

A Corpus Study of Italian Proverbs: implications for lexicographical description

Laura CIGNONI and Stephen COFFEY, Pisa, Italy

Abstract

In this paper we look at the way in which proverbs were found to behave in a corpus of contemporary written Italian. We comment especially on the types of formal variation they were found to undergo, whether in terms of alternative citation forms or context related changes. We also comment on the more significant relationships between variation and textual function. Our findings would suggest that proverbs are far from being invariable units, and that lexicographical description of proverbs could be more complete than it usually is, especially as regards formal variation and syntactic flexibility. We acknowledge the need to compare our findings with data regarding proverb usage in the spoken language, and also the size limitations of the corpus used in the present study.

1 Introduction

Although much has been written about proverbs, from historical, cultural, cognitive and linguistic points of view, they are not dealt with very fully in lexicographical description. In lexicographical works of a general nature, they are sometimes merely listed under a related head word, and at best are given a simple definition. In more specialized works, they may also be described from a historical and social viewpoint.

In this paper we describe some results of a corpus study of Italian proverbs¹, noting especially those points which would seem to be of relevance to lexicographical description.

1.1 The linguistic units discussed in this paper

The present study emerges from a wider study of Italian phraseological units. Some of the latter have been assigned the status of 'proverb' by virtue of their having certain inherent characteristics. These are: (i) they constitute complete units of meaning, syntax and intonation, rather than being lexical units or clauses, (ii) within a given culture, past or present, they are either metaphorical statements about the human condition, or statements about the physical world around us, or statements about how people live together, or recommendations about how one should live, either within oneself or in relation to others, (iii) they are associated with the acquired wisdom of society in general rather than being the thoughts or opinions of a particular person.

Many proverbs display a number of other characteristics, notably metaphor, euphonic features such as alliteration or rhyme, and structural and semantic parallelism. These, however, are tendencies rather than defining features. For overviews of the discussion surrounding the nature of proverbs, see [Arnaud (1991)] and [Mieder (1989), pp.13-27].

1.2 The corpus and search methodology

The corpus upon which we have based our study is the Italian Reference Corpus (IRC) located at the Institute of Computational Linguistics in Pisa. The IRC is a corpus of written language which consisted of approximately 16,000,000 words of text at the time of consultation. Magazines and other periodicals accounted for over 40% of the total and newspapers about 28%, thus making journalism the predominant text type. Fictional works made up almost 11% of the corpus, and books of non-fiction about 12.5%. Finally, there were also a number of technical research reports (7.2%). A major limitation of our study is thus immediately evident: whereas proverbs are intuitively associated with the oral tradition, the corpus we consulted was of written language. Despite this limitation, we feel, nevertheless, that our findings justify the partial nature of the study.

Since there was no fully automatic way of looking for proverbs, we first drew up a list of items from various lexicographical sources, and then proceeded to interrogate the corpus. For each proverb we looked for one or more key words. We learnt from experience that it was counterproductive to look for the co-occurrence of *all* key words since proverbs often occurred in variant forms. For example, in the case of the proverb *Il fine giustifica i mezzi* [The end justifies the means] we carried out a search for the co-occurrence of any two out of the three key words. The 'word family' thus created was *fine + giustifica* OR *fine + mezzi* OR *mezzi + giustifica*. We did not specify the order in which the words within each pair should occur, and we allowed for as many as ten intervening words. Search flexibility of this type allowed us to retrieve many variations, including, from this particular search, *Il fine giustifica i media?* [Does the end justify the media?]².

2 Corpus findings

In all, a total of 248 proverbs were looked for. Of these, 120 (48.4%) were not found in the corpus at all, and for the 128 which were present in the corpus, frequency was generally low. Only 21 proverbs were found more than 5 times, and only 5 more than 10 times. The resulting data, therefore, tells us little about the typical behaviour of individual proverbs. It does, however, allow us to begin building up a picture of the overall way in which Italian proverbs behave in the types of written language being studied. The total number of proverb tokens found was 439.

2.1 The formal variation and flexibility of proverbs

Some proverb definitions include a specific comment underlining the fixedness of the linguistic unit in question. This aspect of proverbs, however, would not seem to apply to the type of written texts which we have been examining. In this section we describe the main ways in which proverbs were found to be subject to formal variation. We distinguish between variation which we refer to as 'proverb-inherent', and variation which is clearly connected with or dependent on the surrounding text.

2.1.1 Proverb-inherent variation

Corpus evidence confirmed two main types of proverb-inherent variation. Firstly, we found cases of alternative canonical forms of a proverb, typically involving slight variations in lexis

or syntax. An example is the proverb *Non è tutt'oro quel che luccica / riluce* [All that glitters / glistens is not gold]. Secondly, a number of proverbs were found in shortened form, for example *Chi di spada ferisce ...* [He who lives by the sword ...] and *Rosso di sera ...* [Red sky in the evening....]. Truncation of this sort usually involved elimination of the second part of proverbs divisible into two parts.

2.1.2 Variation and flexibility directly relatable to context

Every instance of a proverb, excepting metalinguistic ones, will be related in some way, semantically or pragmatically, to its context. The proverb itself, however, may undergo no formal change and remain, from a formal point of view, as a stand-alone unit. An example from the corpus is *In passato si diceva: "Moglie e buoi dei paesi tuoi"* [In the past people used to say 'Wives and oxen from your own village'], where the proverb remains completely intact within quotation marks.

In many cases, however, it was found that the relationship between the proverb and its context entailed a change in the form of the proverb. We will distinguish here between two major categories of change, which depend on the presumed intentions of the writer and the corresponding impact made on the reader. The first is that of slight adaptations of a syntactic, grammatical or semantic nature, where the writer has probably unconsciously adjusted the form of the proverb and the reader may remain unaware of this fact. The second is that of significant lexico-semantic changes in which it is clear that the writer has consciously changed the proverb to create a particular effect.

The first of these two categories includes, but is not limited to, the following features: (i) the proverb becomes part of a dependent clause, for example, *Dunque, tutte queste sono pure coincidenze o vistose prove che ci inducono a pensare che non c'è mai nulla di nuovo sotto il sole?* [So, are all these mere coincidences or are they clear indications that there is never anything new under the sun?]; (ii) the proverb is interrupted by a parenthesis of some sort, for example *La fortuna, si sa, aiuta gli audaci* [Fortune - as is well known - favours the brave]; (iii) there is a change in tense or mood, for example, *... nella convinzione che non vi fosse nulla di nuovo sotto il sole* [... convinced that there was nothing new under the sun]; (iv) a modifier is added or changed, for example, *piangere un po' sul latte versato* [to cry a little over spilt milk].

The second category involves the exploitation of the proverb through radical lexico-semantic change, often to create a humorous effect. The most common way of achieving this was to replace a word inherent in the proverb with another word relevant to the context in which it was being used. An example from the corpus is the proverb *Finché c'è vita c'è speranza* [While there's life there's hope], which is found as *Finché c'è televisione c'è speranza* [Where there's television there's hope]. The amount and type of change which may take place within a given proverb depends on a number of factors. Structural, syntactic, lexical, semantic and euphonic elements within the proverb itself may all have a role in determining how much change is permissible; the only definitive limitation on exploitation is that the original proverb must still remain instantly recognizable. In the following example, not one but two key words have been replaced with words relating to the content of the passage. The original proverb, *L'erba del vicino è sempre più verde* [literally 'Your neighbour's grass is always greener'], is changed to *La moda del vicino è sempre più chic* [Your neighbour's fashion is always chicer]. Sometimes

one half of the proverb was found to be changed. Thus, *Dimmi con chi vai, e ti dirò chi sei* [literally 'Tell me who you go round with and I'll tell you who you are'] is changed to *Fammi vedere il tuo décor, ti dirò chi sei* [Show me your décor and I'll tell you who you are].

The reason why we chose to search the corpus so thoroughly should be clear from the above examples, and hopefully we were able to track down the vast majority of occurrences. It is to be noted that exploitation of this type accounted for no less than 103 of the 439 tokens found.³

2.2 Text type and function

Proverbs were found above all in fiction, newspapers and magazines, with the highest proportion being in the latter. Although it is beyond the scope of the present paper to go into very great detail regarding the textual function of proverbs, we will, however, outline the most salient features. A first observation is that a considerable number of tokens (73) are instances of proverbs functioning as names or titles of some sort. Examples are the book title *Volere è potere* [Where there's a will there's a way] and the film title *Dove c'è guerra c'è speranza* [Where there's war there's hope]. As might be expected, these were found mainly in journalism, though a few were also present in works of non-fiction.

A second observation is that within journalism, where over 80% of the tokens were found, proverbs typically appeared at key points in the text. There was a higher proportion than normal text distribution would predict in graphically evident positions, notably headlines and internal section headings. Proverbs also featured prominently in the summaries following headlines, and in introductory and concluding sentences. These key positions in the articles accounted for 28% of all proverb occurrences in journalism.⁴

2.2.1 Relationships between form, text and function

In order to see whether variation dependent on context (see 2.1.2) was in any way correlatable with text type, we drew up a list of the most commonly found types of variation and compared their expected frequency per text type with their actual frequency. We excluded from our calculations those proverbs which functioned as proper names since, except in one isolated case, any variation present was an intrinsic part of the proper name itself and not in any way context dependent. The features taken into consideration were (i) radical exploitation through lexical substitution, (ii) the addition or change of a modifier, (iii) presence within a dependent clause, (iv) change of tense or mood, and (v) discontinuity. Two of these categories showed up notable differences in the expected number of features per text type. These were radical exploitation and the addition or change of a modifier.

In the case of the former, it was found that magazines contained 77% of all occurrences of radical exploitation, as opposed to the 61% which average distribution would have predicted. Works of fiction had notably fewer examples than would have been expected (3.6% as opposed to 13.1%). With regard to the addition or change of a modifier, newspapers contained more examples than expected (32.3% as opposed to 22.9%) and fiction had fewer (3.4% as opposed to 13.1%).

Since exploitation was the most frequent form of change and was especially associated with magazines, we also examined the relationship between exploitation and its textual function

within magazine articles. It was found to be particularly associated with the types of textually prominent positions described above (see 2.2), 43 out of the 78 occurrences in magazines (55%) occurring in such positions.

3 Discussion and implications for lexicography

From a lexicographical point of view, proverbs are usually considered to be very simple items. Intuitively, they are typically fixed in form, and, being complete units of meaning, slot directly into discourse without any syntactic or other grammatical changes. Our own study suggests that this is not the case, at least in the written language. The number of occurrences of proverbs cited in their basic form, without any variation whatsoever, and not functioning as dependent clauses but as complete sentences, was the relatively low figure of 111 (25%).

We now make some suggestions as to how proverbs could be dealt with more completely in lexicographical description. The ideas will be more readily incorporated in dictionaries of proverbs, as well as electronic dictionaries, where space should not be a problem. We have identified five specific points which we feel are worth considering:

- (i) more attention should be paid to stating accurately the citation form of proverbs. Whereas important lexical variants are sometimes recorded, corpus analysis shows up other less obvious alternations. An example is *Ride bene chi ride ultimo / l'ultimo* [He who laughs last laughs longest], with the optional presence of the definite article.
- (ii) it would be useful to include contextualized examples of proverb usage in order to illustrate the many different types of context-dependent variation to which they are subject and any eventual relationships with text type. Contextualization would also show up ad hoc changes to the citation form, including truncation.
- (iii) where specific proverbs frequently undergo exploitation, this fact should be specifically stated as well as being exemplified. Such would be the case, for example, of the already cited proverb *Il fine giustifica i mezzi* [The end justifies the means]; 4 out of the 9 corpus occurrences involved radical lexical exploitation.
- (iv) some proverbs were found to vary in a very systematic way, and this should be commented on. *Vedere per credere* [Seeing is believing] is a case in hand. Out of a total of 24 tokens, the citation form occurs only 3 times. In the remaining 21 occurrences on all occasions the structure of the proverb remains unchanged, and in all but one case the first verb is substituted by another. A total of 7 verbs are used, the most common being *provare* [to try]. A more accurate statement of the proverb's form would be, on the basis of this corpus evidence, *vedere per credere* or [verb (especially *provare*)] *per credere*.
- (v) some proverbs were found to vary grammatically so frequently that it would seem best to record them as both proverbs and verbal idioms. An example is afforded by the proverb *Tutti i nodi vengono al pettine* [literally 'All the knots come to the comb' = Problems or mistakes will show up in the end]. It occurs 26 times in the corpus, though only once as an unmodified sentence. There are 23 cases of adjectival and/or adverbial modification,

16 of changes in tense and 6 of clause dependency. Very often it seems to be functioning as a sentence length verbal idiom rather than a proverb as such.

4 Concluding remarks

By way of conclusion we will comment above all on the limitations of our study. Firstly, whereas the study we carried out gave valuable information about the general ways in which Italian proverbs behave in written language, the corpus consulted was by no means large enough to determine accurately the citation forms and typical usage of individual proverbs except in a small number of cases. Secondly, it is clear that a thorough analysis of proverbial behaviour and consequent lexicographical description of Italian proverbs will also depend on the availability of very large corpora of spoken language, which for the moment is very much a future prospect rather than a present reality.

Notes

¹We are not familiar with other modern corpus-based studies of Italian proverbs, though Turrini *et al.*, a corpus-based dictionary of idiomatic phrases and expressions, does include some proverbs. For corpus-based discussion of French and English proverbs, see [Arnaud/Moon (1993)].

²The corpus was interrogated using the search program DBT, which is described in Picchi. For further details of how we used DBT to locate proverbs as well as other multiword units, the reader is referred to [Cignoni/Coffey (1995)], available on request from the authors.

³[Mieder (1989), pp. 241-4] discusses the same phenomenon with regard to English proverbs, and [Moon (1998), pp. 170-4] discusses it with regard to English metaphorical expressions in general. Compare also the exploitation of Italian idioms described in [Cignoni/Coffey (1998), pp. 294-5].

⁴For discussion of Italian newspaper headlines see [Dardano (1973), pp. 265-271] and [Lepri (1986), pp. 121-126]. For discussion of proverbs in the American press see [Norrick (1995), pp. 22-4].

References

- [Arnaud (1991)] Arnaud, Pierre (1991). Réflexions sur le proverbe, in *Cahiers de lexicologie*, vol. LIX - II, pp. 6-27.
- [Arnaud/Moon (1993)] Arnaud, Pierre & Moon, Rosamund (1993). Fr. Fréquence et emploi des proverbes anglais et français, in C. Plantin (ed.), *Lieux Communs: Topoi, Stéréotypes, Clichés*. Kimé, Paris, pp. 323-341.
- [Cignoni/Coffey (1995)] Cignoni, Laura & Stephen Coffey (1995). *Looking for preselected multiword units in an untagged corpus of written Italian: maximizing the potential of the search program DBT*. Istituto di Linguistica Computazionale, CNR, Pisa.
- [Cignoni/Coffey (1998)] Cignoni, Laura & Stephen Coffey (1998). A corpus-based study of Italian idiomatic phrases: from citation forms to 'real-life' occurrences, in T. Fontenelle, P. Hiligsmann, A. Michiels, A. Moulin & S. Theissen (eds.), *Euralex '98 Proceedings*, Vol. 1. University of Liège, English and Dutch Departments, pp. 291-300.
- [Dardano (1973)] Dardano, Maurizio (1973). *Il linguaggio dei giornali italiani*. Laterza, Bari.

- [Lepri (1986)] Lepri, Sergio (1986). *Medium e messaggio: il trattamento concettuale e linguistico dell'informazione*. Gutenberg 2000, Torino.
- [Mieder (1989)] Mieder, Wolfgang (1989). *American Proverbs: A Study of Texts and Contexts*. Peter Lang, New York.
- [Moon (1998)] Moon, Rosamund (1998). *Fixed Expressions and Idioms in English: a corpus-based approach*. Clarendon Press, Oxford.
- [Norrick (1995)] Norrick, Neal R. (1995). *How Proverbs Mean: Semantic Studies in English Proverbs*. Mouton, Berlin.
- [Picchi (1991)] Picchi, Eugenio (1991). DBT: A Textual Database System, in Laura Cignoni & Carol Peters (eds.), *Computational Lexicology and Lexicography*, Vol. VII, Istituto di Linguistica Computazionale, CNR, Pisa, pp. 177-205.
- [Turrini *et al.*] urrini, G., Alberti, C., Santullo, M. L. & Zanchi, G. (1995). *Capire L'Antifona: Dizionario dei modi di dire con esempi d'autore*. Zanichelli, Bologna.

