

NOTES ON RESEARCH INTO IDIOMATICITY

JERZY ZYBERT I STANISLAV KAVKA

The great interest in vocabulary issues recently observed in applied linguistics manifests itself in extensive research and a great number of publications. This interest is noticeable particularly in the fields of second language acquisition, language teaching methodology, translation studies, lexicography, etc. It also concerns idiomaticity recognized as occupying a specific place in lexis. The present article is a modest attempt to trace the most significant developments in the field signalled in its title above.

Before going into details with regard to idiomaticity issues we will touch on some relevant questions of a more general nature. To begin with, it seems appropriate to consider a terminological problem: what differences in meaning and use exist between *idiomaticity*, *idiematology*, *idiomatics*, and *phraseology*? Actually, all these terms, and perhaps others, have been used to refer to one and the same area of past and present linguistic interest. Clearly, linguists employ specific terms to express their arguments and to promote their own, often exclusive, understanding of a given idea. Their intention is normally to state precisely and unambiguously what they mean by the terms they choose when referring to the matters they discuss. Unfortunately, the intentions do not always come true. For instance, in this particular case *phraseology* as a term does not cover the vast domain of our present interest since in common understanding it refers only to lexis. However, it may be fully acceptable for language teachers who usually understand phraseology as lists of “useful phrases” for their students to memorize. They may not realise the fact that idiomatic expressions are based on semantic rather than lexical grounds, or that the very term “phraseology” is derived from the base-term “phrase” which for modern linguists has connotations

of reference primarily to grammatical structures. As for the terms *idiomaticity* and *idiomatology*, we can be content with the way of thinking about a parallel to some other terms of an apparently identical morphological structure. We speak traditionally of “phonology”, “morphology”, “philology”, and so on. Therefore the term *idiomatology* makes us regard the discipline as a truly linguistic one, treated as a field of science proper, i.e. one that has its objectives (goals) to probe and also its own methods of investigation. Thus *idiomaticity* (morphologically, like “regularity”, “priority”, etc.) will refer to “a quality” derived, in turn, from an attribution of, say, “constituting, or containing (an) idiom(s)”.¹ Initially we could be happy with a view like that, but at once we feel that the explanation does not encompass everything that we would like to include. Certainly, *idiomaticity* does refer to quality; however, it does not necessarily need to imply that the idiomaticity of an expression depends on its containing of an idiom and the term itself can be used for semantic and structural irregularity of phrasal idioms (Reichstein 1984). In a broader sense of the term it can be said that a given expression is “idiomatic” (or, has “proper idiomaticity”) if it is judged intuitively by native speakers as usual, natural, and commonly acceptable. In this respect an acceptable and concise definition of “idiomaticity” will be one that takes it as a “function of familiarity and frequency of use” (Sonomura 1996).

The term *idiomatology*, first introduced by Grace (1981), was seen as more appropriate in connection with sociolinguistic and pragmatic aspects of language but, as a matter of fact, Grace used it in the sense that can be considered synonymous, at least to some extent, with the generally accepted sense of *idiomaticity*. Grace has compiled an impressive list of unusual linguistic structures that are not normally accounted for in grammar and which he grouped under the label of “idiomatology”. Grace presents several types of idiomatological phenomena that range from many kinds of seemingly arbitrary and unmotivated restrictions, via illogical and semantically anomalous forms, to grammatical exceptions, such as e.g. “fifty-cent cigar”; “by and large”; “I slept late”; “kick the bucket”; “didn’t you know that?”

Grace’s *idiomatology* was soon systematised by Pawley and Syder (1983). In his view fluent and idiomatic control of performance in a language results from the knowledge of “sets of sentence stems” that become “institutionalised” or “lexicalised”, where a “set” is a syntactic unit like a clause whose form and lexical content are fixed. To this he added the notion of *speech formula* which meant a conventional link of a particular formal construction and a particular

¹ This is a definition offered and accepted generally by most linguists interested in *idiomatology*.

conventional idea². Consequently, all genuine idioms are to be seen as speech formulas, but not all speech formulas are idioms. Moreover, in psycholinguistic terms, true idioms are such speech formulas that are semantically “non-compositional” and, to make this view complete, idioms are syntactically “non-conforming”. However, even if “non-conformity” in syntax is understandable fairly well as a grammatical peculiarity of an expression, “non-compositionality” requires some comment since, undoubtedly, there may be idioms, which are at least partly compositional³.

Admittedly, interest in idioms developed rather late in linguistics. Truly, the discipline was unable at its start to handle phenomena which appeared odd, exceptional, not fitting the patterns of grammar. It was only after linguistics consolidated having arrived at descriptions of what is general, common, and perhaps even universal that it was ready to look at phenomena regarded as exceptional. Lots of valuable contributions that dealt with idioms and similar expressions soon appeared. Jespersen (1933) called them *formulas* to show that these exceptional expressions demanded a mental activity that was different from that required for “free expressions”. Looking back, three main periods of idiom investigation can be distinguished: the first one belongs to the very beginning of the 20th century, the second to the 1950s, when the work in the field was resumed thanks to theoretical developments, and the third extending from late 80s through the 90s bringing results of great importance, due to interest in pragmatics and psycholinguistics. However, to date, this period is not history proper yet.

The first period was crowned with Smith’s book entitled *Words and Idioms* (1925) which is a collection of his essays. The longest of them, called simply *English Idioms*, contains the greatest number of examples that Smith was able to put together, file and classify. Smith was influenced by Jespersen and worked within the then favourite tradition of etymology, apparently using lists of idioms that had been compiled mostly by others. Nevertheless, the merit goes to him: his classification of idioms is indeed detailed and elaborate, encompassing every area of origin possible, whether this is sea, war, nature, farming, cattle, birds, etc. He informs us of idioms “from foreign sources”, he deals separately with idioms drawing on the Bible, as well as with Shakespeare’s own original idioms. Smith provides additional information on some of the examples, very often fairly interesting⁴. It is interesting to note that he does not hesitate to admit he is

² Let us notice here that the very term “formula” is widely used by linguists in various subtle meanings and specifications; however, it seems to be a sort of cover term embracing what might simply be called “an idiomatic expression”.

³ Based on a specific research, this issue is discussed in Kavka, S., 2003: Chapter 5.

⁴ For instance, *curry favour*, originally ‘curry Favel’, where ‘curry’ means ‘to rub and clean a horse’, here a horse of certain colour called ‘favel’ and being a symbol of cunning character.

ignorant of a given etymology. It is worth making an observation that many years before Smith, Dr. Samuel Johnson was rather conceited about his knowledge of etymologies and that he disliked idioms as something that sullied language purity. Smith may not be a man of Johnson's reputation, yet his simile is now worth remembering: "*Idioms are like little sparks of life and energy in our speech*".

Roberts (1944) proposed a polar relation between 'discourse' and 'language' – categories "which are expected to conjoin in order to produce the complete sphere of communication" (p. 299). Obviously enough, this relation is very close to the well-known dichotomy between content and meaning of an expression. In other words, recalling de Saussure's view, Roberts' "language" is the psychophysical mechanism, i.e. "langue", which does the expressing through its dynamic aspect of utterance, i.e. "parole". To this end, according to Roberts, idioms belong primarily to discourse and, since they create language, they must also create grammar, which belongs primarily to "language". Hence grammar is viewed as fossil idiom! It is a concept conceived rather broadly, indeed. Yet, what may be appreciated is the fact that all idioms are believed to have originated as innovations of individuals and, using Roberts' words, each idiom is, as a matter of fact, "a mental monument of history" (p. 304). Therefore we can also draw one challenging conclusion worth considering: idioms can, or should, be studied as a source of language change! Hockett was one of the few American linguists who showed some interest in idioms. He offered a very formal definition of the concept of *idiom*. In his *Course in Modern Linguistics* he writes:

[The idiom is...] "any Y in any occurrence in which it is not a constituent of a larger Y", where Y is 'any grammatical form whose meaning is not deducible from its structure'

(Hockett 1958: 172).

The consequences of this definition are fairly complex and far-reaching. It should suffice to note, however, that according to it every single morpheme must be granted idiomatic status because a morpheme has no structure from which we could tell its meaning. Hence, it follows that idioms in Hockett's view are not understood as multilexical units only, contrary to the way we regard them traditionally. The distinction between unilexical and multilexical units seems to be only arbitrary for Hockett, which teachers will appreciate, claiming that both simple words as well as complex units, i.e. traditional idioms and kindred expressions, must be learnt separately. Incidentally, we wish to note that a contemporary psycholinguistic account would put it in similar way, only perhaps

using its own terminology⁵. Unlike Hockett, Malkiel (1993) claims that idioms are always multilexical. His contribution in the field may seem to be of little importance at first sight because he examined only phrases that conventionally linked two items, appearing always in the same order, e.g. *spick and span*; *kith and kin*; *by and large*, etc.; they are called "binomials". Yet the criterion of "irreversibility" which he postulated is something that consequently sets a major problem for the approaches we know of as transformational-generative (TG). Briefly, any two items linked by the conjunction "and" can be swapped, i.e. they are expected to be reversible; yet certain multilexical units of this structure do not allow for reversibility.

The first linguists who tackled idioms within the TG framework were Katz and Postal (1963). They advanced two novel proposals concerning idiomaticity: (1) the division of the lexicon into two parts, namely a lexical part and a phrase-idiom part, and (2) the criterion of non-compositionality. This is what we can probably take for granted nowadays. Their ideas were being elaborated by others, recently also by psycholinguists. What we must point out, however, is their distinction between "lexical idioms" and "phrase idioms", the two types being defined on syntactic grounds. Basically, the former are described as syntactically dominated by one of the lowest syntactic (grammatical) categories, namely by noun, adjective, and verb while the latter, on the contrary, cannot be described this way. Suffice it to compare, e.g., *white lie* and *How do you do?*

The idea of the two types seems important for the reason that "idiomatic expressions" can subsume such lexemes as clichés, compounds, or even phrasal verbs. Admittedly, some linguists exclude compounds altogether (e.g. Balint 1969, who argues that compounds are not phrases), while others treat compounds as minimal idiomatic expressions (Weinreich 1972; Makkai 1972), and still others do not seem to be quite certain about the issue and prefer to introduce a separate category with the label "cross-cutting terms" (Sonomura 1996).

The contemporary perspective on idioms was initiated by Weinreich. He originally noted that an idiom is a complex expression whose meaning is not the sum of its parts and so it cannot be inferred from the meanings of its elements. He also developed a more adequate terminology, proposing that an idiom is a subset of a "phraseological unit", which, incidentally, resembles some later views on lexicon and lexical idioms, for example those of Katz and Postal (1963). In Weinreich's understanding a phraseological unit is an expression with at least one constituent being polysemous as, for instance, in *white lie* where *white* is polysemous in the intended terminological sense.

5 Idioms are believed to be stored in the mental lexicon as 'single lexical units'; other hypotheses work with *Direct Access Model* or *Dual Process Model*. For more information, see Kavka, 2003, Chapter 5.

On the other hand, Weinreich's views on idioms are also comparable with Katz and Postal's on phrase idioms. However, his approach is more elaborate in that he makes clear reference to context: he writes that the selection of subsense in a phraseological unit is determined by context and also that there is a reciprocal contextual selection of subsenses in an idiom. Although the role of context is, undoubtedly, decisive, it is worth noting, however, that ambiguity is not always eliminated in the case of idiomatic expressions. Therefore, if Weinreich referred literal and non-literal (i.e. metaphorical) meaning, then quite a few multiword expressions would have both interpretations (or, readings) – yet they are called idioms. All in all, Weinreich's contribution to idiomaticity studies was very significant at his time when rigor and explicitness was demanded in linguistic studies. Particularly valuable were his notes on idiomaticity in terms of unproductive and semiproductive (syntactic) constructions and on the familiarity of use of idioms.

With the developments in TG linguistics other attempts were made to study idioms within this framework but they did not have a significant impact in the field. It worth mentioning Fillmore et al., whose work was probably the last attempt to accommodate idiomaticity within the TG framework. In their view "...an idiomatic expression or construction is something a language user could fail to know while knowing everything else in the language" (1988: 504).

Idiomaticity issues were of interest also to non-generativists, among them tagmemicists and stratificationists. For example, in Pike's (1967) perspective the idiom is a phrasal unit which he called a "hypermorpheme" and described as a specific sequence of two or more specific morphemes. Thus what we normally refer to as "idiomatic expression" must be a subset of the hypermorpheme. Pike's contribution, however, should be valued first of all for his incorporation of cultural factors into the linguistic theory with his insistence that idiomatic expressions do reflect the culture of a given language community.

Makkai's ideas derive from stratificational approach and therefore it is rather difficult to relate them to non-stratificationist discussions. Still, the data that he collected and most of his terminology on idiomaticity are a significant contribution to the field. For example, he distinguished "lexemic idioms" from "sememic idioms", which are said to be placed in two separate "idiomaticity areas" (i.e. strata, layers). According to Makkai, an idiom is made up of more than one minimal free form has two different characteristics: (1) each "lexon" (i.e. component) can occur in other environments as the realisation of a monolexemic lexeme – hence there are so-called lexemic idioms, e.g., *White House*; *blackbird*, and (2) the aggregate literal meaning, as derived from the respective constituent lexemes, works additionally as the realisation of a sememic network which is unpredictable – hence so-called sememic idioms, e.g.

chew the fat. Another terminological invention is the distinction between the act of "encoding" and the act of "decoding". The former can be illustrated by using proper prepositions. For example, we do not say **with* but rather *at* in *He drove...70 m.p.h.* Actually, Makkai prefers to speak of "phraseological peculiarities" here rather than of idioms. Genuine idioms are based on the act of decoding, and in his truly precise taxonomy they are of various type, such as "lexical clusters", e.g. *red herring*, "tourneures", e.g., *fly off the handle*, etc. We could very well add that all idioms of decoding are simultaneously idioms of encoding, but not necessarily vice versa (Makkai, 1972: 25). Thus, e.g. *hot potato* used in the sense of "embarrassing issue" is idiomatic from the semantic point of view (in terms of so-called sememic idioms), and it is also idiomatic as a peculiar phrase since we do not say **burning potato* or *hot chestnut*. On the other hand, not every act of encoding is idiomatic. According to Makkai, in every natural language there is a sort of middle style, that is to say neutral, devoid of either type of idiom (which, as is known, non-native English speakers are very fond of using).

So far we have looked at idiomaticity issues discussed within theoretical linguistics. However, idiomaticity is also of significant interest to applied linguistics and psycholinguistics. Idioms and idiomatic expressions, as part of vocabulary of a language, are of serious concern in the field of foreign language teaching and learning. Although in some cases idioms do not pose serious problems to learners (Harmer 1998: 156-7), which is the case when certain features are shared by the source and the target language, they, nevertheless, require particular attention in language pedagogy. With regard to learning L2 lexis, including idioms, a number of empirical studies have been conducted to investigate whether there exist idiomatic relationships between L1 and L2 on the one hand, and whether they are susceptible to transfer from L1 in learning L2, on the other. The most notable and best known research related to this issue is, probably, that done by Kellerman (1979, 1983, 1986).

Concerning transfer, it is claimed that transferability of linguistic features depends on the degree of their markedness. Markedness refers to "the idea that some linguistic structures are 'special' or 'less natural' or 'less basic' than others. For example, the use of 'break' in 'she broke my heart' can be considered marked in relation to the use of break in 'she broke a cup'" (Ellis 1994: 713). Markedness is understood in two ways. One is explained in terms of the theory of Universal Grammar, where language rules are either core or periphery: core rules are universal, thus unmarked; peripheral rules are language specific, idiosyncratic, thus marked. The other account of the notion of markedness is that which pertains to language typology – typological universals enable making distinctions between marked and unmarked language features: those found in all

or most natural languages are unmarked, those that are specific to a particular language or found only in a few are marked. The degree of markedness can be determined through reference to a linguistic description or theory (Zobl 1984; Eckman 1985, White 1987)

Recent claims about the relationship between markedness and language transfer hold that learners transfer L1 unmarked forms into their interlanguages more easily than marked forms and that the latter, actually, resist transfer (Ellis 1994). Thus, it should be clear that idioms, which are highly marked, are not transferable. The general conviction that marked features are not transferable or, more precisely, that the more a feature is marked, the less transferable it is sounds reasonable; however, it was challenged as some findings contradicted it, especially that it is rather difficult to assess the degree of markedness of an expression. A way of establishing this degree was proposed by Kellerman. He suggested that markedness and its degree could best be determined by competent language users (native speakers) when asked whether they sense specific features of their language as "infrequent, irregular, semantically or structurally opaque, or in any other way exceptional" (1983: 117).

Kellerman replaced the term "markedness" with "prototypicality". Exploring the concept he attempted to find out what makes some items potentially transferable and others potentially non-transferable. It turned out that speakers have intuitions about their native language and hold specific perceptions of their own L1 structures. The perceptions are largely responsible for transfer and non-transfer of particular structures. These perceptions determine that some L1 structures are vulnerable to transfer and others are not. Kellerman demonstrated this in his well-known "breken" study and later in the "eye" study, both on lexical transfer (Kellerman 1978 and 1986, respectively). The findings that he obtained yielded a forceful claim that "transferability can indeed be established entirely on the basis of the learner's knowledge of his native language, and that the establishment of these probabilities will have validity for any given L2" (Kellerman 1986: 37). In other words, whether or not learners transfer a form depends on their perception of its acceptability in L2, i.e. how marked in L1 it appears to them. Logically, then, transfer of idioms is possible but constrained by the learner's perception of distance between L1 and L2: being less frequent, less transparent they are less transferable than non-idioms that are unmarked.

The natural conclusion that ensues from the above is that idioms, perceived to be specific, marked, idiosyncratic elements of language resist transfer. On the other hand, idioms and idiomatic expressions on the whole, play an important role in everyday language use. This means that L2 learners' communicative competence should include knowledge of them, both receptive and productive, to

make verbal communication with native speakers fully successful. How idioms can be learned and taught is of serious concern to L2 teachers. Admittedly, relatively little has been done to facilitate learning of idioms and this situation calls for further research to elaborate successful learning strategies and improve foreign language teaching in general.

Bibliography

- Balint, A. 1969. "Sector analysis and idioms". *Kivung Quarterly* 14/4, pp. 433-441.
- Davies, A., C. Criper and A. Howatt (eds.) 1984. *Interlanguage*. Edinburgh: Edinburgh University Press.
- Eckman, F. 1985. "Some theoretical and pedagogical implications of the markedness differential hypothesis." *Studies in Second Language Acquisition* 7: 289-307.
- Ellis, R.. 1994. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Fillmore, C.J., P. Kay, J. and O'Connor. 1988. "Regularity and idiomaticity in grammatical constructions: the case of 'let alone'." *Language* 64, pp. 501-538.
- Gass, S. and L. Selinker (eds.) 1983. *Language Transfer in Language Learning*. Rowley, Mass.: Newbury House.
- Grace, G. 1981. *Essay on Language*. Columbia SC: Hornbeam Press.
- Harmer, J. 1998. *How to teach English*. London: Longman
- Hockett, C.F. 1958, 1967. *A course in modern linguistics*. New York: The Macmillan Co.,
- Jespersen, O. 1966. *Essentials of English grammar*. London: Allen and Unwin.
- Katz, J.J. & P.M. Postal. 1963. "Semantic interpretation of idioms and sentences containing them." *M.I.T. Research Laboratory of Electronics: Quarterly Progress Report* 70, pp. 275-282.
- Kavka, S. 2003. *Book on Idiomaticity*. Žilina: EDIS.
- Kellerman, E. 1979. "Transfer and non-transfer: where are we now." *Studies in Second Language Acquisition* 2: 37-57.
- Kellerman, E. 1983. "Now you see it, now you don't" in Gass and Selinker (eds.) 1983.
- Kellerman, E. 1986. "An eye for an eye: cross-linguistic constraints on the development of the L2 lexicon" in Kellerman and Sharwood Smith (eds.) 1986.

- Kellerman, E. and M. Sharwood Smith (eds.) 1986. *Cross-linguistic Influence in Second Language Acquisition*. Oxford: Pergamon.
- Makkai, A. 1972. *Idiom structure in English*. The Hague / Paris: Mouton.
- Malkiel, Y. 1993. *Etymology*. Cambridge, UK: Cambridge University Press.
- Pawley, A. and F. Syder. 1983. "Two puzzles for linguistic theory" in: Richards, J.C. and R.W. Schmidt, (eds.): *Language and Communication*, pp. 191-225, London: Longman.
- Pike, K.L. 1967. *Language in relation to a unified theory of the structure of human behavior*. The Hague: Mouton.
- Reichstein, A.D. 1984. *Nemeckie ustojčivye frazy i ustojčivye oredikativnye edinicy*. Moscow.
- Roberts, M.H. 1944. "The Science of idiom: a methods of inquiry into the cognitive design of language." *Publications of the Modern Language Association* 59, pp. 291-306.
- Smith, L.P. 1925. *Words and idioms*. London: Constable & Co. Ltd..
- Sonomura, O. M. 1996. *Idiomacity in the basic writing of American English*. New York: Peter Lang Publishing, Inc.
- Weinreich, U. 1972. *Explorations in semantic theory*. The Hague / Paris: Mouton.
- White, L. 1987. "Markedness and second language acquisition: the question of transfer." *Studies in Second Language Acquisition* 9: 261-86.
- Zobl, H. 1984. "Cross-language generalizations and the contrastive dimension of the interlanguage hypothesis" in Davies et al. (eds.) 1984.