Formulaic Language in Thought and Word: Vygotskian Perspectives

David Wood
School of Linguistics and Applied Language Studies, Carleton University
Faculty of Education, University of Ottawa
Department of English Education, Naruto University of Education, Japan

Abstract: It has become apparent in recent years that formulaic language, or preferred sequences or chunks of words such as idioms, collocations, and lexical phrases, is of great importance in language acquisition and production. Several theories of cognition can help to explain how these sequences are stored, processed, and retrieved in language development and use. It is possible to draw strong connections between the work on formulaic language and the social learning theories of Vygotsky, particularly as regards the development of thought through egocentric and inner speech. This paper presents a review of the current state of knowledge about formulaic language in acquisition and production, as well as psycholinguistic models of language production. Parallels are drawn to the work of Vygotsky in language and thought development, and, in the end, an integration of the language acquisition, psycholinguistic, and social learning perspectives is presented.

Keywords: formulaic language, Vygotsky, cognition, inner speech.

1. Introduction

While a perception of language as rule-governed, systematic behavior has been dominant in linguistic theory for several decades, there has been an increased interest recently in the nature and role of formulaic language units. A growing body of work suggests that ready-made chunks or preferred sequences of words play a significant part in language acquisition and production. These formulaic language units include fixed phrases and idiomatic chunks such as on the other hand, or all in all, or hold your horses, and longer phrases, clauses, and sentence-building frameworks of words such as the bigger the better, or if X, then Y. Numerous researchers have attempted to define and categorize this generally
David Woods

overlooked aspect of language, and the evidence that formulaic language is basic to language development, processing, production, and learning is growing.

The role of formulaic sequences in language production and acquisition, particularly with regard to spontaneous spoken language, is a topic of investigation which has rich potential for our understanding of how language is produced in real time. It is becoming more and more apparent, based on some thirty years of research, that formulaic sequences are of critical importance in the production of fluent speech and that they play a key role in acquisition. Several theories of cognition can be integrated with knowledge about formulaic sequences to produce models of language production, including controlled and automatic processing (McLaughlin, Rossmann, & McLeod, 1983), declarative and procedural knowledge (Anderson, 1983; Levelt, 1989), and instance theory (Logan, 1988).

The social learning theories of Vygotsky (1978; 1934/1986) have rarely been linked to psycholinguistic models of language production and acquisition, but there are definite connections to be made. Vygotsky’s notions of the development of concepts and of egocentric and inner speech (1934/1986) in particular have strong potential to further illuminate descriptions of language processes. This paper will review research and theory related to formulaic sequences, and draw connections to the work of Vygotsky. In the end we can see that it is possible to integrate ideas of formulaic language, proceduralized knowledge, thought development, and inner speech development.

2. Formulaic Sequences Defined

While various researchers have studied the phenomenon of formulaic language over the years, there is a certain amount of agreement on basic definitions of what constitutes a formulaic sequence and what characteristics such sequences share which makes them distinct. The consensus seems to be that
formulaic language sequences are multi-word units of language which are stored in long-term memory as if they were single lexical units. Wray and Perkins (2000) define formulaic sequences in typical fashion, as multi-word units of language:

A sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (2000:1)

This type of definition is common in the literature. Pawley and Syder (1983) refer to formulas as "sentence stems" which are lexicalized, that is, which are "regular form-meaning pairings." (p. 192) This notion of lexicalization is echoed by Nattinger and DeCarrico (1992) in a highly influential work which focuses on lexical phrases, an alternate term for formulaic language units:

...lexical phrases [are] form/function composites, lexico-grammatical units that occupy a position somewhere between the traditional poles of lexicon and syntax; they are similar to lexicon in being treated as units, yet most of them consist of more than one word, and many of them can, at the same time, be derived from the regular rules of syntax, just like other sentences. Their use is governed by principles of pragmatic competence, which also select and assign particular functions to lexical phrase units (p.36)

The notion that storage of formulaic units is as single lexical units is key to a definition.

4. Formulaic Language and Performance

The importance of formulaic sequences in language performance, particularly speech performance, has been documented quite extensively.
Studies which have investigated the nature of fluency in speech have uncovered a great deal about the strong facilitative role of formulaic sequences in the production of fluent, running speech under the time and attention constraints of real life communication.

Early research in the area of fluency, in the 1980's, produced mention of notions of elements of speech fluency which may be interpreted to imply that formulaic language was of importance. Raupach (1984), in a study of an adult learner of French, found evidence of formulaic constructions contributing to fluency, particularly modifiers and rhetorical organizers.

Formulae of these types likely are generated at many points in the planning and execution of speech. Dechert (1984) observed that the most fluent German students of English, required to retell a narrative in their second language in his study, appeared to establish "islands of reliability" of ideas and language, around which they pieced together a spoken narrative. Sajavaara (1987), in a reflection on a wide range of factors affecting second language speech, observed that a concept or a single lexical item could trigger the release of other lexical items and phrases:

A "word" activates, for example, certain frequent and prefabricated phrases, word combinations, grammatical constraints, selectional restrictions, semantic concepts and fields. (Sajavaara, 1987: 54)

A store of many aspects of conceptual items and links, lexical items, phrases, and patterns of language and ideas can be activated by stimuli in the input or the context. Strings of language can then be generated appropriate to the ideas linked to the stimuli, while more specific items and constructions can be placed with or within the formulae. In this way, fluent speech is generated.

5. The Psycholinguistics of Production
Key to this entire process are the workings of mental processes and skills. The psycholinguistic concepts of automatic and controlled processing provide a conceptual framework which can explain many of the empirical phenomena described above. Anderson’s ACT* model of cognitive development (1983) acknowledges the limited capacity of working or short-term memory to assist in processing knowledge at the speed of real-time speech and posits a mental process he terms “proceduralization” to account for how such extremely rapid processing could occur. McLaughlin, Rossman, and McLeod (1983) distinguish between the two mental processes as they apply to second language processing. Citing Shiffrin and Schneider, they describe memory as “a large collection of nodes that become ‘complexly interassociated’ through learning.” (McLaughlin et al., 1983: 139) These nodes can become activated by either automatic or controlled processing. Automatic processing causes certain nodes to activate every time certain appropriate types of input occur. Consistent activation in a certain pattern by the same type of input over time leads to a learned, automatic process. This process is extremely rapid, and requires little or no effort or attention. On the other hand, controlled processing is described as a response which is not learned, but temporary. Attention by the subject or speaker is required. Only one such sequence of activation of “nodes” can be controlled at any time. These activations, though, can be applied to novel situations. Schmidt (1992: 360) categorizes the two processing styles as differing as to a number of important characteristics:

<table>
<thead>
<tr>
<th>Automatic Processing</th>
<th>Controlled Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>fast and efficient</td>
<td>slow and inefficient</td>
</tr>
<tr>
<td>effortless</td>
<td>effortful</td>
</tr>
</tbody>
</table>
not limited by short-term memory capacity limited by the capacity of short-term memory
not under voluntary control under subject control flexible inflexible inaccessible to introspection at least partly accessible to introspection

Once a response sequence has been automatized, then attention and effort can be used to perform other tasks at the same time.

This distinction between automatic and controlled processing in speech is linked, of course, to the notions of procedural knowledge and declarative knowledge. Procedural knowledge, or knowledge of how to do something, is likely to be automatized much more readily than is declarative knowledge, or knowledge about something.

Logan (1988) posits a process by which automatization or proceduralization occurs, based on memory. Logan’s assumptions about how memories are stored and retrieved have to do with a process of substitution. Unlike Anderson, who held that underlying processes do not change, but simply speed up and form proceduralized chunks, Logan theorizes that the process is replaced over time by a simpler, faster process, which is single-step memory retrieval. Output is determined, then, by a sort of race between the original step-by-step declarative sequence, and memory retrieval. Eventually, memory retrieval dominates, being a more efficient means of producing strings of language under the time and attention pressures of spontaneous speech. Memory retrieval and encoding are natural consequences of attention. Therefore, it is not the internalization of rules which accounts for fluent speech production, but rather, memory retrieval.
6. Formulaic Sequences in Language Acquisition

There is a certain amount of evidence of formulaic sequences being used as a learning strategy in children. It appears that first and second language acquisition in children is largely a function of attending to formulaic sequences in language input, adopting them for use, and later segmenting and analyzing them. Wong-Fillmore (1976) was one of the first to study the language acquisition of a child and find that a process of formulaic chunk acquisition was followed by a process of segmentation or syntactic and semantic analysis and breakdown which fostered development of overall linguistic competence. Hakuta (1974) conducted a 60-week study of language acquisition of a Japanese child and found evidence of initial acquisition of prefabricated chunks later analyzed and used to facilitate overall language development. Hickey (1993), in a longitudinal examination of the acquisition of Irish Gaelic of a child also discovered a role for formulas in acquisition. Again, they were later broken down and analyzed, providing grist for the linguistic competence mill. According to Peters (1983), early on the child develops strategies for extracting meaningful chunks from the flow of conversation. He is able to remember them, compare them phonologically with others, and remember them as new lexical units. They are stored as wholes in the lexicon. Later in his cognitive development, he is able to analyze the stored chunks and then recognize and remember structural patterns and information about distribution classes revealed by the analysis. He is then ready to develop
an ability to utilize lexical and syntactic information already acquired to analyze new chunks in the linguistic environment.

Other researchers have been able to determine that processes related to pragmatic competence are at work when children acquire formulaic sequences. Bahns, Burmeister and Vogel (1986) investigated the second language acquisition of a group of children and found evidence of a formula segmentation process at work. The authors note that it was common to discover exceptionally sophisticated language in stretches of child learner speech in research:

In their attempts to write grammars for different stagers of development, mainly in structural areas like negation or interrogation, child language researchers were very often confronted with utterances of a rather complex nature. The structure of these utterances was somehow "outside" the rules written to account for the bulk of data representing syntactic development for the stage in question. (pp. 696, 697)

In their study, Bahns et al. found a large range of formulas used by the children, accounting for the complex utterances noted by earlier researchers.

7. Vygotsky and Formulaic Language

In *Thought and Language* (1934/1986), Vygotsky puts forth theories of how thought emerges in human development, and how a phenomenon termed "egocentric speech", which evolves into "inner speech" facilitates the growth of cognition. These ideas link to and enrich the ideas elaborated about declarative and procedural knowledge and instance theory.

Vygotsky states that word meaning is the essential form of the connection between thoughts and words. Meaning is essential to a word as an act of thought, and it is specifically a phenomenon of verbal thought or meaningful speech. His studies showed that word meanings are not fixed, but that they
develop, and that thoughts develop and come into existence through words (Vygotsky, 1934/1986).

In Vygotsky’s view (Vygotsky, 1934/1986), there are two planes of speech, the inner, semantic one, and the outer, or phonetic representation of it. A child masters outer speech from word to sentence, from part to whole, and inner speech from whole, meaningful complex to whole. These two processes run in opposite directions. At first, a child’s thought is a vague, shapeless whole, and needs expression in a word. As thought gets more differentiated over time, there is less need for words, but for wholes such as phrases and sentences and texts. As speech progresses to wholes in this way, thought also progresses in the opposite direction, becoming more particular and less holistic. Therefore, the structure of speech is not the same as that of thought. Later in development, grammar precedes logic at times, as ideas such as logical connections among ideas like because and although are used by the child before he really understands the concepts they stand for.

Here, it is clear that there is a link to the research on formulaic sequences in child language acquisition. The sequences are similar to words in that they are lexicalized and stored and retrieved as meaningful wholes. A child would acquire the sequences as chunks, regardless of the individual words they contain, and only later break them down through a process termed segmentation. Child language acquisition research shows that there is a tendency to acquire formulaic sequences at first and use them wholesale, only later appearing able to use pieces of the sequences in other syntactic and semantic circumstances.

Presumably this parallels the development of thought processes in the child, as the sequences can be seen as analogous to words as Vygotsky uses the term. As thought gets more differentiated, then, there is less need for the
formulaic sequences themselves, but for their composite parts, which are more subtle and complex parts of language. As more complex thoughts need expression, a reliance on formulaic sequences is inadequate, and more flexibility of language is required. This might explain the need for segmentation of formulaic sequences which occurs at a certain stage of child language development.

8. Formulaic Language and the Evolution of Human Language

The idea that human language has evolved through a similar process from a prehistorical protolanguage also has parallels in the Vygotskian view of speech and thought development. Wray (1998) elaborates a theory of the evolutionary origins of formulaic sequences in human language and their connection to aspects of acquisition and production. Wray points out that many high level primate species produce holistic utterances which serve to help with social behaviors such as grooming, as well as inciting desired behavior in others. She posits that human ancestors, at an early stage of evolution, developed more complex systems of such vocalizations. Then, as the human brain evolved and cognitive skills developed, analytic, generative language became available, and segmentation of holistic utterances began, but was stymied by the limitations of short-term memory. Thus, formulaic sequences were born of a protolanguage from the earliest stages of human development, and were retained as a feature of language production as analytic language abilities proved to be of limited use in real-time production due to short-term memory boundaries. In a sense, then, holistic utterances are still used to maximize processing space in the brain during productions.

Originally, protolanguage had no grammar, but large numbers of specific utterances used for specific pragmatic purposes, according to Wray (1998). Chimpanzees, for example, use specific vocalizations to mean things, with no
internal morphological structure, producing strings and chunks of sounds functioning as free-standing wholes to achieve a limited number of specific communication goals. They rely on a limited stock of utterances, each of which has a specific communicative purpose related to a limited number of aspects of personal and immediate experience. These utterances use no lexical or semantic system and have no syntactic or morphological structure, and so they have to be learned one by one by members of the group of chimpanzees. Presumably the first humans communicated in the same way, with language perhaps evolving during grooming to talk about other members of the group, to move communication and awareness beyond mere personal experience and to get a second-hand perspective on things (Wray, 1998).

As posited by Wray (1998), the transition from noise and discrete sounds to structured phonetic sequences was likely gradual, developing along with the physical evolution of the speech tract. As well, in the absence of a grammar or formal structure of any sort, each individual would have to memorize all of language piece by piece. However, along with cognitive development over thousands of years, it is probable that rudimentary grammars emerged from the strings of sounds associated with specific meanings. For example, a combination such as stone + hot + food could mean any of a number of things, such as put hot food on the stone, hot food is available at the stone, or sit at the front of the cave tonight, so there would be a need to agree a specific sequence has one particular meaning from the possible choices. This is the beginning of grammar. Wray (1998) theorizes that the emergence of grammar would likely have happened through a gradual linking of various areas of the brain as awareness of various aspects of experience developed, such as thematic analysis, conceptualization of actor, recipient, action. Through the evolution of human consciousness and
neurological development, these awarenesses would then carry over into language so as to express and describe the more complex aspects of reality. The originators of this primitive grammar probably had a brain like ours today. However, according to Wray (1998), the ancient primate “discrete utterance” mode of communication carries over into present-day language in formulaic language, which often tends to defy or thwart the rules of syntax and morphology. Formulas often function as holistic items with their own peculiar structure, as in *by and large* or *as it were*.

Wray’s hypothetical development process of human languages relates to Vygotsky’s ideas of the two planes of speech and the idea that speech and thought develop in opposite directions. In the protolanguage development outlined by Wray, it appears that speech developed along a continuum from the general to the specific, in that whole sound sequences with specific semantic or pragmatic purposes came to be refined and broken into pieces over time to express ever more cognitively challenging and subtle ideas. Both the language and the concepts it expressed developed in the direction of increased subtlety and range. Similarly, Vygotsky observes that, as a child’s thought develops, it gets more differentiated from the “dim, amorphous whole” (Vygotsky, 1934/1986: 219). As this occurs, the need for individual words in child language expands into a need for phrases and sentences expressing more complex and differentiated ideas. As formulaic sequences in language and in human protolanguage are the semantic equivalents of single words, it appears that their segmentation would serve to enhance expression of more subtle ideas and concepts through grammar. Grammar enables a broader range of potential sentences and ideas to be expressed.

9. Formulaic Language from a Vygotskian Perspective
Another key point in Vygotsky's ideas of the development of speech and thought is the development of egocentric and inner speech (Vygotsky, 1934/1986). Both of these forms of speech in children are speech for oneself, while external speech is speech for others. Inner speech, the foundation of thought, is preceded by egocentric speech, a transition from the social connections through language of the child to individual activity and thought development. Egocentric speech, or talking aloud to oneself, serves mental orientation to tasks and conscious understanding of the environment. It especially appears to help the child overcome difficulty, and it increases in frequency when tasks require reflection and focus of consciousness. Egocentric speech is evident in young children who imitate speech sequences and structures observed in adult conversation, and use them to talk to themselves during individual play. Over time, this egocentric speech transforms into inner speech, a more internal, less vocalized, and less readily understood form of self-talk. In the end, inner speech separates completely from speech for others or social speech, and vocalization stops altogether. This is the foundation of thought. Vygotsky draws a parallel to counting, in which a child initially learns to do so with the aid of counting on his fingers. Over time, he is able to count in his head without the use of external aids like fingers, and, in fact, may be able over time to quickly determine numbers of things through much more complex processes than counting. For example, he could count the chunks of five, or by twos, and so on (Vygotsky, 1934/1986: 230).

As a child develops inner speech, various changes occur in his speech for himself, away from the social speech norms he hears around him. A particular syntax develops in inner speech, in which things become abbreviated and subjects are omitted from sentences and only predicates remain. This is similar to face-to-face social speech in which subjects are often omitted as both
interlocutors have shared apperception of who or what the subject is. This is built up in context as communication is negotiated in conversation. With inner speech, however, the “conversation” is purely internal or personal, with the child being both sender and receiver of messages. Inner speech is marked by predication and implied subjects, a phenomenon which develops as the child talks about what he sees, hears, and does, using egocentric speech. He comes to omit objects and focus particularly on these actions, vocalizing less and reducing syntax and sound. Meaning moves to the forefront and the phonological element of speech disappears (Vygotsky, 1934/1986).

Vygotsky (1934/1986) also posits that the semantic aspect of inner speech evolves along with the syntactic changes. Sense is emphasized over meaning as the process goes on. Sense is that aspect of word meaning which is inextricably linked to context, and is, therefore, at the core of what is termed the pragmatic aspect of language usage. Sense in word meaning changes from one context to another. Meaning of words, on the other hand, is a greater generalization, and it remains consistent over all contexts ... word meaning is definable, whereas word sense is much more difficult to translate from thought into language. With the development of word senses over time, agglutination or combination of words into complex whole and webs of sense and meaning occurs in the mind. Vygotsky notes that a single word in the minds becomes saturated with senses, to the degree that transferring it to external speech would require considerable verbal gymnastics. In the final analysis, it appears that inner speech, silent to the world and increasingly separated from the social speech it develops from, is incomprehensible to others (Vygotsky, 1934/1986).

This notion of a progression of abstraction and sense has certain parallels in the constructs of proceduralization and automatization previously discussed. As formulaic sequences, for example, are acquired by a learner over time, the
concepts they represent become embedded in memory. Certainly, when a lexical item like a formulaic sequence is first encountered or dealt with by a learner, it is in the range of declarative knowledge. With repeated exposure to the target language and repeated pragmatic context stimuli which require the use of the formulaic sequence for native-like expression, the formulas becomes a chunk in long-term memory, proceduralized. It is accessed most likely by the instance theory principle of direct memory access, bypassing the laborious and time-wasting process of building it up from its constituent parts.

As a second language learner becomes more skilled and adept at using formulaic sequences, they likely become part of the pragmatic system of senses. This pragmatic system includes the context-sensitive shades of lexical meaning which Vygotsky (1934/1986) posits as a part of inner speech and, ultimately, thought. In this way, like a child developing first language ability, the learner in effect begins to be able to think in the second language, to exhibit the instant and flexible selection of context-appropriate phrases and words to maintain fluent speech. In his model of the development of inner speech Vygotsky (1934/1986) discusses how words become saturated with senses. Similarly, the standard phrases and strings and sentence or utterance frames of the second language likely become saturated with senses too. They become a part of inner speech, which is the foundation of thought. When this occurs alongside the first language inner speech and thought structure of an adult second language learner, two semantic networks or webs of meaning and sense coexist. The interconnectedness of thinking or inner speaking in a first language and in a second language likely facilitates or accelerates the ability of adult language learners to grasp apparently quite complex and abstract ideas through the
medium of the second language. Thus, first and second language speech development occur along similar lines and to similar ends.

Vygotsky goes on to a deeper level of verbal thought beyond inner speech, to thought itself. According to Vygotsky (1934/1986), thought does not need words because it exists in its own structure which does not consist of single units. While thought may not consist of single units, it does require realization in words. To understand one's words, we need to understand his thought and, ultimately, his motivation. Motives engender thoughts, according to Vygotsky, and they are realized first in inner speech, then in word meanings, then finally in uttered words themselves. The movement is outward, from the level of motivation and abstract thought through truncated and sense-saturated inner speech, and finally into the physically articulated external social speech.

This has distinct parallels to the ideas of the cognitive theorists discussed earlier. Automatization or proceduralization of chunks of language can be seen as a part of their transition into inner speech. As motives or contexts or other stimuli prompt the expression of an idea or a thought, the first connection in thought and inner speech is likely to be to memory. Logan (1988), sees memory as the access route and storage system of proceduralized language items and formulas. As memory is stimulated through thought and motive, its links with inner speech enable a formulaic sequence to be uttered rapidly enough to bypass the cumbersome procedure of formulating everything from scratch syntactically. The formulaic sequences permit a broader flexibility of expression and allow a shared social framework to be instantly created, and so they are the vehicles for expression of the semantic senses Vygotsky refers to as a hallmark of inner speech.

10. Towards an Integration of Perspectives
In the end, it is clear that research on formulaic sequences in language and Vygotsky's social learning theories of the development of language share quite a lot. Like the blind men and the elephant, the formulaic language researchers and the Russian developmental psychologists have described part of the phenomenon of language development in different parts. Vygotsky was concerned largely with child language and cognitive development. The formulaic language researchers were focusing on particular language phenomena and their role in acquisition and production. However, the overlapping themes are remarkable.

From Vygotsky we can have a fuller and richer perspective on language in its formulaic aspect. It is clear that formulaic sequences provide a great deal of advantage in the development of inner speech and thought, in that they allow a broader range and subtlety of expression of lexical sense. As well, their segmentation may push forward the differentiation of thought, as the component grammar of formulaic sequences is dissected and used in recombination by the developing child to express more and more complex ideas.

Vygotsky's notions of the evolution of thought through the condensation and agglutination of egocentric speech provide a powerful framework by which we can examine how thinking in a second language may occur. Given that external and egocentric speech become inner speech and saturated with word sense, it follows that thought is affected as well. When one acquires a second language, it may be that second language thought develops through a similar process, and an inextricable intertwining of word senses and proceduralized chunks of language contribute to native-like fluency and control. It may be that procedural knowledge in the first language facilitates the acquisition of procedural knowledge in the second language.
Finally, formulaic sequences provide a broader range of word senses and ways of expressing them. In Vygotsky's model, words become saturated with sense in inner speech, and expression of these senses would require great numbers of words for each one. The use of formulaic sequences allows a broader range of senses to be expressed. As well, it allows for more sense in the first place; a formulaic sequence is a lexicalized item and functions largely as a single word.

References


46
Formulaic Language in Thought and Word: Vygotskian Perspectives


David Wood
School of Linguistics and Applied Language Studies, Carleton University
Faculty of Education, University of Ottawa
Department of English Education, Naruto University of Education, Japan
Email: d.w78@yahoo.com