

A Semantic and Lexical-Based Approach to the Lemmatisation of Idioms in Bilingual Italian-English Dictionaries

Chris Mulhall

Waterford Institute of Technology, Republic of Ireland

The aim of this paper is to propose a new semantic and lexical-based lemmatisation framework for the recording of idioms in bilingual Italian-English dictionaries. Many of the difficulties and inconsistencies characterising the lexicographic treatment of idioms stem from them being viewed as a semantic and lexically homogenous phrasal category. This incorrect generalisation typically motivates the traditional description of idioms as being non-compositional and lexically fixed units. Current bilingual Italian-English dictionaries treat idioms quite unsystematically, mainly due to their reliance on the subjective judgement of lexicographers and generic syntax-based listing strategies. The rationale for pursuing these methods remains unclear, particularly given the availability of substantive semantic and lexical information that could provide a more defined template for determining the position of idioms in a dictionary. This paper looks at two particular aspects of idioms in five current bilingual Italian-English dictionaries: Il Ragazzini (ZIR) (2009), Grande Hoepli Dizionario Inglese (GHDI) (2007), Il Sansoni Inglese (SI) (2006), Oxford Paravia Italian Dictionary (OPID) (2006) and Hazon Garzanti Inglese (HGI) (2009). The first is a semantic-based investigation, which analyses the entry procedures for 150 English and 150 Italian idioms across three categories: pure idioms, figurative idioms and semi-idioms. The second examines the listing strategies for 40 English and 40 Italian idioms with variable verb and noun components. Overall, two particular trends emanate from the analysis. Firstly, the arrangement of idioms is unsystematic and the allocated entry points do not reflect or emphasise their individual semantic or lexical features, which are central to their identity. Secondly, the English-Italian and Italian-English sections of certain dictionaries are disparate in their overall coverage with Italian idioms assigned a greater number of listings. These discrepancies call for a formulaic entry model that eliminates the subjectivity, inconsistency and unsystematic approach currently associated with the treatment of idioms in bilingual Italian-English dictionaries.

1. Theoretical Perspectives on the Lemmatisation of Idioms

The problematic status of idioms as dictionary entries has long been at the heart of lexicographical debate and still remains topical to the present day. Over the course of the last four decades a number of theorists have proposed solutions to this long-standing problem, but few, if any, have ever been truly assimilated into the process of dictionary compilation. Many modern-day monolingual and bilingual dictionaries prefer to sublemmatise idioms in one of their constituent entries largely due to the conventionality of this listing technique and its economisation of space. This method, however, has three particular drawbacks. Firstly, it diminishes the importance of idioms by placing them in a subordinate role to words, which misconstrues their equal status as units of meaning in the lexicon. Secondly, it overlooks specific lexical and semantic features that are central to their identification and help to distinguish them from other phrases. Thirdly, it requires the lexicographer to select, often arbitrarily, one or more lexical element entries in which to record the idiom. Identifying the ‘most important’ or ‘semantically heaviest’ element in an idiom is a largely subjective and complex decision and is practically impossible for those with a non-compositional meaning due to the imposed figurative sense on each lexical component. Lexicographic theorising occasionally puts forward potential solutions to systematise the treatment of idioms and some of these are presented in Table 1.1.

Theorist	Year	Proposed Listing Strategy	Treatment within the Dictionary Structure
Petermann	1983	List in the entry of one notional component with supplementary cross-references at other possible entry points	Sublemma
Tomaszczyk	1986	List in each constituent entry	Sublemma
Burger	1989	List in one constituent entry – Not founded upon semantic criteria	Sublemma
Botha	1992	List in first main component entry	Lemma
Lorentzen	1996	List in noun component entry	Sublemma
Harras & Proost	2005	Designated point of entry determined by the semantic features	Non-Compositional Idioms = Lemma Partially Compositional Idioms = Sublemma

Table 1.1. A Chronological Overview of Proposed Idiom Listing Strategies

Deciding the most appropriate point of entry is central to this debate, but theorists show little consensus on this matter. They diverge on two key points; the selection and number of entry point(s) and the rationale for this choice. The variety of proposed listing strategies gives some insight into the true complexity of the task encountered by lexicographers when recording idioms. For example, Petermann (1983) and Burger (1989) advocate that idioms should only be listed once, but do not make any reference to a specific location. By contrast, Botha (1992) and Lorentzen (1996) propose a first main component entry and noun entry strategy respectively, whereas Tomaszczyk (1986) puts forward a more comprehensive remedy by suggesting that idioms should be listed in all of their constituent entries.

An equally important consideration in the formulation of a listing strategy is the criteria underpinning it. The unconventional and irregular meaning of idioms makes semantics an obvious and valid criterion, but this view is not widely endorsed by theorists. Burger (1989) claims that organising idioms on the basis of their semantic features is unsuitable as it requires dictionary users to be *au fait* with a significant body of semantic information. Conversely, Harras and Proost (2005) advance a semantically-orientated lemmatisation model that aims to accurately reflect the unitary meaning of idioms as well as accentuating their salient features through their assigned point of entry. They propose that non-compositional idioms should be listed separately at the base of their first main component entry to illustrate the semantic mismatch between the lemma as an independent lexical item and idiom constituent. In addition, they suggest partially compositional idioms should be inserted as a sub-entry under their literal component lemma. This is an innovative approach to an old problem, but its failure account for certain semantic features of idioms, such as non-words, as well as overlooking the important phenomenon of lexical variation, means it cannot be considered a truly comprehensive solution. Atkins and Rundell (2008: 168) sum up the complexity of idioms as a lexicographical entry by calling them ‘the most difficult multiword expressions (MWE) to handle in lexicography. In the absence of hard and fast criteria, it is well nigh impossible to be wholly consistent.’ This claim bears some truth as idioms are widely regarded as the most complex lexicographical entry, but it overlooks their distinct semantic and lexical features that are legitimate and accepted criteria and could provide a more solid foundation for constructing a consistent entry model.

2. Defining the Characteristics of Idioms

Linguistic theorising has not yet proffered an all-encompassing definition for idioms, thus they remain a poorly defined set of expressions. One of the factors militating against them being more accurately described is the multiplicity of generalised definitions that fail to convey their actual characteristics. These tend to include generic terms of reference, such as ‘non-compositional,’ ‘semantically opaque’ and ‘fixed unit,’ which, although can be readily applied to a large proportion of idioms, are not representative of the entire category. Idioms are neither semantically nor lexically homogenous; therefore, they must be defined and classified in a way that accentuates their diverse, but salient, characteristics.

Theorist(s)	Semantic Features	Theorist(s)	Lexical Features
Fraser (1970: 22)	A constituent or series of constituents for which the semantic interpretation is not a compositional function of the formatives of which it is composed.	Hartmann & James (1998: 71)	A fixed expression whose overall meaning is not always transparent from the combination of its constituent words.
Fernando & Flavell (1981: 22)	The meaning of an idiom is not the result of the compositional function of its constituent parts, i.e. the syntax is non-correlative and the resultant signification non-literal in terms of the referents denoted by these parts.	Fontenelle (1998: 191)	Idioms are part of the larger class of multiword units. They are frequently described as semantically opaque word combinations.
Pinker (1999: 292)	A phrase whose meaning cannot be predicted from the literal meaning of its parts.	Jackson & Zé Amvela (2007: 243)	A more or less fixed expression with a non-literal meaning.

Table 1.2. Theoretical Perspectives on Idioms

2.1. A Semantic Classification of Idioms

Idioms are frequently referred to as ‘non-compositional expressions,’ but this is misleading as it portrays them as having a universal semantic representation. Originally all idioms were literal expressions; however, the imposition of figurative meanings was not entirely uniform, which resulted in some becoming more idiomatic than others. This stratified idioms into three semantic layers determined by their level of compositionality: pure idioms (non-compositional, semantically opaque), figurative idioms (non-compositional, logically interpretable) and semi-idioms (partially compositional). Although the different semantic subtypes all collectively function as single units of meaning in the lexicon, the above three-way classification categorises idioms on the basis of their inherent semantic features and identifies them as a semantically heterogeneous subgroup. A number of theorists (Fernando and Flavell 1981; Cowie et al. 1993; Mel’čuk 1995; Fernando 1996 *inter alia*) also construct similar analytical frameworks, but dictionaries continue to overlook these valid suggestions when planning their listing procedures for idioms. The following sections discuss the different semantic types of idioms as well one of their more unusual semantic features.

2.1.1. Pure Idioms

Pure idioms are fossilised, non-compositional expressions that are impervious to any lexical, semantic or morphological changes as it results in a loss of their unitary meaning. A classic example of a pure idiom is *to kick the bucket* as it is a semantically opaque and lexically fixed expression. The combination of these characteristics places pure idioms at the top of the scale of idiomaticity.

2.1.2. Figurative Idioms

Figurative idioms also have a non-compositional meaning but there is a more plausible link between their literal and figurative senses than pure idioms. For example, the underlying sense of futility in the phrase *to carry coals to Newcastle* is comprehensible by placing the action in a literal context.

2.1.3. Semi-Idioms

Semi-idioms comprise of at least one literal and one figurative component, which gives them an overall partially compositional meaning. The presence of a literal element makes them both the most interpretable and least idiomatic of all three categories. An example of a semi-idiom is *to cost an arm and a leg* with the verb element, *to cost*, preserving its literal sense when used in this idiomatic expression. On the basis of this empirical analysis it is the verb element that most regularly retains its literal sense with 45/50 English semi-idioms and 47/50 Italian semi-idioms following this trend.

2.1.4. Idioms with Non-words

The large majority of idiom constituents also operate as independent lexical items in the lexicon. There are, however, a very limited number of idioms that contain non-words, also referred to as ‘unique sublexical items’ by Gouws (1991: 87), which occur only in phraseological constructions. Some examples of non-words in English include *amok* (*to run amok*), *cropper* (*to come a cropper*) and in Italian *catinella* (*piovere a catinelle*) and *cilecca* (*fare cilecca*). In many cases, their idiom listing preserves their lexicographic representation as shown in the entry AMOK from the ZIR (2009) (Figure 1.1).

amok /ə' mɒk/ avv. (*solo nella loc.:*) **to run a. 1** essere in preda a furore omicida; essere preso da un raptus **2** (*fig.*) scatenarsi; diventare sfrenato; (*di situazione, ecc*) impazzire.

(*Il Ragazzini* 2009: 48)

Figure 1.1. The entry AMOK in *Il Ragazzini* (2009)

Although non-words are not fully fledged lexical items *per se*, it is important that dictionaries continue using their entry to include their idiom listing. Therefore, this provides legitimate grounds for retaining their lemma status, as well as presenting a comprehensive account of the lexicon as possible.

2.2. Lexical Variation in Idioms

Despite some theoretical assertions to the contrary, idioms cannot be regarded as ‘fixed expressions.’ The rationale for this description is possibly motivated by some theorists considering idioms as stable units that resist any alteration to their lexical or syntactic

structure. This, to some extent, is true, but it cannot be holistically applied to all idioms. Moon (1998: 120) acknowledges that ‘fixedness’ is an important property of idioms, but research shows that it is not universal feature across this category of phrases. Her empirical analysis reveals that 40 percent of fixed expressions undergo lexical variation or strongly institutionalised transformations and around 14 percent show two or more variations on their canonical form. However, similar to the lexicographic neglect of their semantic features, the coverage of lexical variation in dictionaries is also inadequate (Fellbaum 2007). The type of idiom variation most regularly encountered by a lexicographer is of a lexical nature, which fundamentally involves the replacement of one or more of their constituents. This, however, is not a uniform phenomenon and the following sections discuss the various forms of lexical variation possible in idioms.

2.2.1. Content Word Variation

The most common form of lexical variation in idioms occurs through the substitution of one content word, either a verb, noun or adjective, by another. In the majority of cases, the lexical replacement is synonymous, for example, *to raise/up the ante*, *to scream/yell blue murder*, whereas in others the variants show little or no semantic relationship, for example, *to cool/kick one’s heels*, *to burn one’s bridges/boats*. Replacing one content word for another is an accepted transformation in idioms once the lexical alternative is a nominated variant, whether synonymous or not, that preserves the overall figurative representation of the expression. Lexical variability is not restricted to a particular syntactic slot in idioms, but certain elements are more likely to vary than others. Moon (1998) notes that verb components show the most lexical flexibility followed by nouns and then adjectives. However, the phenomenon of lexical variation reduces the idiomaticity of an expression. According to Fernando (1996), the potential variation of an idiom’s lexical components diminishes its status as a composite unit, thus making it less idiomatic than those with immutable lexis.

2.2.2. Idiom Schema Variation

The majority of lexically variable idioms permit lexical substitution in only one of their syntactic slots; either verb, noun or adjective, but a very limited number allow this transformation in both their verb and noun components. This results in some idioms displaying both paradigmatic and syntagmatic variation patterns, which creates what Moon (1998: 161) terms as an ‘idiom schema.’ One of the more regularly cited examples of an English idiom schema is *to add fuel to the fire*, which has four possible verb alternatives, *to fan*, *to fuel*, *to pour oil on* and *to throw oil on*, and one lexical substitute in the noun slot, *flame*, which can also show inflectional variations in this particular expression. An example of an idiom schema in English and Italian can be seen in Figure 1.2.

English Idiom Schema

To add fuel to the fire/flame(s)
 To fan the fire/flame(s)
 To fuel the fire/flame(s)
 To pour oil on the fire/flame(s)
 To throw oil on the fire/flame(s)

Italian Idiom Schema

Aggiungere/Dare esca al fuoco
 Aggiungere/Mettere legna al fuoco
 Buttare/Gettare olio sul fuoco/sulle fiamme
 Soffiare sul fuoco/sulle fiamme
 Versare benzina sul fuoco

Figure 1.2 An English and Italian Idiom Schema

From a lexicographical viewpoint, achieving a consistent and comprehensive coverage of idiom schemas is fraught with difficulty, largely due to the flexibility of both their syntactic and lexical structures. This prevents a lexicographer from pinpointing a specific lexical constituent that is consistently found across each variation, whose entry could accommodate all possible lexicalisations and allow the user to access these in a single look-up.

2.2.3. Morphosyntactic Variation

A small number of idioms exhibit morphosyntactic variation, which involves an inflectional change to their noun components, for example, *to go by the board/boards*, *to rack your brain/brains*, etc. This represents a very marginal form of variation as the lexical structure of the idiom is not modified, apart from showing inflections permitted within the parameters of the phraseological system.

2.2.4. Function Word Variation

Function words, also referred to as prepositional or particle elements, exert no semantic bearing on an idiom, but may also be subject to variation. This applies to a relatively low number of idioms and is afforded minimal lexicographic coverage. Sometimes the substituting function word shows only a minor morphological adjustment from the original, for example, *to go around/round in circles*, whereas in others, it is replaced by a broadly synonymous equivalent, such as *to beat about/around the bush*.

2.2.5. Optional Lexis Variation

Lexical variation in idioms, for the most part, is of a paradigmatic nature; however, in a very small number of cases it involves a change to their syntagmatic lexical structure. This is referred to as optional lexis variation and describes the arbitrary inclusion or exclusion of specific lexis from an idiom, for example, *to scrape (the bottom of) the barrel*, *to go the (full) distance*, etc. Idioms are generally composite syntactic structures that do not permit the insertion of additional lexical elements or modifiers as this regenerates the literal meaning of the other constituent parts. But the optionality of some constituents means that their omission does not affect the idiom's overall signification, but their inclusion can help to intensify or reinforce it.

3. Listing Strategies for Idioms in Current Bilingual Italian-English Dictionaries

Idioms have long been, and still remain, a troublesome lexicographic entry. Throughout recent decades theorists have advanced some possible explanations for this problem, such as the word bias of dictionaries, as well as potential solutions, like recording idioms under each of their lexical components, etc. Dictionaries have never widely endorsed or implemented any of the proposed theoretical solutions, which may suggest that there is no panacea for reconciling the totality of lexicographical difficulties posed by idioms. However, the absence of a methodical listing strategy that encompasses the diverse qualities of idioms is still a persistent weakness of many modern day dictionaries. In their definition of lemmatisation, Hartmann and James (1998) refer to the organisational problems created by MWEs, such as idioms, but 12 years on these still remain current despite technological advances and improved compilation methods in the same period.

‘Lemmatisation is a problem awaiting a comprehensive solution (attempted by computational approaches) in connection with wider tasks such as how to choose a suitable headword from the constituents of a fixed expression, and how to list binomial compounds (under the last or first constituent).’

(Hartmann and James 1998: 83)

Both theorists and lexicographers often lament about the disregard of the style guide by dictionary users, but the poor or non-existent information relating to the organisation of phrases and idioms is an equal criticism of dictionaries and contrasts with their thorough treatment of other MWEs, such as compound nouns. This adds some credence to the belief that dictionaries focus primarily on providing information at word-level with phrases viewed as being of secondary importance. According to Atkins and Rundell (2008), a coherent policy on the treatment of all MWEs is one of the most important functions of a dictionary’s style guide. But a review of current bilingual Italian-English dictionaries reveals that while five of the six provide in-depth information on their treatment of compound nouns, only two indicate their methods for arranging idioms. In some cases, information about the lemmatisation of compound nouns is so detailed that it not only states their entry point, but also addresses the vagaries of whether they are hyphenated or not. These include:

- Il Ragazzini (2009):** These are given under the headword of the first component. Anyone looking for compound words in English must be aware of the fact that the use of the hyphen in English is very uncertain, fluctuating and continuously evolving. Compound words will therefore be found under the headword which corresponds to the first component. If not, they are in fact probably single non-hyphenated words, and as such will be found as separate headwords (ZIR 2009: 4).
- Grande Hoepli
Dizionario Inglese (2007):** *No information included*
- Il Sansoni Inglese (2006):** *Listing policy as per idioms*
- Oxford Paravia
Italian Dictionary (2006):** In the Italian-English section the symbol ♦♦ can be found at the bottom of the entry reserved for Italian **set noun phrases**, always listed under the first headword in the phrase. When there is more than one, they are arranged in alphabetical order based on the second full component (adjective, noun, exclusive of prepositions in between them) (OPID 2006: xv).
- Hazon Garzanti
Inglese (2009)** Compound words are treated as independent entries in alphabetical order, whether they are represented graphically as a single word or two separate words, with or without a hyphen. (HGI 2009: xi)

The SI (2006) is the only dictionary forming part of this empirical analysis to include a concise listing policy for idioms. It states that idioms are located in their first main component entry, which is most often a verb, but defaults to the next noun component if the verb has a high frequency phrasal use. The listing policy in full states that:

The list of phrases and idioms for each word is preceded by the symbol ■. To help find individual phrases given under a headword, the phrases have been put in alphabetical order according to the phrase's most important key word or words, which have been highlighted in a different colour.

The phrases, idiomatic expressions and proverbs that make up the phraseology section are listed under the first key word contained in the expression (be it verb, noun or adjective). Therefore, for example, the proverb *he who pays the piper calls the tune* is found under the verb **pay**, and the phrase *as hard as iron* is listed under the adjective **hard**. Likewise, the Italian proverb *le bugie hanno le gambe corte* is given under **bugia** and the phrase *cavalieri della Tavola Rotonda* under **cavaliere**.

As an exception to this, certain extremely common verbs (*be, can, come, do, get, give, go, have, keep, let, make, must, put, take, will* in English and *andare, avere, dare, dovere, essere, fare, lasciare, mettere, potere, prendere, stare, tenere, venire, volere* in Italian) have been ignored in listing the phrases under the headwords. As a result, the phrase, *to get one's cards* is given under the headword **card**, and *prendere qcu. in castagna* is given under **castagna**.

(*Il Sansoni Inglese* 2006: 14)

The inclusion of this organisational information makes the SI (2006) particularly user-friendly and has two direct benefits. Firstly, it provides a structured template for arranging idioms and eliminates the reliance on the subjective judgement of the lexicographer to determine their entry point. Secondly, it also takes into consideration the syntactic importance of the verb element in phrasal expressions, which is relevant for both English and Italian languages. The OPID (2006) is the only other dictionary to provide information relating to its treatment of idioms, but this is rather ambiguous and does not clarify the number or location of their entries. It states that:

Where there are sections containing idiomatic expressions under the entry they are marked with the special symbol ♦. These also include proverbs, preceded by the label PROV.; they are translated, wherever possible, with an equivalent proverb (even if formally different); otherwise, they are translated literally or explained (with the symbol =). Since more is better than less to a user of a bilingual dictionary, translations of idiomatic expressions can be found under more than one of its components in order to make them easier to find.

(*Oxford Paravia Italian Dictionary* 2006: xv)

The OPID's policy does not appear to take into account that idioms do not all have the same syntactic shape, therefore, expressions with three lexical components have potentially more entry points than those of a more common V+N structure. As a result, it is unclear from the information in the preface of the OPID whether idioms with a longer syntactic structure are assigned more points of entry or not.

4. An Empirical Analysis of Listing Strategies for Idioms in Current Bilingual Italian-English Dictionaries

The following section discusses the results from the analyses of the idioms in the English-Italian sections (Table 1.3) and Italian-English sections (Table 1.4) of the selected bilingual Italian-English dictionaries. Overall, three clear trends emerge from the empirical investigation. Firstly, idioms are an inconsistently and under-treated component of bilingual Italian-English dictionaries with no discernible or systematic pattern to their arrangement. Secondly, certain bilingual Italian-English dictionaries include organisational information pertaining to idioms, but do not always strictly adhere to it when recording them. Thirdly, listing patterns from both the English and Italian samples reveal that Italian idioms are more extensively recorded than English idioms. The totality

of these findings further underlines the absence of and need for a structured recording system for idioms in bilingual Italian-English dictionaries.

Idiom Status	Semantic (n=150)			Lexical (n=40)	
Idiom Type	Pure Idioms (n=50)	Figurative Idioms (n=50)	Semi-Idioms (n=50)	Variable Verb Idioms (n=20)	Variable Noun Idioms (n=20)
Syntactic Position of Constituent Entry	1st/2nd/3rd/Multiple	1st/2nd/3rd/Multiple	Lit/Non-Lit/Mix	Fixed/Variable /Multiple	Fixed/Variable /Multiple
Il Ragazzini (2009)	7/17/0/24 Not Listed = 2	4/11/1/32 Not Listed = 2	6/10/33 Not Listed = 1	2/2/14 Not Listed = 2	5/1/13 Not Listed = 1
Grande Hoepli Dizionario Inglese (2007)	11/37/0/0 Not Listed = 2	13/28/2/6 Not Listed = 1	14/20/15 Not Listed = 1	10/3/5 Not Listed = 2	12/6/2
Il Sansoni Inglese (2006)	27/12/0/3 Not Listed = 8	22/12/2/5 Not Listed = 9	22/18/6 Not Listed = 4	2/13/4 Not Listed = 1	13/2/4 Not Listed = 1
Oxford Paravia Italian Dictionary (2006)	5/39/0/5 Not Listed = 1	4/24/3/17 Not Listed = 2	7/34/7 Not Listed = 2	12/2/5 Not Listed = 1	6/5/8 Not Listed = 1
Hazon Garzanti Inglese (2009)	9/26/0/10 Not Listed = 5	8/11/3/21 Not Listed = 7	9/18/19 Not Listed = 4	5/5/9 Not Listed = 1	7/6/6 Not Listed = 1

Table 1.3. The Lemmatisation of English Idioms in Current Bilingual Italian-English Dictionaries

Il Ragazzini (2009)

Empirical data reveal that the ZIR (2009) adopts a multi-entry strategy when recording the majority of idioms in the semantic and lexical categories under analysis. Its use of this strategy is, however, not absolutely uniform across the three semantic groups, for example, 33/50 semi-idioms are assigned at least two listings compared to 24/50 pure idioms. The trend of assigning multiple entries is also apparent in its recording of lexically variable English idioms with 27/40 entered in two of more places. But these are not consistently distributed with some listed in all their variable component entries and others in both a variable and fixed component entry.

Grande Hoepli Dizionario Inglese (2007)

On the basis of the analysis, a second main component entry is the preferred strategy for recording pure idioms, figurative idioms and semi-idioms in the English-Italian section of the GHDI (2007). This is particularly apparent in both non-compositional groups with 37/50 pure idioms and 28/50 figurative idioms following this method, but the arrangement of semi-idioms shows less predictability. In total 20/50 semi-idioms are inserted in their non-literal component entry, which is also quite regularly the second syntactic component, but this figure is somewhat lower than the two non-compositional groups. A contributing factor to this divergence is the high number of literal component entries (14/50), which may be considered by the compilers of the GHDI (2007) as either having a greater semantic significance or being a more likely consultation point by users. Its treatment of lexically variable English idioms is relatively uniform with the majority of those analysed (ten variable verb/12 variable noun) listed in one or more fixed component entries. While this shows some degree of consistency, it requires the user to

search in different syntactic component entries depending on the type of variation and also know the number and position of fixed elements in order to ascertain and retrieve all the relevant listed information.

Il Sansoni Inglese (2006)

Generally, the organisation of idioms in the SI (2006) conforms to its declared policy of listing them in the 'first main component' entry. This emerges from each of the three semantic categories, which show the verb element as the most regularly used location to include idioms. However, its overall coverage of idioms is not entirely systematic and this can be attributed to two factors; the allocation of multiple entries to some idioms and residual listings from earlier editions of the dictionary. Overall, the entry point for 93 of the 129 listed English idioms is in line with the listing policy, equating to a 72 percent level of conformity. Another noteworthy finding is that 21 of the sampled English idioms are not listed, which also appears to be due to their omission from previous editions. Consistency and efficiency in arrangement are the two main benefits of the SI (2006) idiom listing strategy, but it also has certain limitations. The most evident of these occurs in its treatment of lexically variable idioms. Although a first main component entry strategy is suitable for recording idioms with fixed lexis, it creates some organisational inconsistencies for those with variable components. Listing patterns from the empirical analysis confirm this problem with variable verb idioms generally being recorded under all of their variable components (two or more listings), but variable noun idioms listed only once, typically in the fixed (verb) component entry (single listing). This disparity means that the SI (2006) does not give equal representation to the same idiomatic phenomenon in the English-Italian section.

Oxford Paravia Italian Dictionary (2006)

The empirical analysis of the selected English idioms in the OPID (2006) highlights an anomaly between the listing policy in the preface and their actual arrangement in the dictionary. Contrary to its claim of translating 'idiomatic expressions under more than one of its components,' it assigns only one point of entry to the majority of the sampled English idioms. This emerges quite clearly in the distribution of entries to pure idioms and semi-idioms, but figurative idioms register a comparably higher number of multiple listings, which makes it the most adherent of all three groups to the listing policy. Its coverage of lexical variation is also inconsistent with a fixed entry component the preferred location for variable verb idioms, but no discernible strategy in use for those with variable noun parts. In summary, these trends suggest that the information relating to the organisation of the idioms in the preface of the OPID (2006) is not entirely representative of its positioning of English idioms.

Hazon Garzanti Inglese (2009)

The HGI (2009) is the least systematic of all current bilingual Italian-English dictionaries in its lemmatisation of English idioms. A cross-analysis of entry points in each semantic group reveals a general lack of order in their organisation. For example, listing patterns from the empirical data show that the majority of pure idioms tend to be inserted in their second main component entry, whereas just under half of the figurative idiom sample receive two or more listings. This trend, in itself, shows that the HGI (2009) does not record idioms with a similar semantic representation in the same way. Furthermore, another inconsistency surfaces in the semi-idiom sample with an almost equal use of a

non-literal component and multiple entry strategy. Its treatment of lexically variable idioms is also lacking systematicity with the three possible entry strategies; variable component entries, fixed component entry and a complete recording, all having a similar level of application throughout both variable samples.

Idiom Status	Semantic (n=150)			Lexical (n=40)	
Idiom Type	Pure Idioms (n=50)	Figurative Idioms (n=50)	Semi-Idioms (n=50)	Variable Verb Idioms (n=20)	Variable Noun Idioms (n=20)
Syntactic Position of Constituent Entry	1st/2nd/3rd/Multiple	1st/2nd/3rd/Multiple	Lit/Non-Lit/Mix	Fixed/Variable /Multiple	Fixed/Variable /Multiple
Il Ragazzini (2009)	3/15/3/29	2/7/0/40 Not Listed = 1	4/16/27 Not Listed = 3	4/0/16	2/3/15
Grande Hoepli Dizionario Inglese (2007)	5/31/6/6 Not Listed = 2	4/27/1/17 Not Listed = 1	7/27/13 Not Listed = 3	9/3/8	2/8/10
Il Sansoni Inglese (2006)	24/18/1/6 Not Listed = 1	23/9/0/14 Not Listed = 4	17/14/17 Not Listed = 2	2/10/8	10/2/8
Oxford Paravia Italian Dictionary (2006)	2/9/1/34 Not Listed = 4	0/5/1/44	2/11/34 Not Listed = 3	2/1/17	0/5/15
Hazon Garzanti Inglese (2009)	5/19/3/18 Not Listed = 5	6/11/2/26 Not Listed = 5	7/19/18 Not Listed = 6	5/3/11 Not Listed = 1	3/5/12

Table 1.4. The Lemmatisation of Italian Idioms in Current Bilingual Italian-English Dictionaries

Il Ragazzini (2009)

In terms of providing a comprehensive coverage of idioms, the ZIR (2009) is comparably the best of all current bilingual Italian-English dictionaries. Its largely consistent use of a multiple entry strategy gives an equally robust idiom content in both sections of the dictionary. The empirical evidence shows that it gravitates more towards this listing method for non-compositional Italian idioms, but this trend is reversed for partially compositional Italian idioms, which are recorded less extensively than those in the English sample. One area of the ZIR (2009) that shows some disparity is the coverage of lexically variable idioms. On the basis of this empirical analysis Italian lexically variable idioms are treated more thoroughly and systematically, primarily through the allocation of more points of entry. This is epitomised in the Italian variable noun sample, in which all the analysed expressions receive a full account of their lexical variability.

Grande Hoepli Dizionario Inglese (2007)

Continuing the trend from the English sample, a second main component entry is also the most frequently used strategy by the GHDI (2007) to record Italian idioms in each of the three semantic categories. The consistent use of this method ensures a general uniformity in the organisation of idioms with an identical semantic representation. However, one of the limitations of this strategy is its unsuitability for semi-idioms as it bypasses their verb component, which is predominantly the part that retains its literal meaning and is a more obvious consultation point for users. Generally, English and Italian idioms follow a very similar listing format, but there are some measurable differences in the treatment of those

in the latter category. One of these is the high number of Italian figurative idioms (17/49) assigned two or more entries, which on the basis of other groups is a quite sporadically used listing procedure in the GHDI (2007). In addition, Italian lexically variable idioms are organised less consistently, for example, a single fixed component entry is the most prevalent listing strategy for variable verb idioms, whereas half of those with variable noun elements are listed twice or more.

Il Sansoni Inglese (2006)

The arrangement of Italian and English idioms in the SI (2006) shows a number of parallels, particularly the use of a first main component entry strategy, but the empirical analysis also highlights some contrasting trends. Firstly, a higher number of Italian idioms are recorded (143/150) compared to the English sample (129/150), but they deviate more frequently from the stated listing policy. In total, the entry points for 98 of the 143 listed Italian idioms have a first or subsequent main component entry as detailed in the preface, which is a 68 percent level of adherence. Secondly, the breakdown of entry points in the Italian sample reveals 37 fixed idioms receive two or more listings, which is almost three times the number of English fixed idioms with the same level of coverage (14/129). This is a particularly interesting finding given that the preface of the SI (2006) states that it, for the most part, assigns only one entry per idiom. A recurring pattern across both samples is the disparate treatment of lexically variable verb and noun idioms. Similar to the English sample, half of the analysed variable verb idioms are entered under one or more of their variable components. By contrast, the treatment of those with variable noun constituents is less extensive with ten of the sample accorded a fixed component entry only. The entirety of these anomalies highlight that despite adopting a clear-cut listing strategy, the coverage of both English and Italian idioms in the SI (2006) is not without imperfections and inconsistencies.

Oxford Paravia Italian Dictionary (2006)

Listing patterns from both semantic and lexical categories show that Italian idioms are more comprehensively recorded in the OPID (2006) than English idioms. This emanates clearly from a comparative analysis of both samples, in which the majority of the selected Italian idioms receive two or more entries compared to the predominance of a single listing strategy in the English group. A similar imbalance is also apparent in the coverage of lexically variable idioms with a far higher application of a multiple entry strategy to Italian idioms (32/40) than English idioms (13/40). The utilisation of a multiple entry method in the Italian-English section is more in accordance with the stated idiom entry policy, but it has widespread ramifications for the dictionary. First and foremost, this distorts the overall idiom content of the OPID (2006) as Italian idioms receive a more extensive treatment by virtue of their greater number of listings. Furthermore, it is also advantageous to Italian users as they can potentially consult a higher number of entries to locate the same idiom. Such discrepancies suggest that the OPID (2006) organises and orientates its idiom content more to an Italian usership and that its stated method for arranging idioms is only truly representative of those listed in the Italian-English section.

Hazon Garzanti Inglese (2009)

Similar to its treatment of English idioms, the HGI's organisation of Italian idioms is also characterised by a lack of systematic order. The selection of entry points across the three semantic divisions shows little linearity and the only detectable pattern is the alternation

between a second main component and multiple entry strategy across all three semantic categories. A more general overview reveals the true inconsistency of the HGI's approach to listing Italian idioms; over half of the selected semi-idioms are listed once, the majority of pure idioms receive two entries, but the distribution of single and multiple listings is practically equal in the Italian figurative idiom sample. However, its arrangement of lexically variable Italian idioms follows a more predictable pattern than in the English sample. The majority of analysed expressions in each variable subgroup are accorded two or more listings, but these are not always contiguously ordered. For example, some are found under all variable components, others in all fixed element entries, whilst it also occasionally lists some idioms under their different syntactic parts. On the basis of the empirical evidence obtained from both English and Italian samples, the HGI (2009) offers the least structured coverage of idioms of all current bilingual Italian-English dictionaries.

5. Systematising the Recording of Idioms in Bilingual Italian-English Dictionaries

The two previous sections highlight the lack of consensus among theorists regarding the lemmatisation of idioms and the diverse approaches taken by current bilingual Italian-English dictionaries in dealing with this problem. In formulating the proposed lemmatisation model due consideration has been afforded to a number of relevant factors, such as the syntactic importance of the verb element in English phrases, the tendency of Italian dictionary users to opt for first main component entry when attempting to locate idioms (cf: Atkins and Varantola 1998), the body of empirical evidence obtained, but most importantly the intrinsic semantic and lexical features of idioms. Figure 1.3 presents a paradigmatic overview of the proposed lemmatisation model.

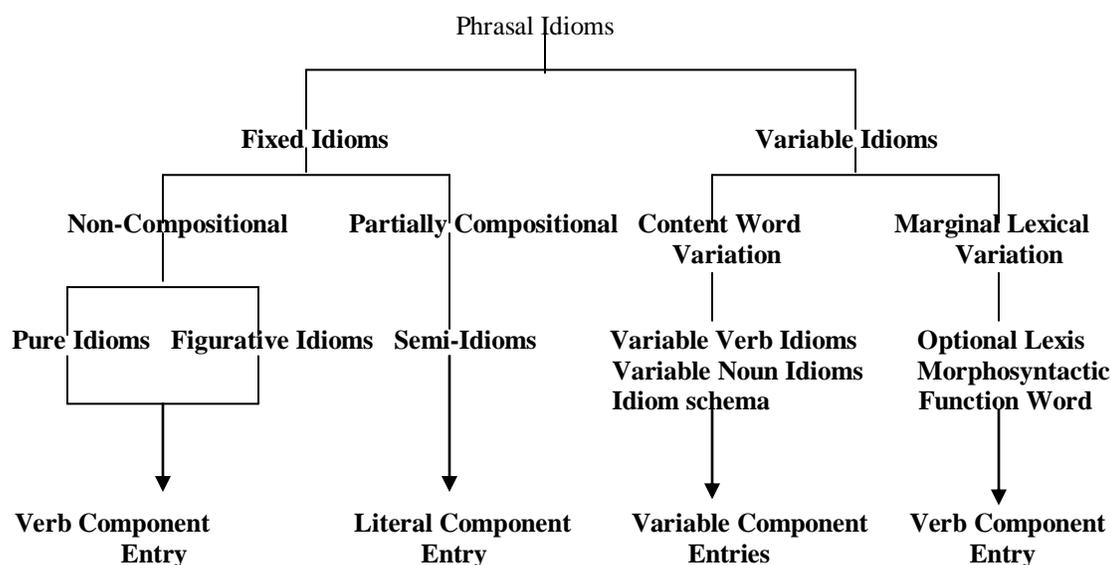


Figure 1.3. A Semantic and Lexical-Based Lemmatisation Model for Recording Idioms in Bilingual Italian-English Dictionaries

The stated entry points in the proposed lemmatisation model are chosen on the basis of the most salient semantic or lexical feature of that particular idiom group, but they are

also aligned to current practice in bilingual Italian-English lexicography, which sees most fixed idioms allocated a single listing and variable idioms assigned two or more. Fixed non-compositional expressions (pure idioms and figurative idioms) are recorded in their first main component entry. The rationale for this choice is supported by several factors; the semantic parity of each lexical part, the syntactic importance of the verb as well as the look-up tendencies of Italian dictionary users. A first main component entry is also the proposed listing procedure for partially compositional idioms as the verb element most often retains its literal meaning in semi-idioms and its inclusion at this point also highlights its equal semantic role as a stand-alone lexical item and idiom constituent. In a small number of semi-idioms, the literal component is either the noun or adjectival element; therefore, in order to maintain an internal organisational consistency a cross-reference is included under the verb component to redirect the user to the full listing at the literal element entry. Along with systematising the overall recording of idioms, the proposed model also aims to highlight their more marginal features. This, in particular, concerns idioms with non-words, which typically occupy the second main component slot. Therefore, idioms with this feature are to be recorded under their non-word lemma with a cross-reference inserted in their verb entry.

Empirical evidence clearly shows that the treatment of lexically variable idioms is unsystematic in current bilingual Italian-English dictionaries. Therefore, the proposed conceptual model offers a more structured listing method that also emphasises the occurrence of lexical variation. Idioms with variable content words are to be recorded in the each variable element entry accompanied by cross-references to guide the dictionary user to other listed lexical varieties of the same expression. Those with more marginal forms of lexical variation do not experience any major change to their lexical structure and as a result can be recorded in their verb component entry detailing the relevant type of variation. The proposed lemmatisation model offers four clear benefits to both compilers and users of bilingual Italian-English dictionaries:

1. It accurately reflects the semantic and lexical status of idioms.
2. It is based on rational and accepted linguistic criteria.
3. It eliminates the reliance on the subjective judgement of a lexicographer to choose a suitable point of entry and provides dictionary users with a consistent listing system.
4. It remains in line with, but is more systematic than, the current methods used in bilingual Italian-English dictionaries to record idioms.

Although the proposed model is discussed only in the context of bilingual Italian-English lexicography, it may have a more general application, particularly for bilingual dictionaries with an English dimension. Idioms can be subdivided and classified according to their semantic and lexical features; therefore this aspect of the analysis is universally applicable. However, the emphasis on the first main component entry for recording non-compositional and partially compositional idioms could possibly preclude a more widespread implementation of this model in bilingual lexicography. While it may be particularly suited to the syntactic structure of Romance languages, such as Italian, French, Spanish, etc, it may be incongruent with others like Danish, Dutch and German, which view the noun or adjective component as having a more important semantic and syntactic role in the expression.

6. Conclusion

Empirical evidence shows that current bilingual Italian-English dictionaries follow a variety of listing strategies for idioms, but these are both inconsistent and disinformative. This problem emanates largely from their ‘one size fits all’ approach to recording phraseology, despite one of these categories, idioms, having distinct semantic and lexical features. It is these characteristics, however, that could provide bilingual Italian-English lexicographers with a more legitimate and solid foundation to formulate a consistent and systematic entry strategy, which eradicates many of the current organisational problems associated with idioms and accurately reflects their salient semantic and lexical traits.

Bibliography

Dictionaries

- [GHDI] *Grande Hoepli Dizionario Inglese*. Milano: Ulrico Hoepli Editore SpA. 2007.
[HGI] *Hazon Garzanti Inglese*. [Nuova Edizione] Milano: Garzanti. 2009.
[OPID] *Oxford Paravia Italian Dictionary* (2006) [Second Edition] Paravia Bruno Mondadori Editori and Oxford University Press. 2006.
[SI] *Il Sansoni Inglese* [Quinta Edizione]. Milano: Edigeo. 2006.
[ZIR] *Il Ragazzini* [Quarta Edizione]. Bologna: Zanichelli SpA. 2008.

Other Literature

- Atkins, B.T.S.; Varantola, K. (1998). 'Language Learners Using Dictionaries: The Final Report on the EURALEX/AILA Research Project on Dictionary Use'. In B.T.S. Atkins (ed.) (1998). *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*. Tübingen: Max Niemeyer Verlag. 21-81.
- Atkins, B.T.S. and Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Botha, W. (1992). 'The Lemmatization of Expressions in Descriptive Dictionaries'. In H. Tommola; K. Varantola; T. Salmi-Tolonen; J. Schopp (eds.). *EURALEX '92 Proceedings I-II*. Tampere: Department of Translation Studies, University of Tampere. 493-502.
- Burger, H. (1989). 'Phraseologismen im allgemeinen einsprachigen Wörterbuch'. In F.J. Hausmann; F. Josef et al. (eds.). *Wörterbücher: Ein internationales Handbuch zur Lexikographie*. Volume I. Berlin/New York: De Gruyter. 593-599.
- Cowie, A.P.; Mackin, R.; McCaig, I.R. (1993). *Oxford Dictionary of English Idioms*. Oxford: Oxford University Press.
- Fellbaum, C. (ed., 2007). *Idioms and Collocations*. London: Continuum.
- Fernando, C.; Flavell, R. (1981). *On Idiom: Critical Views and Perspectives*. Exeter: University of Exeter.
- Fernando, C. (1996). *Idioms and Idiomaticity*. Oxford: Oxford University Press.
- Fontenelle, T. (1998). 'Discovering Significant Lexical Functions in Dictionary Entries'. In A.P. Cowie (ed., 1998). *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press. 189-207.
- Fraser, B. (1970). 'Idioms within a transformational grammar'. In *Foundations of Language* 6. 22-42.
- Gouws, R. H. (1991). 'Toward a Lexicon-based Lexicography'. In *Dictionaries: Journal of Dictionary Society of North America* 13. 75-90.
- Harras, G.; Proost, K. (2005). 'The Lemmatisation of Idioms'. In H. Gottlieb; J.E. Mogensen; A. Zettersen (2005). *Symposium on Lexicography XI, Proceedings of the Eleventh International Symposium on Lexicography*. Tübingen: Max Niemeyer. 277-291.
- Hartmann, R.R.K.; James, G. (1998). *Dictionary of Lexicography*. London: Routledge.
- Jackson H.; Zé Amvela, E. (2007). *Words, Meaning and Vocabulary: An Introduction to Modern English Lexicology*. London: Continuum.
- Mel'čuk, I. (1995). 'Phrasemes in Language and Phraseology in Linguistics'. In M. Everaert; E.J. van der Linden; A. Schenk; R. Schreuder (eds., 1995). *Idioms Structural and Psychological Perspectives*. New Jersey: Lawrence Erlbaum Associates Publishers. 167-232.
- Moon, R. (1998). *Fixed Expressions in English: A Corpus-Based Approach*. New York: Oxford University Press.
- Petermann, J. (1983). 'Zur Erstellung ein - und zweisprachiger phraseologischer Wörterbücher: Prinzipien der formalen Gestaltung und der Einordnung von Phrasemen'. In J. Matesic (ed.). *Phraseologie und ihre Aufgaben*. Heidelberg: Groos. 172-191.
- Pinker, S. (1999). *Words and Rules The Ingredients of Language*. New York: Harper Collins.

Section 9. Lexicological Issues of Lexicographical Relevance

Tomaszczyk, J. (1986). 'The Bilingual Dictionary under Review'. In M. Snell-Hornby (ed., 1986). *ZüriLEX 1986 Proceedings Euralex International Congress*. Zurich: University of Zurich. 289-297.