

Compound Dictionary Entries (The Case of Slovene Noun Phrases)

Vojko Gorjanc¹, Andreja Žele²

¹Faculty of Arts, University in Ljubljana
Aškerčeva 2, SI-1000 Ljubljana
vojko.gorjanc@guest.arnes.si

²Institute for Slovene Language, Slovene Academy of Science
Novi trg 4, SI-1000 Ljubljana
andreja.zele@siol.net

Abstract

The present paper points out the special characteristics of the lexicalised multi-word units and presents the semantic and syntactic standards for determining the degree of permanence and stability of noun phrases in Slovene. Since lexicalisation is linked to the obligatory classifying character of adjectives (primary or secondary/obtained by conversion) which is formally marked in Slovene by the endings *-ni/-i*, the issue of simple adjectival dictionary entries in *-en* and *-ni/-i* can be resolved indirectly, along with the issue of lexicalised multi-word units as dictionary entries in the Slovene language. At the same time, a new question arises: which criteria must a multi-word unit meet to get the status of a dictionary entry. On the basis of the material from the *Dictionary of the Standard Slovene Language (Slovar slovenskega knjižnega jezika)*, which was analysed using traditional methods, and the new textual material from the *Corpus of the Slovene Language, FIDA*, the paper presents standards for determining the permanence of a multi-word unit, and within them, standards for including such multi-word units in the dictionary as entries.

1 Theoretical background

First of all, it is necessary to draw attention to what is special about the lexicalised noun phrases, above all the semantic and syntactic standards for determining the degree of permanence and stability of multi-word units. Lexicalisation is linked to the obligatory classifying character of adjectives (primary or secondary/obtained by conversion) which is formally marked in Slovene by the endings *-ni/-i*, (as opposed to the relation of definiteness in adjectives expressing characteristics *-ni* : *-en*). It is typical/usual of classifying adjectives to occur in the syntactic function of noun modifiers, it is also typical of them to be syntactically and semantically interdependent with the noun; such relations are proved by transformations.

In defining lexicalisation, the established propositions can serve as the starting point [Vidovič Muha 2000]; these propositions repeatedly address the problem of semantic and syntactic delimitation of new words:

- For a lexicalised noun phrase, a semantic and syntactic link between the adjective and the noun is necessary; such a link can only be formed by relational (above all classifying) adjectives;
- Within the framework of the gradual transition from being current to being automatic, it seems reasonable to seek a transition between the phraseological (more general) and the

terminological (more narrow) aspects, as well as to seek the semantic and syntactic point of contact or intersection between them.

Regarding the lexicalisation of multi-word units, it is above all necessary to stress — in connection with adjectives as premodifiers — the distinction between the generic and the converted classifying character of adjectives. It is thus the indirect aim of the present paper to take into consideration, when dealing with adjectival dictionary entries, the classifying character of adjectives as an inherent semantic (dictionary) category of the denotative meaning of classifying adjectives, which, being semantically and morphemically automatized dictionary units, are the most suitable for forming multi-word units. It has also been established that, in addition to the so called permanence, being phraseological and terminological are potential characteristics of all semantic groups of classifying adjectives, e.g. *aliirana država* (Engl. *allied country*), *inicijativni odbor* (Engl. *initiative committee*), *apokrifni spis* (Engl. *apocryphal document*), *disketna enotà* (Engl. *floppy disk drive*) etc. From the morphosyntactic point of view, the classifying adjectives in *-ni* stand out; in Slovene they are homonymous with the morphosyntactically definite adjectives expressing characteristics.

In the treatment of lexicalised multi-word units (above all the more general or popular ones, not the narrow, terminological ones), a sort of a typological overview of adjectival entries (classifying adjectives and adjectives expressing characteristics) from the *Dictionary of the Standard Slovene Language (Slovar slovenskega knjižnega jezika)* obtained through selected adjectival modifiers in noun phrases is taken into account although only as a starting point giving direction [Žele 1998]. The dictionary presentations are out of date from the linguistic point of view (the compiling of the dictionary took 30 years and ended with the fifth volume published in 1991) and need to be updated with new theoretical discoveries and solutions which have already become or are about to become a part of the linguistic practice. One of these new discoveries is the generic definiteness of relational adjectives, which has been proved on many examples; being definite as a special type of adjectives and therefore invariant in their endings, possessive and classifying adjectives do not distinguish the category of morphosyntactic definiteness or the so called syntactic category of adjectival aspect [Tafra 1988], e.g. *makaronovo meso* (Engl. *macaroni with Sauce Bolognese, literary: macaroni meat*), *avgijev hlev* (Engl. *Augean stables*), *damoklejev meč* (Engl. *sword of Damocles*) as opposed to *integralni del* (Engl. *integral part*), *interesna sfera* (Engl. *sphere of interest*) etc.

Fluctuating or even failing to distinguish between the *-i* as a morphosyntactic ending and the *-i* as a semantic and transformational classifying ending is noticeable above all in adjectives with suffixes in *-en*, since this suffix is widely used both in qualitative and non-qualitative adjectives. From the point of view of adjectival meaning, we speak of semantic definiteness which is an inherent characteristic of the so-called relational adjectives (classifying, possessive and material); from the point of view of syntactic use of the suffix of definiteness *-i*, we speak of morphosyntactic definiteness. The category of the so called adjectival aspect is completely neutralised only in classifying adjectives — the denominative classifying adjectives such as *cestni* (Engl. *referring to the road*) cannot express textual definiteness with their primary definiteness (of the linguistic system) — they are therefore an ideal

component of terminological expressions. Thus, morphosyntactic definiteness, i.e. relational semantic or part-of-speech definiteness, is typical of the adjective in a lexicalised multi-word unit as a dictionary entry. Semantic definiteness is generally accompanied by morphosyntactic definiteness, which is always obligatory in terminological expressions (there are exceptions, such as *papinov lonec* (Engl. *Papin's digester*) and *pitagorov izrek* (Engl. *Pythagora's theorem*), *brižinski spomeniki* (Engl. *Freising Manuscripts*) defined as classifying adjectives derived from proper nouns). Because of the constant coexistence of the morphosyntactic and word semantic definiteness in terminological expressions, a clear-cut distinction between the original classifying and the secondary/converted definiteness is of even greater importance.

The formation of relational adjectives, which can implicitly (in their semantic base) contain various relations, ensures a simultaneous motivation. When the relational adjective is derived from an object (e.g. *železen*, Engl. *made of iron*) — and it is, for instance, used in a multi-word unit such as *železna volja* (Engl. *iron will*) where the adjective is outside its relation with the meaning of the headword (*volja*, Engl. *will*) — the relational adjective thus becomes a qualitative adjective, i.e. it loses the close semantic interdependence with the headword and re-establishes it on a second non-disciplinary/metaphorical level, for instance *črnsko delo* (Engl. *hard work*), *človeški odnos/pristop* (Engl. *human attitude /approach*), *krokodilove solze* (Engl. *crocodile tears*) etc. On the other hand, adjectives which are originally qualitative, such as *drag* (Engl. *dear, expensive, precious*), change their primary semantic componentiality in collocations with certain words, e.g. *dragi kamen* (Engl. *precious stone*) — the mutual semantic componential influence changes the meaning of the word *kamen* (Engl. *stone*) into "very hard mineral or crystal".

Otherwise, relational adjectives express their possessiveness and material possessiveness with suffixes such as *-ov/-ev*, *-in*, while general possessiveness is expressed with *-ski/-ški*. Material adjectives show morphosyntactic definiteness only in part, the suffixes used express both "characteristics" and "relations" (*-en*, *-ov/-ev*, *-in*, *-at*, *-ast*, *-ski/-ški*, *-ji*: *jagodni/jagodov sok*, Engl. *strawberry juice*; *gumijast/kovinski izdelek*, Engl. *rubber/metal product*); even though they express semantic definiteness, they only have morphosyntactic definiteness in collocations, such as *kraljevi vrelec* (Engl. *royal spring*), *državni kapital* (Engl. *state assets*). The transformation from possessive into classifying adjectives can be seen in examples such as *sizifovo delo* (Engl. *Sisyphean task*), *faradejeva kletka* (*Faraday cage*), in Slovene, both collocations begin with a small letter in spelling, thus the orthography illustrates the transformation. The transformation from material into classifying adjectives is noticeable in multi-word units (especially in the group of semantically transparent material adjectives — the meaning is closer to classification: *železni križ* (Engl. *iron cross*) as opposed to *železen/železov križ* (Engl. *iron cross, cross made from iron*), *asfaltni sloj* (Engl. *asphalt layer*), *lésni pepel* (Engl. *wood ash*), *lésni izdelek* (Engl. *product made from wood*) as opposed to *lesén izdelek* (Engl. *wooden product*).

The fixed expression is semantically and syntactically confirmed by the possibility of single word replacements, such as *ementalski sir* > *ementalec*, (Engl. *Emmenthal cheese* > *Emmenthal*), *edamski sir* > *edamec* (Engl. *Edam cheese* > *Edam*) etc.

Qualitative adjectives or adjectives expressing characteristics, on the other hand, are originally, in their semantic relations, indefinite from the point of view of their possible headwords. When used attributively or predicatively, the adjectives ascribe, intensify or deny the headwords a certain characteristic, which, however, is not semantically linked to the headwords, for instance: *dober predlog* : *Predlog je dober* : *Predlog ni najboljši*. (Engl. *a good suggestion*: *The suggestion is good*: *The suggestion is not the best possible*.) In certain multi-word units they become definite in their semantic relations or are converted into classifying adjectives, thus losing their original meaning of expressing characteristics, their morphosyntactic indefiniteness and the possibility of gradation, e.g. *črni premog*, *beli/črni gaber*, *dragi kamen* (Engl. *black coal*, *white/black beech tree*, *precious stone*) etc.

2 Simple and compound entries in the general dictionary of the Slovene language

In the general explanatory dictionary, the question of the justification of phraseological and terminological parts of the entries arises for noun phrases and adjectival phrases as dictionary entries. Presenting the most frequent multi-word units as independent dictionary units or entries is more suitable because of their lexemic as well as semantic and syntactic independence, and for practical reasons of clarity of the dictionary.

The dictionary presentation of pragmatic idioms such as *dober dan*, *hvala* (Engl. *hello*, *thank you*) etc. needs to be organised and unified from the communicative and pragmatic perspective. Because of their semantic and syntactic self-sufficiency, they should be independent dictionary entries; the *Collins Cobuild Dictionary* confirms this, as it treats expressions such as *good morning*, *good afternoon*, *good evening*, *good day*, *good night*, *look out*, *hello*, *hail*, *cheers*, *thanks* etc. as independent dictionary entries. In the *Dictionary of the Standard Slovene Language*, the communicative and pragmatic aspects are blurred or taken into consideration only implicitly with explanations such as "a greeting, used to congratulate, used to warn" and with the part-of-speech label interjection or the explanation "used as an interjection". It is certainly not appropriate to deal with expressions such as *dober dan*, *dober večer* (Engl. *hello* – *literary*: *good day* – *good evening*) within the adjectival entry *dober* (Engl. *good*), *lahko noč* (Engl. *good night*, *literary*: *easy night*) within the entry *lahk* (Engl. *easy*, *light*), or *srečno* (Engl. *good luck*, *goodbye*) within the entry *srečen* (Engl. *lucky*, *happy*).

As mentioned above, the overview of adjectival entries for /non/qualitative adjectives in the *Dictionary of the Standard Slovene Language* can serve as a starting point for a typology of potential multi-word units as dictionary entries, since it best reveals the justification of (among other things) the inflectional delineation (with *-ni*) of classifying (semantic) definiteness. The illustrative material of terminological parts of the individual adjectival entries shows that the *Dictionary of the Standard Slovene Language* does not take into consideration the semantic systematisation of relational adjectives, which was made for the Slovene language already in the second half of the 1970s: in the illustrative material certain groups of classifying adjectives are inadvertently omitted. Adjectival entries in the *Dictionary of the Standard Slovene Language*, being essential parts of noun phrases and adjectival phrases in this dictionary, can be grouped into the following categories: (a) two

dictionary entries ending in *-en* with distinct meanings, e.g. *ličén¹ : ličen²* (Engl. *of the cheek : nice, neat*), sometimes the meaning is distinguished by word stress, e.g. *lésen : lesén* (Engl. *referring to wood : made of wood*) (b) several consecutive dictionary entries which solve the problem of semantic definiteness or classifying character of the adjective by mutual comparison, e.g. *déžen : dežéven : dežévnat : déžnat* (Engl. *for protection from rain : referring to rain : rainy : with lots of rain*), (c) within a single entry the meaning is differentiated by stylistic labels, e.g. *sončen* (Engl. *sunny*), *bukov* (Engl. *of the beech tree*), *žolčen* (Engl. *bilious*) etc.; the illustrative materials shows a need for a new/additional entry, e. g. *akustičen, biblijski* (Engl. *acoustic, Biblical*) etc., (d) all of the illustrative material shows a need for reflecting that the adjective is semantically definite or classifying, e.g. *cestni* (Engl. *of the road*), *cevni* (Engl. *referring to the pipe*), *licenčni* (Engl. *licensing*) etc.

(a) The adjectival entry *ličén²*, meaning "referring to the cheek", e.g. *lična kost, lična mišica* (Engl. *cheek-bone, muscle of the cheek*), is a semantically transparent possessive adjective which needs the ending *-ni* and the basic form *lični* (as opposed to the adjectival entry *ličén¹*, a qualitative adjective, meaning "nice, neat, especially in form or production" e.g. *lična hišica*, Engl. *a nice little house*, *lična pisava*, Engl. *neat handwriting*). Because *lični* (Engl. *of the cheek*) is a motivated classifying adjective, multi-word units such as *lična kost* (Engl. *cheek-bone*) are possible; this is confirmed by the fact that the *Collins Cobuild Dictionary* treats it as a separate unit (*cheek-bone*); many such multi-word lexemes are considered compounds in English which is indicated by (at least) the possibility of hyphenation. — In adjectival entries where the meaning is differentiated by word stress, e.g. *lésen* – "referring to wood", thus the correct form is *lésni*, as opposed to *lesén* – "made of wood"; the fact that the former case is a classifying adjective is confirmed by multi-word units such as *lesni črv* (Engl. *wood-worm/woodworm*), *lesna žaga* (Engl. *wood-pile/woodpile*), *lesni ocet* (Engl. *wood vinegar*), *lesna industrija* (Engl. *woodworking industry* – in the English lexeme the indirect purpose is explicitly expressed) as opposed to *lesen križ/okvir/vijak* (Engl. *wooden cross/frame/screw*).

(b) Several consecutive dictionary entries which solve the problem of semantic definiteness by mutual comparison can be found in adjectival entries such as *céven* – "referring to the pipe" in collocations such as *cevni navoj* (Engl. *pipe coil*), *cevna odprtina* (Engl. *pipe opening*) which need the classifying form, *cévni* in the entry; the classifying adjective converted from the qualitative adjective occurs in multi-word units such as *cevni ključ* (Engl. *pipe wrench*); the originally similarity-conveying qualitative adjective or adjective expressing characteristics, *cévast*, meaning "shaped like a pipe" occurs as a classifying adjective converted from a qualitative adjective in set (terminological) expressions such as *cevasti organi* (Engl. *tubular organs*), *cevasta kost* (Engl. *tubular bone*), *cevaste gobe* (Engl. *Boletales, a type of mushrooms*). The originally qualitative adjective, such as *kljukast* (Engl. *hooked*) – with the primary meaning: "shaped like a hook", in expressions such *kljukast nos* (Engl. *hook-nose*) – can, in the fixed expression *kljukasti križ* (Engl. *swastika, literary: hooked cross*), gain only morphosyntactic definiteness.

(c) Differentiating between semantic definiteness and indefiniteness or the classifying character, and thus the different semantic and syntactic uses within the same entry (differences in use are usually additionally marked by the stylistic label *expressive*): thus, in

an adjective such as *bukov*, we have its original meaning "referring to the beech tree", e.g. *bukov gozd* (Engl. *beech forest*), *bukovo pohištvo* (Engl. *beech furniture*), and a change from the relational meanings of the permanent classifying component or material to the temporary characteristic meaning "clumsy, awkward", labelled in the dictionary as expressive, e.g. *bukov človek* (Engl. *clumsy person*). Similarly, *žolčni* (Engl. *bilious*) meaning "referring to bile" in collocations such as *žolčna snov* (Engl. *bilious matter*), *žolčni napad* (Engl. *bilious attack*) as opposed to the meaning of expressing a characteristic "cranky, irritable" in occasional collocations such as *žolčen človek* (Engl. *bilious person*) or *žolčen očitek/pogovor* (Engl. *hostile reproach/discussion*). The fact that adjectives such as *službeni* (Engl. *referring to work*) and *uradni* (Engl. *official*) are originally classifying can be confirmed by multi-word units. A possible meaning of expressing characteristics is linked only to the predicative use in the sense of "showing distance, formality", e.g. *Ta človek je vedno zelo služben/uraden*. (Engl. *This man is always very "official", i.e. formal, distant.*)

(d) All of the illustrative material (for multi-word units) shows that adjectival entries which are semantically definite or classifying are necessary, e.g. *cestni* (Engl. *of the road*) in *cestni ovinek* (Engl. *road curve*), *cestni prah* (Engl. *road dust*), *cestni promet* (Engl. *road traffic*), *cestni delavec* (Engl. *road worker*); *cévni* (Engl. *of the pipe*) v *cevni nastavek* (Engl. *pipe fitting*), *cevni navoj* (Engl. *pipe coil*), *cevni spoj* (Engl. *pipe joint*), *cevni daljnovod* (Engl. *pipeline*), *cevni sistem* (Engl. *pipe system*).

The classifying character of adjectives motivated in the semantic base needs to be determined in the use of foreign words as for instance *licénčni* (Engl. *licensing*), in multi-word units such as *licenčni partner* (Engl. *licensing partner*), *licenčni igralec* (Engl. *licensee*), as opposed to e.g. *likviden* (Engl. *liquid*), in *likviden partner* (Engl. *partner with liquid assets available*). Starting from the classifying character motivated in the semantic base, we can adjust the relation between the Slovene and the corresponding foreign adjectives, e.g. *vpisni rok* (Engl. *time appointed for enrollment*) : *inskripcijski rok* (Engl. *time appointed for registering*), *podredni veznik* (Engl. *subordinating conjunction*) : *hipotaktični veznik* (Engl. *hypotactic conjunction*).

Fixed expressions in which the modifier is indeclinable present a special case; they can be phraseological (*top lestvica* (Engl. *the charts*), *lego kocke* (Engl. *Legos*), *baby hojca* (Engl. *walker*), *talk show*, *jafa pomaranča* (Engl. *Jaffa orange*) etc.) and terminological (*žiro račun* (Engl. *giro account*), *kasko zavarovanje* (Engl. *automobile and liability insurance*), *koli bacil* (Engl. *bacillus coli*) etc.) Nouns in which the premodifier or the postmodifier may be declined are rare, e.g. *jaga baba*, *pehtra baba* (Engl. *mythological creatures in the form of old women*), as opposed to *ježi baba* (Engl. *a mythological creature in the form of an old woman*), *človek žaba* (Engl. *diver, literary: frog-man*) etc.

As far as semantic componentiality of the words or their semantic componential relations in multi-word units is concerned, four types of phraseological (Ph) or terminological (T) multi-word units can be distinguished: a) multi-word units with a denominative classifying adjective: *oglasna deska* (Engl. *notice board*) (Ph), *oglasna tabla* (Engl. *notice board*) (Ph), *centralna kurjava* (Engl. *central heating*) (Ph), *nedeljska maša* (Engl. *Sunday mass*) (Ph), *enojajčna dvojčka* (Engl. *identical twins*) (T), *granatno jabolko* (Engl. *pomegranate*) (T),

edamski sir (Engl. *Edam cheese*) (T); b) multi-word units in which at least one word, usually the headword, is anticipated, others may obtain new metaphorical/metonymic semantic components *javna hiša* (Engl. *brothel*, literary: *public house*) (Ph), *dragi kamen* (Engl. *precious stone*) (Ph), *črna maša* (Engl. *mass for the dead*, literary: *black mass*) (T), *makaronovo meso* (Engl. *macaroni with Sauce Bolognese*, literary: *macaroni meat*) (T); c) multi-word units in which the words acquire new semantic components, but the associative link with the original semantic componentiality is kept: *mrtvi kot* (Engl. *blind spot*) (Ph), *mesečno perilo* (Engl. *menstruation*) (Ph), *blažev žegen* (Engl. *something useless*, literary: *the blessing of St. Blaise*) (Ph); d) multi-word units in which the associative link with the original semantic componentiality of the words is not kept: *adamovo jabolko* (Engl. *Adam's apple*) (Ph), *suha južina* (Engl. *daddy-long-legs*) (Ph), *španski jezdec* (Engl. *knife rest*) (T) etc.

The question which multi-word units can have the status of dictionary entries remains largely unresolved in lexicography [Béjoint 2000]. For noun phrases in the Slovene language, it seems reasonable to follow the standards for choosing compound dictionary entries which have been created for bilingual corpus-based dictionaries according to corpus analysis [Gorjanc and Krek 2001]:

- frequency:
the multi-word unit must be frequent enough, it must not occur less frequently than the simple entries;
- semantic:
the meaning of at least a part of the multi-word unit is not transparent;
the multi-word unit candidates for dictionary entries are those with more than one meaning;
the multi-word unit as dictionary entries should have different strings of collocators, e. g. adjectives and noun complements; adjectives and verbs; noun complements and verbs.

3 Conclusion

The paper attempts to find arguments to introduce multi-word dictionary entries in Slovene. Its focus is limited to typical noun phrases with adjectival modifiers. Using the possible semantic componential relations within these phrases as a basis, it seeks to set out the possible types of dictionary entries thus forming the beginnings of a typology. In the multi-word units in Slovene, the semantic componential role of adjectives, their original qualitative or classifying character etc. is very important too.

References

- [Babić 1986] Babić, S., 1986. *Tvorba rječi v hrvatskom književnom jeziku*, Nacrt za gramatiku, Zagreb.
- [Béjoint 2000] Béjoint, H., 2000. *Modern Lexicography*, Oxford, New York, OUP.
- [Bowker 1996] Bowker, Lynne, 1996. *Towards a corpus-based approach to terminography*, in: *Terminology 1* (3), 27–52.
- [Gorjanc & Krek 2001] Gorjanc, V. & Krek, S., 2001. *A corpus-based dictionary database as the source for compiling Slovene-X dictionaries*, in: *COMPLEX 2001. 6th Conference on Computational Lexicography and Corpus Research*, Birmingham.

- [Hanks 1987] Hanks, P., 1987. Definitions and explanations, in J. Sinclair (ed.) *Looking Up: An Account of the COBUILD Project in Lexical Computing*, London, Collins, 116–136.
- [Hanks 1988] Hanks, P., 1998. Typicality in meaning potentials, in: M. Snell-Hornby (ed.) *ZüriLEX'86 Proceedings*, Tübingen, Francke Verlag, 37–47.
- [Čermák & Blatná 1995] Čermák, F. & Blatná, R. (eds.), 1995 *Manual Lexikografie*, Praha, Nakladatelství H&H.
- [Ray-Debove 1971] Ray-Debove, J., 1971. *Etude linguistique et sémiotique des dictionnaires français contemporains*, The Hague, Mouton.
- [Tafra 1988] Tafra, B., 1988. Razgraničavanje opisnih in odnosnih pridjeva, *Razprave Zavoda za jezik* 14, 185–197.
- [Toporišič 1976] Toporišič, J., 1976. *Slovenska slovnica*, Maribor, Obzorja.
- [Vidovič Muha 1978] Vidovič Muha, A., 1978. Merila pomenske delitve nezaimenske pridevniške besede, *Slavistična revija* 3 (26), 255–275.
- [Vidovič Muha 2000] Vidovič Muha, A., 2000., *Slovensko leksikalno pomenoslovje*, Govorica slovarja, Ljubljana, Znanstveni inštitut Filozofske fakultete.
- [Žele 1998] Žele, A., 1998. Terminološka gnezda v Slovarju slovenskega knjižnega jezika s stališča pomensko-skladenjskih lastnosti pridevnikov (Pridevniške iztočnice na -en), in: M. Humar (ed.) *Zbornik referatov s posvetovanja o slovenskem naravoslovno-tehničnem izrazju*, Ljubljana, ZRC SAZU, 129–137.