Multi-word Lexical Items
and the Advanced Foreign Language User:
Awareness Raising in the Context of Oral
and Written Translation Training

Abstract

The paper has two main objectives: the first is to present an overview of the phenomenon of multi-word items in English, and discuss their prevalence in native speaker usage. The second is to discuss how proficient non-native speakers of English, and in particular those who are training to become translators, may benefit from classroom training that increases awareness of the primary role of chunks and other multiword units in native-like speech. It is argued that classroom training may tend to emphasise grammar rules and lexis over building a repertoire of multi-word items; more practise in this area may improve fluency, conserve energy, and enhance long-term language learning among adult foreign language users with nuanced foreign language performance goals.

Keywords: idiomaticity, formulaicity, foreign language learning, translator training, oral translator training, multi-word units, interlanguage development

1. Grammar and words: what native speakers can do
   (but might not be bothered to)

The language of native speakers is highly formulaic; in their everyday output, native speakers seem to draw upon a stock of stored bits of language in order to perform particular linguistic functions. Among them are idioms (like break the ice), discourse markers (which is to say, you know what I mean, on the one hand, and so forth), collocations (highly significant), multi-word items (wall decor, room number), frames (make a/an X out of), and sayings or proverbs (silence is golden, that’s life for you). These divisions are not universal to the study of formulaic language or that of constructions, but for the purposes of discussion have been simplified (cf. more detailed taxonomies Nattinger and Decarrico 1992; Hudson 1998; Moon 1998).
The tendency for using re-constructed forms in both production and comprehension lies at the heart of communication; Bolinger (1976, 2) once claimed a speaker performs “at least as much remembering as they do putting together” in their language. They have an interest in doing so, as longer units, “ready-made frameworks on which to hang the expression of ideas”, allow one to circumvent the energy-consuming process of producing novel sequences of words “all the way out from ‘S’ [on a traditional grammar tree diagram] every time we want to say something” (Becker 1975, 17). To those interested in building accuracy, toward reaching the highest degree of effective communication, such as translators and in particular, oral translators, this becomes even more crucial. One’s idiomaticity, or command of phraseology, often serves as a yardstick for naturalness, beyond the ability to use grammar rules and an extensive vocabulary.

Native speakers’ reliance on stored, ready-made language units and noticeable choosiness about the forms they accept, had been left largely unaccounted for by the then-mainstream generative paradigm in linguistics when Pawley and Syder (1983) published a paper on what they referred to as “the puzzle of native-like selection”. Rather than assume that one calculates each utterance through a series of complex processes, the authors point out that much of what one actually does with language, and moreover what is perceived as acceptable in language, comprise a paradoxically small fraction of possible linguistic forms. Native speakers have a repertoire comprising thousands, if not hundreds of thousands of lengthy, ready-made items at their disposal (Pawley and Syder 1983). These vary in fixedness, length, and frequency, but may be distinguished by their “minimal unit” status for specific syntactic functions. As well, they function as institutionalized items: “the expression is a conventional label for a conventional concept [...]” (209–213). The authors emphasize that much of what we do with language is highly unoriginal, as compared to our potential for creativity in language.

Once a field of study associated mainly with an account of natural, native fluency, the area of formulaic language has seen increased research in the context of foreign language learning and learner production (Wray 2002; Wray 2013). To sound native-like in using a language suggests producing likely language units, and these may consist of multiple words. The appropriacy of these items is socially and culturally determined, to a great extent. The status of these multiword items as particular social constructs lies at the intersection of a given meaning and its recognizable (familiar, retrievable) form of expression. The process of “association” or “recognition” of language items in culture echoes a definition by Yorio (1980, 434), characterizing a conventional form as being predictable and expected in a given social context. Similarly, Erman and Warren (2000) identify the process of conventionalization as a prerequisite to the preference shown by native speakers for a particular combination of two or more words, over other groupings with equivalent meanings; it would appear to be a factor of restriction on choice. Such expressions are described in Sinclair
as “single choices, even though they might appear to be analysable into segments”.

Proficient speakers can be said to “know what is coming” in many exchanges, due to their socio-cultural experience; indeed, active “anticipation” of language (PL antycypacja), described by Marchwiński as “foreseeing given elements of an utterance on the basis of familiarity with language rules” [orig. przewidywanie określonych elementów wypowiedzi na podstawie znajomości reguł językowych] (Marchwiński 2008, 30 in Pędzisz 2014, 68), is practiced by professional oral translators as a key strategy for saving time and stores of energy. This technique of anticipation should be highlighted frequently during training courses; awareness raising in this area helps one to avoid unwanted borrowings or calques in language B (during retour), while also improving the alertness of the translator in the context of longer utterances in the mother tongue.

An influential definition of lexical items with a multi-word yet unit-like character comes from Wray (2002), who includes a separation from analytic (grammar-mediated) language processing; this gives the definition an explanatory dimension. Formulaic sequences (formulae) are:

continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (Wray 2002, 9)

However, this is essentially an account of language storage, and not necessarily one of a language item’s function. While unambiguously determining the status of an item as prefabricated in a speaker’s lexical repertoire is rather impossible, the definition suggests what purposes a formula/chunk of language may serve to a native speaker in language production: unanalysed and remembered units. The original source of those units (other speakers) is included in a definition by Buerki (2016, 18): “Formulaic sequences are phrases that are conventional pairings of form and unit of meaning in a speech community”.

It is these “conventional” texts, in the form of specific, recurring word combinations which the Chomskyan linguistic competence restraints could not fully account for the existence of without referring to performance data. Phraseological competence requires memory and experience with the ideas and functions most often conveyed through language – thus given easily-recognizable forms, creating a sort of culturally-endorsed shorthand. This “formulaic competence” (Gałkowski 2006, 128) has been characterized as a tripartite construct, consisting of processing (“ability-based”), socio-cultural and lexico-grammatical aspects.

The question arises as to why one would choose formulae (being conventional) in their language output, while having boundless, creative ways of expressing the self (indeed some do so, for artistic ends – easily recognized as such by virtue of their breaking recognizably with our linguistic expectations). We may know that
language allows us the opportunity to behave creatively, though the more likely route is via “the expected” – particularly given that cognitive resources (ours and that of our interlocutors) are precious, and the effort needed to think ahead in a conversation may require mainly energy-conserving strategies (cf. Wray 1990).

Formulae (chunks) may be employed to buy extra time for creating an argument or recalling information; these include well-known discourse organizers (cf. a discussion in Wray and Perkins 2000), such as padding, filler, turn-holders and the like. A speaker’s primary motive in making such moves might include expediency, or convenience. In order to minimise the effort expended during translation jobs lasting many hours, and requiring work in two language systems, it is advisable to use as many ready-made formulae (multi words chunks of language) as possible, for instance.

While our memories and retrieval are quick, we seem to economize on energy by default. So how would these motives change when functioning in a foreign language? The need for expediency would not change, though one’s resources (repertoire) might be undersized, in unexpected places, which may be due to various factors, among them features endemic to classroom training contexts. If foreign language teachers give their adult learners tools for consciously building a repertoire of multi-word units, it could lead to facilitated access and more accurate use of target language items.

2. Grammar and words: the non-native speaker and classroom language learning: advanced level language input and the training for oral and written translation

Foreign language pedagogy more and more frequently includes multi-word, ready-made expressions to enable smoother communication. Indeed, this is precisely what an advanced foreign language user needs in order to perform efficiently – many chunks that may be retrieved and possibly modified with grammar; this entails the opposite approach to teaching rules and inserting lexis, later. While for advanced adult language users a certain amount of commonplace language use – common within a given professional context – is sufficient to meet communication objectives, the broadest possible repertoire of routine language use is crucial to those whose work demands manifold tools for building and augmenting cultural literacy, self-confidence, and decision-making skills. Translator trainees may be seen as a prime example of such a demanding adult foreign language user group.

When a translator’s foreign languages are to be used under time pressure, with high expectations for bi-directional accuracy and fluency on the side of the employer, it will be in one’s personal interest to have as many opportunities to function in phrase-length units as possible (cf. Wray 2002). This type of advanced foreign language user needs situation-dependent language items to flesh out
particular skill sets, as well as a very rich store of multi-word expressions with a general character, such as those frequently used during speeches at conferences or lectures, undergoing simultaneous translations. First and foremost, this must occur more regularly in translator trainees’ foreign language classrooms and practise scenarios; as expressed by Gasek (2012, 281), “Undoubtedly, among the primary objectives of general preparation for oral translators one may include precisely this development of the habit of recognizing reproducible multi-word items” [orig. Niewątpliwie wśród zasadniczych celów ogólnego przygotowania tłumacza ustnego można wymienić właśnie wyrobienie nawyku rozpoznawania odtwarzalnych wielowyrazowców]. This crucial skill enables efficient production when the content of one’s source text is changing dynamically, and in a manner that is often highly formulaic (signals for digressions, apologies, or additional cultural information, etcetera).

If working phrasally is the preferred model, what might get in the way of doing so more often? While native speakers have been argued above to process language more holistically (that is, in longer units) thanks to stores of chunks in memory, adult foreign language users will tend to submit idioms and multi-word items to further analysis. Multi-word units often have meanings or functions that are not readily reached through analysing their elements separately, meaning the rules so often taught and tested in the foreign language classroom may appear rather unhelpful, even resulting in errors. Memorisation, strategies of association, particular experiences with the target language of generalisations about its rules, or other foreign language grammars may all affect learning and later retrieval.

It would appear that while adult learners of foreign languages amass large stores of words and grammar rules, the boundaries of collocations, fixed expressions and idioms may remain opaque, even among users who could otherwise pass as native speakers (examples from writing: *on other hand/*from second side; *the key to become; *the aim of this report is to supply overview for). Unidiomatic but grammatically acceptable forms may also result. Though there are some nearly irreplaceable functional items in language, most can be creatively constructed:

(1a) when a model which is used is not up to date one;
    they keep in rapid touch with everyone and quickly obtain a lot of goods via Internet shops;
    we think here about...

Foreign language users are continually forced to choose between using complex items (such as chunks) which are potentially richer in function (easily recognised by interlocutors), or building structures which may or may not be equal to the particular meaning making task at hand. This constant decision-making and judgment process is one of the key drivers in foreign language development; risk taking, in this case understood as using longer and longer units of language, should
be encouraged, in and out of training contexts. Teaching multi-word units yields benefits to speakers at all levels even while presenting challenges in classrooms dominated by an analytic, grammar-and-lexis approach to language.

Gozdawa-Gołębiowski (2008) has suggested highlighting word-by-word those cognitive scripts (produced in lieu of an existing multi-word form in the foreign language, while possibly having the status of a multi-word item in the interlanguage) in order to show their less than desirable communicative consequences, as well as overtly teaching to draw attention to form and not just the meanings of particular language items and language behaviours – that is, what is achieved through using a given item. The advantages may not immediately be apparent in the training context, when teaching adult foreign language users to attend to “rule-breaking” expressions. Despite that, highlighting message-level meanings, in addition to word meanings, is an essential step in raising speakers’ phrasal repertoire, long-term. It is these skills which translators (particularly oral translators in the domain of simultaneous translation) require when the task demands that one leave the lexis of the original speaker, in search of communicating correct, efficiently chosen multi-word units of meaning while upholding norms of appropriacy; in a discussion about the delicate “balancing act” (Kalina 2005, 771) involved in those high-pressure contexts, where the quality of one’s interpreting service is measured by its attentiveness to accuracy “on behalf of the weaker party” during a particular speaking event (Kalina 2005, 771, referring to Mack 2002, 114), success and quality in interpreting may be argued as “a textual product which provides access to the original speaker’s message in such a way as to make it meaningful and effective within the socio-cultural space of the addressee” (Pöchhacker 2001, 421 in Kalina 2005, 771). The aims of written translations tend to be similar.

3. Retrieval issues from word-by-word analysis

There are cases where a higher degree of correctness is demanded under the added pressure of time constraints, such as when producing an oral translation. Multiple unidiomatic forms could compound difficulties with retrieving the interpreted message, distorting the speaker’s intended style or meaning; examples of this include the following, recorded during training sessions in simultaneous translation at the Institute of Applied Linguistics, University of Warsaw, Poland, in 2019; they represent output during retour translation of a lecture (from the trainee’s native Polish into non-native English, level C1):

\[(2a) \quad \text{*without any doubts, some seem more indispensable [undoubtedly / without a doubt...]}\]
\[(2b) \quad \text{*I had pleasure to work with them on that [I had the pleasure of working with them on that]}\]
Multi-word Lexical Items and the Advanced Foreign Language User...

(2c) *lack of the knowledge holds him from (success) [lack of knowledge holds him back from]

(2d) *in D&R department [in the R&D department]

(2e) *his kid had allergy for dust [...was allergic to.../...had an allergy to...]

(2f) *it is one of this kind, then appear others, so it’s not only one store anymore [the only one of its kind, then others appear, so it’s not the only one / the one and only store]

(2g) *from board of members to the market [from (the) members of the board...]

The reasons for these errors are far from clear: there are many possible interpretations. Each may have been an issue of performance at a given moment in speaking – a retrieval error due to stress, tiredness, or other factors. These approximations may have been creatively constructed. They may also reflect similar constructions, transferred from the L1 (Polish) or another known language system. Alternately, they may have been multi-word items which were encoded incorrectly in the given speaker’s interlanguage systems, as “successful language use does not promote interlanguage growth” (Gozdawa-Gołębiewski 2003, 169). An incomplete formula may fossilise if uncorrected and if it brings “sufficient” communicative effectiveness to the user. Another interpretation concerns L1 transfer (not only negative but also its positive, facilitative effects), particularly concerning function words; foreign language users may not attend to these items (an issue in 2h. and 2i., below). The target expression may not have been encoded in the speaker’s interlanguage as a multi-word unit to start with. Concerning the near-miss character of the examples, errors occur when the application of known rules fails and reliance on content words as primary message carriers proves insufficient.

In terms of written language – in preparing translations or producing other text types, such as those in academic writing courses, the following are practical examples of how classroom language training in fixed expressions (in the context of writing in English as a foreign language) could raise awareness of phrasal equivalents (beyond individual words and rules). The following examples concern language assessment data (cf. Mitchel-Masiejczyk 2012). In the interest of testing how highlighting and awareness raising might contribute to improved results in advanced adult foreign language learners’ written work, the author has compared the results of testing in a module from second-year writing classes at the Institute of Applied Linguistics, University of Warsaw (2018, 2019). In 2018, the author tested two groups of second-year EFL academic writing students (B2) on phrasal equivalents and found that 213 out of 327 answers (65% overall) were incorrect, containing at least one error within a multi-word expression (phrasal verbs, meta expressions or idioms which were taken from texts the students had studied in class). In 2019, the error rate in the module (using identical academic texts and accompanying written exercises) was significantly lower: 114 of 357 test answers contained at least one error (32%). The author attributes that change to adding
a component to the module, where multi-word items are defined, identified in the texts, and highlighted during the writing exercises (such as noting the obligatory inclusion of an article which appears to “break” a known rule, or the presence of a preposition which differs from the Polish phrasal equivalent). Testing, as in the previous year, included translating the same selected expressions. The author found that those taught to treat phrases as wholes seem more likely to study them as multi-word items, and have a better chance of retrieving them accurately during testing.

As for the error types affecting fixed expressions or forms used in lieu of them, let us consider two of the test items, from the group (2019) that had overt instruction about multi-word items:

(2h) A na dodatek, – And on top of [all] that, [Literally: And-on-addition; Prompt: And on ...]
(2i) Ogólnie rzecz biorąc, – On the whole, [Literally: Generally-thing-taking; Prompt: On ...]

In the case of (2h.), 20 of the 51 learners tested answered correctly. However, despite the presence of on in the prompt, eight answered addition and two the addition (*And on addition/*And on the addition). Another tendency in the data is the addition of the by nine people (*And on the top of [all] that). At the top of/on the top of may be frequently encountered in other contexts; knowledge of the rules of articles may have contributed. In the case of (2i.), 26 out of 51 learners reproduced the correct form (seven others also produced the alternate on a general note; variants included *on a/the/ general basis, on average, and on the surface; another group includes on the general, on generally speaking, or the misspelled *on hole; while contexts could be found where they would comprise grammatical forms; they are inappropriate here. Moreover, each variant is likely to result in resource-taxing/time-consuming misunderstanding, which is not in the interest of the user. A longer period of exposure and further awareness-raising exercises, such as those mentioned in the following section, could bring more positive effects.

4. Toward awareness raising among trainees of written and oral translation types (retour – Polish to English)

To those who intend to actively translate from and into a foreign language (as the translation market in Poland may demand both types of services), or teach others a foreign language (all subject to assessment in the professional environment), additional perspectives for enhancing one’s self-study, organizing content, and increasing work efficiency are desirable. To address existing interest in improving accuracy and speed of retrieval in oral and written translation training, the author
carried out a series of workshops with fifth-year students of translation studies (Language A Polish, Language B English) at the Institute of Applied Linguistics, University of Warsaw, in 2018 and 2019. The series aimed to highlight the key role of multi-word expressions in English (and Polish) by examining language in order to identify reoccurring chunks, and to encourage conscious implementation of chunking strategies in language learning. It was postulated that improvements in accuracy and speed of retrieval, as well as a higher degree of naturalness or authenticity would result if chunks of language could be more readily identified in the input, and considered during output.

The participants readily adapted to the concept of looking for multi-word units (having studied “the translation unit”, collocations and phrasal equivalency in other courses, and as highly advanced users of English with frequent exposure to FL content), but claimed they have not heard of the essential role of “chunks other than words” (in the words of Lewis 1993, 186), cultural and pragmatic rituals, or high-frequency frames in which certain words but not others may be interchanged. The idea of viewing multi-word items (chunks) as “[units] of memory organization” or chunking/memorizing longer units to improve “the ability to build up such structures recursively” as put forward by Newell (1990, 7 in Ellis 2003, 76) makes the concept accessible as a strategic behaviour, for accelerating one’s learning process. It may also systematise one’s approach to one’s native language and how it is being used on an everyday basis: examples from Polish, which also abounds in culturally-prescribed formulae, rituals and preferred patterns of co-selection (cf. Galkowski 2006), drove the concept further that multi-word units may function as time-saving devices and relieve cognitive effort, while raising naturalness.

The activities such as those described later in this section were addressed to (and well received by) students who felt they had reached a plateau in their learning, as well as those who were entering the job market and needed a new way to organize existing knowledge to perform better. The notion of “watching for” language units and adjusting the way one learns or memorizes lexis, proved of interest; the participants were also encouraged to eschew memorizing words out of context, which often implied changing one’s habits from nearly two decades of classroom and self-study practices. The practise of observing speakers and trying to identify multi-word items which serve as anchors in their discourse, rather than listening for key words only, was another central concept in the workshop activities. The practise of framing language use in terms of its multi-word items was deemed “very eye-opening”; “a game-changer”; “worth all the work”, and reducing the tendency to “focus on single words and now I’m sure I won’t do only that” (remarks from students’ written feedback on the workshops). In further translator training courses of this type, the priority would lie in improving accuracy at the phrasal level (recognition, skill in choosing equivalents) during simultaneous interpreting.
As presented above, in order to enhance processing speeds and improve competence, it is useful to locate multi-word expressions and then discuss their significance in a given context. The techniques given below are examples of training activities for consideration, many of which are useful to oral and written translation types. They are also intended to encourage self-study among advanced foreign language learners/users. Their applications are not limited to training in English as a foreign language.

Phrase-length situational language

Situational phrases (how do you do, if I were you, it’s just around the corner, would you mind, accept our apologies for) are emphasized more in the curricula of early education but never actually diminish in importance. A training scenario may begin from eliciting a fixed expression through asking focus questions which are highly contextualized, without analysing the individual words which are being used. It is important to emphasize the full expressions, especially those which are sentence length, and not only look at their key words. Equivalents in other known languages should be discussed and compared. A variation on this is to present formulaic replies and reaction language seen in common social contexts, and elicit the most likely questions (preferably also formulaic, highly situational). This exercise may be done at all levels of competence, introducing phrases of increasing complexity or containing particular cultural nuances. For self-training, the process of directing one’s attention to remarks or responses in common social situations found in the input should not be limited to key words only, but focus on looking for phrase length or sentence length items, by default.

Popular culture and its fixed expressions

In a training or self-training context, a language user should note the language of small talk, soap operas, the meta expressions of news casts and interviews, and expressions typical of particular genres in other popular media channels for examples of highly regular language items (minimum two words together) with clear functions. Their relation to maintaining (or violation of) cultural norms should be noted. In classroom settings, learners may use highly conventional situations to write “cultural scripts” for, with an emphasis on function rather than artistic invention; they may compare results and which expressions re-appear in various scripts. An additional step involves adapting the texts to sound as “non-formulaic” (atypical) as possible and discuss the resulting meanings and likely effects. To extend the activity, one may present the non-formulaic forms for assessment by native speakers and ask what was meant; record their formulaic answers. Outside of a classroom, one may compare what they have identified to contexts in language corpora or phrasal dictionary resources.
Multi-word listing practise

It is a common practise at all levels of study to present learners with lists of decontextualised lexis for memorisation; this does not address the existence of preferred units in the target language. Foreign language learners, especially adults, may benefit from learning chunks of natural language rather than single-word items. Outside of the classroom, for self-study, words should be noted down with their co-text wherever possible; again, phrasal dictionaries and corpora may be helpful in supplying likely co-occurring items.

Identifying “odd words out” which are not part of larger units

When analyzing a text in written form in a classroom setting, it is useful to identify units of language which do not seem part of fixed or conventional expressions (start in the native language), and note when writers depart from fixed expressions and what the consequences/likely intentions are. Different genres of texts may be compared in terms of how idiomatic/fixed the language tends to be, and where it is freest.

Building communication around multi-word chunks

When training students for written translation work, one may introduce lessons “writing around” multi-word items (thereby also resisting the urge to break up or analyse these fixed expressions) by giving translation trainees a list of such items, all of which must be included in a short composition, or utterance, as wholes. This encourages the approach of stringing together bundles of lexis to conserve energy and time resources. This may begin from writing form letters, instructions or conveying messages with particular social or cultural significance, issuing complaints or making declarations, and grow in sophistication to include more complex argument types and their discourse markers, or frames for digressions and storytelling. To those training to work in particular subfields of oral translation, with access to glossaries, this is a particularly helpful exercise. The trainees receive lists of expressions; their equivalents may be discussed and later practiced on the basis of recorded speeches. Self-study may include practising the habit of noting where multi-word items occur in the input, along with their functions.

5. Conclusions

It has been argued that multi-word items with idiomatic properties comprise a significant proportion of native speakers’ daily communication in English. Foreign language processing among adults seems to differ from that of native speakers.
in the processing strategies employed; the standard grammar/lexis approach to foreign language teaching in classrooms may have an influence on production. The social and professional needs of advanced foreign language users, such as oral translators, demand additional energy-conserving processing strategies for streamlining message comprehension and formation under time constraints. Certain changes in foreign language teaching procedures should include planning more opportunities for noticing, build skills of anticipation, and provide practice in incorporating more multi-word items. That said, the adoption of a highly lexical or phrasal focus in training advanced foreign language users is not without its difficulties, as the teaching of multi-word items involves frequent highlighting of the items as wholes, as users should not subject these language chunks to extensive internal analysis. Attempts to provide additional training in the identification of multi-word items, such as those presented above, have brought encouraging effects. The incorporation of modules on recognising and learning longer chunks of language would address the unique needs for increased efficiency seen in translator training contexts among advanced foreign language users. This presents a promising pedagogical area as well as a valuable venue for continued research.

References


