# Position of the head in Italian n-n compounds: the case of "mirror compounds"[1]

Jan Radimský (České Budějovice)

**ABSTRACT**

Several authors have recently pointed out that under certain conditions, a language can have both left-headed and right-headed compounding patterns. As contemporary Italian N-N compounds seem to encompass this phenomenon, the present article aims to examine parameters that determine or indicate the head position in these cases. Our hypotheses are based on a qualitative sample drawn from the Gradit dictionary and are further verified by a quantitative study about "mirror compounds" (*radiogiornale* vs. *giornale radio*) on the ItWac corpus. The position of the head in N-N compounds is claimed to show a strong correlation with "tight" and "loose" orthography. In conclusion, possible explanations of this phenomenon as well as open questions for future research are suggested.

**KEYWORDS**

word compounding, N-N compounds, learned compounding, Italian

## 1. INTRODUCTION

The position of the head in compounds has been a widely discussed subject in recent literature. In the eighties, generative grammar presumed that the parameter of directionality — and thus, the location of the head — remains the same in compounding as well as in other morphosyntactic processes of a given language. Accordingly, the position of the head in compounds would be guided by the setting of this parameter and one language could have either only left-headed or only right-headed compounds.[2] Since then several studies have argued that one language can have both left-headed and right-headed compounds: Packard (2000) claims that in Chinese, verbal compounds are right-headed, while nominal compounds are left-headed.[3] Pepper (2010:35–41) has recently shown that in Nizaa (a language spoken in Cameroon), both

---

1   This text was produced as part of the GAČR Grant No. 13–00496S (N-N Compounding in Contemporary Italian).

2   Cf. Scalise (1990) for a detailed discussion of this topic, especially in the chapters concerned with the "Righthand head rule". In his later work, the same author (Scalise, 1994:194) states overtly that at the current state of knowledge, it does not seem possible for a language to form freely both left-headed and right-headed compounds.

3   Cf. a detailed discussion in Scalise S. — Fábregas A. (2010:116–118) and Scalise S. and Masini F. (2012:81–82).

compound types coexist and the position of the head is motivated by the semantic features (semantic class) of the compound.

According to the earlier generative assumptions, contemporary Italian compounding has been described as uniformly left-headed, while the "few" right-headed compounds in use were considered either as relics of Latin morphological processes or as calques from foreign languages (Scalise, 1994:194–195).[4] However, some data suggest that in spite of its *learned* origin, right-headed compounding could represent a vital process in contemporary Italian, entering in competition with the "Romance" left-headed compounding. This point emerges in a particular way for the so-called "mirror-compounds" like *radiogiornale* vs. *giornale radio* (*radio news*). As these compounds have the same internal structure (N-N) and — being often synonymous — the same semantic features, the question arises as to which factor determines or bears the mark of the head position in such a compound.

The paper is structured as follows: in section (2), a recent typology of productive Italian N-N compounds by Sergio Scalise and Antonio Fábregas is presented and discussed in contrast with data drawn from the Gradit dictionary, allowing us to make a hypothesis about the distinction between right-headed and left-headed types. Section (3) is concerned with the "mirror" N-N compounds; the data from the ItWac corpus are used in order to verify the hypothesis presented in section (2). Finally, conclusions are drawn and further research perspectives are proposed in section (4).

## 2. AVAILABLE PATTERNS OF ITALIAN SUBORDINATE N-N COMPOUNDS

One of the most recent typologies of contemporary Italian N-N compounds has been given by Sergio Scalise and Antonio Fábregas (2010). The following table presents only the types considered as productive by these authors.

| Type | Head position | Example | Translation |
|------|---------------|---------|-------------|
| 1 | Right-headed | *gasdinamica* | *gas dynamics* |
| 2a | | *ufficio viaggi* | *travel agency* |
| 2b | Left-headed | *uomo rana* | *frog man* |
| 2c | | *studente lavoratore* | *student worker* |

**TABLE 1** — Current Italian N-N compounds, according to Scalise S. — Fábregas A. (2010:119)

Types (2a-c) are considered as regular and productive patterns, while Type (1) would represent "*only one case of an 'Italian' right-headed compound, which, furthermore, belongs to a specific lexical domain and also has equivalents in English*" (Scalise S. — Fábre-

4   Cf. also Scalise (1990:263), where it is suggested that the possibility of a language — Italian at least — having both right-headed and left-headed compounds is only illusory: right-headed and left-headed compounds would in fact correspond to two different diachronic layers, connected with the basic syntactic order of Latin and Italian respectively.

gas A., 2010:119). In fact, we agree that the *gasdinamica* type could be plausibly interpreted as a new evolution stage of the so-called *learned compounding*, which has become a vital word-formation process in all Romance (and also other European) languages especially in the domain of terminology. But is it really so marginal, from both the quantitative and sociolinguistic point of view?

Let us examine from this angle a sample of 1390 recent Italian compounds collected in Radimský (2006); this sample contains all lemmas marked as "compound" in the *Gradit* dictionary (De Mauro, 1999), first attested between 1980–1999. Considering that *Gradit* is the biggest modern Italian dictionary with an open editorial policy and an accurate treatment of etymology, the sample can be considered as representative of compounds created between 1980 and 1999, including low-frequency items and terminology, but involving obviously only compounds written by convention as one graphic word (i.e. without a blank or hyphen between the internal elements).

As far as the N-N compounds are concerned, the sample contains 9 left-headed and 91 right-headed items, which corresponds to 0.6% and 6.5% respectively. Thus, the left-headed compounds (Radimský J., 2006:90) can hardly be said as representing a vital paradigm: six of them belong to the very productive, but limited, paradigm in *capo-*[5] and among the remaining three, one is in reality probably exocentric (*frescolana*), one has an internal preposition (*punkabbestia < punk a bestia*), and the last (*discogay*) has one element of English origin (*gay*, attested in Italian since 1959 according to the *Gradit*).

On the other hand, the right-headed compounds in the sample seem to be good examples of the *gasdinamica* paradigm. Their complete list is given in the following table.

---

autopattuglia; autosoccorso; aviosuperficie; cicloalpinista; cicloamatore; cicloambientalismo; cicloambientalista; cicloraduno; cineautobiografia; cronoscalata; discobar; discocultura; discopub; eurozona; fotobiografia; fotodisco; fotogiornalismo; fotogiornalista; fotolaboratorio; fotolibro; fotopiano; fotosafari; fotoservizio; fotostoria; infospettacolo; laserchirurgia; macrolinguaggio; macroprogramma; motoalpinismo; mototurismo; ortovivaista; pornoattore; pornocassetta; pornoconsumatore; pornodiva; pornodivo; pornoeroe; pornofan; pornofestino; pornogruppo; pornoromanzo; pornosala; pornotelefonata; pornotelefono; radarriflettente; radarschermo; radarsonda; radiodrammaturgia; radiosveglia; sciescursionismo; sciescursionista; siloporto; spazionave; totogol; turbodiesel; turborazzo; vetrocamera; vetrometallo; vetroresina; videobar; videocamera; videocatalogo; videoconferenza; videocontrollo; videodipendenza; videoenciclopedia; videofilm; videogioco; videogiornalismo; videogiornalista; videoimpaginatore; videoinformazione; videolettore; videolibro; videonoleggio; videopirateria; videoproiettore; videoreporter; videorivista; videorock; videoscrittura; videosimulazione; videosistema; videoterminale; videoverbalizzazione

**TABLE 2** — Right-headed N-N compounds in Radimský (2006)[6]

5   Viz. *capoarea, capocarro, capocommessa, capolevate, capomensa, capozona* (Radimský, 2006:90).

6   The right-headed N-N compounds are not presented explicitly in Radimský (2006), because of a different terminology adopted in that book. We have obtained a list of these items by filtering the electronic version of the complete sample in Radimský (2006:197–255).

Many of these compounds contain a first element etymologically related to a learned stem, but attested in the present day language (generally with a new sense) both as a compound element and as an independent noun. Thus, the paradigm can be observed as a new evolution stage of learned compounding. Nevertheless, it has also to be pointed out that not *all* first elements have learned origin: the first position also contains ordinary Italian nouns of Italian (*vetro, spazio*) or foreign (*sci, radar, laser, disco*) origin or abbreviated Italian nouns (*aviazione > avio, informazione > info*).[7]

From the sociolinguistic point of view, these compounds would be expected to be specialized terms, like *gasdinamica*. However, the Gradit dictionary states that only 42 of them (46%) are of a terminological nature, while 39 belong to the common lexicon (43%), 3 are marked as both common and terminological and 7 would be of rare use ("di basso uso").

These data allow us to make the following assumption: right-headed N-N compounding is becoming a vital word-formation paradigm in contemporary Italian. It can be interpreted as a direct continuation of the learned compounding paradigm (common to most of Standard Average European languages), but it seems to be breaking two traditional limits of the learned compounding, i.e. usage of a limited set of compound elements with learned origin on one hand, and a highly terminological nature of the compounds on the other hand.

How should we interpret the absence of left-headed N-N compounds in this sample? Recent literature[8] clearly shows that left-headed N-N compounds like those in Table 1 belong to vital Italian word-formation paradigms. Therefore, their absence in our sample must be due to the fact that the sample is limited by an orthographic convention: it contains only compounds written as one graphic word. On the other hand, studies which (implicitly) take into account Italian compounds written as two graphic words (both hyphenated and unhyphenated)[9] do not mention any right-headed N-N compounds. These facts suggest that, despite the assumptions shared in Romance word compounding since the pioneer studies of Arsène Darmesteter (1894) and Federico Tolemache (1945), an orthographic convention may play a role as an indicator of internal morphosyntactic structure in contemporary Italian compounds.[10]

Let us now examine this assumption by a quantitative study of the "mirror" N-N compounds.

---

7  We have observed that in these right-headed compounds, some of the first elements appear quite frequently in the first position; a similar phenomenon has been observed by Pepper (2010:38) for left-headed Nizaa compounds. However, neither in Nizaa nor in Italian, the simultaneous presence of left-headed and right-headed compounds can be accounted for by this family size effect (cf. Radimský, 2011).

8  Cf. Baroni, Guevara, Pirelli (2007); Baroni, Guevara, Zamparelli (2009).

9  Cf. Scalise and Masini (2012:79).

10  Since Italian orthography does not have a normalized shape, it is essentially governed by implicit mechanisms based on language in use.

## 3. POSITION OF THE HEAD IN "MIRROR COMPOUNDS"

The so-called "mirror" N-N compounds may not be central in terms of frequency in Italian word formation processes, but they represent an interesting central point of contact between right-headed and left-headed N-N compounding. Migliorini (1963) noticed almost fifty years ago that some compounds could be either left-headed or right-headed, producing "mirror pairs" like *radiogiornale* vs. *giornale radio* (*radio news*) or *radiostazione* vs. *stazione radio* (*radio station*). These pairs are mostly synonymous[11], they have the same internal structure (N-N) but a different position of the head. From the diachronic point of view, the "tight" right-headed member (*radiogiornale*) represents the latest evolution stage of the neoclassical (learned) compound patterns (the above mentioned *gasdinamica* type), while the "loose" left-headed member (*giornale radio*) is a fruit of native Romance morphosyntactic processes.[12]

In these prototypical examples, traditional orthography seems to reflect the position of the head, as the tight member is right-headed and the loose member is left-headed. The crucial question is whether this correlation could be observed as significant in large text corpora and with lower-frequency compounds. In order to perform this verification, we have taken into account a basic set of nominal "mirror effect items" (like *radio* in the above-mentioned example) identified in previous studies (Radimský 2007, 2011): *audio, auto, avio, bici, foto, laser, meteo, moto, porno, radar, radio, turbo* and *video*. Then, the compounds these items allow to form were observed in the ItWac corpus.

Hypothesizing that tight N-N compounds are systematically right-headed (RH) and loose N-N compounds are left-headed (LH) allows us to assume that only two of the four possible combinations are predicted, as shown in Table 3:

| Predicted forms | | Non-predicted forms | |
|---|---|---|---|
| Tight right-headed | Loose left-headed | *Loose right-headed | *Tight left-headed |
| *audio***cassetta** | **cassetta** *audio* | * *audio* **cassetta** | ***cassett(a)**audio* |

**TABLE 3** — Predicted and non-predicted forms

In quantitative terms, non-predicted forms should have a significantly lower frequency of occurrence than the predicted ones. The proportion between the two predicted forms should thus be accidental (cf. Radimský 2007, 2011 for the last point).

11    The synonymy is frequent, but not necessary: *audiocassetta* and *cassetta audio* are synonymous, while *radiocomando* and *comando radio* not necessarily (both terms can mean *remote control*, but *comando radio* could also mean *order given through a radio transmitter* or it can refer to a special type of control of a car radio receiver, situated near the steering wheel; cf. Radimský, 2011:120–121). However, the synonymy of a mirror-pair is not required in our analysis as far as all the interpretations attested respect the predicted position of the head.

12    Cf. Migliorini (1963:33–34), Antonelli (1996:292–293), Iacobini (2004:75), Radimský (2007), Radimský (2011).

Nevertheless, the verification cannot be only mechanical, because some compounds can be interpreted in "both" directions, which means that the apparently "non-predicted" combinations are in reality predicted, because they bear a different interpretation. This is the case of *fotoarchivio* and *foto archivio* in the following example:[13]
*fotoarchivio* — *archive of photographs* (Gradit): tight, right-headed.
*foto archivio* — *photograph of/from an archive* (interpretation supported by occurrences from the Itwac corpus): loose, left-headed.

The verification of this second property presupposes a qualitative interpretation of each occurrence in context.

These hypotheses have been tested on the Itwac corpus by assembling all predicted and unpredicted compounds (types) for each "mirror effect element" and calculating the sum of the frequencies of these types. The frequency lists were obtained from the Itwac corpus by the following CQL queries.

| Type of compound | | Query (for the item *auto*) |
|---|---|---|
| A | Tight right-headed | [lemma="auto..*" & tag="NOUN"] |
| B | Loose left-headed | [tag="NOUN"] [word="auto"] |
| C | Loose right-headed | [word="auto"] [tag="NOUN"] |
| D | Tight left-headed | [lemma="..*auto" & tag="NOUN"] |

**TABLE 4** — Queries

Since these frequency lists contained a lot of noise, they were filtered automatically using the Marco Baroni's "Morph-it!" database of inflected Italian nouns. In this way, the filtered frequency lists contain only words (word forms) that can be a possible n-gram combination of a nominal "mirror item" (such as *audio, auto*, etc.) with any Italian noun.[14] As the "tight" forms (type A and D) also contain hyphenated forms, we have calculated their rate.

The overall results based on the filtered frequency lists are given in Table 5.

In total, about 384 872 tokens (from 13 961 different word forms) were taken into account. These data obviously still contain some noise, because it was not possible to control all 13961 word forms individually. However, the overall rate of non-predicted tokens was of 3.23% only, which means that the general tendency reveals a strong support for the initial hypothesis.

---

13  In this respect, it is interesting to observe the orthography of the non-predicted form *fototessera* (*passport-format picture*, tight and left-headed). According to Zingarelli's dictionary (2011), the tight orthography has been lexicalized since 1966, but in the Itwac corpus, there is a strong concurrence between this lexicalized tight form (307 sg. + 119 pl.) and the predicted, but non-lexicalized loose form (337 — *foto tessera* and 79 — *foto-tessera*).
14  The list of nominal word-forms from Baroni's database contained 32.779 words which graphically differ one from another.

| Item | A | | B | Total A+B (predicted) | | C | D | | Total C+D (non-predicted) | | Rate of non-predicted |
|------|-----|--------|-------|-------|---------------------|-----|-----|--------|-------|------------------|------|
| | All | Hyphen. | | Total | Rate A — hyphen. | | All | Hyphen. | Total | Rate D — hyphen. | |
| **audio** | 2987 | 828 | 15752 | **18739** | 27.72% | 371 | 29 | 25 | **400** | 86.21% | 2.09% |
| **auto** | 134528 | 7840 | 9978 | **144506** | 5.83% | 2946 | 260 | 251 | **3206** | 96.54% | 2.17% |
| **avio** | 642 | 8 | 89 | **731** | 1.25% | 31 | 0 | 0 | **31** | — | 4.07% |
| **bici** | 163 | 65 | 370 | **533** | 39.88% | 115 | 0 | 0 | **115** | — | 17.75% |
| **foto** | 35740 | 632 | 1483 | **37223** | 1.77% | 1200 | 23 | 19 | **1223** | 82.61% | 3.18% |
| **laser** | 201 | 43 | 5064 | **5265** | 21.39% | 284 | 17 | 12 | **301** | 70.59% | 5.41% |
| **meteo** | 292 | 112 | 3988 | **4280** | 38. 6% | 77 | 6 | 4 | **83** | 66.67% | 1.90% |
| **moto** | 21666 | 127 | 783 | **22449** | 0.59% | 930 | 8 | 4 | **938** | 50.00% | 4.01% |
| **porno** | 4973 | 117 | 2501 | **7474** | 2.35% | 242 | 11 | 10 | **253** | 90.91% | 3.27% |
| **radar** | 55 | 24 | 4036 | **4091** | 43.64% | 97 | 37 | 29 | **134** | 78.38% | 3.17% |
| **radio** | 28806 | 916 | 14282 | **43088** | 3.18% | 1538 | 174 | 136 | **1712** | 78.16% | 3.82% |
| **turbo** | 1745 | 70 | 321 | **2066** | 4.01% | 126 | 6 | 6 | **132** | 100.00% | 6.01% |
| **video** | 64218 | 2645 | 17763 | **81981** | 4.12% | 3758 | 160 | 135 | **3918** | 84.38% | 4.56% |
| Total | 297745 | 13427 | 76410 | 374155 | 4.51% | 11715 | 731 | 631 | 12446 | 86.32% | 3.23% |

**TABLE 5** — Rate of predicted and non-predicted tokens in Itwac[15]

It is also interesting to observe the rate of hyphenated forms for Types A and D: while for Type A, the overall rate of hyphenated forms is only of 4.51%, it reaches 86.32% for Type D. However, Type D (tight left-headed) is very marginal in quantitative terms, as shown in Table 6.

| Type of compound | | Frequency | Rate |
|---|---|---|---|
| A | Tight right-headed | 296016 | 76.91% |
| B | Loose left-headed | 76410 | 19.85% |
| C | Loose right-headed | 11715 | 3.04% |
| D | Tight left-headed | 731 | 0.19% |
| Total | | **384872** | 100.00% |

**TABLE 6** — Rate of compound types, based on frequency count

Since the frequency lists with all the corresponding types are too long and would exceed the space of the present text, a shortened frequency list of compounds is given as an example by *video* in Appendix. The example reveals at first sight that the paradigm of predicted forms has a significantly more vital curve.

---

15   A = tight right-headed (*audiocassetta*); B = loose left-headed (*cassetta audio*); C = loose right-headed (*audio cassetta*); D = tight left-headed (*cassetta/a/udio*).

## 4. CONCLUSION

A quantitative verification of current typologies of Italian compounds suggests that the traditional "tight" left-headed N-N compounding (as in *pescecane*) is no longer an available morphological process. Instead, a new right-headed N-N pattern (as in *gas-dinamica*) has been developing as a direct continuation of learned compounding patterns. Data from the Gradit dictionary show that the latter type seems to be breaking the two traditional limits of the learned compounding, since it ceases to combine only a set of compound elements of learned origin on one hand, and it begins to produce a significant number of non-terminological compounds on the other hand. This tendency should be explored more deeply on large language corpora and by taking into account especially forms that are neither of learned origin nor of a terminological nature (e.g. *pescaturismo* — litt. "fishing tourism").

In this state of affairs, Italian seems to have both left-headed and right-headed vital N-N compounding patterns. Therefore, the question arises as to which factor determines or bears the mark of the head position in such a compound. Our research shows that in the Itwac corpus, the location of the head in N-N "mirror" compounds (*radiogiornale* — *giornale radio*) presents a high correlation with the orthography, as the right-headed N-N compounds appear as "tight" and left-headed N-N compounds as "loose".

As the Italian orthography of neologisms is steered by implicit language rules rather than by explicit rules of a language policy, we think that the orthography of N-N compounds could reflect the fact that right-headed and left-headed N-N compounds correspond to different mental word-formation models in the mind of language users. There are also reasons to think that it may reflect a possible phonetic distinction between "tight" and "loose" compounds. According to a traditional point of view expressed in the classical handbook of Žarko Muljačić (1969:474–475), the difference would involve the relative strength of the two stresses concerned, as shown in Table 7.

| Tight compound (= right-headed) | Loose compound (= left-headed) |
|---|---|
| fotoarchivio | archivio foto |
| [ˌfotoarˈchivjo] | [arˈchivjoˈfoːto] |

**TABLE 7** — Phonetic difference between tight and loose compounds

While tight compounds would have one main stress and one secondary stress, loose compounds would have two main stresses, which could also have an impact on other correlated phenomena such as vowel length. Such a view is supported for instance by Sugeta (1989), whilst Nespor (1993) claims that there is no significant difference between the two stresses. However, neither of these hypotheses is based on material evidence from real phonetic research[16] and the most recent sources (Krämer M.,

---

16   This view is supported by my personal communication with Marina Nespor.

2009:194–195) reveal that the question of the "secondary stress" in Italian compounds and (free) noun phrases is far from being understood at the current state of knowledge. The question thus remains to be further verified by phoneticians.

## REFERENCES

Antonelli, G. (1996) Sui prefissoidi dell'italiano contemporaneo, *Studi di lessicografia italiana* XIII, Firenze: Accademia della Crusca, 253–293.

Baroni, M., E. Guevara and V. Pirelli (2007) NN Compounds in Italian: Modelling category induction and analogical extension, *Lingue e linguaggio* 2/2007, Bologna: Mulino, 263–290.

Baroni, M., E. Guevara and V. Pirelli (2009) Sulla tipologia dei composti N+N in italiano: principi categoriali ed evidenza distribuzionale a confronto, in Benatti, R., G. Ferrari and M. Mosca (eds), *Linguistica e modelli tecnologici di ricerca* (Atti del 40esimo Congresso della Società di Linguistica Italiana), Roma: Bulzoni, 73–95.

Baroni, M., E. Guevara and R. Zamparelli (2009) The dual nature of deverbal nominal constructions: Evidence from acceptability ratings and corpus analysis, *Corpus Linguistics and Linguistic Theory* 5/1, Berlin: Mouton de Gruyter, 27–60.

Darmesteter, A. (1894) *Traité de la formation des mots composés dans la langue française*, Paris: Émile Bouillon, 2e édition revue, corrigée et en partie refondue.

De Mauro, T., ideato e diretto da (1999) *Grande dizionario italiano dell'uso (Gradit)*, CD-ROM, Torino: Utet.

Gaeta, L. (2010) Synthetic compounds, in Scalise, S. and I. Vogel (eds), *Cross-Disciplinary Issues in Compounding*, Amsterdam: John Benjamins, 219–236.

Grossmann, M. and F. Rainer (eds), (2004) *La formazione delle parole in italiano*, Tübingen: Max Niemeyer Verlag.

Iacobini, C. (2004) Composizione con elementi neoclassici, in Grossmann, M. and F. Rainer (eds), *La formazione delle parole in italiano*, Tübingen: Max Niemeyer Verlag, 69–95.

ItWac, Electronic web corpus of Italian. Available at: http://www.sketchengine.co.uk/.

Krämer M. (2009) *The Phonology of Italian*, Oxford, Oxford University Press.

Masini, F. and S. Scalise (2012) Italian compounds, *Probus* 1/24, Berlin: Mouton de Gruyter, 61–91.

Migliorini, B. (1963) *Saggi sulla lingua del novecento*, Firenze: Sansoni.

Muljačić, Ž. (1969) *Fonologia generale e fonologia della lingua italiana*, Bologna: Mulino.

Nespor, M. (1993) *Fonologia*, Bologna: Mulino.

Packard, J. L. (2000) *The Morphology of Chinese: A Linguistic and Cognitive Approach*, Cambridge: Cambridge University Press.

Pepper, S. (2010) *Nominal Compounding in Nizaa — A Cognitive Perspective*, MA dissertation, London: University of London.

Radimský, J. (2006) *Les composés italiens actuels*, Paris: Cellule de recherche en linguistique.

Radimský, J. (2007) Un confixe autonome est-il un substantif épithète? Remarques à propos de la directionnalité dans la morphologie lexicale, *Écho des études romanes* III/1–2, České Budějovice: Jihočeská univerzita, 109–122. Available at: www.eer.cz.

Radimský, J. (2011) Konfixace v současné italské slovotvorbě (Confixation in contemporary Italian word formativ), *Časopis pro moderní filologii* 93, 104–124.

Scalise, S. (1986) *Generative morphology*, Dordrecht: Foris.

Scalise, S. (1990) *Morfologia e lessico*, Bologna: Mulino.

Scalise, S. (1994) *Morfologia*, Bologna: Mulino.

Scalise, S. and A. Bisetto (2005) The classification of compounds, *Lingue e Linguaggio* 4/2, 319–332.

Scalise, S. and A. Bisetto (2008) *La struttura delle parole*, Bologna: Mulino.

Scalise, S. and A. Fábregas (2010) The head in compounding, in Scalise, S. and I. Vogel

(eds), *Cross-Disciplinary Issues in Compounding*, Amsterdam: John Benjamins, 109–126.

Scalise, S. and F. Masini (2012) Italian compounds, *Probus* 1/24, 61–91.

Sugeta, S. (1989) Il sintagma nominale del tipo 'parola-chiave' in italiano e nelle lingue romanze, in Foresti, F., E. Rizzi and P. Benedini (eds), *L'italiano tra le lingue romanze — atti del 20. congresso internazionale di studi*, Roma: Bulzoni, 195–212.

Tollemache, F. (1945) *Le parole composte nella lingua italiana*, Roma: Edizioni Rores.

Zingarelli, N. (2011) *Lo Zingarelli 2011, vocabolario della lingua italiana,* CD-ROM, Bologna: Zanichelli.

## APPENDIX — FREQUENCY LIST (SHORTENED)
## OF PREDICTED AND NON-PREDICTED TYPES WITH *VIDEO*

| Predicted forms | | | | Non-predicted forms | | | |
|---|---|---|---|---|---|---|---|
| Tight right-headed (A) | | Loose left-headed (B) | | Loose right-headed (C) | | Tight left-headed (D) | |
| Word | Fq. | Word | Fq. | Word | Fq. | Word | Fq. |
| videogiochi | 13547 | scheda video | 2454 | video conferenza | 367 | installazione-video | 6 |
| videogioco | 5839 | schede video | 1050 | video giochi | 228 | film-video | 5 |
| videogame | 4523 | home video | 936 | video sorveglianza | 192 | sala-video | 5 |
| videocassette | 3860 | qualità video | 511 | video cassette | 174 | terminali-video | 5 |
| videoclip | 3774 | file video | 435 | video cassetta | 127 | cassette-video | 4 |
| videocassetta | 3732 | formato video | 406 | video arte | 124 | intervista-video | 4 |
| videocamera | 3108 | memoria video | 296 | video conferenze | 123 | data-video | 3 |
| videocon-ferenza | 2760 | montaggio video | 291 | video proiettore | 113 | documentario-video | 3 |
| video-sorveglianza | 2099 | risoluzione video | 268 | video gioco | 109 | post-video | 3 |
| video-registratore | 1938 | immagini video | 244 | video scrittura | 101 | proiezione-video | 3 |
| videocamere | 1649 | riprese video | 238 | video proiezioni | 86 | proiezioni-video | 3 |
| videoteca | 1321 | segnale video | 219 | video messaggio | 83 | radio-video | 3 |
| videoscrittura | 1195 | sezione video | 217 | video camera | 82 | riprese-video | 3 |
| videoterminali | 987 | rumore video | 213 | video installazioni | 81 | rockvideo | 3 |
| videogiocatori | 805 | proiezioni video | 208 | video comunicazione | 43 | tele-video | 3 |
| video-conferenze | 781 | acquisizione video | 202 | video proiezione | 43 | uomo-video | 3 |
| video-registratori | 679 | produzione video | 197 | video camere | 42 | aula-video | 2 |
| videopoker | 599 | contenuti video | 187 | video installazione | 39 | catalogo-video | 2 |
| videoterminale | 592 | presentazione video | 171 | video intervista | 37 | contatto-video | 2 |
| video-proiettore | 536 | proiezione video | 171 | video interviste | 34 | flashvideo | 2 |
| videogiocatore | 500 | modalità video | 161 | video reportage | 32 | fumettivideo | 2 |
| video-comunicazione | 495 | ripresa video | 151 | video lezioni | 28 | messaggio-video | 2 |
| videotelefono | 391 | sequenze video | 151 | video messaggi | 27 | pagina-video | 2 |

| Predicted forms | | | | Non-predicted forms | | | |
| Tight right-headed (A) | | Loose left-headed (B) | | Loose right-headed (C) | | Tight left-headed (D) | |
| Word | Fq. | Word | Fq. | Word | Fq. | Word | Fq. |
|---|---|---|---|---|---|---|---|
| videoteche | 352 | uscita video | 145 | video chiamate | 26 | pagine-video | 2 |
| video-registrazione | 343 | traccia video | 126 | video telefono | 25 | pop-video | 2 |
| video-clip | 302 | compressione video | 125 | video box | 23 | scheda-video | 2 |
| videoarte | 263 | sala video | 119 | video prodotti | 21 | schede-video | 2 |
| video-proiezione | 235 | servizi video | 118 | video telefonata | 21 | televideo | 2 |
| videoproiettori | 214 | produzioni video | 117 | video telefonate | 21 | televisione-video | 2 |
| video-proiezioni | 205 | clip video | 113 | video artisti | 20 | album-video | 1 |
| video-registrazioni | 168 | registrazione video | 109 | video controllo | 19 | artisti-video | 1 |
| videotelefoni | 167 | segnali video | 104 | video disco | 17 | auto-video | 1 |
| video-conferenza | 164 | supporto video | 103 | video proiettori | 17 | box-video | 1 |
| video-sorveglianza | 150 | riproduzione video | 97 | video performance | 16 | braccio-video | 1 |
| videotelefonia | 144 | registrazioni video | 95 | video registrazione | 16 | carta-video | 1 |
| video-installazioni | 126 | collegamento video | 90 | video server | 16 | cellulare-video | 1 |
| video-messaggio | 121 | presentazioni video | 90 | video telefonia | 16 | cinema-video | 1 |
| videointervista | 115 | immagine video | 87 | video milioni | 14 | comando-video | 1 |
| videonoleggio | 113 | parte video | 87 | video artista | 12 | danza-video | 1 |
| video-installazione | 110 | cassette video | 84 | video dipendenza | 12 | dati-video | 1 |
| videomusica | 103 | schermo video | 81 | video girl | 12 | dato-video | 1 |
| video-dipendenza | 101 | sequenza video | 77 | video web | 12 | dichiarazione-video | 1 |
| videotelefonini | 100 | documenta-zione video | 73 | video blog | 11 | documenti-video | 1 |
| video-games | 99 | sistema video | 72 | video film | 11 | | |
| videocitofono | 85 | contributi video | 69 | video giochini | 11 | | |
| videogrammi | 84 | cattura video | 68 | video operatori | 11 | | |
| video-arte | 83 | poker video | 68 | video rap | 10 | | |
| videochiamate | 78 | installazioni video | 67 | video spot | 10 | | |