

Idioms

Idioms are a class of multi-word units ‘which pose a challenge to our understanding of grammar and lexis that has not yet been fully met’ (Fellbaum et al. 2006: 349). They are commonly believed to be qualitatively different from ‘normal’ language, but the precise nature of this difference can be elusive. Even amongst idiom scholars, it is difficult to find a consensus as to what precisely is, or is not, an idiom, because of the heterogeneity of the class.

There is widespread agreement on one general principle: an idiom is an institutionalised expression whose overall meaning does not correspond to the combined meanings of its component parts. However, this criterion can be said to apply to a wide range of **phraseological structures**, such as **collocations**, **formulaic greetings**, **clichés** and other **conventionalised expressions** which, although idiomatic to some extent, are not idioms in the strict sense. The challenge for idiom researchers is therefore to formulate a definition which is flexible enough to include all known idioms, yet exclude non-idioms such as those mentioned above.

An idiom is composed of two or more constituent parts, generally deemed to be words, although Hockett (1958: 177) admitted phonemes as constituents and Makkai (1972: 58) morphemes. Despite appearances to the contrary, each of these words does not contribute to the overall meaning of the phrase, which operates as if it were a lexical item in its own right and expresses a semantically complete idea which may be quite independent of the mean-

ings of its components. The reasons for this semantic anomaly derive mainly from the fact that an idiom is not built up word by word, according to the grammar of the language, but is a **non-compositional phrase** which is learned, stored and recycled as a single chunk.

Current psycholinguistic views support the argument in favour of considering idiom as a type of ‘long word’ whose meaning is accessed directly and not through prior **decomposition** or **analysis** of the constituents (Gibbs 1994, 2002). However, when an idiom is encountered for the very first time, language-users have no choice but to decipher its meaning from the meaning of the constituents, usually doing so by taking into account the most salient meanings first (Giora 1997, 2002; Peleg and Giora 2001). That this tactic enjoys a limited success rate is due to the difficulty in identifying which meaning of polysemous components is relevant and the extent to which the idiom is semantically motivated or transparent.

The ease with which an idiom can be interpreted is based on its level of semantic transparency as well as truth conditions and other contextual cues. A **transparent** idiom yields its meaning easily, because there is a straightforward connection between the phrase and the intended meaning. For example, *not see the wood for the trees* (‘to lose oneself in details and fail to see the larger picture’) requires little semantic re-elaboration; it is therefore located towards the transparent end of the scale. On the other hand, an expression which has a more arbitrary relationship with its meaning, such as *to go cold turkey* (‘suddenly stop taking a drug that you have become addicted to’), can be described as

unmotivated or **opaque**. The transparency or opacity of an idiom cannot be measured in absolute terms, as it is affected by the individual's real-world knowledge, awareness of cultural norms, and general familiarity with the phrase.

The more closely the wording of an idiom reflects a real-world situation, the easier it is to interpret: *make one's blood boil* reflects the heat felt in the body when enraged; *to bite the hand that feeds you* can easily be connected to ingratitude. In much the same way, an idiom which refers to a culturally familiar situation poses little difficulty to interpretation: knowledge of team sports reveals the principles of equality and inequality respectively encoded in *a level playing field* and *move the goalposts*. It is also true that an idiom which is familiar to the hearer is perceived as being more transparent than one which is not so familiar, regardless of its real-world or cultural relevancy: *like a red rag to a bull* ('a provocation') is much less frequent than *make sb see red* ('provoke or anger sb') (Philip 2000), and therefore requires more effort in decoding. Finally, it is worth noting that, as with all figurative language, even transparent idioms pose problems for language learners who, lacking the necessary linguistic and cultural knowledge to decipher them, are apt to interpret them literally.

While some idioms dovetail into our conceptual system, not all do, and one well-documented feature of idioms is their adherence to, or violation of **truth conditions**. When a phrase alludes to events or situations that cannot possibly occur in the real world, a literal interpretation is incongruous: human blood is always red (*blue blooded*), kitchen implements do not speak to each other (*the pot calling the kettle black*), and animals do not fall from the sky as precipitation (*rain cats and dogs*). In situations such as these, the only way to make sense of the meaning is to treat the expression as idiomatic. Not all idioms violate truth conditions, and many phrases can, at least theoretically, be read literally or figuratively depending on which interpretation best fits the context in which the phrase appears.

A great deal of psycholinguistic literature deals with the effects of context on the interpretation of phraseological homophones – idioms which can have both literal and idiomatic readings. Here context is textual, not pragmatic,

and is characterised by **biasing contexts** designed to sway the reader's interpretation towards an idiomatic or a literal meaning (for an example of this, see Giora and Fein 1999: 1605). Outside experimental conditions, contextual cues are particularly important in determining the meaning of idioms whose literal and figurative meanings are either not well established or occur with relatively low frequency: the phrase *cherry picking* may be used literally or figuratively, but its location in a text on blue-chip business would be incongruous if read literally, thus triggering its idiomatic reading ('being selective').

Recent corpus-based research into homonyms suggests that context is less crucial than previously believed, and that one of the possible readings usually predominates. According to Hoey (2005: 82ff.), it can be argued that language users will avoid using a familiar idiom in a context where it could be interpreted literally, preferring instead to paraphrase or use an alternative expression. Thus, under normal communicative conditions, a person who is literally skating on ice which is thin would not be described as *skating on thin ice*; and if a person who hits a bucket with their foot is described as having *kicked the bucket*, humour automatically ensues because of the clear mismatch between the more familiar, idiomatic meaning and the literal description of events.

Idioms are learned and reused as single lexical items, yet they are not single words. While the **canonical form** of an idiom (the citation form used for dictionary definitions) is fixed for the purposes of language description, the reality of language in use is that most idioms can undergo a controlled amount of variation to their typical realisation. There is some divergence in opinion on this point between theoretical and descriptive studies on idioms. Pre-corpus scholars defined idioms as being **fixed** or **frozen** in form, in reference to the fact that they resist morpho-syntactic change; now it is more common to find them described as **stable** (Čermák 1988) or of **limited flexibility** (Barkema 1996: 128). This difference in terminology is due to the fact that much pre-corpus literature on idioms deals only with what is theoretically possible, with the result that the categories and principles devised, while extremely detailed and rigorous, fail to reflect adequately the attested behaviour of idioms in

use. Successive studies informed by corpus data, notably Moon (1998), have challenged the notion of fixity in light of the observation that most idioms do in fact allow variation to occur, so long as some vestige of the canonical form survives.

Demonstrating the syntactic and semantic stability of idioms has been one of the prime considerations of figurative language scholars, especially those working within the generativist tradition. Idioms are said to be **transformationally deficient**, and in order to prove the case that non-canonical realisations of idioms cause their meaning to revert to literal, they can be subjected to a series of tests. The tests adopted fall into two broad categories: lexical and grammatical. The lexical tests include the augmentation test (addition of lexical constituents), the elimination test (deletion of constituents), the substitution test (replacing a constituent by a semantically-related word), the permutation test (rearranging constituents whose order is fixed). The grammatical tests include blocking of predication, blocking of the formation of comparative and superlative forms of adjectives, blocking of nominalisation and blocking of passivisation (Gläser 1988: 268–9). As Gläser explains, '[a]s soon as these practical procedures are followed, the resulting construction will be grammatically correct and empirically sensible, but it will cease to be an idiom' (1988: 268).

Transformation tests do not stand up well to empirical scrutiny. Even before the widespread use of computer corpora, criticisms were levelled against this method of idiom classification, because it fails to look beyond the tested phrase and compare its behaviour to similar structures or semantically related language items. Chafe (1968: 122) argues that the blocking of passivisation can be explained by the underlying meaning of an idiom, not its idiomaticity. Citing *kick the bucket*, he points out that the literal equivalent *die* would similarly fail the passivisation test (**to be died*). The other transformation tests do little better, and are of limited relevance to those idioms which have no literal homonym (*hue and cry*, *in fine fettle*, *run amok*).

The availability of large, electronically searchable linguistic corpora has allowed idiom scholars to put transformations and other theoretical considerations to the test. Corpus-based

studies illustrate that lexical variation in idioms is a widespread phenomenon, not one restricted to the creation of special linguistic effects such as punning, humour and irony. In Moon's (1998) study of fixed expressions and idioms in a 18-million-word corpus, attested lexical and morpho-syntactic variation is described in detail (1998: 75–174). Moon reports that that approximately 40 per cent of the idioms and other fixed phrases studied occurred in a variant form (1998: 120). However, the larger the corpus is, the more variation occurs; in some cases the canonical form can be outnumbered by its variants (Philip 2008: 103).

Even if idioms are not fixed, they do have a stable form which is learned as a multi-word lexical item. This canonical form is subject to exploitation in the normal course of language use, and so idioms can appear with lexical and grammatical alterations, in truncated and augmented forms, and in phrases which merely allude to the original: 'Talk about Mr Pot and Mr Kettle?' (*the pot calling the kettle black*; Philip 2008: 103). The rules governing such exploitations have yet to be determined, but are believed to be predominantly conceptual and semantic in nature.

It has been established that figurative expressions are not merely colourful add-ons to the lexicon, but that they contribute to its **evaluative inventory** (Carter 1997: 159). Simply put, idioms have a literal counterpart in the language, but this counterpart is not a true synonym because it fails to express the evaluative meaning encoded in the idiom. Čermák (2001: 13) notes that 'idioms are a primary means for the expression of positive and negative attitudes', but goes on to lament the fact that little research has been carried out into the matter.

Idioms resist pigeon-hole definitions because they constitute a heterogeneous class of anomalous lexical items. In order to understand them fully, it is necessary to understand better the mechanisms at work in 'normal' language, and here, too, corpus analysis is challenging traditional descriptions. Idioms are less fixed than used to be believed, 'normal' language less free.

G. P.

Suggestions for further reading

- Fernando, C. and Flavell, R. (1981) 'On Idiom: Critical Views and Perspectives', *Exeter Linguistic Studies*, vol. V, Exeter: University of Exeter.
- Makkai, A. (1972) *Idiom Structure in English*, The Hague: Mouton.
- Moon, R. (1998) *Fixed Expressions and Idioms in English*, Oxford: Oxford University Press.

The International Phonetic Alphabet

The International Phonetic Alphabet is a means of symbolising the segments and certain non-segmental features of any language or accent, using a set of symbols and diacritics drawn up by the **International Phonetic Association (IPA)**. It is one of a large number of phonetic alphabets that have been devised in Western Europe over the centuries, but in terms of influence and prestige it is now the most highly regarded of them all. Hundreds of published works have employed it. It is used throughout the world by a variety of professionals concerned with different aspects of speech, including phoneticians, linguists, dialectologists, philologists, speech scientists, speech and language therapists, teachers of the deaf, language teachers, and devisers of orthographic systems.

Its origins lie in the alphabet (or rather alphabets) used by the forerunner of the IPA, the Phonetic Teachers' Association, founded in 1886 by Paul Passy (1859–1940), a teacher of modern languages in Paris. Since then, a number of slightly differing versions of the alphabet have been published at irregular intervals by the IPA. The latest was published in November 2005. Four versions of the alphabet can be found in publications since 1951: 'revised to 1951', 'revised to 1989', 'revised to 1993, updated to 1996' and 'revised to 2005'. All are available in near-A4-size chart form (see the reproductions in Figures 1–4).

The 2005 chart is freely downloadable from <http://www.langsci.ucl.ac.uk/ipa/ipachart.html>.

Braille versions of the alphabet have been proposed at various times, but there is as yet no standard one. An additional alphabet, ExtIPA (Extensions to the IPA), for the symbolisation of forms of disordered speech was formally adopted by the Association in 1994.

Detailed guidance on the manner in which the alphabet is used can be found in another of the Association's publications, the *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet* (1999). This is a large-scale revision of *The Principles of the International Phonetic Association* (1949). The guiding principles for the symbolisation of sounds have remained essentially, though not entirely, the same as those that the Association drew up and publicised as early as August 1888.

The aim of the notation is to provide the means for making a phonemic transcription of speech, or, in the original words of the Association, 'there should be a separate letter for each distinctive sound; that is, for each sound which being used instead of another, in the same language, can change the meaning of a word' (Phonetic Teachers' Association 1888). Thus, the distinction between English *thin* and *sin* can be indicated by the use of θ and s for the first segment in each word. It is often the case, however, that by the use of symbols, with or without diacritics, an allophonic as well as a phonemic [see PHONEMICS] notation can be produced. So, for example, the labiodental nasal in some English pronunciations of the /m/ in *symphony* can be symbolised allophonically as [m̥] since the symbol exists to notate the phonemic difference between that sound and [m] in a language like Teke, a language of Central Africa. Nevertheless, the phonemic principle has sometimes been set aside in order to allow the notation of discernible allophonic differences within a single phoneme. Thus, far greater use is made in practice of the m̥ symbol for notating the labiodental nasal allophone of /m/ or /n/ in languages like English, Italian, and Spanish than for showing the phonemic contrast between /m/ and /m̥/.

It is sometimes assumed that, since the alphabet is designated as *phonetic*, it should have the capacity to symbolise *any* human speech sound. This is not, nor has it ever been, the purpose of the alphabet. Its prime purpose is to handle the