

# Annotation guidelines for inherently reflexive verbs

## PARSEME shared task on automatic detection of verbal MWEs

Marie Candito, Carlos Ramisch, Silvio Ricardo Cordeiro, Carla Parra Escartín, Carlos Herrero, Verginica Mititelu, Johanna Monti, Manfred Sailer, Agata Savary

Version 2 - 18 July 2016

**DISCLAIMER**: this guide was written for German (DE), Spanish (ES), French (FR), Italian (IT), Portuguese (PT), Polish (PL) and Romanian (RO). This version was adapted from [version 1](#).<sup>1</sup>

**Reflexive clitics (RCLI)** are clitic pronouns that refer to the subject of the verb, like *oneself* in English. They are very common in many languages and play several semantic roles depending on the context, as detailed below. **Reflexive verbs (REFLV)**, sometimes also called pronominal verbs, are formed by a full verb combined with a RCLI, although the clitic does not always have a reflexive meaning.

REFLV can be categorized into different classes, some of which should be annotated as verbal MWEs. Namely, we will only annotate a REFLV as an **inherently reflexive verb (IRefIV)** when (a) it never occurs without the clitic, or (b) the REFLV and non-reflexive versions have clearly different senses or subcategorization frames. We start by listing rapidly the various categories of **REFLV** before providing tests to decide whether to annotate a given occurrence as **IRefIV**.

- **Inherently reflexive ⇒ ANNOTATE as IRefIV**

- The verb without the RCLI does not exist
  - DE *sich schämen* (be ashamed), *sich wundern* (wonder)
  - ES *suicidarse* (suicide), *abstenerse* (to abstain)
  - FR *s'évanouir* (faint), *se suicider* (suicide)
  - IT *suicidarsi* (suicide), *arrabbiarsi* (get angry)
  - PL *dowiedzieć się* (find out), *bać się* (be afraid)
  - PT *queixar-se* (complain), *abster-se* (abstain)
  - RO *a se teme* (be afraid)
- The verb without the RCLI does exist, but has a very different meaning
  - DE *sich enthalten* ≠ *enthalten* (abstain ≠ contain), *sich (um etw.) handeln* ≠ *handeln* (be ≠ handle)
  - ES *recoger* ≠ *recogerse* (gather ≠ go home), *empeñar* ≠ *empeñarse* (pawn ≠ insist)
  - FR *s'apercevoir* ≠ *apercevoir* (realize ≠ see), *s'agir* ≠ *agir* (be ≠ act)
  - IT *riferire* ≠ *riferirsi* (report, tell ≠ refer)
  - PL *znajdować* ≠ *znajdować się* (find ≠ be), *radzić* ≠ *radzić sobie* (advise ≠ manage)
  - PT *encontrar-se* ≠ *encontrar* (be ≠ meet), *referir-se* ≠ *referir* (concern ≠ refer)

<sup>1</sup> Color codes in this document still need to be adapted to the main guidelines.

- RO *a îndura ≠ a se îndura* (*suffer ≠ have the heart*)
- **Reciprocal ⇒ NOT ANNOTATED**
  - The RCLI has a sense of mutuality:
    - DE *sich küssen* (*kiss each other*), *sich treffen* (*meet each other*)
    - ES *besarse* (*kiss each other*), *verse* (*see each other*)
    - FR *s'embrasser* (*kiss each other*), *se rencontrer* (*meet each other*)
    - IT *baciarsi* (*kiss each other*)
    - PL *całować się* (*kiss each other*), *spotykać się* (*meet each other*)
    - PT *cumprimentar-se* (*greet each other*), *ver-se* (*see each other*)
    - RO *a se saluta* (*greet each other*)
- **Reflexive ⇒ NOT ANNOTATED**
  - The RCLI marks the reflexive or reciprocal construction, that is, the clitic plays the role of *self* in English
    - DE *sich waschen* (*wash oneself*), *sich kratzen* (*scratch oneself*)
    - ES *mirarse* (*look at oneself*), *vestirse* (*dress oneself*)
    - FR *se laver* (*wash oneself*), *se parler* (*talk to oneself*)
    - IT *lavarsi* (*wash oneself*), *vestirsi* (*dress oneself*)
    - PL *myć się* (*wash oneself*), *drapać się po głowie* (*scratch oneself on the head*)
    - PT *apressar-se* (*hurry oneself*), *vestir-se* (*dress oneself*)
    - RO *a se spăla* (*wash oneself*)
- **Body part, also called possessive reflexive ⇒ NOT ANNOTATED**
  - Specific type of reflexive use in which the direct object is a body part or, more generally, an inalienable part of the subject
 

Note: the RCLI regularly alternates with a non-reflexive dative clitic, but does not alternate with an indirect PP

    - DE *sich das Bein brechen* (*lit. RCLI the leg break, 'break one's leg'*)
    - ES *rascarse el brazo* (*lit. scratch.RCLI the arm, 'scratch one's arm'*)
    - FR *se gratter la tête* (*lit. RCLI scratch the head, 'scratch one's head'*)
    - IT *grattarsi la testa* (*lit. RCLI scratch the head, 'scratch one's head'*)
    - PL *myć sobie nogi* (*lit. wash RCLI.DAT the feet, 'wash one's feet'*)
    - PT impossible, uses possessive instead
    - RO *a-și rupe mâna* (*RCLI.DAT break arm, 'break one's arm'*)
- **Middle with preverbal subject, also called synthetic passive ⇒ NOT ANNOTATED**
  - The clitic marks a regular syntactic alternation for transitive verbs. Just like in regular passive alternation, the direct object of the transitive version appears as the subject of the REFLV version, and thus the verb agrees with the subject.
  - Differently from inchoative (see below), the subject of the transitive version is absent in the REFLV version but it exists necessarily, though it is underspecified
    - DE *die Häuser verkaufen sich gut* (*lit. the houses sell RCLI well, 'the houses sell well'*)
    - ES *las casas se venden bien* (*lit. the houses RCLI sell well, 'the houses sell well'*)
    - FR *les pots se vendent bien* (*lit. the pots RCLI sell well, 'the pots sell well'*)
    - IT *le case si affittano* (*lit. the houses RCLI rent, 'the houses are rented'*)
    - PL *domy dobrze się sprzedają* (*lit. houses sell.PL RCLI well, 'houses sell well'*)

- PT *as casas se vendem bem* (lit. *the houses RCLI sell well, 'the houses sell well'*)
  - RO *casele se vând bine* (lit. *houses-the RCLI sell well, 'houses sell well'*)
- **Middle with postverbal subject, also called synthetic passive ⇒ NOT ANNOTATED**
  - In some languages, middle alternation with preverbal subject sounds unnatural and middle alternation with postverbal subject is preferred. Depending on the languages, it is viewed as a postverbal subject (ES, PL, PT, RO) or as an object which agrees with the unaccusative verb form (IT). Middle alternation with postverbal subject is impossible in FR and DE.
  - Note: To simplify the tests, we treat the postverbal argument as a subject in all the following examples
    - ES *se alquilan casas* (lit. *RCLI rent houses, 'people rent houses'*)
    - IT *si affittano case* (lit. *RCLI rent houses, 'people rent houses'*)
    - PL *dobrze sprzedają się te domy* (lit. *well sell RCLI these houses, 'these houses sell well'*)
  - Note: Polish is a relatively free word-order language and a postverbal subject is a regular (even if stylistically marked) alternation
  - PT *alugam-se casas* (lit. *rent-RCLI houses, 'people rent houses'*)
  - RO *se vând bine apartamentele din blocurile noi* (lit. *RCLI sell well apartments-the from blocks-the new, 'Apartments from new blocks sell well'*)
- **Impersonal ⇒ NOT ANNOTATED**
  - The RCLI marks an impersonal verb alternation possible for various transitivity classes, depending on the language: only transitive verbs (FR), only intransitive verbs with manner adjuncts (DE), preferably intransitive but tolerated for transitive verbs (PT), either transitive or intransitive verbs (IT, ES, RO, PL)
  - There is no noun phrase before the verb (empty subject slot), the presence of the RCLI indicates a verb interpreted with a generic and underspecified subject
  - The verb is in third person singular, even when the object is plural
    - DE *hier tanzt es sich gut* (lit. *here dances it RCLI well, 'people dance well here'*)
    - ES *se busca a actores* (lit. *RCLI searches to actors, 'people look for actors'*)
    - ES *se trabaja mejor aquí* (lit. *RCLI works better here, 'people work better here'*)
    - FR *il se dit des bêtises* (lit. *it RCLI says silly things, 'people say silly things'*)
    - IT *si lavora troppo* (lit. *RCLI works too much, 'people work too much'*)
    - IT *si affitta molte case* (lit. *RCLI rents many houses, 'people rent many houses'*)
    - PL *za dużo się pracuje* (lit. *too much RCLI works, 'people work too much'*)
    - PL *bzdury się opowiada* (lit. *nonsense RCLI tells, 'people tell nonsense'*)
    - PT *dorme-se muito* (lit. *sleeps-RCLI much, 'people sleep a lot'*)
    - PT *conta-se histórias* (lit. *tells-RCLI stories, 'people tell stories'*)
  - Note: transitive impersonal is considered wrong by traditional PT grammars but it is found in corpora
  - RO *se lucrează până târziu*. (lit. *RCLI works until late, 'people work until late'*)
  - RO *se citesc multe romane* (lit. *RCLI read many novels 'people read many novels'*)
- **Inchoative ⇒ NOT ANNOTATED**
  - Similar to middle, but the RCLI marks a less productive syntactic alternation:
    - the direct object of the transitive version appears as subject of the REFLV
    - the subject of the transitive version is not only absent, it is also semantically unclear or nonexistent
      - DE *die Tür öffnet sich* (the door opens)

- ES *la puerta se abrió* (the door opened)
- FR *la porte s'est subitement ouverte* (the door suddenly opened)
- IT *la porta si apre* (the door opens)
- PL *drzwi się otwierają* (the door opens)
- PT *o vaso se quebrou* (the vase broke)
- RO *mașina s-a stricat* (the car broke down)

## Tests for annotating inherently reflexive verbs

We suggest some tests to decide when to annotate an IRefIV. Before describing the specific tests, we provide a synthetic decision tree that guides their application:

### → APPLY test 1 - Inherent

- ↳ if YES ⇒ ANNOTATE as IRefIV
- ↳ if NO ⇒ APPLY test 2 - Diff-sense
  - ↳ if YES ⇒ ANNOTATE as IRefIV
  - ↳ if NO or UNSURE ⇒ APPLY test 3 - Diff-subcat
    - ↳ if YES ⇒ ANNOTATE as IRefIV
    - ↳ if NO ⇒
      - ↳ if verb has no subject ⇒ APPLY test 4 - Impersonal
        - ↳ if YES ⇒ DO NOT ANNOTATE
        - ↳ if NO ⇒ ANNOTATE as IRefIV
      - ↳ if verb has a subject ⇒ APPLY test 5 - Middle-Inchoative
        - ↳ if YES ⇒ DO NOT ANNOTATE
        - ↳ if NO APPLY test 6 - Reflexive
          - ↳ if YES ⇒ DO NOT ANNOTATE
          - ↳ if NO or UNSURE ⇒
            - ↳ if subject is SINGULAR ⇒ APPLY test 7 - Reflexive-mutual
              - ↳ if YES ⇒ DO NOT ANNOTATE
              - ↳ if NO ⇒ ANNOTATE as IRefIV
            - ↳ if subject is PLURAL ⇒ APPLY test 8 - Reciprocal
              - ↳ if YES ⇒ DO NOT ANNOTATE
              - ↳ if NO ⇒ ANNOTATE as IRefIV

The tests below should discriminate between IRefIV (which are verbal MWEs) and other REFLV cases (which are regular and productive constructions). Language-specific sections of the guidelines should decline/translate the tests below.

### 1. Inherent

Does the verb only exist with the RCLI and never occurs without it?

- **If YES: annotate as IRefIV**
  - DE *sich schämen* - \**schämen* (be ashamed), *sich wundern* - \**wundern* (wonder)
  - ES *suicidarse* - \**suicidar* (suicide), *abstenerse* - \**abstener* (abstain)
  - FR *s'évanouir* - \**évanouir* (faint), *se suicider* - \**suicider* (suicide)

- IT *suicidarsi* - \**suicidare* (suicide)
    - PL *dowiedzieć się* - \**dowiedzieć* (find out), *bać się* - \**bać* (be afraid)
    - PT *queixar-se* - \**queixar* (complain), *abster-se* - \**abster* (abstain)
    - RO *a se teme* - \**a teme* (be afraid)
  - **IF NO: next test**
2. **Diff-sense**
- Given the same verb without the RCLI, are all of its meanings clearly different from the REFLV form?
- **If YES: annotate as IRefIV**
    - DE *sich verstehen* ≠ *verstehen* (get along well ≠ understand)
    - ES *recoger* ≠ *recogerse* (pick up, gather ≠ go home)
    - FR *s'apercevoir* ≠ *apercevoir* (realize ≠ see), *s'agir* ≠ *agir* (be ≠ act)
    - IT *riferire* ≠ *riferirsi* (report, tell ≠ refer)
    - PL *znajdować się* (to find oneself, to be)
    - PT *encontrar-se* ≠ *encontrar* (be ≠ meet), *referir-se* ≠ *referir* (concern ≠ refer)
    - RO *a se îndura* ≠ *a îndura* (have the heart to ≠ suffer)
  - **If NO or you are unsure: next test**
3. **Diff-subcat**
- Given the same verb without the RCLI and with the same/similar meaning as the REFLV form, does it have a different subcategorization frame?
- **If YES: annotate as IRefIV**
    - DE *X verliert sich in Y*    *X verliert Y* (lit. X loses RCLI in Y, X loses Y)
    - ES *X se olvidó de Y*    *X olvidó Y* (lit. X RCLI forgot of Y, 'X forgot Y').
    - FR *X se confesse de Y*    *X confesse Y* (lit. X RCLI confesses of Y, 'X confesses Y')
    - IT *X si è dimenticato di Y*    *X ha dimenticato Y* (lit. X RCLI forgot of Y, 'X forgot Y')
    - PL *X tłumaczy się z pomyłek*    *X tłumaczy pomyłki* (lit. X explains SELF of errors, 'X explains errors')
    - PT *X se esqueceu de Y*    *X esqueceu Y* (lit. X RCLI forgot of Y, 'X forgot Y')
    - RO *X se gândește la Y*    *X gândește că Y* (X RCLI thinks of Y    X thinks that Y)
  - **If NO: next test**
4. **Impersonal, applicable when no preverbal noun phrase and verb in 3<sup>rd</sup> person singular**
- When you replace the RCLI by an underspecified subject such as *one* or *people*, does the sentence keep its meaning?
- **If YES: do not annotate as verbal MWE**
    - DE *hier tanzt es sich gut*    *hier tanzen die Leute gut* (people dance well here)
    - ES *se duerme mucho*    *las personas duermen mucho* (people sleep a lot)
    - ES *se busca a actores*    *la gente busca a actores* (people look for actors)
    - FR *il se dit des bêtises*    *les personnes disent des bêtises* (people say silly things)
    - IT *si dorme molto*    *le persone dormono molto* (people sleep a lot)
    - IT *si affitta molte case*    *le persone affittano molte case* (people rent many houses)
    - PL *pracuje się za dużo*    *ludzie pracują za dużo* (people work too much)
    - PL *opowiada się bzdury*    *ludzie opowiadają bzdury* (people tell nonsense)
    - PT *dorme-se muito*    *as pessoas dormem muito* (people sleep a lot)

*conta-se histórias* as pessoas contam histórias (people tell stories)

- RO *se lucrează până târziu* lumea lucrează până târziu (people work until late)  
*se citesc multe romane* lumea citește multe romane (people read many novels)

○ **If NO: annotate as IRefIV**

**5. Middle-Inchoative, applicable when there is a subject (either pre- or postverbal)**

When you move the subject to the object position, remove the RCLI and add a generic subject (*people, somebody*), thus building a transitive version, does it imply the REFLV version? In other words, *people/somebody V [to] X* *X REFLV*?

Note: logic implication is unidirectional. The REFLV version does not necessarily imply the transitive version

Note: depending on the meaning, either use *someone* or *people*: *people* has to be used when the middle version describes a generic habit

○ **If YES: do not annotate as verbal MWE**

- DE *man kann die Häuser gut verkaufen* die Häuser verkaufen sich gut (people can sell the houses well the houses sell well)  
*jemand öffnet die Tür* die Tür öffnet sich (somebody opens the door the door opens)
- ES *la gente cuenta historias* se cuentan historias (people tell stories stories are told)  
*alguien abrió la puerta* la puerta se abrió (somebody opened the door the door opened)
- FR *on vend bien ce produit* ce produit se vend bien (people sell this product well this product sells well)  
*quelqu'un ouvre la porte* la porte s'ouvre, (somebody opens the door the door opens)
- IT *qualcuno vende bene questo prodotto* questo prodotto si vende bene (someone people sells this product well this product sells well)  
*qualcuno apre la porta* la porta si apre (somebody opens the door the door opens)
- PL *ktoś sprzedaje te domy* te domy się sprzedają (somebody sells these houses these houses sell well)  
*ktoś otwiera drzwi* drzwi się otwierają (somebody opens the door the door opens)
- PT *alguém conta histórias* contam-se histórias (somebody tells stories stories are told)  
*alguém acalmou o menino* o menino se acalmou (somebody calmed the boy down the boy calmed down)
- RO *cineva spune glume.* se spun glume (somebody tells jokes jokes are told)  
*cineva a deschis u a* u a s-a deschis (somebody opened the door the door opened)

○ **If NO: next test**

**6. Reflexive, applicable when there is a subject**

When you replace the RCLI by *oneself only* or *to oneself only*, does it imply the REFLV version? In other words, *X V [to] himself only* *X REFLV*?

Note: The logical implication operator is unidirectional: The REFLV version does not necessarily imply that you apply

the action on yourself only

Note: The reflexive test is sometimes hard to apply, and can be substituted by the reflexive-mutual test

- **If YES: do not annotate as verbal MWE**
  - DE *Paul kratzt nur sich selbst*    *Paul kratzt sich* (*Paul scratches himself*)
  - ES *Pablo se lava a sí mismo*    *Pablo se lava* (*Paul washes himself*)
  - FR *Paul ne soigne que lui-même*    *Paul se soigne* (*Paul heals himself*)  
*Paul ne parle qu'à lui-même*    *Paul se parle* (*Paul talks to himself*)
  - IT *Paolo cura solo se stesso*    *Paolo si cura* (*Paul heals himself*)  
*Paolo parla solo a se stesso*    *Paolo si parla* (*Paul talks to himself*)
  - PL *Paweł leczy tylko siebie*    *Paweł leczy się* (*Paul heals himself*)
  - PT *Paulo só lava a si mesmo*    *Paulo se lava* (*Paul washes himself*)
  - RO *Paul se spală doar pe sine.*    *Paul se spală.* (*Paul washes himself.*)
- **If NO or if you are unsure:**
  - The subject is singular: test Reflexive-mutual
  - The subject is plural or coordinated (*Bob and Alice*): test Reciprocal

## 7. Reflexive-mutual, applicable when the subject is singular

Is a reciprocal version possible? Namely: Is it acceptable to replace the singular subject by a plural and add *each other* to the REFLV form without changing the REFLV's meaning?

Note: The sentence changes its meaning, but the REFLV is still used in the same sense

- **If YES: do not annotate as verbal MWE**<sup>2</sup>
  - DE *Paul wäscht sich - Sie waschen sich gegenseitig / einander* (*they wash each other*)
  - ES *Pablo se lava - ellos se lavan mutuamente / los unos a los otros* (*they wash each other*)
  - FR *Paul se lave - ils se lavent mutuellement / les uns les autres* (*they wash each other*)
  - IT *Paolo si lava - essi si lavano reciprocamente / l'un l'altro* (*they wash each other*)
  - PT *Paulo se lava - eles se lavam mutuamente / uns aos outros* (*they wash each other*)
  - PL *Paweł się myje - oni myją się nawzajem* (*they wash each other*)
  - RO *ei se spală - ei se spală unul pe altul* (*they wash each other*)
- **If NO: annotate as IRefIV**

## 8. Reciprocal, applicable when the subject is plural

Is it possible to remove the RCLI and replace the coordinated subject (*A and B*) or plural subject (*A.PL*) by a singular subject (*A* or *A.PL*) and a singular object, often introduced by *to/with* (*B* or *A.PL*), without changing the REFLV's meaning? That is:

- Coordinated subject: *A and B PronV*    *A V [to/with] B and B V [to/with] A?*
- Plural subject: *A.PL PronV*    *A.PL V [to/with] A.PL?*

Note: even though the equivalent formulation sounds weird, it keeps the sentence's meaning. Alternatively, one can try to add the modifier *mutually/each other* to the REFLV version without changing its meaning.

- **If YES: do not annotate as verbal MWE**

---

<sup>2</sup> Caution: the test applies only if test 2 has failed. For example, for *X se marie* (*X gets married*) in French, it is odd though possible to say *X et Y se marient mutuellement* (*X and Y marry each other*), but this does not mean *X se marie*, because it is only possible if *X* and *Y* are marriage officiants.

- DE *Paul und Anna umarmen sich*    *Paul umarmt Anna and Anna umarmt Paul*  
(*Paul and Anna hug each other*)  
*die Affen kratzen sich*    *die Affen kratzen die Affen* (*the monkeys scratch each other*)
  - ES *Pablo y Ana se abrazan*    *Pablo abraza a Ana and Ana abraza a Pablo* (*Paul and Ann hug each other*)  
*los niños se abrazan*    *los niños abrazan a los niños* (*the children hug each other*)
  - FR *Paul et Anne s'embrassent*    *Paul embrasse Anne and Anne embrasse Paul*  
(*Paul and Ann kiss*)  
*les jours se suivent*    *les jours suivent les jours* (*the days follow each other*)
  - IT *Giovanni e Anna si baciano*    *Giovanni bacia Anna and Anna bacia Giovanni*  
(*John and Ann kiss*)  
*i giorni si seguono*    *i giorni seguono i giorni* (*i giorni seguono l'un l'altro*)
  - PL *Paweł i Elena się całują*    *Paweł całuje Elenę and Elenę całuje Paweł* (*Paweł and Elena kiss*)
  - PT *João e Ana se beijam*    *João beija Ana and Ana beija João* (*John and Ann kiss*)  
*os presos se agrirem*    *os presos agrirem os presos* (*the prisoners aggress each other*)
  - RO *Ion și George se salută*    *Ion îl salută pe George and George îl salută pe Ion*  
(*Ion and George greet each other*)  
*participan ii se salută*    *participan ii îi salută pe participan i.* (*the participants greet each other*)
- If NO: annotate as IRefIV

## Problematic cases and remarks

### Polysemy

Keep in mind that both simple and reflexive verbs can have several senses. In test 2, we ask that ALL senses you can think of are different from the REFLV form in the given context. For example, French verb *trouver* can mean *to find something*, *to have an opinion about something*, *discover something*, etc. But it has a totally different and unrelated meaning of *to be (located at)* in the sentence *L'église se trouve à Paris* (*the church is located in Paris*)<sup>3</sup>. It should thus be annotated as a MWE. As the REFLV is polysemous itself, it should NOT be annotated as IRefIV in sentences like *Elle se trouve grosse* (*she finds herself fat*) where it means *have an opinion about (herself)*, equivalent to the non-reflexive version.

### Clitics position and concatenation

In some languages the clitics are joint with the verb, sometimes using a hyphen but not always. When there is no hyphen, the REFLV will probably be tokenized as a single token in the corpus.

<sup>3</sup> Although one can imagine that the locative meaning derives from a middle alternation *one can find the church in Paris*, but still the current meaning of *se trouver* is now far from a middle, since there is no interpretable additional agent argument.



- In French, orthography and pronunciation rules require that the clitic is concatenated with the verb and its last vowel is replaced by an apostrophe (*liaison*)
  - FR *s'abstenir* (*abstain*)
- In Spanish and Italian, the clitic can appear concatenated after the verb in some verbal forms (e.g. infinitives, gerunds)
  - ES *enamorarse* (*fall in love*)
  - IT *alzarsi* (*get up*)
- In Portuguese, there are always hyphens for postponed clitics (enclisis), but in conditional tense the clitic is in the middle of the verb (mesoclis), separating the root from the suffix
  - PT *queixar-se-ia* (*would complain*)
- In Romanian the clitic and the verb are either separate or have a hyphen between them:
  - RO *se aude un clopot* (*RCLI hears a bell, 'a bell is heard'*)
  - RO *s-aude un clopot* (*RCLI-hears a bell, 'a bell is heard'*)

The current annotation format allows annotating a single token as a MWE if it is a multiword token. Therefore, it should be annotated as an MWE.

## Overlap ID - IRefIV

Some idiomatic constructions include reflexive clitics. Two cases are possible:

- If a syntactically comparable literal construction is impossible or the REFLV would not be annotated in syntactically comparable literal constructions, annotate only the ID
  - DE *sich (über etwas) im Klaren sein (dass S)* (*lit. RCLI about s.th. in.the clear be, 'be aware of s.th./that S'*) - *\*sich in N sein, dass*
  - ES *darse cuenta de* (*realize*) - *\*darse N de* for any noun N  
*meterse en líos* (*get in trouble*) - REFLV not annotated in literal equivalents like *meterse en una tienda* (*get in a store*)
  - FR *se rendre compte de* (*realize*) - *\*se rendre N de* for any noun N  
*s'arracher les cheveux* (*lit. RCLI tear the hair, 'worry'*) - REFLV not annotated in literal equivalents like *s'arracher un ongle* (*tear oneself's nail*)
  - IT *rendersi conto di* (*realize*) - *\*si rende N di* for any noun N  
*si strappa i capelli* (*lit. RCLI tear the hair, 'worry'*) - REFLV not annotated in literal equivalents like *strapparsi un unghia* (*tear oneself's nail*)
  - PL *zdawać sobie sprawę z* (*realize*) - *\*zdawać sobie N z N*
  - PT *dar-se mal* (*fail*) - *dar-se* ADV intransitive is acceptable only for antonym *bem* (*well*)  
*meter-se numa fria* (*lit. get-RCLI in a cold, 'get in trouble'*) - REFLV not annotated in literal equivalent like *meter-se numa cabine* (*get into a cabin*)
  - RO *el își dă seama* (*he realizes*) - REFLV not annotated in literal equivalent like *el își dă banii pe cărți* (*he spends his money on books*)  
*se duce pe apa sâmbetei* (*RCLI go on water-the Saturday-of, 'to get lost'*) - REFLV not annotated in literal equivalent *se duce pe apa Bistriei* (*he goes on the river Bistrița*)
- If the REFLV would be annotated as IRefIV in syntactically comparable literal constructions, annotate both the IRefIV and the ID as embedded MWEs (rare)
  - PT *virar-se nos trinta* = *virar-se* (*get by*) ≠ *virar* (*turn/become*)

- PL rozlatywać się w proch (lit. scatter itself into dust, 'disappear')

## Overlap LVC - IRefIV

It is rare, although possible, to find light verb constructions in which a reflexive clitic changes the original meaning significantly, thus characterising an IRefIV:

- DE *Fragen stellen* (ask questions), *sich Fragen stellen* (doubt/hesitate)
- FR *poser des questions* (ask questions), *se poser des questions* (doubt/hesitate)

In this case, the whole construction, including the verb, the noun and the reflexive clitic, must be annotated as ID, since there are two syntactic arguments:

- DE *sich Fragen stellen*
- FR *se poser question*

Notice that annotating only the verb and the RCLI as IRefIV would be wrong, since it will have a completely different meaning without the noun, sometimes even coinciding with another IRefIV like the FR example below:

- DE *sich stellen* (surrender)
- FR *se poser* (sit/lay down)

## Dative clitics and double clitics

In some languages, e.g. Polish, clitics inflect for case. While most cases of IRefIV seem to be restricted to the accusative case (PL *bać się* 'to be afraid'), other cases can appear in IRefIV, e.g. PL *radzić sobie* (lit. to advise oneself.DAT 'to manage'). Some expressions have double clitics, e.g. PL *przyglądać się sobie* (lit. to observe RCLI.acc RCLI.DAT, 'to observe each other'), PL *radzić sobie z sobą* (lit. to advise RCLI.DAT with RCLI.INST 'to manage with oneself'), . Here, only the first two words belong to the IRefIV.

## Non-reflexive clitics

This category does not cover other types of pronouns and clitics. They are covered by regular ID tests and should be annotated as such. Examples of constructions that should be annotated as ID rather than IRefIV include:

- FR *l'emporter* (win), *s'en aller* (leave), *en avoir marre* (be fed up), *il y avoir* (exist)
- IT *prender-ci* (make the right choice), *prender-le* (be beaten)