 2006). Many of the most frequently used words are ambiguous in their homophony and polysemy (e.g. *to, too, two; there, their, they're; I, eye, aye*): there are a large number of meanings to hang onto a limited number of short sounds. This pattern generalizes across languages: the greater the number of monosyllabic words in the lexicon of a language, the greater the degree of homophony (Ke 2006). Ambiguity is a loss of communicative capacity that arises if individual sounds are linked to more than one meaning, as in homophony and polysemy. If the absence of word ambiguity is a mark of evolutionary fitness, then word formation provides an exponential increase in fitness with length (Nowak et al. 2002; see also Chapters 12 and 13 of this volume).

4.2.2 Change Affects Learning

4.2.2.1 Low-Salience Cues Are Poorly Learned

The Rescorla–Wagner (1972) model of associative learning described in Section 4.1.2 summarizes how low-salience cues are poorly learned. As described in Section 4.1.4, grammatical morphemes and functors are low in all three aspects of salience.

The associative learning phenomenon of 'blocking' entails that the effects will be even greater for second language learners than for first language learners (Ellis and Sagarra 2010, 2011). Blocking occurs in animals and humans alike (Kruschke and Blair 2000; Kamin 1969; Mackintosh 1975; Kruschke June 2006). It describes how learners' attention to input is affected by prior experience (Shanks 1995; Rescorla and Wagner 1972; Wills 2005). Knowing that a particular stimulus is associated with a particular outcome makes it harder to learn that another cue, subsequently paired with that same outcome, is also a good predictor of it. The prior association 'blocks' further associations.

Consider the learning of tense morphology. Infants are learning meanings at the same time as words, and children learning their native language only acquire the meanings of temporal adverbs quite late in development. But adults, with their experience of the world and of their native language, know a variety of pragmatic and lexical means for expressing temporal reference (serialization: presenting events in their order of occurrence; adverbials, e.g. *soon, now*; prepositional phrases, e.g. *in the morning*; calendric reference, e.g. *May 12, Monday*, etc.) (Schumann 1987). Thus, adult language learners' expression of temporality exhibits a sequence from pragmatic to lexical to grammatical devices, and the earlier, other means, block the acquisition of the later, morphosyntactic ones: 'Whereas all learners apparently achieve the pragmatic and lexical stages of development, fewer learners achieve the morphological stage of development' (Bardovi-Harlig 2000). Lexical and serialization strategies for expressing temporal reference are salient, constant and simple to apply. Morphological cues to tense are non-salient, they often vary by person and number and typically there are additional irregularities. If, in expression, adult learners can get their message across by using these simpler strategies, they have achieved their goal. In the words of Simon (1962), they have 'satisficed' rather than 'optimized', using the minimum necessary level of formal accuracy to achieve their communicative intention, whereas optimizing upon native-like accuracy would be beyond their current cognitive bounds. Good enough (for the naturalistic world), but not perfect (for the more formal criteria of schooling).

4.2.2.2 Homophonous (Low Form-Function Contingency) Forms Are Poorly Learned

The learning of associations between cues and outcomes is a function of their contingency (Rescorla 1968; see Chapter 8 of this volume). The more reliably a cue predicts an outcome, the better the association is learned. The contingency as measured using, for example, ΔP, the one-way dependency measure of the directional association between a cue and an outcome, predicts difficulty of learning (Shanks 1995) in a wide variety of human and animal learning. This relationship is at the core of connectionist (Chater and Manning 2006; Christiansen and Chater 2001) and competition (MacWhinney 1987a) models of language learning.

Consider an ESL learner trying to learn, from the naturalistic input, the interpretation of *–s* at the ends of words. Plural *–s*, third-person singular present *–s* and possessive *–s* are all homophonous with each other as well as with the contracted allomorphs of copula and auxiliary *be*. Therefore, if we evaluate *–s* as a cue for one of these functional interpretations in particular, it is clear that there are many instances in which the cue is there but that outcome does not pertain; ΔP is, accordingly, low. Evaluate the mappings from the other direction as well: plural *–s*, third-person singular present *–s* and possessive *–s* all have variant expression as the allomorphs [/s/, /z/, /ez/]. Therefore, if we evaluate just one of these, say /ez/, as a cue for one particular outcome, say plurality, it is clear that there are many instances of that outcome in the absence of the cue; ΔP is concomitantly reduced. Thus, a contingency analysis of these cue-interpretation associations suggests that they will not be readily learnable (Ellis 2006a, 2006b; Goldschneider and DeKeyser 2001).

This is just one illustration of the general case. The ambiguity of grammatical functors out of context, their homophony and polysemy that results from high frequency of usage, erosion, desemanticization and extension, entails that they are low-contingency constructions that are difficult to learn.

4.2.2.3 Compounded Prejudices: Low Salience and Low Contingency

These simple analyses have profound consequences. If, as Herron and Bates (1997) demonstrated, fluent native speakers can only perceive grammatical functors from the bottom-up evidence of input 50 percent of the time compared to open-class words, how can language learners hear them, thence to learn their function? If the functions themselves are multiple, ambiguous and redundant, they are difficult to learn for these reasons, too.

4.2.3 Salience and Adult Naturalistic Second Language Acquisition

Consider the following sample of ESL writing, a classic piece from Lightbown and Spada (1999: 74–75) that has introduced many students to the study of SLA. It is a piece of writing intended to describe the cartoon film *The Great Toy Robbery*, written by an ESL French-speaking secondary school pupil: 'During a sunny day, a cowboy go in the desert with his horse. He has a big hat. His horse eat a flour. In the same time, Santa Clause go in a city to give some surprises.' It illustrates a classic ESL difficulty – the omission of third-person singular present tense *-s*. Third-person present *-s* and possessive *-s* are the latest-acquired functors in the morpheme order studies (Goldschneider and DeKeyser 2001; Bailey et al. 1974).

This is a specific example of the more general phenomenon that, although naturalistic second language learners are surrounded by language, not all of it 'goes in', and SLA is typically much less successful than L1A. This is Corder's distinction between input, the available target language, and intake, that subset of input that actually gets in and is utilized by the learner in some way (Corder 1967). Schmidt (1984) described a naturalistic language learner, Wes, as very fluent, with high levels of strategic competence but low levels of grammatical accuracy: 'using 90 percent correct in obligatory contexts as the criterion for acquisition, none of the grammatical morphemes counted has changed from unacquired to acquired status over a five-year period" (Schmidt 1984: 5). These grammatical functors abound in the input, but, as a result of their low salience, the low contingency of their form-function mappings and adult acquirers' learned attentional biases and L1-tuned automatized processing of language, they are simply not implicitly learned by many naturalistic learners whose attentional focus is on communication.

4.2.4 Language Use Causes Language Change

Linguistic evolution proceeds by natural selection from among the competing alternatives made available from the idiolects of individual speakers, which vary among them (Croft 2000; Mufwene 2001, 2008). Since adults are typically less successful than children at language learning, language use by a high proportion of adult language learners typically means simplification, most obviously manifested in a loss of redundancy and irregularity and an increase in transparency (Trudgill 2002b: chapter 7; Trudgill 2002a). The 'Basic Variety' of interlanguage (Klein 1998; Perdue 1993) shows similarities with pidgins (Schumann 1978) because pidgins are the languages that result from maximal contact and adult language learning (McWhorter 2001). Veronique (1999, 2001) and Becker and Veenstra (2003) detail many parallels between the grammatical structures of French-based creoles and the Basic Variety of interlanguage of learners of French as a second language, particularly in the 1:1 iconicity of their mapping of function and form (Andersen 1984), their controller-first, focus-last constituent ordering principles, their lack of verbal morphology and the order of development of their means of temporal reference. Some creoles evolve as the complexification of pidgins resulting from the habitual use by children who learn them as their native tongue. Others, such as the Atlantic and Indian Ocean French-related creoles developed from the interactions of adults speakers of nonstandard varieties of the target language and non-natives (Mufwene 2001). Creoles have systematic grammar, but not so many syntactic features as do languages like West African Fula, with its 16 grammatical genders, or morphophonological features such as the complex system of consonant mutations of Welsh, or phonological features like the tonal languages of South East Asia, all of these being languages that have had much longer to evolve their grammatical elaborations and diachronically motivated, but synchronically obscure, irregularities. Creoles typically have little or no inflection, they have little or no tone distinguishing words or expressing grammar, and their prefix/suffix + root combinations are semantically predictable (McWhorter 2001, 2002: chapter 5).

McWhorter argues that the older a language is, the more complexity it has; that is, the more it overtly signals distinctions beyond strict communicative necessity. The most elaborate languages in these respects are those older, more isolated languages that are spoken by groups of people whose interactions are primarily with other speakers of the language, and which thus are learned as native languages by children whose plastic brains are ready to optimally represent them. But their linguistic complexities pose great difficulties to second language learners, prejudiced by L1 transfer, blocking and entrenchment. So some languages are easier for adults to learn, in an absolute sense, than others: 'If one were given a month to learn a language of one's choice, I think one would select Norwegian rather than Faroese, Spanish rather than Latin, and Sranan rather than English' (Trudgill 1983). It is no accident that Faroese, as a low-contact language not subject to adult language learning, has maintained a degree of inflectional complexity which Norwegian has lost. Stasis allows a language, left to its own devices, to develop historical baggage – linguistic overgrowths that, however interesting, are strictly incidental to the needs of human exchange and expression (McWhorter 2001, 2002, 2004). In the same way that, in nature, niche-stability during the flat periods of punctuated evolution allows the continuation of elaborate vestigial forms while competition selects them out, so in language, isolation allows the slow accretion of complexity and its maintenance, while large amounts of external contact and adult language learning select out the less functional linguistic overdevelopments.

Consider again the case in point of third-person present *-s*. It weaves through this stream like a yellow rubber duck, illuminating the flow of the English language and SLA wherever it bobs. English is no longer a language spoken primarily as an L1. The 375 million L1 speakers are in a very definite minority compared to the 750 million EFL and 375 million ESL speakers (Graddol 2000). This preponderance of adult language learning of English is changing its nature. Seidlhofer (2004: 236) describes these changes as English is used across the world as a Lingua Franca. First and foremost on her list of observables is 'dropping' the third-person present tense *-s* (as in "She look very sad")'.

So languages are 'streamlined' when history leads them to be learned more as second languages than as first ones, which abbreviates some of the more difficult parts of their grammars (McWhorter 2004). As complex, adaptive systems, languages emerge, evolve and change over time (Larsen-Freeman 1997; Ellis and Larsen-Freeman 2006; Beckner et al. 2009; Croft 2000). Just as they are socially constructed, so too are they honed by use in social interaction. They adapt to their speakers. Because children are better language learners than adults, languages that adults can learn are simpler than languages that only children can learn. Second language acquisition by adults changes the very nature of language itself, in ways that are understandable in terms of the psycholinguistics of salience and general principles of associative learning (for examples in the history of English, see Schneider 2011; Schreier and Hundt 2013).

4.3 Psycholinguistics Meets Historical Linguistics

4.3.1 The Potential

'The mechanisms and principles involved in grammaticalization conform to a complex process of coding and organization of language which is universally applicable to describe the evolution of grammatical forms' (Wischer and Diewald 2002). These universal processes emerge from dynamic processes of cognition and diachrony – 'For a theory of grammaticalization, it is both unjustified and impractical to maintain a distinction between synchrony and diachrony' (Heine et al. 1991: chapter 9) – and of usage and social-interaction – 'Grammar is not absolutely formulated and abstractly represented, but always anchored in the specific form of an utterance . . . Its forms are not fixed templates, but are negotiable in face-to-face interaction in ways that reflect individual speakers' past experience of these forms, and their assessment of the present context' (Hopper 1998: 142).

Complexity arises in systems via incremental changes, based on locally available resources, rather than via top-down direction or deliberate movement toward some goal (see, e.g., Dawkins 1985). Similarly, in a complex systems framework, language is viewed as an extension of numerous domain-general cognitive capacities such as shared attention, imitation, sequential learning, chunking and categorization (Bybee 1998; Ellis 1996; Beckner et al. 2009).

Language is emergent from ongoing human social interactions, and its structure is fundamentally molded by the pre-existing cognitive abilities, processing idiosyncrasies and limitations and general and specific conceptual circuitry of the human brain. Because this has been true in every generation of language users from its very origin, in some formulations language is said to be a form of cultural adaptation to the human mind, rather than the result of the brain adapting to process natural language grammar (Deacon 1997; Schoenemann 2005; Christiansen and Chater

2008). Recognition of language as a complex adaptive system allows us to better understand change across levels and timescales (MacWhinney and O'Grady 2015). That is why enterprises like this, which bring linguists and psychologists together, have potential.

Language and usage are like the shoreline and the sea.

4.3.2 *The Difficulties*

Of course there are difficulties. Our methods are different. We come from different theoretical traditions. Language on the page, the necessary focus for historical language change, is a far cry from patterns of online language processing in discourse. Indeed, written language bestows salience on matters that are not so in spoken language, and has the power to freeze or slow language change. We cannot track online consciousness on the page. Surprisal, as well as the other operationalizations of salience described in this chapter, are hard to identify in print. Nor do children learn language from books. But scholars of language change, as well as cognitivists, agree that usage matters. We share a focus on corpus analysis. And in this volume, we focus upon the same phenomena. Something useful has to come from this.

4.4 Conclusion

It is always entertaining to hear the reports from the American Dialect Society (ADS), *Merriam-Webster* or the *Oxford English Dictionary* for candidates in their 'Word of the Year' (WOTY) or 'Phrase of the Year'. The new arrivals are current, creative and fun. Favorites of mine include *Recombobulation area* (an area at airport security in which passengers who have passed through screening can get their clothes and belongings back in order), *gate lice* (airline passengers who crowd around a gate waiting to board), *Dracula sneeze* (covering one's mouth with the crook of one's elbow when sneezing, as in popular portrayals of the vampire Dracula, in which he hides the lower half of his face with a cape), *selfie* (a self-portrait photograph, typically taken with a hand-held digital camera or camera phone), and *omnishambles* (a situation that has been comprehensively mismanaged). Words or phonemes can be blended into a portmanteau, with two meanings packed into one word (e.g. *motel*, blending *motor* and *hotel; electrocute*, blending *electric* and *execute*). They thrive on witty analogy, 'the fuel and fire of thinking' (Hofstadter and Sander 2013). The ADS WOTY for 2013 was *because* (introducing a noun, adjective or other part of speech (e.g. 'because reasons', 'because awesome')), a language change involving the loss of functors (e.g. 'because *of* reasons', 'because *it's* awesome'). What is unusual is that the passing of these functors was noticed at all. WOTY typically focuses upon the shock of the new, rather than the decline of the old.

At the start of a Linguistic Cycle, highly salient, new constructions enter a language. They are psychophysically intense. They are full of meaning. They are unique in their interpretation. Their novelty charms and surprises.

At the end of a cycle, grammatical constructions exit: by dint of frequency, they have shortened and become psychophysically slight; by dint of shortening, they have become homophonous with low contingency between form and function; by dint of habitual overuse, they have become semantically bleached. Fluent language users expect them to be there and perceive them through expectation. Novice language learners, especially second language learners, tend neither to notice them, nor to understand their function. So morphemes leave the language.