

Annotation of lexical combinations in LEKO

Andrea Abel, Christine Konecny, Erica Autelli, Lorenzo Zanasi

Edited by: Jennifer-Carmen Frey & Aivars Glaznieks

Version 1.0, May 2022

This description of lexical annotations in LEKO is a scaled-down version of the original annotation scheme of the LEKO project and should describe and clarify the annotations available in the corpus search interface of the LEKO corpora at <https://commul.eurac.edu/annis/leko>.¹

The purpose of the LEKO project was to analyse “lexical combinations” or formulaic sequences (FS), often also referred to as multiword expressions, phrasemes, phraseologisms, or phraseological units², used by L2 learners of Italian. For this purpose, all occurrences of FS in the texts of the LEKO corpora have been annotated as `fs_type`, indicating the type of FS (cf. chapter 1).

Some occurrences have additionally been annotated as `error` if the FS was constructed or used incorrectly (cf. chapter 2). Error annotations contain a lexical error typology. A target hypothesis for the error can be found in an additional `targethyp` annotation, given for every error annotation.

In general, annotations span the whole FS if written in a continuous way; if written in a discontinuous way, all the individual FS elements are annotated.

1 Occurrences of Formulaic Sequences (FS)

The annotation of FS in the LEKO corpora is based on Burger’s functionally motivated “basic classification” of phrasemes, distinguishing between *referential*, *structural*, and *communicative* phrasemes (2007: 36ff.). However, we split referential phrasemes into two subgroups depending on their idiomatic or non-

¹ We reduced the available annotations in order to account for and simplify search options in the ANNIS-search interface.

² Following Burger (2007: 36) we consider the term *phraseme* as a synonym to *phraseologism*.

idiomatic status. Additionally, we include the type `notexistent` to annotate uncommon lexical combinations in the target language that are, for example based on transfer.

Hence, `fs_type` annotations can have the following values:

- `refphras_notid` (referential phrasemes, non- or semi-idiomatic)
- `refphras_id` (referential phrasemes, idiomatic)
- `commphras` (communicative phrasemes)
- `structphras` (structural phrasemes)
- `notexistent` (uncommon lexical combinations, used instead of existing FS and typically based on transfer)

All FS types also have subtypes that are indicated in an additional annotation layer with the name of the superordinate FS type.³ Figure 1 shows an overview of the FS types and their subcategories annotated in the LEKO corpora. In the sections 1.1 – 1.4 below, we give a detailed description of all categories and subcategories.

The `fs_type` annotation is used no matter if the FS occurs in a semantically or formally correct or appropriate way. If there is an error in the use of the FS, one or more additional `error` annotations have been added to the FS span (cf. chapter 2).

Although FS are in general combinations of two or more (orthographic) words, in some cases also single words have been annotated as FS, as for instance in:

- unconventional or erroneous uses of expressions that should be FS. For example, there are cases such as *carini* instead of *miei cari*. These are considered as FS (with the target language sequence in mind) and thus annotated using the respective tags.
- communicative phrasemes where single words (e.g. *Ciao!*, *Cavolo!*) have similar functions as pre-fabricated sequences. Such cases are considered as FS and thus annotated.

³ As Burger points out (2007: 53), it often turns out to be useful for the analysis of text corpora not to refer to a classification based on only one single classification criterion. Thus, the subcategories to further categorize this basic classification follow functional/pragmatical, semantic, syntactic, and structural criteria. The aim of the adopted approach is trying to describe aspects that are difficult and/or important for language learners, especially for German speaking learners of Italian as L2. For difficulties in learning collocations and using them correctly cf. e.g. Deveci (2004) or Hausmann (1984); for the differences between the various FS types as to language reception and production cf. e.g. Ettinger (2013) or Konecny/Autelli (2013).

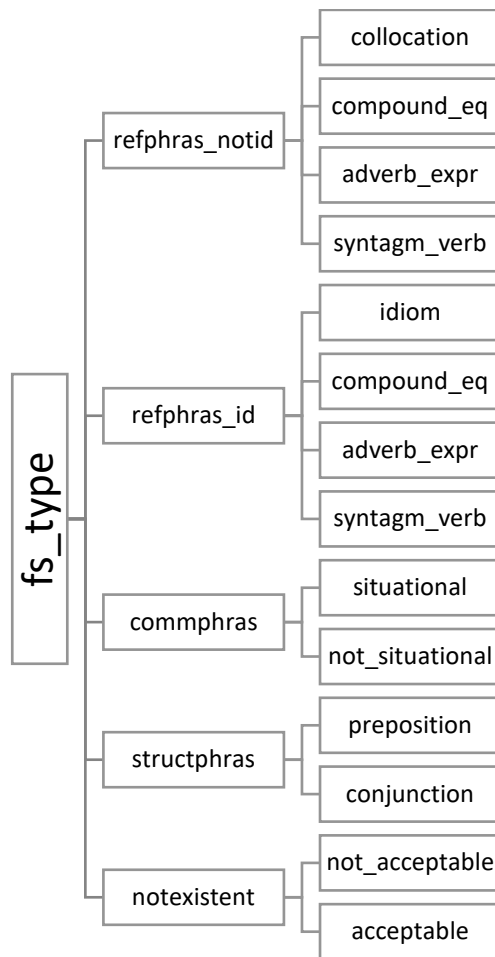


Figure 1: FS types and their subtypes annotated in the LEKO corpora

1.1 Referential phrasemes (**refphras_notid/** **refphras_id**)

Referential phrasemes are FS that refer to objects, processes, or circumstances of the extralinguistic reality. We divide referential phrasemes first into idiomatic (**refphras_id**) and non- or semi-idiomatic (**refphras_notid**) referential phrasemes. On a second layer, we classify them into collocations/idioms, compound equivalents, syntagmatic verbs, or adverbial expressions. Non-idiomatic referential phrasemes (**refphras_notid**)

1.2 Non-idiomatic referential phrasemes (**refphras_notid**)

Non-idiomatic referential phrasemes **refphras_notid** are further divided into 4 subcategories:

- **collocation** (collocations)
- **compound_eq** (compound equivalents)

- `adverb_expr` (adverbial expressions)
- `syntagm_verb` (syntagmatic verbs)

>> `fs_type= "refphras_notid"`

1.2.1.1 Collocations (**collocation**)

According to Paquot/Granger (2012: 136), a collocation is a:

[...] lexically constrained combinations that allow for limited substitution within a particular grammatical construction (e.g. verb-object, adverb-adjective, or adjective-noun). For example, in English you can *perform a task* or *do a task*, but you cannot *make a task*. These lexical restrictions are arbitrary and differ markedly from one language to another. [...] Restricted collocations are opposed to free combinations which are only constrained by semantic or grammatical constraints (e.g., you can *drink coffee* and any other liquid, but you cannot drink solids like bread).

The most typical characteristics of collocations are that they are:

- lexically constrained combinations;
- not (or only minimal/partially) idiomatic;
- not whole sentences.

Collocations are typically binary constructions, i.e. consisting of two elements: 1) the *base* and 2) the *collocator* (cf. Hausmann's approach, e.g. Hausmann 1984; 2004).

In cases of doubt, we adopted a contrastive DE-IT perspective when deciding whether an expression is a collocation or not, trying to translate it literally or looking it up in a bilingual dictionary and checking if it is listed as a phraseological unit or not (e.g. German-Italian: Paravia DIT). Furthermore, combinations that are rule-based were not annotated as FS: e.g. *avere un problema* → no collocation; *andare al supermercato* → collocation (choice of preposition not predictable/rule-based). In cases of doubt, we referred to the monolingual dictionary Treccani or – if the information given in Treccani was not clear or insufficient – to the online version of the dictionary Il Sabatini/Coletti.

As to the parts of speech and the syntactic functions of the elements of a collocation, typical combinations are: noun (subject) – verb, verb – noun (object), adverb – adjective, adjective – noun etc.

If in Treccani a combination is indicated as *fig.*, the whole combination is considered as an idiom instead of a collocation (cf. Nesselhauf 2005: 54).

Note:

Starting from a broad concept of collocation, we also annotate the following phenomena as collocations:

- **Kinegrams:** *storcere il naso* (in the primary meaning of 'to wrinkle one's nose')
- **Words related by implication:** *il cane abbaia, naso camuso*
- **Semi-Idioms:** *numero verde*
- **Comparative frozen sentences:** *buono come il pane, fedele come un cane*
- **Stretched verb constructions / sequences with delexicalized verbs:** *fare la spesa, prendere una decisione*
- **Loose collocations:** *stare sulla {in} spiaggia*

Corpus examples from LEKO_Kolipsi: *sono andato al supermercato, vado nei ferie, fare giri con le biciclette, fare la spesa, è venuto in mente, stare sulla spiaggia, fare le vacanze, andare in vacanza, andare in montagna, ...*

```
>> refphras_notid_type= "collocation"
```

1.2.1.2 Compound equivalents (**compound_eq**)

We annotate only nominal *compound equivalents* that are similar to collocations. (Nominal) non-idiomatic *compound equivalents* are nominal expressions consisting of a noun that functions as head of the whole phrase and another part which functions as modifier of the head noun. The second part may be another noun, an adjective, or a prepositional phrase.

If at least one element is transparent/used in its primary meaning and interpretable as *base*, the meaning of the whole expression is transparent and (semi-)deducible from the meanings of the single elements. Other compounds instead, such as *acqua passata* (lit. "past water", Engl. *water under the bridge*, Germ. *Schnee von gestern*) and *casco blu* (lit. "blue helmet", which means 'UN soldier'), function in a similar way as idioms. They are annotated as *refphras_id* (chapter 1.1.2.2).

Note:

While in the case of *collocations* the *collocator* usually refers to a specific quality or characteristic of the *base* (e.g. *curva a gomito*, *febbre da cavallo*, *amico per la pelle* etc.), in the case of *compound equivalents* the determinant element typically serves to indicate a certain subclass or type of the referent indicated by the determined element (e.g. *occhiali da sole* = special type of *occhiali*, *letto a castello* = special type of *letto* etc.). This distinction, however, is very delicate and not always possible (cf. *birra chiara/scura*, where it is not clear if it is a characteristic of the beer to be 'chiara' or 'scura' or if the adjectives refer to a subclass). Usually, *compound equivalents* can be translated into German with morphological compounds, e.g. *Skipiste*, *Stockbett*, *Schlafwagen* etc.

Corpus examples from LEKO_Kolipsi: *vacanze di Natale*, *ufficio turistico*, *hotel wellness*, *negozio di frutta e verdura*, ...

```
>> refphras_notid_type= "compound_eq"
```

1.2.1.3 Syntagmatic verbs (**syntagm_verb**)

Syntagmatic verbs, i.e. verbs consisting of two words, are constructions which have a compositional meaning (cf. Simone 1997; Venier 1996), and in which one of the words is usually an adverb indicating the concrete or abstract direction of an action.

Corpus examples from LEKO_Kolipsi: *è andata via*, *butato via*, *andava (le righe) su e giù*, *aveva messo (tutto) fuori*, *pensarci sopra*, *che si aveva portato dietro (in una tasca)*, ...

```
>> refphras_notid_type= "syntagm_verb"
```

1.2.1.4 Adverbial expressions (**adverb_expr**)

Adverbial expressions which function as non- or semi-idiomatic referential phrasemes are prefabricated adverbial expressions consisting of more than one word that have a (more or less) transparent meaning.

Non-idiomatic or semi-idiomatic pair formulas corresponding to adverbial expressions are also included.

The single words of the expression are usually used in the same way and order.

Note:

If they are fully idiomatic, they are classified as idiomatic referential phrasemes (cf. chapter 1.1.2.4).

If an adverbial expression or pair formula itself is part of a collocation or a compound equivalent, it is annotated as collocation or compound equivalent, e.g. *raccontare per filo e per segno* (collocation), *ragazza alla pari* (compound equivalent).

Category expressions such as the following are excluded:

- *due giorni fa, una settimana fa, più tardi, all'inizio, tutto il giorno, alle dieci*
→ because they follow certain rules applicable to more combinations.
- *la sera, il venerdì, l'anno dopo, l'estate scorsa*
→ rules valid for more than one word combination (wrong use is due to wrong use of single words
– this does not regard the level of lexical combinations in the strict sense of the term)

In contrast, an example like *oggi come oggi* is included in the category, because it is a unique combination, not explainable by rules, that must be learned separately as a lexical unit.

Corpus examples from LEKO_Kolipsi: *l'altra volta (che), prima di tutto, piano piano, ...*

```
>> refphras_notid_type= "adverb_expr"
```

1.2.2 Idiomatic referential phrasemes (refphras_id)

Idiomatic referential phrasemes are referential phrasemes that have gained idiomatic (figurative) meaning.

We divide this category into:

- Idioms
- Idiomatic compound equivalents
- Idiomatic syntagmatic verbs
- Idiomatic adverbial expressions

```
>> fs_type= "refphras_id"
```

1.2.2.1 Idioms (**idiom**)

An idiom is a combination of words that has a figurative meaning, i.e. its entire meaning is not compositional and does not correspond to the sum of the meanings of the single constituents (although in some cases an idiom might be transparent, i.e. the “image” it is based on might be clear; cf. Burger 2007: 98ff.).

According to Burger (2007), idioms are:

- lexically constraint combinations
- fully idiomatic (note that only partly idiomatic FS are annotated as collocations)
- not complete sentences

Distinction from collocations: We used the monolingual dictionary Treccani as a reference for unclear cases, annotating words as idiomatic when they are marked with *fig.* (figurative meaning) in the dictionary. If there is no reference to a figurative meaning (of a specific word) in the dictionary, the word is NOT idiomatic but has distinct meanings that are listed separately (cf. Nesselhauf 2005: 54). We annotate semi-idioms (in which at least one content-relevant word is used in its literal meaning) as collocations and not as idioms (e.g. *numero verde*; cf. chapter 1.1.1.1).

A sequence of words may have an idiomatic as well as a free/literal meaning. In this case it is annotated as an idiom only if it is used in its idiomatic meaning (e.g. *jemandem einen Korb geben / dare un bidone a*

qualcuno: literal meaning = when you really give a basket to someone, idiomatic meaning = ‘to refuse someone’).

In addition to idioms in a strict sense, we included some closely related concepts that are listed and described below.

- **idiomatic kinegrams:** *storcere il naso* (in the figurative meaning of ‘to look down one's nose at so./sth.’)
- **idiomatic (verbal) multiword expressions:** *rendersi conto, farsi vivo*

Corpus examples from LEKO_Kolipsi: *rendersi conto, mi sono resa conto, farsi vivo, ...*

```
>> refphras_id_type= "idiom"
```

1.2.2.2 Compound equivalents (**compound_eq**)

The difference to non- or semi-idiomatic compound equivalents (cf. chapter 1.1.1.2) is the idiomaticity of the sequence, i.e. in this category (1.1.2.2) we only annotate nominal compound equivalents with an idiomatic meaning. Verbal compound equivalents are not annotated in order to be consistent with their non-idiomatic counterparts.

Corpus examples from LEKO_Kolipsi: *sacco di cose, un prezzo unico, ...*

```
>> refphras_id_type= "compound_eq"
```

1.2.2.3 Syntagmatic verbs (**syntagm_verb**)

The difference to non- or semi-idiomatic syntagmatic verbs (cf. chapter 1.1.1.3) is the idiomaticity of the sequence, i.e. in this category (1.1.2.3) we only annotate syntagmatic verbs with an idiomatic meaning. Syntagmatic verbs are idiomatic when they don't have a compositional meaning. They can, however, be transparent if they are based on comprehensible metaphors (cf. Venier 1996: 153).

Note:

We did not consider complex verbs consisting of verb + adjective such as *far chiaro*.

Corpus examples from LEKO_Kolipsi: *l'ha mandata giù, ...*

```
>> refphras_id_type= "syntagm_verb"
```

1.2.2.4 Adverbial expressions (**adverb_expr**)

See description for non/semi-idiomatic adverbial expressions. The difference to the latter (cf. chapter 1.1.1.4) is the idiomaticity of the sequence. As with non/semi-idiomatic adverbial expressions, this category also includes pair formulas if they function in a similar way as idioms.

Note:

If an adverbial expression or pair formula itself is part of an idiom, it is annotated as idiom and NOT as idiomatic adverbial expression, e.g. *essere al verde, pagare in contanti*.

Corpus examples from LEKO_Kolipsi: *a questo punto*

```
>> refphras_id_type= "adverb_expr"
```

1.3 Communicative phrasemes (**commphras**)

Communicative phrasemes (cf. *routine formulas*, Burger 2007: 56) fulfil specific tasks for the creation, realization, and conclusion of communicative acts. As pragmatic devices they help to perform recurring communicative acts (i.e. communicative routines). In many cases, they lost their literal meaning, i.e. they are semantically bleached, but usually not idiomatic. In contrast to collocations or idioms, they often have a higher degree of variability regarding their form; thus, sometimes different grammatical or lexical variants are possible. Formally, they can be parts of sentences as well as whole sentences.

Communicative phrasemes include expressions like the following:

- **Routine formulas:** *Come stai?, Non c'è male., Tutto a posto., Non so che dirti ...*
- **Guide of attention / exclamations:** *Guarda..., Ascolta...*
- **Greetings, farewell, addressing:** *Gentile... / Caro... / Ciao, Cordiali saluti*
- **Metacommunicative comments:** *come dico sempre ...*
- **Expression of meaning:** *secondo me ..., a mio avviso...*
- **Indication of vagueness:** *per quanto ne so io ...*
- **Expression of astonishment/surprise:** *Accipicchia!, Cavolo!*
- **Pair formulas** which correspond to communicative phrasemes: *cin cin, punto e basta*
- **Common places:** *Quel che è fatto è fatto.*
- **Dicta and citations** (Germ. "geflügelte Worte"): *Non ha prezzo.* (advertisement for Master card)
- **Proverbs** (Ital. "proverbi", Germ. "Sprichwörter"): *Chi dorme non piglia pesci., A caval donato non si guarda in bocca., Si impara per la vita (non per la scuola).*

Note:

There may be cases of communicative phrasemes where single words fulfil the same function as a pre-fabricated sequence (e.g. *Ciao!* or *Cavolo!*, cf. Stein 2007: 226).

When a phraseme can be annotated either as collocation/idiom or as communicative phraseme, the main function of the phraseme (referential or communicative) was chosen.

Following Stein (2007: 226ff.), we split communicative phrasemes into:

- `situational` ones, i.e. communicative phrasemes that are bound to specific situations
- `not_situational` ones, i.e. communicative phraseme that are not situationally bound

```
>> fs_type= "commphras"
```

1.3.1 Situationally bound communicative phrasemes (**situational**)

Situationally bound communicative phrasemes refer to a specific pattern of action (cf. Germ. *Handlungsmuster*) while fulfilling social functions. In other words, they serve to fulfil specific speech acts in a conventional way. Examples are the start or end of a conversation, e.g. greeting, farewell, addressing, or apologies, formulas of approval/affirmation, wishes etc.

Corpus examples from LEKO_Kolipsi: *era lo stesso, non c'era niente da fare, Che ne pensi? È tutto d'aposto, Come stai?, Spero bene!, Io sto bene., Cari saluti e auguri!, va bene? ...*

```
>> commphras_type= "situational"
```


1.3.2 Not situationally bound communicative phrasemes (**not_situational**)

Not situationally bound communicative phrasemes fulfil primarily text organising (and interaction organising) functions. They serve to create metacommunicative references (text internal references), to guide attention, to ensure understanding, facilitate interaction, manage communication, establish relations, or to comment utterances.

Corpus examples from LEKO_Kolipsi: *secondo me, per me, come sai, per me è lo stesso, ...*

```
>> commphras_type= "not_situational"
```

1.4 Structural phrasemes (**structphras**)

Structural phrasemes have the only function of creating relations within a text (whereas collocations/idioms refer to specific entities or actions of reality, and communicative phrasemes have a function regarding communicative activities, cf. Burger 2007: 36). Following Burger, we assume that these relations are of grammatical nature (created e.g. by connectors, prepositions consisting of more than one word etc.).

Further, we divide structural phrasemes into polirematiche preposizionali and polirematiche congiunzionali (terms adopted also by Voghera 2004, 67f.).

```
>> fs_type= "structphras"
```

1.4.1 Polirematiche preposizionali (**preposition**)

These are “structural phrasemes in the form of polylexical prepositions” (Konecny/Abel/Autelli/Zanasi 2016: 539; cf. also Voghera 2004: 67f.).

Corpus examples from LEKO_Kolipsi: *in mezzo a, al prezzo di, nel cuore di, nei pressi di, per colpa di, ...*

```
>> structphras_type= "preposition"
```

1.4.2 Polirematiche congiunzionali (**conjunction**)

These are “structural phrasemes in the form of polylexical or correlative conjunctions or connectors” (Konecny/Abel/Autelli/Zanasi 2016: 539; cf. also Voghera 2004: 68).

Corpus examples from LEKO_Kolipsi: *dato che, visto che, da parte mia, d'altra parte, ...*

```
>> structphras_type= "conjunction"
```

1.5 Non-existent phrasemes (**notexistent**)

Non-existent phrasemes are uncommon and not codified lexical combinations that are usually based on language transfer, which are used instead of an existing FS or word. We divide this annotation category into not acceptable and acceptable combinations based on the judgement of the annotators.

```
>> fs_type= "notexistent"
```

1.5.1 Not acceptable

Corpus examples from LEKO_Kolipsi: *viene anché completamente diverse* (Germ: *es kommt alles ganz anders*) instead of *sarà qualcosa di completamente diverso*, *il foglio di cassa* (Germ: *der Kassenzettel*) instead of *lo scontrino*

```
>> notexistent_type= "not_accetable"
```

1.5.2 Acceptable

Corpus examples from LEKO_Kolipsi: *andiamo shopping* (Germ: *shoppen gehen*) instead of *andiamo a fare shopping*, *così lo ho dentro nella mia testa* (Germ: *etwas im Kopf haben*) instead of *così me lo immagino*

```
>> notexistent_type= "accetable"
```

2 Lexical errors (error)

Error tags were assigned according to the forms incorrectly produced by a learner. Furthermore, we adopted a criterion of philological fidelity to the source text when reconstructing the target form. This means that in case the target form can be associated to two or more alternative forms, the expression formally (and semantically) most similar to the original one was chosen.

The annotations of lexical errors include:

- semantic errors, regarding the denotation (*semdenot*), connotation (*semconn_attr*) or register (*semconn_use_formal*) of the used FS;
- formal errors (*form*), when the FS is used formally incorrectly, for example if it is incomplete, disrupted, or otherwise altered from its target language form.

Semantic error annotations indicate that the FS is used in a semantically wrong way in a given context and therefore affects the comprehensibility of the proposition.⁴ Cases in which only a part of the FS is changed or omitted are considered formal errors.

An overview of error annotations for FS and their subcategories can be found in figure 2.

⁴ This may affect especially idioms as they usually have a meaning as if they were a single word (cf. Burger 2007: 31). As far as communicative phrasemes are concerned, their meaning can be characterized as their social function (e.g. *Ciao*) or their text or interaction organizing function (e.g. *in altri termini, come detto sopra, sai cosa*) (cf. Stein 2007: 228, 230); thus, semantic errors are errors regarding the function of FS, e.g. if *in altri termini* is used and no explanation follows, the text organizing function doesn't work (semantic error).

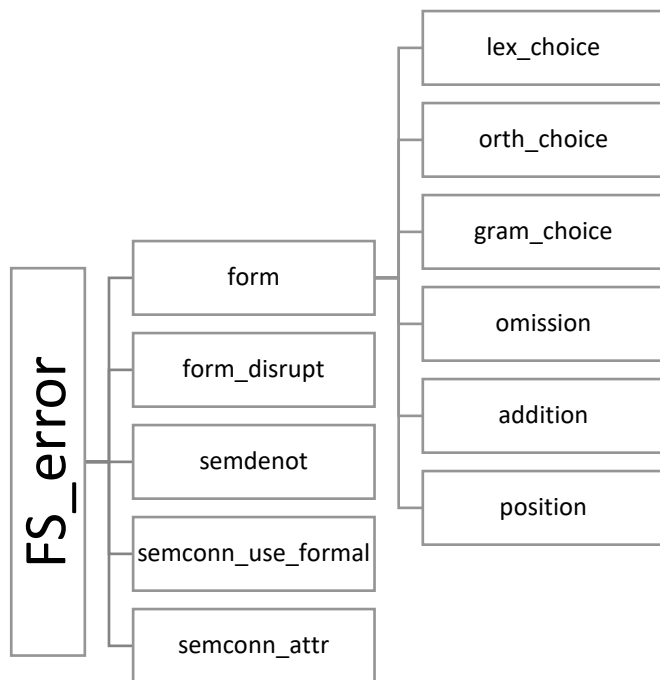


Figure 2: FS error types and their subcategories

2.1 Formal errors (**form**)

Formal errors are incorrect regarding formal elements, i.e., the target form is modified in some incorrect way. We distinguish the following cases.

- omission
- addition
- lex_choice
- orth_choice
- gram_choice
- position

```
>> error= "form"
```

2.1.1 FS is incomplete, omission of one or more words (**omission**)

Omission errors are annotated on the token preceding the missing element.

Corpus examples from LEKO_Kolipsi: *tutti due* instead of *tutti e due*, *tuo amico* instead of *il tuo amico* (as greeting)

```
>> form_error= "omission"
```

2.1.2 Addition of one or more words (**addition**)

Addition errors are annotated on the added token(s).

Corpus examples from LEKO_Kolipsi: *ti aspetto alla tua risposta* instead of *aspetto la tua risposta*, *per tante volte* instead of *tante volte*

```
>> form_error= "addition"
```

2.1.3 Substitution/wrong choice of one or more words (**lex_choice**)

This annotation is used when a lexical part of a FS (one or more words) is substituted with another/others. Such cases are annotated with the error tag **lex_choice**. Substitutions like these may also be connected to a mixture of parts of different FS and hence to phenomena of contamination (cf. Germ. **aus der Reihe fallen* as result of a mixture of *aus der Reihe tanzen* and *aus dem Rahmen fallen*).

Corpus example from LEKO_Kolipsi: *un giro nella bicicletta* (instead "of un giro in bicicletta") *In aspetta di una sua risposta* (instead of "in attesa di una Sua risposta")

```
>> form_error= "lex_choice"
```

2.1.4 Substitution of orthographic features (**orth_choice**)

This annotation is used when orthographic features of a FS are substituted with others, e.g. use of a small instead of a capital letter. Such cases are annotated with the error tag **orth_choice**.

Corpus example: *fare delle passeggiate, verdura fresha*

```
>> form_error= "orth_choice"
```

2.1.5 Substitution of morpho-grammatical features (**gram_choice**)

This annotation is used when morpho-grammatical features of a FS are substituted with others, e.g. use of plural form instead of singular, definite article instead of indefinite article etc. Such cases are annotated with the error tag **gram_choice**. With this tag, we annotated also grammatical errors that are context-dependent, e.g. choice of the wrong tense or concordance errors.

Corpus examples from LEKO_Kolipsi: *cosa raccontami?* instead of *cosa mi racconti?*, *fare belle passeggiata* instead of *fare belle passeggiate*

```
>> form_error= "gram_choice"
```

2.1.6 Wrong position of one or more words (**position**)

With this tag we annotate permutations of word order. The annotation span is the sequence containing the wrongly placed token(s).

Corpus examples from LEKO_Kolipsi: *novità splendide* instead of *splendide novità*, *per la mia parte* instead of *da parte mia*, *avere un tempo bello* instead of *aver un bel tempo*

```
>> form_error= "position"
```

2.2 Completely disrupted form of an FS (form_disrupt)

This tag is used if a FS is disrupted by so many formal errors that it is rather difficult or impossible to reconstruct the FS and to define what is omitted, substituted etc. (i.e. the target language dimensions are not identifiable or traceable anymore).

The FS is semantically correct and acceptable in the specific context although, due to the accumulation of errors, its comprehensibility may be hampered or difficult.

Corpus examples from LEKO_Kolipsi: *poi ha fatto Maria suicido* instead of *poi Maria si è suicidata*, *abbiamo arrampicato tante montagne* instead of *ci siamo arrampicati su tante montagne*

```
>> error= "form_disrupt"
```

2.3 Semantic errors regarding denotation (semdenot)

Semantic denotation errors regard the following two cases (cf. Konecny/Autelli/Zanasi/Abel 2018: 121):

- existent FS to which is assigned a wrong (denotative) meaning, sometimes (but not necessarily based on transfer from other languages), e.g. *far sicuro che* (cf. Germ. *sicherstellen, dass*) instead of *assicurarsi che*;
- non-existent FS (cf. chapter 1.4), whose use is often based on transfer from other languages.

Corpus examples from LEKO_Kolipsi: *lo ho pensato un pò e mi sono messo d'accordo che* instead of [...] *e sono giunto alla conclusione che*; **puoi prenderti un bagno di sole* (cf. Germ. *ein Sonnenbad nehmen*) instead of *puoi abbronzarti*;

```
>> error= "semdenot"
```

2.4 Semantic errors regarding connotation (semconn_att)

The tag `semconn_att` is used if the FS does not have an appropriate connotative meaning⁵ with regard to the speaker's/writer's (emotional) attitude. Such attitudes might be the following:

- negative, pejorative, derogatory
- ironic
- euphemistic
- jocular
- others

In order to check whether such a connotative meaning of the FS exists, we used the monolingual dictionary Treccani as a reference tool.

Corpus example from LEKO_Kolipsi: *vorrei salutare con onore*

⁵ For the categorisation of connotative meanings cf. Schippan (1992: 155f.). Note that the two categories "attitude" and "context of use" identified by Schippan have been put together and thus correspond to one single category in this contribution.

```
>> error= "semconn_att"
```

2.5 Semantic errors regarding register (`semconn_use_formal`)

`semconn_use_formal` is used if a FS has not an appropriate (connotative) meaning in a specific context. It refers to diasystematic dimensions such as the following:

- diaphasic dimension (regarding style/register, e.g. elevated, poetic, formal, informal, colloquial, vulgar, slangy, ...)
- diastratic dimension (e.g. technical terms, youth language)
- diachronic dimension (e.g. archaic)
- diatopic dimension (dialectal; pluricentric: standard varieties, ...).

Corpus example from LEKO_Kolipsi: *cordiali saluti* in a letter to a friend

```
>> error= "semconn_use_formal"
```

3 References

- Burger, Harald (2007): *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Berlin: Schmidt [Grundlagen der Germanistik; 36].
- Deveci, Tanju (2004): "Why and How to Teach Collocations." *English Teaching Forum* 42, 16-20.
- Ettinger, Stefan (2013): "Aktiver Phrasemgebrauch und/oder passive Phrasemkenntnisse im Fremdsprachenunterricht. Einige phraseodidaktische Überlegungen." In: Gonzáles Rey, María Isabel (ed.): *Phraseodidactic Studies on German as a Foreign Language. / Phraseodidaktische Studien zu Deutsch als Fremdsprache*. Hamburg: Dr. Kovač [Lingua – Fremdsprachenunterricht in Forschung und Praxis; 22], 11-30.
- Hausmann, Franz-Josef (1984): "Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen." *Praxis des neusprachlichen Unterrichts* 31/4, 395-406.
- Hausmann, Franz-Josef (2004): "Was sind eigentlich Kollokationen?" In: Steyer, Kathrin (ed.): *Wortverbindungen – mehr oder weniger fest*. Berlin: de Gruyter, 309-334.
- Konecny, Christine; Autelli, Erica (2013): "Learning Italian phrasemes through their conceptualizations." In: Konecny, Christine; Hallsteinsdóttir, Erla; Kacjan, Brigita (eds.): *Phraseologie im Sprachunterricht und in der Sprachendidaktik / Phraseology in language teaching and in language didactics*. Maribor: Mednarodna založba Oddelka za slovanske jezike in književnosti, Filozofska fakulteta [Zora; 94], 117-136.
- Konecny, Christine; Abel, Andrea; Autelli, Erica; Zanasi, Lorenzo (2016): "Identification and Classification of Phrasemes in an L2 Learner Corpus of Italian." In: Corpas Pastor, Gloria (ed.): *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives (Full papers)*. Geneva (Switzerland): Tradulex, 533-542.
- Konecny, Christine; Autelli, Erica; Zanasi, Lorenzo; Abel, Andrea (2018): "**Queste vacanze divantano il martello!* Transferphänomene beim Gebrauch formelhafter Sequenzen im Italienischen seitens deutschsprachiger L2-Lerner/innen und Möglichkeiten ihrer Klassifizierung." In: Zybatow, Lew; Petrova, Alena (eds.): *Sprache verstehen, verwenden, übersetzen. Akten des 50. Linguistischen Kolloquiums, Innsbruck, 03.-05.09.2015*. Frankfurt a.M. etc.: Lang [Linguistik International; 42], 115-126.

- Nesselhauf, Nadja (2005): *Collocations in a Learner Corpus*. Amsterdam / Philadelphia: Benjamins [Studies in Corpus Linguistics; 1].
- Paquot, Magali; Granger, Sylviane (2012): "Formulaic Language in Learner Corpora." *Annual Review of Applied Linguistics* 32, 130-149.
- Schippan, Thea (1992): *Lexikologie der deutschen Gegenwartssprache*. Tübingen: Niemeyer.
- Simone, Raffaele (1997): "Esistono verbi sintagmatici in italiano?" In: De Mauro, Tullio; Lo Cascio, Vincenzo (eds.): *Lessico e grammatica. Teorie linguistiche e applicazioni lessicografiche. Atti del convegno interannuale della Società di Linguistica Italiana. Madrid, 21-25 febbraio 1995*. Roma: Bulzoni [Pubblicazioni della SLI; 36], 155-170.
- Stein, Stephan (2007): "Mündlichkeit und Schriftlichkeit aus phraseologischer Perspektive." In: Burger, Harald; Dobrovolskij, Dmitrij; Kühn, Peter; Norrick, Neal R. (eds.): *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung. / Phraseology. An International Handbook of Contemporary Research*. Vol. 1. Berlin: De Gruyter [Handbücher zur Sprach- und Kommunikationswissenschaft; 28], 220-236.
- Venier, Federica (1996): "I verbi sintagmatici." In: Blumenthal, Peter; Rovere, Giovanni; Schwarze, Christoph (eds.): *Lexikalische Analysen romanischer Sprachen*. Tübingen: Niemeyer [Linguistische Arbeiten; 353], 149-156.
- Voghera, Miriam (2004): "Composizione: Polirematiche." In: Grossmann, Maria; Rainer, Franz (eds.): *La formazione delle parole in italiano*. Tübingen: Niemeyer, 56-69.