Lang. Teach. (2013), **46.3**, 316–334 © Cambridge University Press 2013
doi:10.1017/S0261444813000013

Research Timeline

Formulaic language

Alison Wray Cardiff University, UK
wraya@cardiff.ac.uk

Creating a timeline for formulaic language is far from simple, because several partially independent lines of research have contributed to the emerging picture. Each exhibits cycles of innovation and consolidation over time: domains take a leading role in developing new knowledge and then fall back, while another area comes to the fore. Thus, some of the first observations about formulaic language, back in the nineteenth century, were in the clinical domain of aphasia studies. By the early to mid twentieth century it was theories of language structure that had most to say, until eclipsed by the Chomskian model, which saw little significance in lexicalised units larger than the word (an issue discussed by Jackendoff 2002; see table entry). Meanwhile, changes in language teaching methodology in the mid to late twentieth century increasingly urged teachers to ask how adult learners could best master multiword strings to improve fluency and idiomaticity – a question still asked today. By the end of the twentieth century, new technological advances revealed frequency in usage as a probable agent of formulaicity, and these chimed with new models of lexical knowledge based on neural pathways and networks that could be strengthened by repeated exposure. Drawing on these models, we have seen, as we move into the twenty-first century, the development of new approaches to modelling language as a system – emergent grammars, including Construction Grammar – that are more accommodating of large, internally complex units. And finally, as we gradually understand more about how the brain accesses and retrieves linguistic material, we are seeing a resurgence of interest in formulaic language in neurological and clinical contexts.

One important feature of the way an area of investigation develops over time is that some of the motivations for arguing a position become less relevant – perhaps even incomprehensible – to later readers. Some of the battles fought in the literature have now been won or lost, and no longer have to be re-fought. Thus, it is striking how many of the earlier works on this timeline are significant milestones because they argue that multiword strings might be a unit of lexical processing. They were written at a time when most considered that to be unlikely, so the assertion required courage and evidence to sustain. Yet researchers entering the arena now are generally very comfortable with this idea, and may puzzle about why it was such a big deal in the past. While we all progress our research by building on what has gone before, perhaps a word of caution is in order. It is not impossible for an assertion to become received wisdom just because it is repeated a lot – we should not forget the potential power of the Bellman's Rule of Three in Lewis Carroll's *The Hunting of the Snark*: 'What I tell you three times is true'. It behoves every researcher to ask from time to time 'how do we know this is

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

true?' and to look for evidence for the claim. The way we conceptualise formulaic language is highly contingent on particular cultural and contextual priorities and assumptions, which we use to impose order on phenomena that, being dually rooted in the human brain and in human society, are far too complex for us to engage with in any other way. The next generation of theories about formulaic language will develop when researchers take a step back and unpick some of those assumptions and priorities, in order to re-evaluate the now accumulated evidence. That work will then establish the significant new milestones in our developing understanding of the phenomenon.

Whilst mindful of the second language acquisition (SLA) focus of this journal, I have been keen not to present only the story of developments in that domain, though there would have been enough material for me to do so. The wider story outlined above seems to me crucial in evaluating the significance and usefulness of the L2-oriented research. That is, one cannot answer the question 'how can multiword strings be most effectively taught to learners?' unless one appreciates what a learner's memory and grammar are likely to be able to do with them. One cannot answer the question 'which multiword strings should be prioritised in teaching?' unless one understands what they are for in communication, and how they are managed in the linguistic system.

Before proceeding further, we need a working definition of formulaic language. There is none that is theory- or method-neutral, and there is not even consensus across domains of investigation about how open one should be to other definitions alongside one's own. Keeping focus on the issues most relevant to language teachers and learners, a plausible entry point is that 'formulaic language' refers to sequences of words that are in some regard not entirely predictable, whether on account of a meaning that is wildly or subtly different from the words they contain, a function that is only achieved with the whole expression, or features of structure such as morphology or word order that are non-canonical. For as long as language learning is construed or practised as the assemblage of single words using repeatable rules, formulaic language, under this definition, will be inherently troublesome to learners.

Most researchers will be comfortable to a greater or lesser extent with such a definition, but it does not by any means exhaust or adequately reflect some aspects of the potential broader phenomenon. For instance, it does not of itself address the frequency relationships between the elements in a wordstring, nor directly engage with the possibility that they are 'holistically' processed. It does not specifically allow for the flexibility of a partly fixed and partly open framework, nor does it help us understand how either the speaker or the language might acquire and retain examples of formulaic language. Finally, it has nothing directly to say about whether, or why, formulaic language should be resistant to language damage and loss. Yet all of these features, and others, are prevalent in the wider discussion of formulaic language that the timeline below represents. To my mind, it is the entire set of elements that make formulaic language as a whole interesting enough to care about researching – they require one to look for deep and sustainable answers to what might, within a single domain, seem rather simple questions.

The six themes into which I have organised the timeline references are far less discrete than the taxonomy might imply, but I have resisted extensive cross-labelling, since it should indeed be possible to track each line of the story more or less independently, particularly if

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

those texts given multiple theme labels are used to touch base with how each might interact with the others.

The items on the cutting-room floor are many – I had to be brutal to keep to the required length of list. However, I hope that, with the exception of very recent works, there will be few omissions to which the interested reader cannot find reference via the items that I have included. Inevitably, this particular selection is personal. It is biased by my own interests, and no doubt compromised by my ignorance of some works that readers may feel should be here. At the very least, I hope this timeline will act as a springboard for new discussion about what else should be in the story, and what impact the foregrounding of different historical and current contributions might have on what we understand formulaic language to be.

Themes

A. Theory: Processing, lexis and grammar
Papers under this theme examine (a) the form, meaning and function of formulaic language, locating it, explicitly or implicitly, within theoretical accounts of how language works more generally; (b) how formulaic language is processed and stored.

B. Clinical: Language disorders
This theme recognises the curious resilience of formulaic language in some kinds of language and communication disorder, including several types of aphasia and Alzheimer's Disease. It is also a feature of autism, though none of the papers in this timeline directly addresses that. For a brief overview, see Wray (2008a).

C. Development: First language acquisition
If we accept that formulaic language forms part of the adult speaker's lexical inventory, how does it get there? First language acquisition research helps tease apart two possible origins: we acquire small units and learn to build them up into chunks, or we acquire large units and gradually break them down.

D. Learning and teaching: SLA
Two sub-themes combine here: how L2 learners of different ages manage (or somehow fail) to acquire a reliable knowledge of the formulaic language inventory of the target, including the role of rules and creativity; and what ideally should be taught to learners to support their effective learning of useful multiword strings, in the interests of comprehensibility and fluency.

E. Culture: Oral traditions, social roles and cultural variation
In the literary domain, formulaicity is recognised as featuring in memorised and repeated texts – something that links with the clinical research, since rhymes and songs learned in childhood seem particularly resistant to loss. Research into how we engage creatively with memorised material, as in oral epics, informs our understanding of how memory works for multiword strings, and relates also to how memorisation is used for learning in general education and L2 classrooms. Meanwhile, the social roles that formulas play help us to understand why they persist and how they are acquired.

F. Text: Corpora

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

The relationship between formulaic language and the recurrent wordstrings found in corpora is complex, because the latter are an expression of absolute frequency, whereas some types of formulaic language have low frequency but high salience or uniqueness for their function. Models of language processing (theme A) that favour frequency of exposure as a key determinant of language learning (themes C and D) are likely to draw on corpus data for indications of what will be encountered.

Acknowledgements

This paper has been much improved as a result of insightful comments from Graeme Porte, Kon Kuiper and an anonymous reviewer.

Reference

Wray, A. (2008a). Formulaic sequences and language disorders. In M. Ball, M. R. Perkins, N. Müller & S. Howard (eds.), *Handbook of clinical linguistics*. Oxford: Blackwell, 184–197.

ALISON WRAY is a Research Professor in Language and Communication at Cardiff University, UK. Her research characteristically synthesises evidence from across domains, to create models of how language works, before testing them empirically. Her contribution to research into formulaic language spans some 15 years, and includes scores of papers and two monographs (2002, 2008b). The former won the 2003 book prize of the British Association for Applied Linguistics. She founded the international Formulaic Language Research Network (FLaRN), and has co-authored two textbooks for students: *Projects in linguistics* (3rd edn, Taylor and Francis, 2012), with A. Bloomer, and *Critical reading and writing for postgraduates* (2nd edn, Sage Publications, 2011), with M. Wallace.

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
1874	Hughlings Jackson, J. (1874/1958). On the nature of the duality of the brain. In J. Taylor (ed.) (1958), *Selected writings of John Hughlings Jackson, Vol II*. London: Staples Press, 129–145.	As a doctor examining patients with brain damage, **Hughlings Jackson** formed the view that some language was 'automatic' and 'non-propositional', and was processed by the right hemisphere rather than the linguistically dominant left. Technological advances have increasingly shown that the right hemisphere does indeed play a role in the holistic aspects of language (see VAN LANCKER[1] 1987, WRAY 2002, 2008a[2]).	**B**
1937	Firth, J. R. (1937/1964). *The tongues of men* and *Speech*. London: Oxford University Press.	**Firth**'s famous quote, 'you shall know a word by the company it keeps' is from later writing[3], but these earlier works mark his realisation that language as a syntagmatic system significantly influences how paradigmatic patterns are manifested. This view of language fell out of fashion in mainstream linguistics in the later twentieth century but has powerfully resurfaced in emergent grammar models.	**A**
1960	Lord, A. (1960/2000). *The singer of tales*. Cambridge, MA: Harvard University Press.	Building on the work of Milman Parry, **Lord** lays out the processes by which oral epic poetry was memorised and delivered, from ancient Greek to twentieth-century Serbo-Croatian performers. The formula – with a particular definition and set of functions in this context – was central to the capacity to recall all the material while still being creative.	**E**
1976	Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum* 1, 1–14.	**Bolinger** held fast to a more holistic view of language than some contemporaries, proposing that 'our language does not expect us to build everything starting with lumber, nails, and blueprint, but provides us with an incredibly large number of prefabs' (p. 1). He explained the absence in English of **sometime else* and **an extended time ago* as being 'not because the generative mechanism is lacking' but because 'we have not heard it done. We have no memory of it' (p. 4).	**A**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
1976	Gleason, J. B. & S. Weintraub (1976). The acquisition of routines in child language. *Language in Society* 5, 129–136.	This charming study examines how young American children become socialised into using the appropriate cultural routines during the annual Hallowe'en 'trick or treat' visits to neighbours' houses. The authors show that, in contrast to other elements of first language acquisition, explicit instruction plays a significant role and the child does not need to understand the internal composition of the phrase: 'performance is all that matters. . . with routines, it doesn't matter what you think as long as you perform at the right moment' (p. 134).	**C, E**
1976	Ferguson, C. A. (1976). The structure and use of politeness formulas. *Language in Society* 5, 137–151.	Appearing in the same issue as GLEASON & WEINTRAUB (1976), this paper examines the ubiquity of politeness routines across different languages, and considers the importance of their social role.	**E**
1976	Wong Fillmore, L. (1976). *The second time around: Cognitive and social strategies in second language acquisition.* Ph.D. thesis, Stanford University (www.Proquest.com or http://disexpress.umi.com/dxweb).	This thesis, and an associated shorter paper in 1979[4], presents one of the most thorough accounts of children's naturalistic second language acquisition. **Wong Fillmore** observed five children (5;6 to 9 years) over one school year, and showed how the individual's capacity for social interaction played a key role in determining the speed and effectiveness of their *ab initio* acquisition of English. Picking up and reusing not yet fully understood multiword units was instrumental in bootstrapping to nativelikeness.	**D**
1979	Coulmas, F. (1979). On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics* 3, 239–266.	**Coulmas** argues that one cannot understand, use or translate routine formulae without understanding the 'cognitive system of beliefs, wants, wishes, preferences, norms and values' (p. 239) that underpin their appropriate use. See also his edited collection of 1981, which includes a useful introductory chapter.[5]	**A**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
1980	Peters, A. M. (1980) *Units of language acquisition*. Cambridge: Cambridge University Press.	**Peters** describes a child whose approach to first language acquisition was holistic rather than based on simple words and rules. Her subject repeated long wordstrings, complete with intonation, with vague filler syllables to mark material he did not know. She proposes that his acquisition entailed identifying loci of paradigmatic variation in input, and only introducing variation at those points. Peters' work directly inspired the *needs only analysis* model of WRAY (2002).	**C**
1981	Nelson, K. (1981). Individual differences in language development: Implications for development and language. *Developmental Psychology* 17.2, 170–187.	**Nelson** argues that first language acquisition can favour either of two styles, one of which is a 'gestalt' or holistic phrase-based approach. Nelson's work influenced accounts of L1 acquisition by Bates and others who proposed that grammar knowledge emerges as a consequence of working with multiword units of input.	**C**
1983	Pawley, A. & F. H. Syder (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (eds), *Language and communication*. New York: Longman, 191–226.	In what is perhaps the single most cited paper on formulaic language, the authors ask how native speakers know which subset of the grammatically possible formulations of an idea is the idiomatic, nativelike one, and how speakers succeed in speaking so fluently, given the demands of online processing. Their solution is that lexicalised sentence stems are used as the basic units of language, edited to be appropriate to the specific context.	**A**
1984	Raupach, M. (1984). Formulae in second language speech production. In H. Dechert, D. Möhle & M. Raupach (eds.), *Second language productions*. Tübingen: Gunter Narr, 114–137.	**Raupach** examines the ways in which native and (German) non-native speakers of French produce spoken language, and identifies short phrases such as *je ne sais pas* and *je crois que* – produced formulaically – as a way of sustaining fluency and buying time during planning. He terms these expressions 'islands of reliability'.	**D**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
1987	Sinclair, J. McH. (1987). Collocation: Progress report. In R. Steele & T. Threadgold (eds.), *Language topics Vol I: Essays in honour of Michael Halliday.* Amsterdam: John Benjamins, 319–331.	In this report, **Sinclair** introduces for the first time his *open choice* and *idiom* principle, the latter founded on the claim that 'a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments' (p. 320).[6]	**A**
1987	Van Lancker, D. (1987). Nonpropositional speech: Neurolinguistic studies. In A. W. Ellis (ed.), *Progress in the psychology of language, Vol III.* Hillsdale, NJ: Lawrence Erlbaum, 49–118.	A remarkably informative and engaging paper drawing together the evidence for formulaic language as a feature of disordered language, and locating it within a broader discussion of how language is processed by the left and right hemispheres of the brain.	**B**
1988	Fillmore, C. J., P. Kay & M. C. O'Connor (1988). Regularity and idiomaticity in grammatical constructions: The case of 'let alone'. *Language* 64.3, 501–538.	**Fillmore et al**.'s paper is a significant marker in the development of Construction Grammar. An account designed to explain the exceptional formulations in language (idioms) is shown to be equally effective in explaining the regular patterns. The authors argue for the simultaneous operation of morphosyntactic, semantic and pragmatic patterns that can apply above the lexical level.	**A**
1988	Gatbonton, E. & N. Segalowitz (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly* 22.3, 473–492.	A significant contribution to practical classroom L2 teaching, **Gatbonton & Segalowitz** demonstrate ways of incorporating multiple rehearsals of useful multiword strings into learning activities as a genuinely communicative act. The rationale is that such rehearsal promotes fluency and familiarity with expressions that can be used more generally.	**D**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
1992	Nattinger, J. R. & J. S. De Carrico (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.	**Nattinger & De Carrico** explain the linguistic rationale for focussing on lexical phrases in language teaching, and then present practical approaches for doing so. The theoretical framework examines these phrases on both the formal and the functional dimensions.	**D**
1993	Lewis, M. (1993). *The lexical approach*. Hove, UK: Language Teaching Publications.	Addressing language teachers, **Lewis** proposes a method for teaching English that is based on the combination of chunks to produce output: words, collocations, fixed expressions and semi-fixed expressions. Although not the only method to do this, it has been one of the most influential.	**D**
1993	Hickey, T. (1993). Identifying formulas in first language acquisition. *Journal of Child Language* 20, 27–41.	**Hickey** examines the approaches taken by previous researchers to identifying formulaic language in first language acquisition data, and operationalises for her own data a set of criteria that capture the key features.	**C**
1993	Plunkett, K. (1993). Lexical segmentation and vocabulary growth in early language acquisition. *Journal of Child Language* 20, 43–60.	**Plunkett** examines the timing of the vocabulary spurt as a reflection of how the individual child learns to segment input. His data confirm the findings of NELSON (1981): the child's preference for large or small units is determined by the nature of carer input.	**C**
1994	Nunberg, G., I. A. Sag & T. Wasow (1994). Idioms. *Language* 70.3, 491–538.	The authors argue that idioms fall into types, some of which are more productive than others. By proposing that this flexibility is driven by semantics, not grammar, they challenged key assumptions of the then dominant generative grammar model.	**A**
1994	Williams, E. (1994). Remarks on lexical knowledge. *Lingua* 92, 7–34.	**Williams** (see also Disciullo & Williams, 1986[7]) defines the lexicon as 'the repository of forms about which something special must be learned' (p. 8), thus raising the question of what in a language needs to be learned outright, rather than derived by rule. In his view, not only idioms but also abstract structures can be stored in the lexicon.	**A, C**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
1995	Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics* 16.2, 180–205.	This paper was a true landmark, in surveying a range of existing research into formulaic language in the L2 context, opening the way for others to engage with and extend the discussion of underlying patterns and issues. **Weinert**'s plea for attention to how best to define and identify formulaic language continues to resonate today.	**D**
1995	Goldberg, A. (1995). *Constructions*: *A Construction Grammar approach to argument structure*. Chicago, IL: University of Chicago Press.	One of the milestones in the development of Construction Grammar, in which **Goldberg** argues that language forms are located on a grammar–lexis continuum and consist of Constructions: form–meaning correspondences of various sizes from morphemes to multiword strings, including frames that are partly fixed and partly open to the insertion of other Constructions. An important insight is that Constructions can carry meaning even in the absence of specified lexical items. For a shorter introduction to this theory, see Goldberg (2003), and for a revised account of the model, Goldberg (2006).[8]	**A**
1995	Rubin, D. C. (1995). *Memory in oral traditions*. Oxford: Oxford University Press.	**Rubin** reviews the role of memory in the transmission of oral culture, with theoretical modelling of how information must be packaged (including formulaically) to ensure it can be reliably recovered.	**E**
1995	Grace, G. W. (1995). Why I don't believe that language acquisition involves the construction of a grammar. *Ethnolinguistic notes* Series 4, number 1, http://www2.hawaii.edu/~grace/elniv1.html	**Grace**'s contribution to our current understandings of language has been rather under-recognised. In this note, he observes that 'our knowledge of language should be thought of as essentially taking the form of a repertoire of linguistic expressions' (p. 2). In another essay[9] he argues that our linguistic theories have been too heavily influenced by local cultural and linguistic features to capture the general principles of language.	**A**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
1996	Kuiper, K. (1996). *Smooth talkers.* Mahwah, NJ: Lawrence Erlbaum.	A valuable contribution using evidence from auctioneers and sports commentators to support PAWLEY & SYDER's (1983) claim that we construct spoken output from pre-formed chunks. **Kuiper** shows how formulaic material sustains fluency and how hearers who are initiated into the cultural practice can recognise figure and ground so that the most important information is easily identified. **Kuiper** has made a sustained and significant contribution to formulaic language research; Kuiper (2009)[10] presents a further collection of his studies.	**A**
1996	Aijmer, K. (1996). *Conversational routines in English: Convention and creativity*. London: Longman.	Using data from the London-Lund corpus, **Aijmer** shows how different pragmatic functions are achieved in real conversation, with a useful introduction which locates them in theory and teaching.	**A**
1998	Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach.* Oxford: Clarendon Press.	An account of idioms and their variations in an 18 million-word corpus. Alongside a detailed description of many examples, there is extensive discussion of the significance of the findings for our understanding of creativity and meaning-making, and a recognition that frequency is not enough to explain our knowledge of idioms.	**F**
1998	Cowie, A. P. (1998). *Phraseology: Theory, analysis and applications.* Oxford: Clarendon Press.	Of particular significance in this collection is the coverage – all too rare – of the theoretical approaches to phraseology taken by researchers in eastern Europe (including the former Soviet Union), most notably Igor Mel'čuk.	**A**
1999	Perkins, M. R. (1999). Productivity and formulaicity in language development. In M. Garman, C. Letts, B. Richards, C. Schelletter & S. Edwards (eds.), *Issues in normal and disordered child language: From phonology to narrative*. Special Issue of *The New Bulmershe Papers*. Reading, UK: University of Reading, 51–67.	**Perkins** proposes that 'the adult language system tends towards formulaicity and . . . the default mode of language processing is holistic' (p. 52). He suggests that first language acquisition begins and ends with an emphasis on formulaicity, with an interim period of analysis that creates the productive system. Part of the transition to adult knowledge is finding a balance between the two, such that any idea can be expressed, but without unnecessary processing effort.	**C**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
1999	Myles, F., R. Mitchell & J. Hooper (1999). Interrogative chunks in French L2: A basis for creative construction? *Studies in Second Language Acquisition* 21, 49–80.	This paper, one of several from the authors' project, reports on a longitudinal study of secondary school children learning French, and shows how they used rote-learned wordstrings as the basis for developing a more flexible capacity for expression.	**D**
2000	Hunston, S. & G. Francis (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.	A highly significant study that demonstrates the power of corpus research for informing linguistic theory. Drawing on the data gathered for the COBUILD dictionaries, the authors reveal how grammatical patterns are associated with particular semantic sets. The observations offer independent confirmation of the patterns described in Construction Grammar (e.g. GOLDBERG 1995).	**A**
2002	Jackendoff, R. (2002). What's in the lexicon? In S. Nooteboom, F. Weerman & F. Wijnen (eds.), *Storage and computation in the language faculty*. Dordrecht: Kluwer, 23–58.	This very important essay reflects **Jackendoff**'s concerns over a number of years regarding Chomsky's (1965)[11] conception of the lexicon as containing only items that cannot be constructed by rule (see also WILLIAMS 1994) . He proposes a larger lexicon that contains, amongst other things, phrasal structures of the kind found in Construction Grammar (see GOLDBERG 1995). His argument is based on the premise that the best theory of grammar will reflect how we compute language in real time.	**A**
2002	Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24, 143–188.	This paper is representative of **Ellis**' outstandingly thorough and scholarly contribution to research into formulaicity from the perspective of frequency, particularly in relation to second language acquisition (see also SIMPSON-VLACH & ELLIS 2010). His view, that 'formulas are lexical chunks that result from binding frequent collocations' (p. 155), emphasises the potential for small units that frequently combine to become chunked (and is therefore in contrast to PETERS' 1980 and WRAY's 2002 position that formulaic language is usually broken down, not built up).	**D**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
2002	Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.	This book set out to draw together most of the different strands of evidence for formulaic language and develop a model that could account for them coherently. The model was subsequently tested empirically and further discussed in WRAY (2008b).	**A-D, F**
2003	Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.	It is difficult to pick just one item from **Tomasello**'s significant contribution to usage-based accounts of first language acquisition. This book presents a detailed and coherent account of how children extract form–meaning chunks (often several words long) from their input and use them as building blocks for their own output.	**C, A**
2004	Schmitt, N. (ed.) (2004). *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins.	This collection presents a number of new methodologies for capturing elements of the essence of formulaic language, including measurements of processing in real time, approaches to identification, and units of recall.	**A**
2004	Ji, F. (2004). *Linguistic engineering: Language and politics in Mao's China*. Honolulu, HI: University of Hawai'i Press.	This fascinating and unusual study examines a real-life version of Orwell's novel *1984* from Mao's China, where the regime enforced the memorisation and repetition of approved slogans and formulas, in order to change people's beliefs, thoughts and values. WRAY (2008b) uses **Ji**'s study as a lynchpin for her discussion of whether formulaic language does indeed control thought, concluding, with **Ji**, that ultimately it does not.	**A**
2005	Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Abingdon, UK: Routledge.	**Hoey** argues that grammar is an emergent property of the patterns into which words fall. Although his account has similarities with other emergent grammar models (e.g. FILLMORE ET AL. 1988, GOLDBERG 1995), for him the fundamental unit of language is the word (rather than the construction, which can be larger or smaller), and is primed for the selection of collocates.	**A**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
2005	Dąbrowska, E. & E. Lieven (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16.3, 437–474.	This research presents convincing evidence that children are able to build up their own novel utterances using lexicalised frames (constructions) that they have heard in their input, enabling them to produce quite complex sentences by means of very simple insertion operations that embed one fully or partly lexicalised string into another. See also Bannard & Lieven's paper in CORRIGAN ET AL. (EDS) 2009, Vol. II.	**C**
2006	Boers, F., J. Eyckmans, J. Kappel, H. Stengers & M. Demecheleer (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research* 10, 245–261.	The issues of how formulaic sequences can best be taught, and whether doing so has any real influence on fluency, are persistent questions in L2 research, and the focus of many M.A. and Ph.D. dissertations. This study explores both. Although it is not surprising that learners whose attention was drawn to formulaic sequences ended up producing more of them, the method employed here will be a valuable point of reference for scholars wishing to run similar investigations.	**D**
2007	Burger, H., D. Dobrovol'skij, P. Kühn & N. R. Norrick (eds.). (2007). *Phraseologie/Phraseology: Ein internationales Handbuch zeitgenössischer Forschung* (2 volumes). Berlin: de Gruyter.	This is a two-volume collection that examines multiword strings from the continental research perspective. A number of significant pieces can be found in these volumes, broadening the perspective on how wordstrings are formed, processed and used. Some papers are in English, some in German.	**A, C, D, E, F**
2007	Wray, A. & G. W. Grace (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117.3, 543–578.	This paper extends the discussion of formulaic language into an exploration of how present-day languages in literate, internationally oriented societies might feature less formulaicity than is 'natural' in evolutionary terms – that is, that features of modern culture distort the balance of lexical unit sizes.	**A, E**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
2007	Ding, Y. (2007). Text memorization and imitation: The practices of successful Chinese learners of English. *System* 35.2, 271–280.	Interest in formulaic language from Chinese scholars has been significant, in part because of the country's huge commitment to the effective teaching of foreign languages, especially English, superimposed on an educational system that values repetition and memorisation as a method of accurate learning. **Ding** reports the role of these methods in high-achieving L2 English users.	**D**
2007	Kuiper, K., M.-E. van Egmond, G. Kempen & S. A. Sprenger (2007). Slipping on superlemmas: Multiword lexical items in speech production. *The Mental Lexicon* 2.3, 313–357.	**Kuiper et al.** examine two corpora of naturally occurring speech errors (in English and Dutch) involving formulaic language and collocations, in order to test two competing models of how formulaic expressions are processed during speech production.	**A**
2008	Wray, A. (2008b). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.	The book aims to extend understanding of formulaic language by examining the implications of the WRAY (2002) model for wider theory, and by comparing and contrasting lines of empirical evidence. It ends by identifying a range of questions that remain unanswered.	**All**
2008	Granger, S. & F. Meunier (eds.) (2008). *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins.	This is one of two significant collections of essays deriving from an international conference hosted in 2005 by **Granger**'s team in Belgium. It examines the nature of multiword sequences from a range of perspectives, with an impressive line-up of authors.	**A, E, F**
2008	Meunier, F. & S. Granger (eds.) (2008). *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins.	This second volume from the Belgian team is focussed specifically on questions around language teaching and learning, with a strong emphasis on corpus approaches to identifying multiword strings.	**D**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
2009	Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics* 14.3, 275–311.	This paper is chosen as representative of **Biber**'s high quality research on the distribution of 'lexical bundles' (high frequency wordstrings) in corpora, particularly their different distribution across registers and genres. The relationship between lexical bundles and other types of formulaic language remains difficult to pin down, but the frequency-based approach has considerable advantages for many types of research into formulaicity.	**F**
2009	Corrigan, R., E. A. Moravcsik, H. Ouali & K. M. Wheatley (eds.) (2009). *Formulaic language, Vol I (Distribution & historical change) and Vol II (Acquisition, loss, psychological reality, and functional explanations)*. Amsterdam: John Benjamins.	These two volumes capture the state of the art as presented at a landmark conference in Milwaukee in 2007. They contain numerous chapters that really deserve their own entries in this timeline, including a retrospective by one of the giants of formulaic language research, Andrew Pawley (see also PAWLEY & SYDER 1983).	**All**
2009	Cheng, W., C. Greaves, J. McH. Sinclair & M. Warren (2009). Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams. *Applied Linguistics* 30.2, 236–252.	This paper offers a method for computational examination of the richer set of information about a word that arises from observing its collocates in their different patterns of occurrence. The authors use this method to explore an idea proposed by Sinclair: that a word and its collocate jointly create new meaning for themselves by co-occurring, and deserve to be viewed as a semantic unit in their own right (a 'meaning shift unit').	**A**
2010	Simpson-Vlach, R. & N. C. Ellis (2010). An academic formulas list: New methods in phraseological research. *Applied Linguistics* 31.4, 487–512.	This is a very important paper for two reasons. Firstly, the Academic Formulas List offers a valuable resource to those interested in the mastery of formulaic sequences in academic English as an L2. Secondly, the authors' method for identifying and classifying the examples demonstrates how the sometimes conflicting priorities of frequency, salience and function can be effectively reconciled.	**D**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
2010	Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.	**Bybee** is another researcher whose contribution to formulaic language research is difficult to capture in just one entry. This book draws together ideas that she has developed over several years into an account of how language comes to take the form it does. A central contribution to our understanding of formulaicity is her evidence that not only formulaic word strings but also elements of grammar can arise through fusion caused by frequent use.	**A**
2010	Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.	A thorough and highly scholarly examination of the evidence from psychology, linguistics and neuroscience regarding the nature of fluent spoken production, in the context of how a second language can be mastered.	**A, D**
2010	Tremblay, A. & H. Baayen (2010). Holistic processing of regular four-word sequences: a behavioural and ERP study of the effects of structure, frequency and probability on immediate recall. In D. Wood (ed.), *Perspectives on formulaic language*. London: Continuum, 151–173.	For many years claims have been made that formulaic sequences are holistically processed, but there has been little direct evidence that they actually are. This study aimed to compare the nature of the brain activity when formulaic and matched non-formulaic wordstrings were processed. While there is more to do in this research, and the authors themselves admit that it cannot differentiate between holistic retrieval and fast sequential retrieval, it is an exciting and significant addition to the portfolio of investigations into formulaic language.	**A**
2010	Lin, P. M. S. (2010). The phonology of formulaic sequences: A review. In D. Wood (ed.), *Perspectives on formulaic language*. London: Continuum, 174–193.	A valuable overview of the phonological features of formulaic sequences, with an eye to how they can be identified in spoken data.	**D**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
2010	Davis, B. H. & M. Maclagan (2010). Pauses, fillers, placeholders and formulaicity in Alzheimer's discourse: Gluing relationships as impairment increases. In N. Amiridze, B. H. Davis & M. Maclagan (eds.). *Fillers, pauses and placeholders*. Amsterdam: John Benjamins, 189–215.	This is a study of how repetition and formulaic fillers are used to patch up the deteriorating communication system in people with Alzheimer's disease. It sheds light on how the roles played by formulaic language in normal discourse might determine the communicative resources available to a person under extreme cognitive pressure.	**B**
2011	Siyanova-Chanturia, A., K. Conklin & N. Schmitt (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research* 27.2, 251–272.	This study uses eye-tracking technology to examine the patterns of fixation by the eyes on formulaic and matched non-formulaic strings. Although the findings are equivocal, eye-tracking is an interesting method for examining how we attend to formulaic language in written text.	**A, D**
2011	Wray, A. (2011). Formulaic language as a barrier to effective communication with people with Alzheimer's Disease. *Canadian Modern Languages Review* 67.4, 429–458.	The repetitive nature of output by people with dementia is well known. **Wray** seeks to open up discussions about the role that formulaic language more generally plays in sustaining and impeding communication in people with Alzheimer's, and also the impact that it can have on carers.	**B**

https://doi.org/10.1017/S0261444813000013 Published online by Cambridge University Press

YEAR	REFERENCE	ANNOTATION	THEME
2012	*Annual Review of Applied Linguistics* 32 (2012). Topics in formulaic language. Cambridge: Cambridge University Press.	It is a significant milestone for formulaic language research that a volume of ARAL has been dedicated to it. The papers, many written by authors featured elsewhere in this timeline, review the most recent developments in different sub-areas, putting down valuable markers for future research.	**A-D, F**

[1] Authors' names are shown in small capitals when the study referred to appears in this timeline.

[2] Wray, A. (2008a). Formulaic sequences and language disorders. In M. Ball, M. Perkins, N. Müller & S. Howard (eds.), *Handbook of clinical linguistics.* Oxford: Blackwell, 184–197.

[3] Firth, J. R. (1957). *Papers in linguistics 1934–1951*. London: Oxford University Press, 11.

[4] Wong Fillmore, L. (1979). Individual differences in second language acquisition. In C. J. Fillmore, D. Kemper & S.-Y. W. Wang (eds.), *Individual differences in language ability and language behavior*. New York: Academic Press, 203–228.

[5] Coulmas, F. (ed.) (1981). *Conversational routine: Explorations in standardized communication situations and prepatterned speech*. The Hague: Mouton.

[6] See also Sinclair, J. McH. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

[7] DiSciullo, A. M. & E. Williams (1986). *On the definition of word*. Cambridge: Cambridge University Press.

[8] Goldberg, A. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Science* 7.5, 219–224; Goldberg, A. (2006). *Constructions at work*. Oxford: Oxford University Press.

[9] Grace, G. W. (2002). Collateral damage from linguistics? 3: The role of culture-centrism. http://www2.hawaii.edu/~grace/elniv23.html

[10] Kuiper, K. (2009). *Formulaic genres*. Basingstoke: Palgrave Macmillan.

[11] Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.