This article was downloaded by: *[Royal Holloway University]*
On: *30 January 2011*
Access details: *Access Details: [subscription number 917204860]*
Publisher *Psychology Press*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100704

The acquisition of morphological knowledge investigated through artificial language learning

Marjolein Merkx[a]; Kathleen Rastle[a]; Matthew H. Davis[b]
[a] Department of Psychology, Royal Holloway, University of London, Egham, UK [b] MRC Cognition & Brain Sciences Unit, Cambridge, UK

Accepted uncorrected manuscript posted online: 20 November 2010

First published on: 24 January 2011

To cite this Article Merkx, Marjolein , Rastle, Kathleen and Davis, Matthew H.(2011) 'The acquisition of morphological knowledge investigated through artificial language learning', The Quarterly Journal of Experimental Psychology,, First published on: 24 January 2011 (iFirst)

To link to this Article: DOI: 10.1080/17470218.2010.538211

URL: http://dx.doi.org/10.1080/17470218.2010.538211

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
i*First*, 1–21

Psychology Press
Taylor & Francis Group

The acquisition of morphological knowledge investigated through artificial language learning

Marjolein Merkx[1], **Kathleen Rastle**[1], and **Matthew H. Davis**[2]

[1]Department of Psychology, Royal Holloway, University of London, Egham, UK
[2]MRC Cognition & Brain Sciences Unit, Cambridge, UK

Affix knowledge plays an important role in visual word recognition, but little is known about how it is acquired. The authors present a new method of investigating the acquisition of affixes in which participants are trained on novel affixes presented in novel word contexts (e.g., *sleepnept*). Experiment 1 investigated the role of semantic information on affix acquisition by comparing a form-learning condition with a condition in which participants also received definitions for each novel word. Experiment 2 investigated the role of long-term consolidation on affix acquisition by comparing knowledge of learned affixes two days and nearly two months after training. Results demonstrated that episodic knowledge of affixes can be acquired shortly after a single training session using either form or semantic learning, but suggested that the development of lexicalized representations of affixes requires the provision of semantic information during learning as well as a substantial period of offline consolidation.

Keywords: Morpheme acquisition; Word learning; Orthography; Semantics; Consolidation.

The ability to generalize knowledge from a limited set of exemplars is at the very heart of our remarkable language abilities and could even be considered as the hallmark of human language acquisition. We are able to express and understand a limitless range of ideas by combining knowledge of a finite set of individual words with knowledge of a small set of syntactic constraints. How is it that we acquire the atomic elements of language that allow us to generate a near-infinite number of possible utterances despite seldom encountering these elements in isolation? Though critically important in the early years of life, the acquisition of individual lexical units and their combinations remains central to the use of language throughout adulthood, as we continue to encounter and produce new words and ideas.

In no domain is this linguistic productivity more evident than in morphology. The vast majority of English words are built by combining a small set of stems (e.g., *kind*) in highly predictable ways with prefixes (e.g., *unkind*) and suffixes (e.g., *kindness*). Critically, our experience with these individual exemplars allows us to abstract

---

Correspondence should be sent to Kathleen Rastle, Department of Psychology, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK. E-mail: Kathy.Rastle@rhul.ac.uk

The research in this article was supported by research grants from the Leverhulme Trust (F/07 537/AB) and the Economic and Social Research Council (RES–062–23–2268) awarded to Kathleen Rastle and Matthew Davis. We are grateful to Jay Rueckl, Arty Samuel, Nicolas Dumay, Lee Wurm, and Jo Taylor for helpful comments on an earlier version of this manuscript.

http://www.psypress.com/qjep

DOI:10.1080/17470218.2010.538211

knowledge about the components of these words (e.g., *-ness*, *un-*) for use in the interpretation and creation of new words (e.g., *unfaxable*; Plaut & Gonnerman, 2000). The power of these generalizations is demonstrated by the fact that up to 70% of words entering the language are created through new combinations of existing morphemes (e.g., *bioweapon*, *arborist*, *therapize*; Algeo, 1991). Indeed, new morphemic endings are even being added to the language (e.g., following the Watergate controversy, *-gate* has become a suffix that can be added to noun stems to denote scandals such as *Monicagate*, *Sachsgate*, and *Climategate*). However, despite the flexibility with which we use abstract knowledge of morphemic units, very little is known about the mechanisms that underlie the acquisition of this knowledge.

The research presented in this article investigates the processes that underlie the acquisition of morphological knowledge using an artificial learning paradigm in which adult participants are trained on novel affixes (e.g., *-nept*) presented in novel word contexts (e.g., *sleepnept*, *buildnept*). Following training with these novel words, participants are tested in various ways to establish the nature of any stored representations of the affixes. Our studies aim to determine not only whether participants can discover new morphological units presented in the context of novel words (i.e., that they develop some representation of the new affixes in spite of the fact that they receive no training on the affixes in isolation), but also whether these learned units come to behave like genuine affixes in supporting the recognition of new words in speeded contexts (i.e., that participants activate these morphemic representations automatically in the analysis of exemplars that they have not previously encountered, as has been shown for the recognition of novel complex words like *quickify*; Meunier & Longtin, 2007). Such representations would be deemed to have become *lexicalized* (following Gaskell & Dumay, 2003) or *engaged* within the lexical system (following Leach & Samuel, 2007).

In studying the processes though which new morphemic units become lexicalized, our investigation addresses two further theoretical issues. First, we investigate the nature of the input required for the development of affix representations, asking whether the lexicalization of new morphemic units can occur on the basis of form information alone or whether the provision of semantic information during learning is necessary. Second, we investigate the time course of the development of affix knowledge, asking whether there is a temporal dissociation between the formation of episodic representations of the new affixes and the development of lexicalized representations that support the analysis of new exemplars, similar to that recently proposed in relation to the acquisition of new words (see Davis & Gaskell, 2010). We discuss these theoretical issues in turn after first considering the wider literature relevant to the discovery of lexical units within longer sequences of linguistic information.

Discovering parts within wholes: Evidence from studies of speech segmentation

In contrast to studies of the acquisition of whole word knowledge in which listeners merely have to learn items presented one at a time (e.g., Gaskell & Dumay, 2003; Leach & Samuel, 2007), the central problem of affix acquisition is that language users must discover linguistic units that they are never exposed to in isolation. For example, in becoming familiar with words like *listener*, *attacker*, and *defender*, language users might learn that there is a unit {-er} that carries an agentive meaning and turns verbs into nouns (such that if it were attached to a nonsense verb like *varb*, it would mean “someone who varbs”). Because written and spoken words are not marked for morphemic boundaries, the discovery of these morphemic units poses a substantial computational challenge. However, as highlighted by Rastle and Davis (2008), an analogous problem is faced in the development of word knowledge in infancy (i.e., infants must acquire knowledge about individual words based on exposure to connected speech that does not have reliable cues to word boundaries). Morphological theorists might therefore profit from conceptualizing their problem in the context of this broader literature.

This kind of conceptualization is especially apt in the case of our work, because researchers interested in the question of how words are discovered from sequences of continuous speech have long used artificial language methods to test their theories in controlled laboratory environments (e.g., Dahan & Brent, 1999; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). Participants in these studies are typically exposed to a defined repertoire of syllables concatenated to form multisyllabic utterances (Dahan & Brent, 1999) or a continuous speech stream (e.g., Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996). The key questions posed by these researchers are (a) to what extent do participants show knowledge of the individual units ("words") contained within the longer spoken utterances, and (b) on what basis are they able to discover these units? For example, Saffran, Newport, et al. (1996) exposed adults to a continuous stream of spoken syllables, in which statistical information provided the only reliable cue to segmentation. Specifically, they had constructed their speech stream such that the sequential probabilities of adjacent syllables *within* their designated "words" were higher than the sequential probabilities of adjacent syllables *between* their designated "words", thus emulating natural speech. Following a familiarization phase, participants were given a forced-choice test asking them to decide which of two sequences of syllables constituted a unit in the new language. Results showed that participants chose the designated "words" more often than sequences of syllables that occurred equally often but that straddled "word" boundaries, thus demonstrating that they had segmented the "words" from the continuous speech stream on a statistical basis. Though this is just an example of the kind of questions asked in this literature, these results could have interesting implications for the acquisition of morphological knowledge, particularly because it is known that sequential probabilities tend to be higher *within* morphemes than *across* morpheme boundaries (Rastle, Davis, & New, 2004).

However, one difficulty with trying to apply methods from artificial language studies of speech segmentation to questions concerning morpheme acquisition is that it is not clear precisely what the participants are learning in these studies. The problem is that these studies typically use tests such as nonspeeded two-alternative forced choice (e.g., Saffran, Newport, et al., 1996) or recognition memory (Dahan & Brent, 1999), which may not be particularly successful in establishing that units discovered in the speech stream are represented as lexical knowledge. For example, Dahan and Brent exposed participants to spoken utterances like /difenu/ along with longer strings like /koSedifenu/ and then presented them with a recognition memory test asking whether particular sound sequences had occurred during the familiarization phase. They found that participants were better able to recognize targets like /koSe/ than targets like /koSed/ (both heard during the familiarization phase as part of /koSedifenu/ though never in isolation), implying that they had formed some representation of /koSe/ through the segmentation of /koSedifenu/ based on the trained lexical item /difenu/. However, though the results of these tests confirm that participants have been able to segment the designated units from longer sequences, they do not establish the nature of the resulting representations. Indeed, there is plenty of evidence to suggest that acquiring factual knowledge about a particular phonological form (e.g., that this item occurred during the familiarization phase) is not a sufficient condition for lexicalization (e.g., Gaskell & Dumay, 2003; Leach & Samuel, 2007). For example, research on the acquisition of spoken words has demonstrated that while adult participants can remember encountering novel words like *cathedruke* in a recognition memory task immediately after familiarization, these new words do not appear to compete with known words like *cathedral* (a signature of lexicalization) until some days after initial exposure (Gaskell & Dumay, 2003).

Recent work by Fernandes and colleagues (Fernandes, Kolinsky, & Ventura, 2009) has gone some way to addressing this problem. They presented adult participants with sequences of continuous speech (as in, e.g., Saffran, Newport, et al., 1996), in which the designated "words" were cohort competitors of existing words and could be segmented from the speech stream on the basis of information about sequential probabilities. Critically, in addition to demonstrating a preference for the designated "words" in a two-alternative forced-choice task, they also established that those designated "words" exerted an inhibitory influence on real words from the same cohort in auditory lexical decision. These results thus suggested that the units discovered from the longer sequences had been lexicalized, although this competitive effect was observed only when the segmentation cues (wordlikeness and sequential probabilities) suggested the same parsing. This work is related to the problem of affix acquisition because it suggests that adults may be able to discover morphemic units that they are never exposed to in isolation and that these units might become lexicalized. However, investigating the acquisition of morphemic knowledge requires us to go one step further than this, in establishing whether the learned morphemic units *themselves* can be identified sufficiently rapidly that they support the recognition of new exemplars (e.g., having been trained on *sleepnept* and *buildnept*, we ask how a learner deals with novel items such as *sailnept* or *parbnept*).

Three theories of affix acquisition

One key aim of this research is to elucidate the role of semantic information in the lexicalization of new morphemic units. In the process, we can distinguish between three theories proposed by Rastle and Davis (2008) concerning the discovery of morphemic units from exposure to complex words. The first two of these theories were derived directly from the speech segmentation literature discussed above and posit that affix representations can be acquired on the basis of information about the orthographic forms of words alone. The first theory (*morpheme boundary detection*) suggests that readers analyse the sequential probabilities of letter combinations to detect morphemic boundaries and thereby identify affix units (as in the work on speech segmentation described above by, e.g., Saffran, Aslin, et al., 1996; see also Christiansen, Allen, & Seidenberg, 1998; Elman, 1990, for relevant computational simulations). The second theory (*morpheme chunking*) suggests that morpheme acquisition arises because affixes are frequent letter combinations that occur in a combinatorial manner (i.e., they occur with many familiar stems). Rastle and Davis (2008) argued that this combinatorial property of affixes should provide for highly efficient segmentation through chunking as demonstrated in similar theories of speech segmentation (e.g., Brent, 1997; Brent & Cartwright, 1996; see also Dahan & Brent, 1999) and visual word recognition (Davis, 1999, 2010). In contrast to these two form-based theories, the third theory described by Rastle and Davis proposes that higher level regularities between form and meaning facilitate lower level orthographic learning of affixes, suggesting that morpheme learning is *semantically driven*. By this account, semantic knowledge of complex words permits the identification and acquisition of orthographic affix knowledge by reinforcing the preferred orthographic alignment of complex words into their constituents. Based on this theory, then, those affixes that occur in transparent contexts and have a consistent meaning should be easiest to learn as they enable readers to use semantic information to identify orthographic regularities.

These three theories lead to different hypotheses about the discovery of morphemic information through exposure to pseudowords containing novel morphological units (new affixes). Does affix learning take place based on form cues alone? Or is the discovery of these morphemic units dependent on or enhanced by the provision of supporting semantic information? Though research suggests that semantic information plays a small or negligible role in the initial stages of morphological decomposition (e.g., Rastle & Davis, 2008; Rastle et al., 2004), previous research has shown that semantic information may be very important for the lexicalization of newly learned words (Leach & Samuel, 2007; Tamminen & Gaskell, 2008; though see Dumay, Gaskell, & Feng, 2004) and that pseudowords characterized by a systematic relationship between orthography and meaning are easier to learn and identify than pseudowords that lack these systematic relationships (Rueckl & Dror, 1994). These findings thus raise the possibility that the *acquisition* of morphological knowledge may require semantic information. The experiments presented in this article address this issue by investigating the acquisition of new affixes under form-learning and semantic-learning conditions.

Temporal dimensions of affix acquisition

The other aim of this work is to explore the time course of the acquisition of morphemic knowledge and, in particular, to establish whether there is a temporal dissociation between the formation of episodic and lexical representations of new affixes. Initial theories of lexical organization were divided as to whether the lexicon was composed of entirely abstract representations (Gaskell & Marslen-Wilson, 1997; McClelland & Elman, 1986; Norris, 1994) or purely episodic representations (Goldinger, 1998). However, recent studies of perceptual and word learning increasingly point to a hybrid or complementary learning system in which initial learning is achieved using context-specific episodic representations that are combined and consolidated (perhaps during nocturnal sleep; Dumay & Gaskell, 2007) into abstract lexical representations (see, e.g., Davis & Gaskell, 2010; Goldinger 2007). Thus, although the long-term goal of affix acquisition is the development of abstract lexical representations that support the many linguistic functions that depend on affix knowledge, these representations may not be immediately apparent following initial learning. Complementary learning theories would propose that language learners proceed by first acquiring multiple forms of knowledge, including specific episodic representations of the orthographic and phonological forms of affixes, their meanings and syntactic functions, and knowledge of how these different representations are modified in specific contexts. It is only once these fragmented, episodic representations of individual encounters with specific words that contain unfamiliar morphemic units have been learned and consolidated that learners should acquire stable lexical representations (see, e.g., Davis & Gaskell, 2010).

Thus the distinction between episodic representations of specific encounters with novel words and the development of stable lexical representations of new linguistic units suggests two stages of word learning. While these accounts remain theoretically underspecified at present, empirical methods of distinguishing between these two forms of knowledge have been proposed in various word-learning studies. For example, Gaskell and Dumay (2003) showed rapid initial acquisition of novel spoken words (e.g., *cathedruke*), sufficient to pass simple recognition memory tests (such as distinguishing learned items from untrained foils like *cathedruce*). However, as mentioned previously, when word learning was assessed using a lexical competition test, they found evidence for lexical storage only after a period of consolidation. Behavioural studies have also shown changes in the degree of lexical influence on phonological category boundaries (Leach & Samuel, 2007) and changes to the speed of production of novel words (Davis, di Betta, MacDonald, & Gaskell, 2009) following offline consolidation, with this latter change associated with modifications to cortical representations of novel words that can be observed using functional magnetic resonance imaging (fMRI; Davis et al., 2009). If these findings extend to the acquisition of affix knowledge, then we may find that participants develop episodic representations of the novel affixes relatively rapidly, but that the emergence of full lexical representations may require substantial offline consolidation that arises over a longer period. By testing participants both two days and two months after learning, we sought to assess the impact of long-term consolidation on the emergence of abstract, lexical representations of affixes.

Studying affix acquisition using artificial language learning

In the experiments reported here, we used a word-learning paradigm to teach participants novel affixes in a controlled laboratory setting. Word learning has been used successfully to examine the acquisition of phonological (e.g., Davis et al., 2009; Gaskell & Dumay, 2003; Leach & Samuel, 2007; Tamminen & Gaskell, 2008), orthographic (e.g., Bowers, Davis, & Hanley, 2005), and semantic (e.g., Clay, Bowers, Davis, & Hanley, 2007; Leach & Samuel, 2007; Rueckl & Dror, 1994) representations of new words. Here we demonstrate that these laboratory-based word-learning methods can be adapted to examine the acquisition of morphemic knowledge. Such experiments provide exquisite experimental control of both the stimuli and the information available to learners, thus permitting careful investigation of the influence of linguistic properties (e.g., differences between affixes with and without consistent meanings) that would be difficult or impossible to manipulate in existing languages.

Constructing artificial affixes

Participants in our experiments were taught novel affixes (e.g., -*nept*, -*ane*) in novel word contexts (e.g., *sleepnept*, *kickane*). Like real affixes (e.g., -*er* as in *painter*), these morphemic units never surfaced in isolation, only in combination with familiar stems. Further, each of our affixes occurred with many different stems, as is the case for real affixes (e.g., the affix -*age* occurs with numerous stems such as *block*, *drain*, *post*, and *wreck*). Participants in our form condition were thus provided with eight novel affixes (e.g., -*nept*), each of which occurred with eight familiar stem morphemes (e.g., the affix -*nept* occurred with the stems *sleep*, *build*, *chop*, *float*, *talk*, *climb*, *dress*, and *steal*; the affix -*ane* occurred with a different set of eight stems, etc.). The fact that each affix occurred in the context of several familiar stems provided the only indication of morphological structure in this condition. Our semantic-learning condition combined the contextual cues of the form-learning condition with a semantic component. Unlike nonmorphological endings, affixes convey meanings through their combination with the meanings of the stems to which they are attached. For example, the word *cloudless* is a semantically transparent combination of the stem *cloud* and the affix -*less*, which means “lacking [stem]”. In order to simplify the training paradigm in this initial investigation of affix learning, we opted to use definitions that were both transparent and consistent (related to the meaning of the stem and based on a consistent affix meaning). Thus, for example, *sleepnept* was “The hourly rate for taking a nap in an airport bed”, and *buildnept* was “The extra costs involved in constructing a house on stilts” with -*nept* relating to a cost in both cases.

Testing episodic and lexical knowledge

The acquisition of affix knowledge was assessed using three different test tasks. In order to determine whether participants had formed episodic representations of the novel affixes, we used a forced-choice recognition memory task that asked not only whether participants could recognize the words they had learned but also whether they could reject complex words containing one familiar unit (a trained stem with an untrained affix or an untrained stem with a trained affix). These tests allowed us to examine affix and stem learning, respectively. We further examined responses to stimuli in which learned affixes and learned stems were recombined (e.g., testing on *kicknept* or *sleepane* following training with *sleepnept* and *kickane*). Because whole-word episodic representations should clearly mark these items as unfamiliar, the presence of false-positive errors in this condition may indicate that the novel affixes are becoming represented as distinct units, separate from the stems to which they attach.

Like Gaskell and Dumay (2003; also, e.g., Fernandes et al., 2009) we view performance in tasks probing automatic lexical processing as a more sensitive test of lexicalization than performance in nonspeeded recognition memory tasks. Thus, in order to assess whether participants had formed lexicalized representations of the novel affixes, we also used a speeded lexical decision task in which the novel affixes were paired with new stems. This task required participants to make a "yes" response only if they encountered a real English word; the critical stimuli containing the novel affixes required a "no" response. Previous research has shown that morphologically structured nonwords that contain existing affixes are particularly difficult to reject in lexical decision (e.g., *clatment:* Caramazza, Laudanna, & Romani, 1988; Crepaldi, Rastle, & Davis, 2010; Laine, Vainio, & Hyönä, 1999; Wurm, 2000). This finding is typically interpreted as demonstrating that participants have lexical representations of affixes that influence the recognition of visually presented letter strings. Thus, our prediction was that if participants had lexicalized representations of the novel affixes, then similar difficulties would be observed in rejecting untrained nonword stimuli containing those affixes.

Finally, as an initial investigation of the acquisition of knowledge about the meanings of affixes and as a check on our manipulation of learning condition, we conducted a definition selection task for those participants in the semantic-learning condition. This nonspeeded task required participants to choose between two definitions both for learned words (e.g., *sleepnept*) and for untrained words comprising an existing stem plus a trained novel affix (e.g., *sailnept*). For these untrained words, only one of the definitions was consistent with the learned affix meaning (e.g., so if *-nept* referred to a cost then a consistent meaning for *sailnept* would be "The hourly cost of learning how to navigate a yacht" as opposed to "A person who excels in open sea catamaran racing"). The observation that participants performed well on this task would indicate that they were able to extract consistent affix meanings from the trained stimuli for use in the interpretation of new exemplars without those meanings being explicitly provided during training.

EXPERIMENT 1

Method

Participants

Participants were 32 native English speakers from Royal Holloway, University of London. Half of these participants were assigned to the form-learning condition, and half were assigned to the semantic-learning condition. Participants all had normal or corrected-to-normal vision and were free from any known language impairments. They were paid for their time and travel expenses over the two sessions.

Materials

Learning phase. The critical stimuli in the learning phase of the experiment were nonwords consisting of an existing stem and a novel affix.

Sixteen novel affixes were created from existing word endings, none of which were words or affixes in their own right. The novel affixes were selected on the basis of four structural types (CVCV, VCV, CVCC, and VCC; C = consonant, V = vowel) and four vowels (A, E, U, and O). Of the novel affixes, eight were learned by participants in Group A, and the other eight were learned by participants in Group B. Those novel affixes learned by Group A were used as untrained control affixes in the test tasks for Group B and vice versa. In this way, trained and untrained novel affixes were counterbalanced between participant groups. Table 1 contains the novel affixes with the "A" and "B" labels showing how they were divided between participant groups.

Eight different sets of eight existing word stems were selected. Each of the eight stems in a set was paired with one of the novel affixes in Group A and one in Group B (e.g., the stem *sleep* occurred with the affix *-nule* for participants in Group A and with the affix *-nept* for participants in Group B), thus creating 64 novel words for each participant to learn. Stems were monosyllabic

Table 1. *The novel affixes by group and structural type, Experiment 1*

Group	*CVCV*	*VCV*	*CVCC*	*VCC*
A	nule	ane	halk	uck
A	tege	ose	lomb	esh
B	labe	ude	tund	aph
B	hoke	ete	nept	ort

Note: C = consonant. V = vowel.

Table 2. *The definition types and example definitions, Experiment 1*

Stem	*Type*	*Example word*	*Example definition*
verb	place	kickort	A large field used by footballers to practise penalties
noun	place	cointund	The factory in which the twenty pence piece is produced
verb	tool	pourlabe	A bottle cap used for decanting exact measures of a liquor
noun	tool	wheathoke	A harvesting tool used by farmers in the Middle Ages
verb	person	sleepnept	A participant in a study about the effects of napping
noun	person	rugete	A person who imports and sells handmade Indian carpets
verb	cost	leapesh	The cost of having a stuntman jump out of a building
noun	cost	bombaph	The cost of buying enough explosives to blow up a car

monomorphemic content words between three and five letters in length. Half of the novel affixes used in each participant group were paired with noun stems, and half were paired with verb stems.

Audio files were created for each of the novel words. These were recorded by a female native English speaker and were edited to a consistent duration of 1,500 ms.

Definitions were then created for each novel word for the semantic-learning condition. These definitions were formulated by combining a consistent affix definition with a semantic reference to the stem. In order to ensure that the definitions themselves did not act as a segmentation cue, none of the stems were used in the definitions, and each novel word was provided with two definitions, which were counterbalanced between participant groups to mitigate the possibility that some definitions were easier to learn than others. This counterbalancing was separate from the counterbalancing of the affixes so that Group A1 and Group B1 saw the same definition (Definition 1) for *sleepnule* and *sleepnept*, respectively, while Groups A2 and B2 saw Definition 2 for *sleepnule* or *sleepnept*. Table 2 lists the affix meaning types as well as an example of each type worked into a definition. These definition types were based on affix meanings that occur in English including a place (e.g., −*ery* in *bakery*, *nunnery*), a tool (e.g., −*er* in *cooker*, *eraser*), a person (e.g., −*ist* in *cyclist*, *racist*), and a cost (e.g., -*age* in *postage*, *corkage*).

Definition length was controlled for the number of words (9−11) and syllables (13−17) in the definition. For the learning phase, all definitions were recorded by the same female native English speaker who recorded the novel words. The audio files of the definitions lasted between 3,500 ms and 4,500 ms.

Test phase. The test phase of the experiment included a lexical decision task, a recognition memory task, and the definition selection task.[1] The trained and untrained novel affixes used in these tasks were the same as those selected for the learning phase of the experiment, and the stems chosen were monosyllabic monomorphemic stimuli between three and six letters in length.

1. *Recognition memory.* The stimuli for the recognition memory task consisted of all 64 learned words (e.g., *sleepnept*) as "yes" responses and three different types of "no" response. The "no" response stimuli included 32 trained stem + untrained novel affix items (four trained stems for each of eight untrained novel affixes, e.g., *sleeptege*), 32 untrained stem + trained novel affix items (four untrained stems for each of eight trained novel affixes, e.g., *fruitnept*), and 64 recombinant pairs consisting of trained stem + trained novel affix combinations, which did not occur during training (eight stems for each of eight trained novel affixes, e.g., *sleephoke*).

---

[1] One additional nonspeeded morphological segmentation test was conducted but is not reported for reasons of brevity.

2. *Lexical decision.* The stimuli for this task consisted of 288 letter strings: 144 familiar English words for which a "yes" response was expected, and 144 nonwords for which a "no" response was expected. No trained novel words were used. The critical stimuli in this task were the "no" responses, which were of three different types: nonword stem + trained novel affix (e.g., *morknept*), nonword stem + untrained novel affix (e.g., *fushnule*), and nonword stem + existing affix (e.g., *clatment*). The existing affixes used were individually matched on length and structure to the trained and untrained novel affixes. The nonword stems were selected from the English Lexicon Project website (Balota et al., 2002) and were between three and five letters in length, had between 10 and 20 orthographic neighbours, and had a positional bigram frequency between 500 and 1,500. They were also matched across conditions on these three factors. Each type of "no" response contained 48 items (six stems for each of the eight affixes of each type), and both participant groups saw exactly the same stimuli (with the stimuli counting as nonword stem + trained novel affix for Group A, counting as nonword stem + untrained novel affix for Group B, and vice versa).

   The lexical decision "yes" responses were real English words that were the same across participant groups. They consisted of two different types (with 72 words of each type): existing complex words (e.g., *duckling*) and noncompound bisyllabic words containing embedded monosyllabic words (e.g., *kidney*, which contains the embedded word *kid*). For the existing complex words, 12 existing affixes were selected, of which 6 were three letters in length and started with a vowel, and 6 were four letters in length and started with a consonant, thus matching the trained and untrained novel affixes in length and type of starting letter. Across "yes" and "no" responses, the lexical decision stimuli were matched on average word length.
3. *Definition selection*. The definition selection task consisted of forced-choice judgements about the meanings of 128 novel words. Of these, 64 were words presented during the learning phase (e.g., *sleepnept*), and 64 paired the learned novel affixes with untrained word stems (e.g., *sailnept*). These untrained word stems had not been used previously and followed the verb or noun preference for each novel affix established in the learning phase. For the novel words presented during the learning phase, participants were forced to choose between two definitions that had been created for those novel words (and which had been counterbalanced across participants). For the untrained novel words, two definitions were created for each stimulus. These definitions were based on the same two definitions for each affix as those employed in the learning phase, so that a definition consistent with each affix meaning for each semantic-learning group was created. In this way, the same stimuli could be shown to both participant groups, with the counterbalancing condition dictating which of the two definitions formed the correct response.

Procedure

The experiment took place on two separate days with one nonexperiment day in between to allow time for consolidation. The learning phase took place on Day 1, while the test tasks were carried out on Day 3. On Day 3, participants performed the lexical decision task followed by the recognition memory task. Participants in the semantic-learning condition then performed the definition selection task.

All parts of the experiment were performed individually on a computer, with responses being made on either the keyboard (for the learning phase) or a button box (for the test tasks). Stimulus presentation and data recording were controlled by the DMDX software (Forster & Forster, 2003). The learning phase on Day 1 lasted around an hour for participants in the form-learning condition and around two hours for participants in the semantic-learning condition. However, the time participants spent looking at the novel words was the same for both learning conditions; the additional time needed for semantic learning was spent listening to the definitions of the novel words, which contained neither the novel words nor any part of the novel words. The test phase on Day 3 lasted around one hour.

Learning phase. Participants were presented with each novel word individually on the screen (in lower case) in white letters on a black background. Each word appeared on the screen for 43 ms before the audio file of the word started and remained on the screen for the 1,500 ms of the audio file. The screen then went blank, at which point participants were instructed to type the word they had just seen. Participants in the semantic-learning condition then heard the audio file of the definition of the word. Pressing the enter key took the participant to the next word.

The learning phase consisted of 12 cycles of all 64 words so that each item was seen 12 times, and each affix was seen 96 times (768 exposures in total). The order of the items was randomized in each cycle.

Test phase

1. *Recognition memory.* During the recognition memory task, a fixation cross (+) appeared on the screen followed by the letter string (in lower case), which remained on the screen until the participant responded. Participants were asked to decide whether each item was one of the words they had learned during the learning phase of the experiment.
2. *Lexical decision.* For the lexical decision task, participants were instructed to indicate with a speeded button press whether each letter string was a real English word or not (using the dominant hand for "yes" responses). Participants were asked to respond as quickly and as accurately as possible. Each letter string appeared on the screen following a fixation cross (+), and remained on the screen until the participant responded. Participants were given eight practice trials prior to starting the main experiment.
3. *Definition selection*. During the definition selection task, each stimulus with its two possible definitions was shown on the screen until the participant responded by picking one of the definitions. Participants were told to select the definition they had learned for the trained novel words and the definition they thought was most suitable for the untrained novel words.

Results

Recognition memory data were subject to signal detection analysis (see Snodgrass & Corwin, 1988). Data from the lexical decision and definition selection tasks were analysed using mixed effects models (Baayen, 2008), with linear mixed effects analyses being used for lexical decision response times and logit analyses being used for categorical measures (Jaeger, 2008).

Recognition memory

In order to correct for any response biases in recognition memory, we computed hit and false-alarm rates for each test condition to derive signal detection measures of performance (d'). We computed a measure of successful stem recognition by calculating the difference between the z-transformed proportion of correct responses to learned items (hits) and the z-transformed proportion of incorrect "yes" responses to items with untrained stems (false alarms). The signal detection measure derived from the comparison of hit and false-alarm rates therefore indicates the participant's ability to recognize trained stems and reject untrained stems while removing an overall bias towards accepting or rejecting items (i.e., criterion shifts). Measures of affix recognition (comparing correct "yes" responses with false alarms to items with untrained affixes) and whole word knowledge (comparing correct "yes" responses with "yes" responses to recombinant pairs) were calculated in the same way. Average d' values and percentage of correct responses are shown in Table 3. Since d' values can only be computed by combining information across different items (i.e., hits and false alarms), we used conventional analyses of variance (ANOVAs) by participants in analysing these data as well as one-sample t tests (comparing d' values to zero to test for above-chance performance in each condition).

An ANOVA was performed on the d' values with learning type (form or semantic) as the between-participant variable and knowledge type (affix, stem, or whole word) as the within-participant variable. The ANOVA showed a main effect of knowledge type, $F(2, 60) = 158.31$, $MS = 17.21$, $p < .001$, no main effect of learning type, $F(1, 30) = 0.22$, $MS = 0.30$, $p > .6$, and no interaction between these factors, $F(2, 60) = 2.24$, $MS = 0.24$, $p > .1$. Post hoc comparisons of the effect of knowledge type revealed that affix recognition was better than stem recognition, $t(31) = 3.75$, $p < .01$. This is unsurprising because each affix was seen 96 times during training, while each stem was only seen 12 times. These tests also revealed that whole word knowledge was significantly worse than both affix recognition and stem recognition—compared to stem recognition, $t(31) = 16.79$, $p < .001$; compared to affix recognition, $t(31) = 13.97$, $p < .001$—indicating that participants had particular difficulty rejecting recombinant pairs. Finally, the ANOVA showed a significant intercept, indicating that overall performance was better than chance, $F(1, 30) = 247.74$, $MS = 337.50$, $p < .001$, with post hoc analyses of the individual d' measures showing that participants performed better than chance after both form and semantic learning for each type of knowledge (all $p < .001$).

Lexical decision

Lexical decision responses were trimmed at 1,900 ms (removing 0.84% of the data). Table 4 shows the mean reaction times and error rates (%) for the "no" responses. The "yes" responses in this task reflected performance on filler stimuli so were not analysed.

Consistent with the existing literature (e.g., Caramazza et al., 1988; Crepaldi et al., 2010), rejection latencies for nonwords containing existing affixes (e.g., *clatment*) were slowed relative to control nonwords without morphological

Table 3. *Recognition memory: Percentage of correct responses and d' measures, Experiment 1*

	Item type	*Semantic learning*	*Form learning*
% Correct	Learned words (yes)	79.4	84.6
	Untrained stem + trained affix (no)	86.9	78.1
	Trained stem + untrained affix (no)	91.2	87.9
	Recombinant pair (no)	58.8	44.2
d'	Stem knowledge	2.25	1.98
	Affix knowledge	2.42	2.50
	Whole word knowledge	1.13	0.98

Table 4. *The mean reaction times and error rates of the lexical decision "no" responses, Experiment 1*

	Semantic learning		*Form learning*	
Item type	*RT*	*ER*	*RT*	*ER*
Nonword stem + trained novel affix	722	4.7	672	2.0
Nonword stem + untrained novel affix	720	1.4	677	2.1
Nonword stem + existing affix	767	5.9	733	8.1
Effect of learning: Existing affixes	47	4.4	56	6.0
Effect of learning: Trained affixes	2	3.3	−5	−0.1

Note: ER = error rate (in %). RT = reaction time (in ms).

structure. Mixed effects models with affix type (existing affix/untrained novel affix) as a fixed factor and participants and items as random factors showed that nonwords containing existing affixes were responded to more slowly and with more errors than nonwords containing untrained novel affixes (reaction times: $t = 4.93$, $p < .001$; errors: $z = 5.42$, $p < .001$).

For the nonwords that did not contain existing affixes, we examined lexical decision performance using a mixed effects model that included learning type (form/semantic) and affix type (trained novel affix/untrained novel affix) and their interaction as fixed factors and participants and items as random factors. The analysis of reaction times showed no effects of affix type or interactions between affix type and learning type ($t < 1$ for all). However, the analysis of error data did show an interaction between learning type and affix type ($z = 2.58$, $p < .01$), reflecting the fact that items with trained novel affixes yielded more errors than items with untrained novel affixes after semantic learning ($z = 3.55$, $p < .001$) but not after form learning ($z < 1$).

Definition selection

Data from the definition selection task showed that participants in the semantic-learning condition had learned the definitions given for the trained novel words well, with performance averaging 94.1% correct and with all participants scoring over 80% correct. Unsurprisingly, a mixed effects logit model of responses to learned novel words (containing no fixed factors and participants and items as random factors) showed that participants performed better than chance on these items ($z = 13.48$, $p < .001$). Participants also showed an impressive ability to generalize affix definitions to the untrained novel words (e.g., *sailnept*), selecting the correct definition for 72% of trials, which was significantly better than chance in a similar mixed effects logit model ($z = 4.1$, $p < .001$).

Discussion

Experiment 1 compared the impact of form and semantic training on the acquisition of novel affixes using a recognition memory task, a speeded lexical decision task, and a nonspeeded definition selection task. Data from the recognition memory task showed that a single learning session was sufficient for participants to discover the novel affix units even in the absence of explicit training on those units. This is demonstrated by participants' substantial difficulties in both learning conditions in rejecting recombinant pairs such as *sleephoke* that incorrectly combined trained stems and trained affixes. These recombinant items were entirely unfamiliar as whole forms and therefore would have been rejected with ease if participants had failed to discover the affix units.[2] The definition selection task administered in the semantic-learning condition also provided evidence not only that participants had discovered the novel affix units but also that they were able to extract consistent affix meanings from them for use in the interpretation of previously unseen exemplars (i.e., generalization).

However, despite this evidence from recognition memory and definition selection that the novel affixes were represented as isolated units, we found only limited evidence to suggest that these affix representations were involved in automatic lexical processing. There was no evidence that stimuli with trained affixes influenced lexical decision latencies or error rates following form training. Further, though participants in the semantic-learning condition had difficulty rejecting items consisting of trained affixes in the lexical decision task, this effect of affix knowledge was apparent only in the accuracy data. Though these data are suggestive that the novel affixes had

---

[2] One possible objection to this interpretation is that we did not include even numbers of "yes" and "no" responses in our recognition memory task (there were only 64 "yes" responses possible), and this may have inflated the number of false alarms as participants may have expected half of the items to require "yes" responses. However, such an increase of false alarms should have applied across the board unless the recombinant pairs were somehow more familiar than would be expected based on their whole-form familiarity.

become lexicalized, this conclusion would be safer if we had also observed an effect on response times.

EXPERIMENT 2

Experiment 1 demonstrated that novel affixes can be learned using a basic word learning paradigm, thus providing us with a basis for further research into affix acquisition. However, while Experiment 1 demonstrated that participants had formed *some* representation of the novel affixes, evidence that these morphemic representations influenced automatic lexical processing was limited. Experiment 2 sought to address this issue by making two changes to the experimental design of Experiment 1. The first change addressed the possibility that the training procedure used in Experiment 1 did not engage participants sufficiently for lexicalization to occur. For this reason, a more active learning procedure was adopted in Experiment 2 that alternated between a relatively passive study task (similar to Experiment 1) and a more difficult test task in which participants had to type the novel word that corresponded to a particular definition (semantic-learning condition) or complete an orthographic fragment of one of the novel words (form-learning condition). This procedural alteration also had the advantage of ensuring that participants in the form and semantic conditions spent the same amount of time in the learning phase (something that we did not achieve in Experiment 1). The second change addressed the possibility that testing two days after training did not provide sufficient time for lexicalization to occur. While previous word-learning studies have shown lexicalization within 12 hours (Dumay & Gaskell, 2007), it is possible that the lexicalization of affix representations requires a longer period of consolidation. In Experiment 2, participants were therefore tested both two days and several weeks after initial learning—a manipulation that also allowed us to examine the longevity of the learned affix representations.

Method

Participants

Participants were 48 native English speakers from Royal Holloway, University of London, 41 of whom also participated in the retest which took place approximately 50 days after the original training session (mean = 49 days; *SD* = 12 days). Half of the participants were assigned to the form-learning condition (21 of whom participated in the retest) and half to the semantic-learning condition (20 of whom participated in the retest). Participants were free from any language or visual impairments and were paid for their time and travel expenses over the three sessions.

Materials

Learning phase. The learning phase of this experiment consisted of the same 64 novel words as those used in Experiment 1. However, for the semantic-learning condition of this experiment, we sought to make the relationship between novel word and stem more salient by including the actual stem in the definition (hence, *sleepnept* was "The hourly rate travellers pay to sleep in an airport bed" rather than using the synonym "nap" as previously). The same four definition types were used as before (a place, a tool, a person, or a cost), and assignment of definitions to affixes was again counterbalanced between participant groups.

Test phase. The test phase consisted of the same tasks as those in Experiment 1.

1. *Recognition memory.* The recognition memory task stimuli were identical to those used in Experiment 1.
2. *Lexical decision.* The lexical decision stimulus set from Experiment 1 was modified to include both word and nonword stems.[3] As in Experiment 1, no trained novel words were used. "No" responses consisted of four sets of 48 items each: nonword stem + trained novel affix (e.g., *morknept*), nonword stem + untrained novel affix (e.g., *fushnule*), word

---

[3] One of the reviewers of a previous version of this manuscript suggested that our learning effects might be stronger in the context of word stems, hence this alternation.

stem + trained novel affix (e.g., *stopnept*), and word stem + untrained novel affix (e.g., *trustnule*). Lexical decision "yes" responses consisted of 96 monomorphemic bisyllabic words with neither an embedded word nor an affix ending (e.g., *stomach*) and 96 noncompound bisyllabic words containing embedded monosyllables (e.g., *kidney*). Nonword stems were selected from the English Lexicon Project website (Balota et al., 2002) and were matched across conditions for length (from 3–5 letters), orthographic neighbours (10–20), and positional bigram frequency (500–1,500). Both participant groups saw the same stimuli though trained novel affixes for Group A were untrained affixes for Group B and vice versa.

3. *Definition selection*. The stimuli for the definition selection task consisted of the same stimulus types as those in Experiment 1.

Procedure

Participants were trained on the first day of the experiment, and initial testing was conducted two days later. During testing, participants first performed a lexical decision task and later a recognition memory task. Participants in the semantic-learning condition then performed the definition selection task. Approximately two months later, we contacted the participants and asked them to return for a second testing session. Participants were not warned that there would be a retest after the initial testing session. The retest included the same test tasks performed at the first testing session. The learning phase on Day 1 lasted around two hours for both learning conditions, and the test tasks lasted around 90 minutes.

Learning phase. Semantic learning consisted of study and verification tasks. Three study blocks were followed by one verification block, making a four-block set. Learning consisted of three of these four-block sets. For the study task, participants saw a novel word and its definition on the screen while listening to a spoken token of the word. When ready, participants pressed a key to clear the screen before typing the word using the keyboard. This task was self-paced, with a minimum presentation duration for the word and its definition of 1,500 ms. During the verification task, a definition appeared on the screen, and participants were asked to type the novel word to which it belonged. Following their response, participants were provided with the correct word and definition before the next trial began.

Form learning consisted of study and fragment completion tasks. Once again, participants received three study blocks followed by one fragment completion block, and this cycle was repeated three times over the course of learning. The study task was identical to that in the semantic-learning condition except that participants did not receive a definition. During the fragment completion task, a fragmented string consisting of letters and underscores appeared on the screen. Two letters of the stem and two letters of the affix were always present (e.g., s _ _ _ p n _ p _ for *sleepnept*). Participants were told that each underscore represented a missing letter and that the string represented one of the words learned during the study task. Participants were asked to type the novel word that the string represented. Following their response, participants were provided with the correct answer.

For both learning conditions, each block of training (study, verification, and fragment completion) consisted of all 64 words, making a total of 12 cycles of all 64 words (nine study blocks and three verification/fragment completion blocks). The order of the items was randomized within each block.

Test phase. The procedure for the test tasks was the same as that used in Experiment 1.

Results

Only the data from the 41 participants who took part in both the Day 3 testing session and the retest were analysed. The same methods of analysis were used as those in Experiment 1.

Table 5. *Recognition memory: Percentage of correct responses and d' measures, Experiment 2*

		Semantic learning		*Form learning*	
	Item type	*Day 3*	*Retest*	*Day 3*	*Retest*
% Correct	Learned words (yes)	86.5	75.7	91.0	85.6
	Untrained stem + trained affix (no)	98.1	88.0	91.4	82.4
	Trained stem + untrained affix (no)	98.9	95.9	92.6	92.1
	Recombinant pair (no)	77.4	68.0	59.3	54.3
d'	Stem knowledge	3.25	2.06	3.04	2.41
	Affix knowledge	3.31	2.63	3.15	2.89
	Whole word knowledge	2.09	1.23	1.79	1.39

Recognition memory

Hit rates (for "yes" responses) and false-alarm rates for each of the different foil conditions were used to calculate signal detection measures of recognition memory (d') for stem recognition, affix recognition, and whole word recognition (shown with percentage correct responses in Table 5). A three-way ANOVA with knowledge type and testing day as within-participants variables and learning type as a between-participant variable showed an interaction between testing day and learning type, $F(1, 39) = 5.89$, $p < .05$. This interaction indicated that, although recognition memory scores decreased over time following both form learning, $F(1, 20) = 11.80$, $p < .01$, and semantic learning, $F(1, 19) = 35.17$, $p < .001$, the decrease was smaller after form learning, possibly indicating the effectiveness of the fragment completion task in supporting form-based learning processes. Despite this decline in recognition memory performance, all d' measures indicated better than chance performance for both learning types in both testing sessions (all $p < .001$), indicating that participants retained good explicit knowledge of the learned stems, affixes, and combinations approximately two months after learning.

Lexical decision

Lexical decision responses were trimmed at 1,500 ms (removing 0.64% of the data[4]). Lexical decision performance was analysed using a mixed effects model that included learning type (form/semantic), affix type (trained novel affix/untrained novel affix), stem type (word/nonword), and testing day (Day 3/retest) and their various interactions as fixed factors and participants and items as random factors. This analysis revealed no interaction between stem type and affix type in either the reaction time ($t < 1$) or the error data ($z = 1.04$; $p > .3$), indicating that stem type did not modulate the learning effect. Table 6 thus shows mean reaction times and error rates (%) for "no" responses collapsed across stem type for both testing days.

The analysis of the reaction time (RT) data revealed a three-way interaction between learning type, affix type, and testing day ($t = 1.97$, $p < .05$). This interaction reflected the fact that a significant effect of affix type was observed at retest in the semantic-learning condition ($t = 2.88$, $p < .01$) but not in the form-learning condition ($t = 1.59$, $p > .1$), while no effects of affix type were observed in the RT data on Day 3 for either learning condition ($p > .1$ for both,

---

[4] Data trimming followed Rastle, Davis, Marslen-Wilson, and Tyler (2000) and Rastle et al. (2004), whereby outliers were detected by visual inspection of the reaction time (RT) distribution averaged over all conditions prior to computation of condition means or other statistics. The criterion for removal of individual data points was set for each experiment separately to ensure that only data points outside of the expected distribution were excluded and that less than 1% of the data points were removed. The thresholds used were roughly equivalent to a four-standard-deviation cut-off, which showed the same pattern of significance.

Table 6. *The mean reaction times and error rates of the lexical decision "no" responses, Experiment 2*

	Semantic learning				*Form learning*			
	Day 3		*Retest*		*Day 3*		*Retest*	
Item type	*RT*	*ER*	*RT*	*ER*	*RT*	*ER*	*RT*	*ER*
Nonword/word stem + trained novel affix	641	4.5	644	5.8	610	2.9	586	2.8
Nonword/word stem + untrained novel affix	645	3.3	629	3.6	605	2.9	580	1.5
Effect of learning	−4	1.3	15	2.2	5	0.0	6	1.3

Note: ER = error rate (in %). RT = reaction time (in ms).

consistent with the null effect on RTs observed in Experiment 1).

The analysis of errors did not show the same three-way interaction between learning type, affix type, and testing day ($z < 1$). However, looking at the semantic-learning and form-learning conditions separately, we see results that both replicate and extend the findings of Experiment 1. Specifically, in the semantic-learning condition there was a main effect of affix type ($z = 3.28$; $p < .01$) that did not interact with testing day ($z < 1$). This main effect indicated that items containing trained affixes were more error prone than items containing untrained affixes both on Day 3 ($z = 1.91$; $p = .057$, as in Experiment 1) and at retest ($z = 3.40$; $p < .001$). In contrast, there was an interaction between affix type and testing day in the form condition ($z = 2.22$; $p < .05$), as items containing trained affixes were more error prone than items containing untrained affixes at retest ($z = 2.91$; $p < .01$) but not on Day 3 ($z < 1$).

Definition selection

Day 3 data from the definition selection task replicated Experiment 1 in showing that participants had learned the trained definitions well (average: 97.7%; $z = 13.02$, $p < .001$) and could generalize the learned affix definitions to untrained novel words (average: 92.3%; $z = 9.40$, $p < .001$). Performance for both trained and untrained items decreased over time (trained: $z = 5.69$, $p < .001$; untrained: $z = 4.91$, $p < .001$), but was still better than chance at retest (trained: 92.8%; $z = 10.27$, $p < .001$; untrained: 86.8%; $z = 8.09$, $p < .001$).

Discussion

Experiment 2 sought to replicate and extend the findings of Experiment 1 by using an enhanced affix learning procedure, by making minor changes to the definitions in the semantic-learning condition, and by adding a retest several weeks after training. Data from the first testing session (which took place two days after training) demonstrates the effectiveness of the first two of these changes. Participants in Experiment 2 had higher recognition memory scores in both the form- and semantic-learning conditions than in Experiment 1 (all d' comparisons: $p < .05$), and those participants in the semantic-learning condition displayed a larger learning effect in definition selection than did those in Experiment 1 (learned words: $z = 3.03$, $p < .01$; untrained novel words: $z = 4.32$, $p < .001$). Data from the second testing session (which took place approximately 50 days after training) showed that these learning effects were long lasting. These findings suggest that participants had developed representations of the novel affixes and their meanings (for those in the semantic condition) that were sufficiently robust to persist over several weeks.

However, in spite of the methodological alterations to this experiment, data from the lexical decision task on Day 3 replicated Experiment 1 in showing only limited evidence that the representations of the novel affixes influenced automatic lexical processing. Though there was an affix learning effect on the accuracy data in the semantic condition (as in Experiment 1), once again there was no effect on response times. Somewhat remarkably, for those participants in the semantic-learning condition, such learning effects *were* apparent at retest several weeks after training, despite the fact that these participants received no further training and were not aware that they would be retested. This interference effect becomes particularly compelling when examining the *change* in performance from Day 3 to retest. In looking at the data in this manner, what is apparent is that performance generally gets better between Day 3 and the retest, presumably as a result of familiarity with the task: Participants in the semantic condition get 16 ms faster in rejecting nonwords with untrained affixes (645−629 ms); participants in the form condition get 25 ms faster in rejecting nonwords with untrained affixes (605−580 ms); and participants in the form condition get 24 ms faster in rejecting nonwords with trained affixes (610−586 ms). However, in contrast to this average improvement of 22 ms in rejection latency, participants in the semantic condition get 3 ms *slower* in rejecting nonwords with trained affixes between Day 3 and the retest. Similarly, while participants in the first three conditions described above show an average increase in the accuracy with which they reject nonwords between Day 3 and the retest, participants in the semantic condition get markedly worse in rejecting nonwords with trained affixes over this time period.[5]

Overall, these data show unambiguously that affixes learned in laboratory conditions can become lexicalized such that they are used in the online interpretation of untrained novel words (as is the case for existing affixes known to participants; e.g., Caramazza et al., 1988; Crepaldi et al., 2010). Semantic information about the novel affixes appears to play an important (or perhaps crucial) role in this lexicalization process, and certainly we have no evidence from this study that mere form learning gives rise to lexicalized affix representations. Finally, this lexicalization process seems to require quite a lengthy period of offline consolidation (sometime between two days and two months after training), though appears to give rise to robust representations that persist over time.

GENERAL DISCUSSION

Morphemes provide the building blocks of meaning in written and spoken language. They allow us to understand a limitless range of words based on knowledge of a relatively small number of elements and also provide the primary means for lexical productivity (Algeo, 1991). Priming experiments have revealed that skilled readers access morphemic information in the very earliest stages of visual word recognition (see Rastle & Davis, 2008), and that this information plays an integral role in the semantic analysis of unfamiliar morphemic combinations (de Vaan, Schreuder, & Baayen, 2007; Meunier & Longtin, 2007). However, despite the central role that morphemic knowledge plays in our use of language, very little is known about how this knowledge is acquired.

Our work sought to advance this area by introducing a new laboratory method for studying the acquisition of affix knowledge. Participants in our experiments were trained on novel affixes (e.g., *-nept*) presented in novel word contexts (e.g., *sleepnept*) and were then tested in various ways on trained and untrained novel stimuli (e.g., *sailnept*). Our experiments assessed not only whether participants could discover the affix units despite never being exposed to them in isolation (cf. Dahan & Brent, 1999; Saffran, Newport, et al., 1996) but also whether these representations would become lexicalized such that they generalized to the analysis of new stimuli in a speeded task. Our experiments further sought to establish the role of semantic information and offline consolidation on this lexicalization process by (a) comparing performance under both semantic-learning and form-learning conditions; and (b) comparing performance both two days and nearly two months after training.

Results of our experiments demonstrated for the first time that participants can acquire knowledge of novel affixes after a single training session in which the only cues to morphological decomposition are those implicitly provided by the novel stimuli to which participants are exposed.

---

[5] We are grateful to Arty Samuel for suggesting this alternative way of examining the data.

Participants in both learning conditions displayed good recognition memory performance, and their difficulty in rejecting recombinant pairs such as *sleephoke* indicates that their judgements were based on morphemic (as opposed to whole word) representations. However, we found little evidence at initial testing that these morphemic representations were impacting on automatic lexical processing, further supporting the dissociation observed by Gaskell and Dumay (2003) between episodic and lexicalized knowledge. It was not until the retest that the novel affixes showed unambiguous evidence of behaving like existing affixes, in slowing the rejection of previously unseen nonwords that contained the novel affixes (e.g., *stopnept*). This effect on response times strongly suggests that the representations of the novel affixes had become lexicalized, though this lexicalization process appears to have required a period of consolidation lasting between two days and nearly two months.

Critically, response time effects indicative of lexicalization of novel affixes were observed only in the semantic-learning condition. Though participants in the form-learning condition appeared to learn the novel affixes well (as established through their recognition memory performance), the only indication that these affix representations influenced automatic lexical processing was an accuracy effect on lexical decision in the retest of Experiment 2. Performance in this condition differed substantially from that in the semantic-learning condition, in which accuracy effects on lexical decision were apparent in every testing session, and in which response time effects on lexical decision were observed in the retest. Our findings therefore raise the possibility that, while participants can discover affix units without being exposed to them in isolation, the lexicalization of these representations may require semantic information. This conclusion is consistent with the *semantically driven* theory of morpheme acquisition advanced by Rastle and Davis (2008) as well as an increasing body of evidence suggesting that semantic information plays an influential role in the acquisition of new words (e.g., Leach & Samuel, 2007; Rueckl & Dror, 1994; Tamminen & Gaskell, 2008). It will be important in future work to establish whether the lexicalization of affix representations can *ever* occur in the absence of semantic information. Such work would require comparison of multiple different forms of affix learning as well as tests of the impact of affix learning in multiple test tasks.

Though we observed evidence consistent with the lexicalization of affix representations, it remains unclear why this process took such an extended period (certainly more than a couple of days), particularly in relation to work on learning new words presented as part of a continuous speech stream (Fernandes et al., 2009) that showed evidence for lexicalization shortly after training. However, one difference between our work and that of Fernandes et al. (also e.g., Gaskell & Dumay, 2003) is that our test of lexicalization involves presentation of nonword stimuli that contain the learned affix (e.g., *sailnept*, which includes an untrained stem paired with the learned affix -*nept*), whereas their test of lexicalization involved presentation of a cohort competitor of the learned stimulus instead of the stimulus itself (e.g., *cathedral*, for the learned stimulus *cathedruke*).

It seems possible that strong episodic representations of the learning experience could facilitate rejection decisions for stimuli containing the novel affixes, thus counteracting any difficulties in rejecting these items as a result of the formation of lexicalized representations. If so, then this may explain why response time effects in lexical decision emerged only after episodic knowledge had deteriorated significantly. However, this account seems potentially at odds with Dumay and Gaskell's (2007) work showing a positive relationship between overnight improvement in free recall and the size of the postsleep competition effect (i.e., greater competition observed with enhanced free recall), though Dumay and Gaskell speculate that free-recall performance following sleep may have a lexical component. Further research will be necessary to assess whether lexicalization of affix representations can be observed in other tasks without such a protracted period of time between training and testing and indeed whether, as suggested by complementary learning system accounts (e.g., Davis & Gaskell, 2010), episodic and lexicalized representations of new lexical knowledge are necessarily in opposition during learning (i.e., lexical knowledge emerges only when episodic representations degrade).

Perhaps the most important of our findings, however, is that we have shown for the first time that participants can discover sublexical information about novel stimuli presented to them and use this in the online interpretation of new exemplars. Following semantic training, participants used lexical representations of the novel affixes in the analysis of untrained novel stimuli (e.g., *stopnept*) presented in the context of a speeded lexical-processing task. Furthermore, though our definition selection task was non-speeded, participants showed a striking ability to select the correct definition of novel stem-plus-affix combinations by using affix meanings acquired during training. To our knowledge, this sort of meaning-based generalization has not been shown in artificial language learning experiments previously, and it sets the stage for more detailed investigations of the acquisition of lexicalized semantic and syntactic knowledge in morpheme learning. Overall, like other demonstrations of linguistic generalization in laboratory studies (e.g., Taylor, Plunkett, & Nation, in press), our findings suggest that the ability to generalize to novel combinations of linguistic elements following limited training with specific exemplars can be readily mimicked in the laboratory.

In conclusion, these findings demonstrate the potential of using an artificial learning paradigm in the investigation of the acquisition of affix knowledge. The paradigm introduced in this article permits the development of episodic and lexicalized affix representations that are well established and long lasting, and which are used in the interpretation of new exemplars. This method thus provides an empirical foundation for future studies in which other aspects of learned affixes are manipulated so as to provide a unique source of information on the nature of form and semantic representations of morphemic units. For example, future research might tease apart the roles of orthographic and phonological information on affix learning (as these sources of information were provided simultaneously in our experiments). Previous word-learning experiments have shown that the presentation of phonology strengthens orthographic representations of novel words (e.g., McKague, Davis, Pratt, & Johnston, 2008), and it is likely that it aided affix learning here. Similarly, one might use this learning paradigm to assess the role of familiar (word) stems on affix learning. The two form-based learning theories proposed by Rastle and Davis (2008) differ in whether they postulate a role for stem knowledge (morpheme chunking) or not (morpheme boundary detection). These two theories can probably be distinguished by using similar tests of affix learning following presentation of novel words that contain word stems (e.g., *sleepnept*) and novel words that contain nonword stems (e.g., *pleepnept*). We thus anticipate that, as for other domains of language (e.g., spoken words, Gaskell & Dumay, 2003; speech segmentation, Saffran, Newport, et al., 1996), significant theoretical progress will be made through studies using these artificial analogues of natural language.

Original manuscript received 3 November 2008
Accepted revision received 22 October 2010

REFERENCES

Algeo, J. (1991). *Fifty years among the new words: A dictionary of neologisms.* Cambridge, UK: Cambridge University Press.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge, UK: Cambridge University Press.

Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., et al. (2002). *The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords*, Retrieved from

Washington University Web site: http://elexicon.wustl.edu.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition*, *97,* B45–B54.

Brent, M. R. (1997). Towards a unified model of lexical acquisition and lexical access. *Journal of Psycholinguistic Research*, *26,* 363–375.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61,* 93–125.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28,* 297–332.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13,* 221–268.

Clay, F., Bowers, J. S., Davis, C. J., & Hanley, D. (2007). Teaching adults new words: The role of practice and consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33,* 970–976.

Crepaldi, D., Rastle, K., & Davis, C. J. (2010). Morphemes in their place: Evidence for position-specific identification of suffixes. *Memory & Cognition*, *38,* 312–321.

Dahan, D., & Brent, M. R. (1999). On the discovery of novel wordlike units from utterances: An artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology: General*, *128,* 165–185.

Davis, C. J. (1999). *The self-organising lexical acquisition and recognition (SOLAR) model of visual word recognition*. Unpublished doctoral dissertation, University of New South Wales, Sydney, Australia.

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review*, *117,* 713–758.

Davis, M. H., di Betta, A. M., MacDonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, *21,* 803–820.

Davis, M. H., & Gaskell, M. G. (2010). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364,* 3773–3800.

de Vaan, L., Schreuder, R., & Baayen, R. H. (2007). Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon*, *2,* 1–23.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18,* 35–39.

Dumay, N., Gaskell, M. G., & Feng, X. (2004). A day in the life of a spoken word. *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 339–344). Mahwah, NJ: Lawrence Erlbaum Associates.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14,* 179–211.

Fernandes, T., Kolinsky, R., & Ventura, P. (2009). The metamorphosis of the statistical segmentation output: Lexicalization during artificial language learning. *Cognition*, *112,* 349–366.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*, *35,* 116–124.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of new words. *Cognition*, *89,* 105–132.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24,* 380–396.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105,* 251–279.

Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. *Proceedings of the International Congress on Phonetic Sciences XVI* (pp. 49–54). Germany: Saarbrücken.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59,* 434–446.

Laine, M., Vainio, S., & Hyönä, J. (1999). Lexical access routes to nouns in a morphologically rich language. *Journal of Memory and Language*, *40,* 109–135.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55,* 306–353.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18,* 1–86.

McKague, M., Davis, C., Pratt, C., & Johnston, M. B. (2008). The role of feedback from phonology to orthography in orthographic learning: An extension of item-based accounts. *Journal of Research in Reading*, *31,* 55–76.

Meunier, F., & Longtin, C.-M. (2007). Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, *56,* 257–471.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52,* 189–234.

Plaut, D. C., & Gonnerman, L. M. (2000). Are nonsemantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processing*, *15,* 445–485.

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23,* 942–971.

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, *15,* 507–537.

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11,* 1090–1098.

Rueckl, J. G., & Dror, I. (1994). The effect of orthographic−semantic systematicity on the acquisition of new words. In C. Umiltà & M. Moscovitch (Eds.), *Attention and Performance XV*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274,* 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35,* 606–621.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117,* 34–50.

Tamminen, J., & Gaskell, M. G. (2008). *Novel words entering the mental lexicon: Is meaning necessary for integration?* Poster presented at the meeting of the Experimental Psychology Society, Cambridge, UK.

Taylor, J., Plunkett, K., & Nation, K. (in press). The influence of consistency, frequency and semantics on learning to read: An artificial orthography paradigm. *Journal of Experimental Psychology: Learning, Memory and Cognition*.

Wurm, L. H. (2000). Auditory processing of polymorphemic pseudowords. *Journal of Memory and Language*, *42,* 255–271.