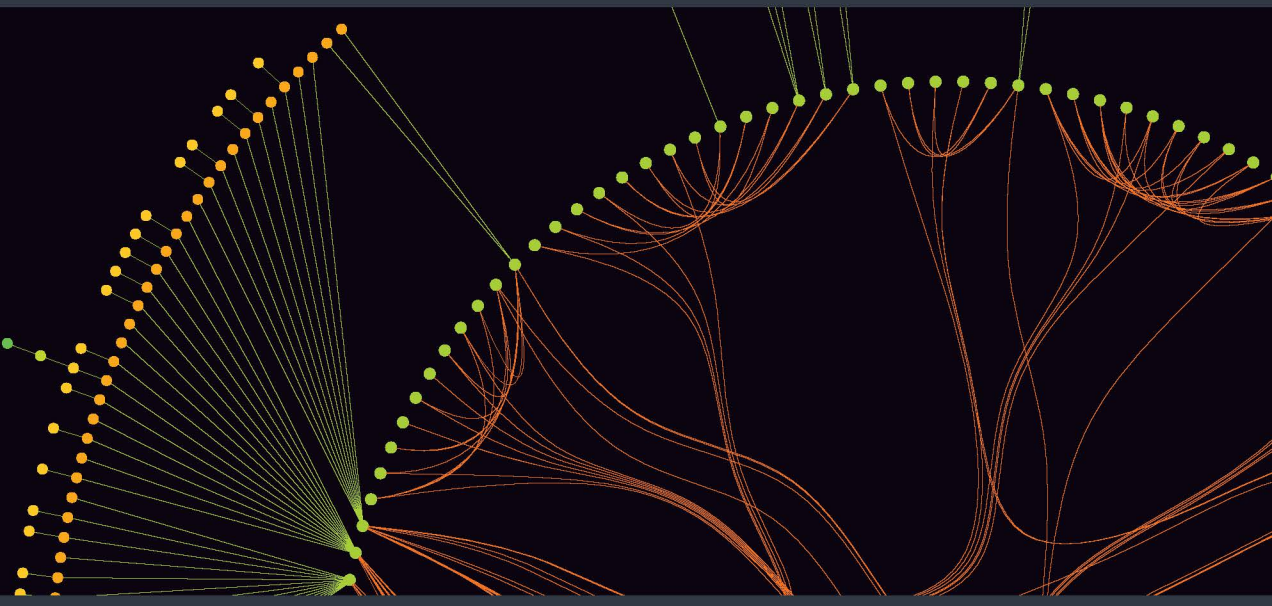
Łódzkie Studia z Językoznawstwa Angielskiego i Ogólnego
Łódź Studies in English and General Linguistics
7

Piotr Pęzik

Facets of prefabrication

Perspectives on modelling and detecting phraseological units



Facets of prefabrication

Perspectives on modelling and detecting phraseological units

Łódzkie Studia z Językoznawstwa Angielskiego i Ogólnego
Łódź Studies in English and General Linguistics

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

Łódzkie Studia z Językoznawstwa Angielskiego i Ogólnego
Łódź Studies in English and General Linguistics
7

Piotr Pęzik

Facets of prefabrication

Perspectives on modelling and detecting phraseological units

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO
Łódź 2018

Piotr Pęzik – University of Łódź, Faculty of Philology, Institute of English Studies
Department of English Language and Applied Linguistics
Corpus and Computational Linguistics Laboratory, Pomorska 171/173 St., 91-137 Łódź

REVIEWER
Barbara Lewandowska-Tomaszczyk

INITIATING EDITOR
Urszula Dzieciątkowska

TYPESETTING
Munda – Maciej Torz

TECHNICAL EDITOR
Leonora Wojciechowska

COVER DESIGN
Katarzyna Turkowska

Cover Image: Visualization of a collocational subsumption graph generated
with the yEd Graph Editor

Published by Łódź University Press
First edition. W.08156.17.0.M

Publisher's sheets 13.3; printing sheets 16.75

ISBN 978-83-8088-973-6
e-ISBN 978-83-8088-974-3

Łódź University Press
90-131 Łódź, 8 Lindleya St.
www.wydawnictwo.uni.lodz.pl
e-mail: ksiegarnia@uni.lodz.pl
phone. (42) 665 58 63

Table of contents

1. From novelty to prefabrication

This is the essence of the language instinct: language conveys news.
(Pinker 2007: 84)

It is evident that rote recall is a factor of minute importance in ordinary use of language.
(Chomsky 1964: 8)

(…) Speakers do at least as much remembering as they do putting together (…)
(Bolinger 1979: 97)

Compositionality and novelty in language

Novelty and prefabrication are some of the basic aspects of language use. As illustrated in the first two quotes in the epigraph to this chapter, the ability to achieve novelty in natural language, taken to mean the ability to create lexically, syntactically and semantically new expressions, is sometimes described as one of its fundamental characteristics. The underlying intuition of these claims is that in most communicative contexts speakers and writers take full advantage of the freedom to build linguistic expressions by adhering to a small number of syntactic rules and constructing complex phrases, clauses, sentences and texts. The meaning of such novel units of language is assumed to be motivated largely by the conventional or contextual meanings of their constituents. The view that such infinitude and novelty constitute the essence of our linguistic experience is partly challenged in the observation made by Bolinger (1979), who emphasizes that even at the level of syntactic clauses, and certainly at the level of phrases, language is as often reused, reproduced and recalled from memory as it is spontaneously generated and that *"a great*

deal of what we have been regarding as syntactic will have to be put down as morphological" (Bolinger 1979: 97). In other words, phrase and even sentence structures are often as formulaic and idiosyncratic as combinations as combinations of lexical morphemes in many complex words. In the most general terms, the present volume concerns itself with these two different perspectives on how language is produced and understood.

The potentially unlimited novelty of language is often formally related to the property of compositionality. Although compositionality is not a universally accepted view of how linguistic expressions can acquire an infinite number of complex meanings, in formal semantics, the so-called Principle of Compositionality remains a highly influential and widely researched idea.[1] In its basic form the Principle states that the meanings of complex expressions are fully and formally determined by the meaning and structure of their constituents (Szabó 2013) and their syntactic arrangement. Compositionality has been described as "a fundamental presupposition of most contemporary work in semantics" (Szabó 2013) and "a widely acknowledged cornerstone for any theory of meaning" (Werning 2012: 633). Together with its underlying syntactic mechanisms such as recursion, the Principle of Compositionality is often found in formal accounts of the freedom to string words together and form grammatically valid expressions, which in all likelihood, have never been uttered or written before in a particular language. As a general idea, it is sometimes traced back to Frege's famous observation that "with a few syllables (human languages) can express an incalculable number of thoughts" and that "this would not be possible, were we not able to distinguish parts in the thought corresponding to the parts of a sentence" (Frege 1923).[2] It is not difficult to explain why compositionality presents itself as a methodologically attractive prospect; if all but a handful of the most obscurely idiomatic linguistic expressions were compositional and derivable from their atomic components, then a rather elegant methodology of describing and explaining the structure of language would be possible. Its theoretical appeal would lie in formal verifiability and parsimony of description, whereby a finite set of rules and structural configurations could be used to account for the full productivity of linguistic expression and the ability to understand grammatically valid propositional language, including "sentences appearing for the first time in the history of the universe" (Pinker 2007: 9).

As already signaled, apart from semantic aspects of compositionality, linguistic novelty has also been defined in purely syntactic terms as the ability to

1 See Werning (2012) for a current review of research in this area.

2 Janssen (2001), however, argues against this interpretation of Frege's observation.

generate an infinite number of grammatically valid sentences which are unique combinations of lexical and syntactic constituents. This type of formal-syntactic novelty of language has remained an important point on the theoretical agenda of the generative grammar tradition. It has been conjectured that the emphatic assertion of the infinite novelty of linguistic expression by the early generative community was part of the general objection to associationist psychology and Skinnerian behaviorism (Pullum and Scholz 2010). By presenting the so-called *Infinitude Claim*, i.e. the claim that the collection of all well-formed linguistic expressions is an infinite set (ibid.) as one of the axioms of linguistic theory, generativists could justify their wholesale rejection of behaviorist linguistic models and emphasize the need for a new theory of language which could account for its potentially unrestrained productivity. The emphasis on novelty as opposed to "rote recall in ordinary use of language" is evident in the following assertions:

> The central fact to which any significant linguistic theory must address itself is this: a mature speaker can produce a new sentence of his language on the appropriate occasion, and other speakers can understand it immediately, though it is equally new to them. (Chomsky 1964: 1)

> It is evident that rote recall is a factor of minute importance in ordinary use of language, that a minimum of the sentences that we utter is learnt by heart as such – that most of them, on the contrary, are composed on the spur of the moment and that one of the fundamental errors of the old science of language was to deal with all human utterances, as long as they remain constant to the common usage, as if something merely reproduced from memory (Chomsky 1964: 8), (Paul 1886).

Generativists have sometimes hinged the seemingly unrestrained formal novelty of linguistic expression upon the syntactic mechanism of recursion. For instance, according to Yang (2006), "many have argued that the property of recursive infinity is perhaps *the* defining feature of our gift for language". It is therefore understandable that recursion has survived as one of the longest-standing members of the shrinking set of 'linguistic universals' and that it has recently been reaffirmed as "the only uniquely human component of the faculty of language" (Hauser, Chomsky, and Fitch 2002).[3]

[3] Some confusion over its exact role and universality stems from the parallel use of two definitions of recursion in linguistics. The first, more formally obliging meaning of recursion is "the self-embededness of syntactic constituents" through syntactic constructs and operations such as possessives, conjunction or clausal

Admittedly, there are many theoretical objections to the role of compositionality and the universality of recursion in human languages. In psycholinguistic terms, the actual productive potential of recursion is limited by the cognitive analogue of computational "stack-overflow errors", which may occur when recursive functions run out of address space due to the lack of a proper termination condition. For example, in natural languages, recursion of depth 6 or greater is rare for complex clauses, probably because of memory constraints and the cognitive effort required to produce and process such recursive structures (Baggio, van Lambalgen, and Hagoort 2012). Also, the Infinitude Claim is logically problematic in its appeal to mathematical induction and formally controversial in its definition of generative grammars (Pullum and Scholz 2010). In fact, more extreme claims have been made about the limited need for subordination and other recursive mechanisms in real-time conversational language, as opposed to written language. Pawley and Syder (2000), for example, put forward the One-Clause-At-A-Time Hypothesis according to which casual conversational language is more naturally described as a chain of clauses than as a set of full syntactic sentences. Also, the universality of recursion across world languages has recently been called into question (Everett 2005). Claims have also been made that strict recursion is not even a uniquely human mechanism and that "acoustic patterns defined by a recursive, self-embedding, context-free grammar" can be recognized by some species of birds (Gentner et al. 2006). At the very least, there is no compelling evidence for accepting recursion as an essential, indispensable mechanism of all human languages.

Despite these ongoing debates about the validity of the Principle of Compositionality and its syntactic enabling mechanisms, it is impossible to deny that, in one sense or another, new meanings or new formulations of familiar meanings can be expressed, at least partly, by combinations of simple and yet meaningful

---

complementation. More specific definitions of this type of recursion may also require that recursion occur over constituency rather than dependency structure, cf. Evans and Levinson (2009). The second meaning of recursion is sometimes loosely defined as the general compositionality of syntactic elements (Nevins et al. 2009), "where recursion is both the recipe for an utterance and the overarching process that creates and executes the recipes" (Coolidge, Overmann, and Wynn 2011). Strict syntactic recursion of the first type, at least theoretically illustrates the freedom of generating unique sentences by embedding and appending constituents of the same type within and next to each other, thus ensuring "no non-arbitrary upper bound to sentence length" (Hauser, Chomsky, and Fitch 2002). This theoretical property of grammar is known as the *No-Maximal-Length Claim*.

linguistic elements, be they morphemes, words or phrases. The vast combinatorial potential of language has achieved the status of a commonsensical view with practical implications. It is, for example, reflected in popular attempts to define plagiarism, which derive from the assumption that an unintentional repetition of even a relatively short passage of written academic text is virtually impossible. Similarly, it is extremely rare for two independent translations of the same passage of text from a relatively unrestrained genre to be identical.

While these examples show that speakers and writers naturally achieve formal novelty in everyday language use, the full implications of this observation are less clear. For example, one may wonder whether the *potentially* unlimited novelty of syntactic sentences inevitably leads to the conclusion that prefabrication, reproduction and memory are of "minute importance in ordinary language use". The evidence presented in this volume and in many other phraseological studies suggests that such a conclusion would be largely unwarranted. There is a growing body of research showing that it is not only complex, multimorphemic words, but also phrases, clauses and chains of syntactic dependents that seem to be recalled from memory rather than spontaneously composed. Even Chomsky's claim about the "zero probability of normal sentences"[4] can be easily challenged by the existence of a considerable number of discourse-specific utterances and sentence-like formulas in conversational language which are recurrent, highly institutionalized, largely petrified and thus most likely recalled from memory rather than recomposed every time they are used.

In accepting the combinatorial potential of natural language grammars, one should not fail to notice that not all sentences are created equal. As noted by Pawley (2009), formal grammars are often overly "egalitarian" in that they grant a similar status to a "nonce sentence" on the one hand, and "a much-cited proverb or a standard form of words for performing an apology, a compliment or a marriage ceremony", on the other. Needless to say, the functional, psycholinguistic and pragmatic status of these two types of sentences may be entirely different. Therefore, the question about the role of memory and the incidence of prefabrication in language use remains open and well-worth investigating. It is not invalidated by the existence of formal properties of languages such as compositionality and recursion, simply because at different structural and functional levels of language, speakers seem to trade novelty and uniqueness for prefabrication, in order to achieve native-like fluency and intelligibility.

4 *The vastness of the set of sentences from which normal discourse draws will yield precisely the same conclusions; the probability of 'normal sentences' will not be significantly different from zero.* (Chomsky 1978: 36)

Memory and prefabrication

As noted by Bolinger (1979), Pawley and Syder (1983) and others, our freedom to compositionally generate a wide range of equivocal expressions seems to be heavily restricted. To put it differently, it tends to be carefully utilized in most registers of linguistic communication. For what we traditionally refer to as words, non-compositionality is a long-recognized phenomenon. Word structure was among the earliest explicitly described linguistic observations, with evidence of compositional morphological analysis found on clay tablets from Ancient Mesopotamia (Haspelmath and Sims 2010). The extent to which the meaning of complex words is motivated by their constituent morphemes has traditionally been a central issue of derivational morphology. Historically, the compositionality of words in the sense of "a one-to-one relationship between form and meaning" is attributed to Humboldt (Zwanenburg 1995), while the distinction between the way simple and complex words receive their meanings can be found in Saussure (1915), (Hoeksema 2000), (aap van Marle 1990). Although the suitability of morphological models is sometimes directly judged by their "maximal compositionality" (Myers 2007), the widespread non-compositionality of complex words is widely acknowledged in derivational morphology. Multimorphemic words, whose meaning is only partly motivated by the meanings of their constituent morphemes are so common that it is impossible to dismiss them as isolated, idiosyncratic exceptions (cf. Haspelmath and Sims 2010: 62). As summarized by Spencer "sometimes, we must recognize meaningless morphemes which nonetheless combine to form meaningful words" (Spencer 1994: 73).

Recurrence of composite units of meaning is also a central issue in diachronic morphological theories. For example, according to Bauer (1983: 45–49), there are three major stages in the formation of complex word forms, namely:

a) Nonce-formation, i.e. the spurious composition of complex words. Such words are "new" at least in the sense of being used independently by independent speakers. The word form *dollarless* (Langacker 2008) is an example of a compositional nonce-formation.
b) Institutionalization, which occurs "when a nonce formation starts to be accepted by other speakers as a known lexical item" (Bauer 1983: 48). One of the implications of institutionalization is the reduction of a word's possible ambiguity. For example, the adjective *penniless* is an institutionalized synonym of the above-mentioned nonce-formation.

c) Lexicalization, which occurs when a lexeme acquires a form which could not have resulted from the application of productive morphological rules. The hyphenated form *dead-broke* could serve as an example of this type of lexicalization.[5]

Anttila (1989: 349) observes that even the most obviously historically related word forms seem to be "stored separately", which leads to the conclusion that "memory or brain storage is on a much more extravagant scale than we would like to think". Although similar processes can be observed in the formation of language units larger than multimorphemic words, linguistic theories have not always done proper justice to the constant interplay of compositional novelty and prefabrication of phrases, clauses, sentences and other lexically and syntactically complex structures. In theories whose general appeal and descriptive adequacy is judged by the focus on analycity, parsimony, generative power and symbolic minimalism, a great deal of emphasis is placed on the compositional aspects of language use. This has led to the overshadowing of our tendency to reuse the same realizations of highly schematic linguistic patterns as an optimization facilitating linguistic communication. The centrality of syntax in linguistic theories of the latter part of the 20th century has contributed to the view that idioms, set phrases and other recurrent word combinations characterized by partial or complete semantic opaqueness, syntactic ill-formedness or simply by significant levels of reuse are peripheral to the core interests of theoretical linguistics and that subtle manifestations of prefabrication such as open collocations are essentially indistinguishable from phrasal nonce-formations (Pawley 2009).

Although Burger (2007: 90) notes that "from a semantic point of view, it does not make much sense to separate phraseology from word formation", it remains generally true that idiomaticity and semantic opaqueness are much more easily recognized in morphological theories than in accounts of phrase and sentence formation. Hoeksema (2000) speculates about two possible reasons for this tendency:

> Idiomaticity is a very common thing and as many linguists have pointed out, it is more common in complex words than in phrases, perhaps because words (being generally shorter) can be listed more easily in the lexicon. Others (Anttila 1985) have suggested that words are inherently different from phrases in that the connections between their parts are tighter and that this more easily leads to loss of compositionality. (Hoeksema 2000: 856)

5 Bauer provides more prototypical examples of fully blended compound words, such as *butterfly* or *blackmail*.

To summarize, it is commonly accepted that many multimorphemic words seem to be holistically retrieved from memory. However, it is much less generally agreed or even recognized at all to what extent similar restrictions on compositionality operate at the level of phrases, clauses and other multiword combinations. For a number of reasons, the issue of choosing between new and conventionalized units of language becomes more subtle when it comes to explaining how we combine words into structurally larger syntactic and semantic units. Perhaps the most important of these reasons is the above-mentioned emphasis on compositional, formally elegant accounts of phrase and sentence structure in which phraseological idiosyncrasies are viewed as peripheral to "the essence of the language instinct" (Pinker 2007).

Recognition of prefabrication

In contrast to the abovementioned views on language production, which predominated in the mainstream linguistic theory of the second half of the 20th century, the issues of language reuse and prefabrication are no longer widely perceived as peripheral to the core of linguistic theory. Recent years have seen a revival of interest in the role of formulaicity and phraseological patterning in some of the key areas of research on language acquisition and processing. Although these perspectives may still be unlinked and seemingly unrelated, evidence from a variety of linguistic and cognitive studies suggests that reuse is not a marginal feature of language whose by-products in the form of linguistic prefabricates such as idioms or restricted collocations can be relegated to a static lexicon which serves merely as "a repository of idiosyncrasies" (Atkins, Levin, and Zampolli 1994: 18) or a "ragbag of irregularities" (Greenbaum 1974: 79). It seems to be more widely recognized that between the level of morphologically complex words and full syntactic sentences, achieving native-like fluency requires sticking to established phraseological units, highly recurrent formulaic expressions, stock phrases and recognized ways of saying things which often undergo semantic bleaching and syntactic petrification and which could only potentially have a vast number of alternative, grammatically viable wordings.

Outside of traditional phraseology, there has been an increasing interest in models of language use which directly recognize the key role of memory, prefabrication and recall in the processing and production of composite lexical, syntactic and semantic units. To name just a few examples:

- In contradiction to the long-predominant view that compositionality is key in optimizing online generation and processing of language, a number of psycholinguistic studies have indicated that a large repository of formulaic sequences stored in long-term memory plays a crucial role in maintaining normal rates of fluency and comprehension (Wray 2002).
- In the influential paradigm of Cognitive Grammar (Langacker 2008), "automization" is regarded as a common cognitive mechanism which may lead to the "entrenchment" of any linguistic structure, thus making it a holistically retrieved and processed entity despite its original compositionality. More generally, the so-called 'usage-based' theories of language "recognize that human beings learn and use many relatively fixed, item-based linguistic expressions (...) which, even when they are potentially decomposable into elements, are stored and produced as single units" (Tomasello 2000: 61). Also, within the so-called "constructionist" approach to grammar there has been a growing recognition for the role of long-term memory in the acquisition and use of constructions, which are defined as any linguistic patterns whose "form or function is not strictly predictable from their component parts or from other constructions recognized to exist". Moreover, such patterns are considered to be "stored as constructions even if they are fully predictable as long as they occur with sufficient frequency." (Goldberg 2006: 5).
- Corpus linguistic studies, especially in the so-called neo-Firthian tradition with their "emphasis on syntagmatic aspects of lexis (...), and stretches of language that constitute indivisible meanings and which display degrees of semantic transparency or opacity and degrees of syntactic productivity" (Malmkjær 2009: 351) have revealed high levels of phraseological patterning and reproduction of structural units such as phrases and entire clauses (Altenberg 1998), (Pęzik 2013), (Pęzik 2014) leading to the recognition of the "underlying rigidity of phraseology, despite a rich superficial variation" (Sinclair 1991: 110).
- Cognitive psychology theories such as the Instance Theory (Logan 1988) emphasize the role of automaticity in achieving high performance at various cognitive tasks. With regard to language as a cognitive faculty, morphological and lexical nonce-formations are claimed to be processed "strategically", while previously encountered lexical compositions are believed to be stored and retrieved as "past solutions" or instances, thus enhancing fluency and comprehension rates (Oppenheim 2000: 221). The general importance of automatic (as opposed

to controlled) processing in the process of comprehending language has been confirmed in numerous other studies as well, cf. (Favreau and Segalowitz 1983) (Hahne and Friederici 1999).

– In the field of natural language processing, there has been a growing appreciation of what computational linguists generally call "recurrent multiword expressions" as one of the missing links and "a pain in the neck" for formal modeling and processing of human languages (Sag et al. 2002). Also, probabilistic models of language, which essentially capture some of the prefabrication and selectional predictability of phrases and other undifferentiated word sequences have been successfully applied in the area of machine translation (Koehn 2010) and automatic speech recognition (Jurafsky 2009) resulting in huge improvements in robustness over rule-based models. To a large extent, these developments derive from Information Theory (Shannon 1948) with its n-gram approximations of language structure and the subsequent data-driven models of linguistic communication inspired by this early work.

These various theories and strands of research have not as yet consolidated into a self-contained discipline with prefabrication as "a precise object" of study. However, taken together, they can be seen as "a repertoire of interests that is not as yet completely unified" (Eco 1976: 7).[6] Some of these interests largely stem from the intuition that the overstatement of the role of compositionality and novelty in language has left unexplained two important issues, which Pawley and Syder (1983) call "the two puzzles for linguistic theory". The first of these puzzles is reflected in the observation that despite the apparently infinite productivity of language, in many communicative contexts, the native-like selection of a sentence, a clause or a phrase from the full range of its grammatically valid paraphrases is often relatively restricted and predictable. In this sense, a large number of sentences and many syntactic types of phrases used by native speakers of a language seem to be more often recalled from memory than composed or recomposed on the spur of the moment. The second puzzle derives from the fact that speakers (of many languages) can produce several words per second in a normally-paced conversation (de Bot 1992). Without recognizing the crucial role of long-term memory in language production, it could be impossible to explain such levels of temporal fluency.

6 To be clear, this is how Eco describes the status of translation studies. The status of prefabrication studies is similar in that they have not developed into a separate discipline outside of phraseology.

To summarize, one of the noticeable changes which seems to have occurred in different areas of linguistic theory over the recent decades consists in the increasing recognition of the extent to which our potential to produce "novel language" is restricted by our preference for "prefabricated language" at different structural levels of language and across the full variety of communicative contexts, registers and modes of expression. There is a general intuition which seems to justify the study of linguistic prefabrication from a variety of perspectives: native-like use of language is not only grammatical. It is also idiomatic. Language is not only and perhaps even (as will be implicitly argued in this volume) not primarily generated, composed and "put together". It is also largely "remembered", both holistically and associatively. The incidence of phraseological prefabrication, which goes by many different names, such as idiomaticity, non-compositionality, prefabrication (Siyanova-Chanturia and Martinez 2014), cognitive entrenchment and automaticity (Langacker 2008) is the general subject of this study.

Facets of prefabrication

The present volume attempts to bring together some new perspectives on the role of frequency, distributional binding and memory as possible factors determining the incidence of prefabrication in language production and reception. It investigates the distribution and properties of different phraseological units, which range from self-evident prefabrications such as idioms to more subtly prefabricated open collocations and multiword chains of binary collocations. What makes many open and restricted collocations as well as larger collocational chains particularly relevant to the debate about the levels of compositionality and novelty in language is the fact that they often constitute borderline cases of linguistic prefabrication. If we accept that such items are indeed largely prefabricated and more adequately described as units recalled from memory rather than independently recomposed in discourse, then we should consequently revise our estimations of the overall incidence of prefabricated language in actual use. For example, one of the hypotheses explored in this study is that the upper bound of prefabrication in language seems to be much higher than some traditional models of phraseology would define it, especially if we recognize certain subtle types of contextually stereotyped collocational chains as phraseological prefabrications. Their conventionality is often contextual rather than formal

and triangulations of methods and approaches are required to identify them as at least partly prefabricated.

The investigations presented in this volume build upon current developments in phraseological research and formulaicity studies. In addressing the role of prefabrication and recall in linguistic communication from a corpus-based perspective, I refer to linguistic theories which reconcile linguistic novelty, compositionality, propositionality or analycity on the one hand, with formulaicity, automaticity, reuse and recomposition, on the other. Throughout this study I make use of collections of annotated reference corpora and complementary sets of experimental data. I also apply a variety of corpus analysis methods and natural language processing techniques including syntactic parsing and automatic extraction of phraseological units to investigate the distributional aspects of phraseology. The combination of tools, resources and methods is meant to provide a fresh perspective on the role and scope of prefabrication and recall as one of the basic aspects of linguistic communication.

Chapter 2 of this volume focuses on some of the basic characteristics of collocability and phraseological prefabrication. More specifically I review a number of distributional, semantic and psycholinguistic features of collocations as phraseological units. The chapter introduces an important methodological distinction between three aspects of linguistic prefabrication: recurrence, recall and recomposition. It also proposes the notion of "stereotyped recurrence" as a common characteristic of open and open-ended collocations without which it would be difficult to regard them as instances of language reuse.

Chapter 3 discusses methods of extracting collocations from corpora and proposes a syntax-based approach to phraseology extraction based on identifying recurrent, multi-element chains of syntactic dependencies, or "catenae" (Osborne et al. 2012). The method is subsequently used to generate automatic combinatorial dictionaries from reference corpora of English. In contrast to many positional and some relational methods of collocation extraction which are restricted to binary word combinations, the approach makes it possible to extend the analysis of phraseological prefabrication to units which consist of multiple lexical and grammatical collocations, i.e. collocational chains and other types of collocational catenae. It also provides a way of recursively investigating the so-called external and internal valency of idiomatic expressions through data structures called "subsumption graphs". Apart from discussing the relevance of the extracted database of potential phraseological units to estimating the incidence of phraseology in naturally occurring discourse, the chapter also discusses its applications in foreign language lexicography and phraseodidactics.

Chapter 4 focuses specifically on collocational chains as a subtle, largely transparent and yet prevalent aspect of linguistic prefabrication. Selected entries drawn from the combinatorial dictionary generated in Chapter 3 are used to design a questionnaire-based study of collocational chains. The recall of these structures is validated against distributional evidence from large reference corpora and corpus-based combinatorial dictionaries. Some methodological conclusions are also drawn from this experiment, including the limited application of elicited, in vitro linguistic data in the study of prefabrication.

Chapter 5 reports the results of a corpus-based study of "phraseology markers", which are defined as a set of conventional expressions used to explicitly indicate the occurrence of phraseological units in naturally occurring discourse. They are argued to bring insights into the popular perception of prefabrication and conventionality by non-expert speakers and writers of English. The evidence from the use of phraseology markers is considered to be methodologically different both from aggregations of corpus data explored in Chapter 3 and elicited experimental data analyzed in Chapter 4.

Chapter 6 concludes the volume with some remarks on the multifaceted nature of prefabrication and further directions of research which were merely indicated or not considered at all in the present volume. Among those are applications of phraseology extraction and detection in phraseostylistics and the role of prefabrication in translational equivalence as well as in studies of second language acquisition.

It is hoped that the distributional, experimental and formal perspectives on or "facets" of phraseological prefabrication investigated in this volume will bring its readers closer to the conclusion that language reuse is of huge, rather than "minute importance" in ordinary use of language and that as speakers and writers we seem to do much more "remembering" and less "putting together" than some formal theories of syntax or semantics would have us believe in abstraction from naturally-occurring language data.