

***Prendere il toro per le corna o lasciare una bocca amara?* – The Treatment of Tripartite Italian Idioms in Monolingual Italian and Bilingual Italian-English Dictionaries**

Chris Mulhall
Waterford Institute of Technology
cmulhall@wit.ie

Abstract

This paper undertakes an empirical investigation into the treatment of tripartite Italian idioms in selected monolingual Italian and bilingual Italian-English dictionaries. Tripartite idioms are phrasal constructs typically arranged into one of the three following syntactic forms: V+N+N, V+ADJ+N or V+N+ADJ. From an organisational viewpoint, these idioms are somewhat more problematic for lexicographers due to the presence of a third lexical constituent. Current lexicographical practice adopts a largely subjective approach to dealing with idioms, which for the most part are V+N forms, therefore those with a wider syntactic form require more considered decision making when determining their point(s) of entry in a dictionary. Certain tripartite idioms also collapse into binomial (N+N) or nominal-adjectival forms (ADJ+N/N+ADJ), thus placing a question mark over the necessity to record their verb element or not. Together, these issues contribute to making tripartite idioms one of the most acutely difficult phrasal categories for lexicographers. This paper examines the entry points of 100 tripartite Italian idioms in selected monolingual Italian and bilingual Italian-English dictionaries published within the last decade. Firstly, it discusses relevant theoretical viewpoints relating to the lemmatisation of idioms over the last 30 years. Thereafter, it focuses on the notion of an idiom as a lexicographical entry in consideration of its varied intrinsic features and how these are productive in selecting an appropriate entry point. The presentation of the empirical data in the following section is stratified according to the publishing house of the analysed dictionaries and gives a comparative discussion on their respective approaches to dealing with tripartite idioms. Finally, the last section unifies the theoretical arguments and practical approaches and proposes a theoretical framework model for lemmatising tripartite idioms to offer a more coherent and consistent platform for their organisation in monolingual Italian and bilingual Italian-English dictionaries.

Keywords: Italian Lexicography; Tripartite Idioms; Lemmatisation

1 The Lemmatisation of Idioms: Theoretical Viewpoints

Lexicographical theorising over the last 30 years has regularly put forward proposals to remedy the problematic issue of how best to position idioms within a dictionary text. In their entirety, these arguments aim to achieve a more considered and structured lexicographical coverage of idioms, but fail to reach a consensus on the most suitable approach. Certain theorists advocate embedding idioms within the microstructure but diverge significantly on the number and position of their assigned entry points, for example, there is little theoretical agreement on a single listing strategy: Petermann (1983) (notional point of entry); Burger (1989) (unspecified entry point); Lorentzen (1996) (noun entry point); Mulhall (2010) (lexico-semantic entry point). Contrastingly, Tomaszczyk (1986) suggests entering idioms under each of their constituent's lemmas. The unitary semantic function of idioms equates them to having a word-like function in the lexicon, which Al-Kasimi (1977), Gouws (1991) and Botha (1992) argue is a substantive rationale for their lemmatisation in a dictionary. This, as Gouws (1991:86) states, prevents an 'ambiguous reading' of an idiom's lexical status by dictionary users. Lemmatising idioms accurately portrays their semantic status in the lexicon, but its lexicographical practicality remains questionable and untried. Harras and Proost (2005) advance a bespoke entry model for idioms, configuring their point of entry in accordance with their semantic features; resulting in semantically opaque idioms being lemmatised and semantically interpretable idioms sub-lemmatised. Adopting this particular method would move dictionaries to a more semantic-based lemmatisation model for idioms but is potentially anomalous given its proposal of different organisational principles for the same category of phrases.

An overview of the proposed entry methods reveals that semantics motivates many decisions relating to the most appropriate entry point with Mulhall (2010) taking into account that any lemmatisation model must also incorporate the notion of lexical variability, which, according to Moon (1998) can occur in between 12 to 40 percent of idioms. Another characteristic feature of idioms that is less topical in lexicographical debates is that of syntactic form. Therefore, considering the different semantic, lexical and syntactic properties of idioms may offer a more robust decision-making platform for identifying their most suitable entry point in a dictionary.

The majority of Italian idioms fall into the standard syntactic category of Verb (V) + Noun (N) with potential syntactic expansion to V/V+N or V+N/N if lexical variation is permissible in either the verb or noun component. In such cases, a lexicographer only has two (or possibly three) available point of entry options. A more problematic subset is that of tripartite idioms, some of which may have two constituents of the same word class (V+N+N) or contain three distinct word classes with two different syntactic structures (V+ADJ+N or V+N+ADJ). Tripartite idioms are particularly challenging for lexicographers on a number of levels. Firstly, the subjective identification of the most important or prominent element becomes more complicated due to the presence of a third lexical component. In the case of V+N+N idioms, lexicographers favouring a noun-based listing model must, from the outset, decide whether the first or second noun element is the most appropriate entry point. Secondly, the

V+ADJ+N and V+N+ADJ categories contain a small number of expressions that retain their idiomatic identity in a nominalised form, for example, *a gonfie vele*, *duro d'orecchio*, *il nodo Gordiano*, *la pecora nera*, etc. Important issues arise from these syntactically truncated forms; such as the necessity to record their verb element(s) or not and the importance of the adjectival element in their syntactic binding and lexical identity. The entirety of these issues and the failure of lexicographical practitioners to adopt and integrate recent theoretical suggestions or propose alternative practical solutions contribute to the long-term status of idioms as arguably the most problematic dictionary entry.

2 Redefining Idioms for Lexicography

Theoretical linguistics offers a multitude of rich and varied definitions for an idiom, but these typically comprise singular, one dimensional features, such as 'fixed expression' or 'non-compositional' or describe it through vague terms of references, such as 'relatively fixed expression'. Idioms, by their nature, are a linguistic concept; therefore such definitions may not offer the requisite scope to give lexicographers a wider understanding of their form and behaviour to deal with them accurately in the context of a dictionary. Therefore, to ensure a more representative lexicographical treatment of idioms, it is important to factor in their three most salient characteristics; semantics, lexis and syntax. In consideration of these characteristics, Table 1 sets out three feature-specific maxims to reconsider idioms as both a lexicographic entry and linguistic unit.

Feature	Definition
Semantics	Idioms are a semantically complex and compositionally divergent subset of expressions. This often results in a clear semantic disconnect between the lemma as a stand-alone lexical unit and as an idiom constituent.
Lexis	Idioms display different layers of lexical fixity. Their potential variability necessitates a dictionary entry strategy that not only recognises variable expressions but also records them in a consistent and representative way.
Syntax	Idioms are syntactically heterogeneous. Therefore, the number and word class of idiom constituents within any given idiomatic frame may potentially influences the number and position of allocated entry points in a dictionary.

Table 1: Lexicographical Definitions of Idiom Features.

Redefining the notion of an idiom as a linguistic unit and a dictionary entry, the following operational definition is proposed to achieve a more holistic and tailored lexicographic treatment of idioms based on their inherent features:

Idioms are a category of multi-lexical, syntactically diverse expressions showing various degrees of semantic compositionality, some of which contain lexical constituents that can substituted for idiomatic equivalents.

This broader definition encapsulates the most salient characteristics of idioms; accurately portraying their status in the lexicon and describing features that are potentially influential in their lexicographical description and organisation.

3 Organisational Approaches to Idioms in Monolingual Italian and Bilingual Italian-English Dictionaries

An analysis of monolingual Italian and bilingual Italian-English dictionaries from the eighteenth century onwards shows that idioms, as a lexicographic entry, failed to gain a centrality and an organisational foothold in the design and content of these reference works. This, in part, can be traced to their perceived linguistic impurity in eighteenth and nineteenth century Italian society, which resulted in their limited coverage in mainstream dictionaries. A change in this outlook came in the early twentieth century due to new linguistic models, burgeoning dictionary content and a reformatted microstructure. But in many dictionaries idioms still remained a peripheral entity; a notable exception to this was the *Sansoni-Harrap Standard Italian-English Dictionary* [SHSIED] (1970-1975), which made the singular attempt of this era to systematise the coverage of idioms, but its subjective,¹ rather than substantive, criteria failed to address this problem for dictionary users. Decisions about restructuring the organisation of idioms and providing this information to users remain relevant, but overlooked, in modern day lexicographical practice. This paper aims to provide further evidence identifying the need for a lexicographical reform, at least in an Italian context, in the approach to organising idioms. The research sample includes 100 tripartite Italian idioms; subdivided into the following syntactic categories: 50 V+N+N, 30 V+N+ADJ and 20 V+ADJ+N. Selecting and organising the empirical sample brought to light two recurring trends: the prominence of the V+N+N syntactic structures within the Italian tripartite subset and the presence of a high frequency verb² (HFV) element in 39 of 100 expressions. To gain a comparative insight into any converging or diverging entry strategies based on the expectations of Italian speaking users a monolingual Italian and bilingual Italian-English dictionary from three publishing houses formed part of the research corpus. Selecting different dictionaries from the same publishing houses allowed an interesting exploration into ascertaining whether or not certain publishing houses follow any systematic procedure when attempting to record tripartite idioms in their monolingual and bilingual reference works. The dictionaries used are as follows: *Il Sansoni Inglese* [ISI] (2006); *Il Sabatini-Coletti* [ISC] (2007); *Il Ragazzini* [ZIR] (2009); *Lo Zingarelli* [ZLZ] (2009); *Hoepli Dizionario Inglese* [HDIN] (2007) and the *Hoepli Dizionario Italiano* [HDIT] (2008).

1 “The phrases, idiomatic expressions, proverbs, etc., that make up the phrase section are generally found under the first important word in the phrase” (SHSIED 1970: viii).

2 The following Italian verbs occur with a high frequency across a number of different phrasal and idiomatic expressions: *andare, avere, dare, essere, fare, mettere, prendere, stare, tenere, venire*.

Dictionary Publishing House	Sansoni		Zanichelli		Hoepli	
	Il Sansoni Inglese	Il Sabatini-Coletti	Il Ragazzini	Lo Zingarelli	Hoepli Inglese	Hoepli Italiano
V+N+N (N=50)						
Verb Entry	17	2	4	2	5	2
N1 Entry/N2 Entry	13/0	8/2	15/3	9/3	19/8	7/3
Double Noun Entry	0	12	7	11	2	13
Verb/Noun Combination Entries	12	19	14	20	4	20
Listed as a Nominalised Idiom	2	2	3	3	5	2
Showing a different Verb Element	1	3	2	0	2	2
Not Listed	5	2	2	2	5	1
V+ N+ADJ (N=30)						
Verb Entry	2	1	0	0	4	1
Noun Entry	15	1	0	7	11	5
Adjective Entry	0	5	11	0	3	1
Verb/Noun/Adjective Combination Entries	6	16	11	14	5	14
Listed as a Nominalised Idiom	1	3	7	4	2	6
Showing a different Verb Element	4	2	0	0	1	1
Not Listed	2	2	1	5	4	2
V+ADJ+N (N=20)						
Verb Entry	3	0	3	0	2	0
Adjective Entry	9	1	7	1	0	1
Noun Entry	0	4	0	5	9	3
Verb/Noun/Adjective Combination Entries	3	8	6	6	4	9
Listed as a Nominalised Idiom	1	0	1	2	0	1
Showing a different Verb Element	1	1	0	0	1	2
Not Listed	3	6	3	6	4	4

Table 2: Empirical Data on the Entry Points of Italian Tripartite Idioms in Monolingual Italian and Bilingual Italian-English Dictionaries.

3.1 Il Sansoni Inglese (2006) and Il Sabatini-Coletti (2007)

An often lamented failing of dictionaries is their failure to provide any guidance to users about the exact location of idioms. A notable exception in this regard is ISI (2006), which in the preface clearly informs users where to locate such expressions with accompanying examples. In contrast, its monolingual equivalent, the ISC (2007), takes a different approach, instead exemplifying certain idioms without explicitly indicating their position in the dictionary (see Table 3).

Il Sansoni Inglese (2006)	Il Sabatini-Coletti (2007)
<p>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 <i>he who pays the piper calls the tune</i> is found under the verb pay and the phrase <i>as hard as iron</i> is listed under the adjective hard. Likewise, the Italian proverb <i>le bugie hanno le gambe corte</i> is given under bugia and the phrase <i>cavalieri della Tavola Rotonda</i> under cavaliere.</p> <p>As an exception to this, certain extremely common verbs (<i>be, can, come, do, get, give, go, have, keep, let, make, must, put, take, will</i> in English and <i>andare, avere, dare, dovere, essere, fare, lasciare, mettere, potere, prendere, stare, tenere, venire, volere</i> in Italian) have been ignored in listing the phrases under the headwords. As a result, the phrase, <i>to get one's cards</i> is given under the headword card, and <i>prendere qcu. in castagna</i> is given under castagna.</p> <p>(<i>Il Sansoni Inglese</i> 2006:14)</p>	<p>All'interno delle parole piene che lo sviluppano, ma con grande evidenza, sono tratte anche le unità polirematiche grammaticali, ossia le locuzioni che hanno valore di preposizione, di congiunzione o di congiunzione testuale (<i>a conti fatti, a costo di, modo che, nella misura in cui...</i>).</p> <p>Ben diverso è il caso delle espressioni idiomatiche, tutte di senso figurato, che appartengono alla lingua comune e sono in genere ben familiari ai parlanti (<i>essere un pozzo di scienza; dare carta bianca; andare per le lunghe; tendere la mano; voltare pagina; cambiare registro</i>). Come appare evidente, queste fanno nesso fisso con un verbo).</p> <p>(<i>Il Sabatini-Coletti</i> 2007:16)</p>

Table 3: Organisational Criteria for Idioms in Sansoni Publishing House Dictionaries.

Like the diversity of their organisational approaches to these entries, the empirical analysis also reveals disparities in the treatment of the same syntactic idiomatic categories in the ISI (2006) and the ISC (2007). On a general level, this divergence can be measured through the number of assigned entry points; for example, the ISI (2006) allocates a single entry to the 61/100 expressions in contrast to a multiple listing strategy favoured by the ISC (2007) for 64/100 expressions. A possible explanation for this different approach is the strict adherence by the ISI (2006) compilation team to inserting idioms under the first key word, but the application of this method is not entirely rigid. For example, data from the empirical sample reveal that 21/100 expressions are listed twice or more and 54/100 expressions are recorded directly in line with the information given in the preface. From an Italian speaking user perspective, locating tripartite idioms with a high frequency verb may prove more labourious in the ISI (2006) than in its monolingual equivalent, thus requiring the dictionary user to engage in the subjective assessment of whether the noun or adjectival element can be considered as the first key word.

3.2 Il Ragazzini (2009) and Lo Zanichelli (2009)

Listing patterns for tripartite idioms found in the two dictionaries from the Zanichelli publishing house reveal a largely unstructured arrangement, a problem exacerbated by the lack of any information detailing their location. This omission is problematic for users, but is, to a certain degree, offset by the multiplicity of idiom listings in both the ZIR (2009) and the ZLZ (2009). This pattern is apparent across all analysed tripartite syntactic groups in the monolingual version, in particular the V+N+N category with 33/50 expressions recorded twice or more. Similar patterns emerge in the ZIR (2009), but on a lesser scale, with a considerable number of V+N+N (23/50) and V+N+ADJ (15/30) expressions listed twice or more. Another overlapping feature of both empirical samples is the comparably higher number of tripartite idioms recorded in nominalised forms in Zanichelli dictionaries; accounting for 11/100 in the ZIR (2009) and 9/100 in the ZLZ (2009). The removal of the verb element is an inherent feature of certain tripartite Italian idioms, but it negates the objective of a dictionary, which is to record lexical items in their fullest and most descriptive forms. Furthermore, the ZLZ (2009) contains the equal lowest coverage of the analysed expressions with 13/100 not recorded.

3.3 Hoepli Inglese (2007) and Hoepli Italiano (2008)

Both Hoepli-published dictionaries follow individually contrasting, but somewhat consistent, approaches in their treatment of tripartite Italian tripartite idioms. The use of a noun-based entry strategy emerges clearly from the analysis of the three syntactic categories in the HDIN (2007), but its consistency is diluted by scattered recording of similar expressions under alternative entries. A noun entry strategy features most prominently in the V+N+N group (27/50), but the division of this into 19/27 under N1 and 8/27 under N2 is a microcosm of the internally inconsistent approach to their general recording. Keeping a consistency with the other monolingual dictionaries, the HDIT (2008) favours a multi-entry strategy for tripartite idioms, in particular V+N+ADJ (14/20) and V+ADJ+N (9/20) forms, but a similar incongruity to that found in the V+N+N group in the HDIN (2007) resurfaces in the HDIT (2008). In this case, 20/50 binomial idioms are found under a verb and noun element(s) with 13/50 inserted under both noun components. The rationale for recording under both noun elements may be explained by the presence of a HFV element, but this does not extend to the entire subset with *cercare un ago nel pagliaio*, *dire peste e corna di qualcuno*, *finire in una bolla di sapone* and *tirare sassi in piccionaia* not inserted in their verb entries.

4 Conclusion

Idioms appear to remain on the periphery of lexicographical importance, at least in an Italian context. Empirical data from monolingual Italian and bilingual Italian-English dictionaries reveal an in-

consistent, unscientific approach to the coverage of tripartite idioms. Generally, idioms tend to be defined by their problematic status rather than their linguistic uniqueness, thus veiling their rich lexical, semantic and syntactic features. This requires a more analytical look at how these characteristics can be influential in systematising the overall accessibility of idioms for dictionary users as well as resolving a perpetual practical difficulty. Figure 1 presents an alternative entry model for tripartite Italian idioms, whether as a verb phrase (VP) or a nominalised form, on the basis of their syntactic composition.

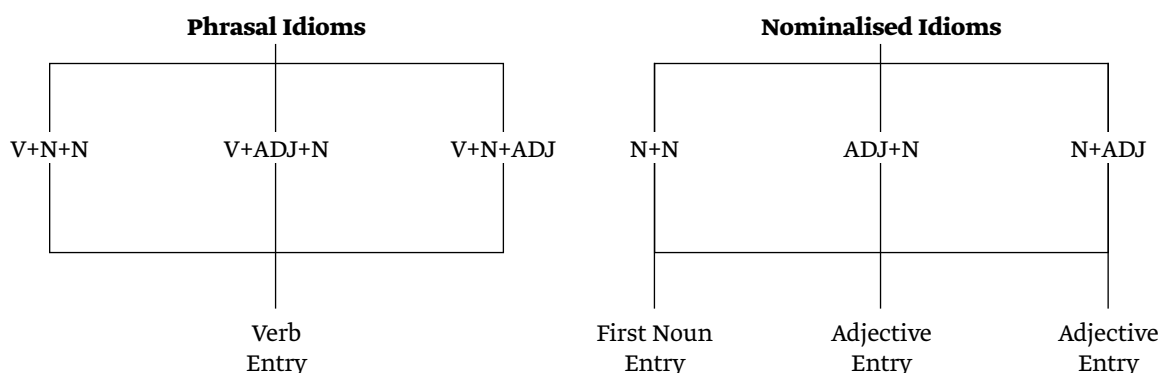


Figure 1: A Theoretical Framework for the Lexicographical Treatment of Tripartite Italian Idioms.

A universal verb entry strategy for tripartite idioms with HFV elements contrasts with current lexicographical practice as HFV entries are overpopulated and thus are considered too long for including such expressions. Tripartite idioms are generally VP structures, thus giving the verb element an elevated syntactic importance and identity for dictionary users. Its syntactic position at the head of the expression also increases its probability as a likely point of consultation for users. Nominalised forms of tripartite idioms present a greater organisational challenge, which unlike their VP equivalents, requires a more tailored entry model with due consideration afforded to syntactic order and word class. Therefore, recording N+N structures in their N1 entry and both ADJ+N and N+ADJ types in their adjective entries. These choices are predicated on the following pertinent criteria; the N1 element assumes the role of the syntactic head in binomial idioms, whereas the idiomaticity of those containing adjective and noun element is preserved by the retention of the adjective, compare, for example, the disparate meanings of *giocare a carte* and *giocare a carte scoperte* due to the presence of the idiomatically-inducing adjectival element *scoperte*. In conclusion, the multifaceted nature of idioms is complex, but also provides a substantive platform for choosing their most appropriate point of entry in a dictionary. The current systemic failure of dictionaries to address this issue reinforces the notion of idioms being subservient to words in both the lexicon and lexicography. Therefore, understanding and reprioritising the notion of an idiom and its associated features is an important objective for lexicographical practice in the twenty-first century.

5 References

Monolingual Italian Dictionaries

- [ISC] *Il Sansoni Coletti Dizionario della Lingua Italiana* (2007). Milano: RCS Libri S.p.A.
[ZLZ] *Lo Zingarelli* (2009). Bologna: Zanichelli.
[HDIT] *Grande Dizionario Hoepli Italiano* (2008). Milano: Ulrico Hoepli Editore S.p.A.
Bilingual Italian-English Dictionaries
[ISI] *Il Sansoni Inglese* (2006). Milano: Edigeo.
[ZIR] *Il Ragazzini* (2009). Bologna: Zanichelli.
[HDIN] *Grande Dizionario Hoepli Inglese* (2007). Milano: Ulrico Hoepli Editore S.p.A.
[SHSIED] *Sansoni-Harrap Standard Italian and English Dictionary* by Vladimiro Macchi, Four Volume, Firenze-Roma: Sansoni Editore, 1970-1975.

Other Publications

- Al-Kasimi, A. (1977). *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill.
Botha, W. (1992). The Lemmatization of Expressions in Descriptive Dictionaries, in H. Tommola, K. Varantola, T. Salmi-Tolonen and J. Schopp (eds.). *EURALEX '92 Proceedings I-II*. Euralex International Congress, Tampere, August 4-9, 1992. Department of Translation Studies, University of Tampere. pp. 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. pp. 593-599.
Gouws, R. H. (1991). Toward a Lexicon-based Lexicography. *Dictionaries: Journal of Dictionary Society of North America*, **13**. pp. 75-90.
Harras, G. and Proost, K. (2005). The Lemmatisation of Idioms in H. Gottlieb, J.E. Mogensen and A. Zettersten (2005) (eds). *Symposium on Lexicography XI, Proceedings of the Eleventh International Symposium on Lexicography*. Copenhagen, May 2-4 2002 Tübingen: Max Niemeyer. pp. 277-291.
Lorentzen, H. (1996). Lemmatization of Multi-word Lexical Units: In which Entry? in M. Gellerstam, J. Järborg, S.G. Malmgren, K. Norén, L. Rogström and C.R. Pappmehl (eds.), *EURALEX '96 Proceedings I-II*. Seventh Euralex International Conference, Göteborg, August 13-18, 1996, Department of Swedish, Göteborg University. pp. 415-421.
Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus Based Approach*. Oxford: Clarendon Press.
Mulhall, C. (2010). A Semantic and Lexical-Based Approach to the Lemmatisation of Idioms in Bilingual Italian-English Dictionaries in A. Dykstra and T. Schoonheim (eds.) (2010) *Proceedings of the XIV Euralex International Congress*. pp. 1355-1371.
Tomaszczyk, J. (1986). The Bilingual Dictionary under Review in M. Snell-Hornby (1986) (ed.) *ZüriLEX 1986 Proceedings, Euralex International Congress*. Zurich, 9-14 September, 1986 University of Zurich. pp. 289-297.

