Annotation Model:

Criteria for Linguistic Categorization in the Database "Adjective-Adverb Interfaces in Romance"

Katharina Gerhalter

2nd Version: January 2021

(revised and modified version, 1st publication: September 2020)

Project Team:

Martin Hummel (Project leader, Department of Romance Studies)

Katharina Gerhalter (Data collection and annotation, Department of Romance Studies)

Gerlinde Schneider (Data Modelling, Centre for Information Modelling - Austrian Centre for Digital Humanities)

Christopher Pollin (Data Modelling, Centre for Information Modelling - Austrian Centre for Digital Humanities)

Table of contents

1. Introduction

The examples in the database *Adjective-Adverb Interfaces in Romance* (=aaif) are tagged according to a shared annotation model, using the same categories for different Romance languages and corpora. Thus, the annotation model offers a cross-linguistic categorization based on the **forms, functions and meanings** of the heterogeneous and multifunctional word-class *adverb*, which could also be applied to other languages besides Romance (e.g. for English). The following description of the tags and their definitions in the aaif-annotation model aims at a transparent and operational systematization. However, not all categories are annotated in all subcorpora, since they were created for different purposes: the metadata-description of each subcorpus (see menu item Corpora) specifies which types of adverbs were compiled, which word-classes (also called *parts of speech*) were tagged and which categories were annotated in a given subcorpus. Hence, the following description is an extensive repertoire of the possibilities offered by the aaif-annotation tool.

For each attested adverb in the database, a **short example phrase** is tagged. The short example phrases are shown in the Search Interface and can be read in a broader context (possibly also the whole full-texts) via the menu item Corpora. Every tagged example (i.e. every short example phrase) in the corpus contains at least one annotated adverb (see section 2). Since this is the main tag of the annotation model and the only obligatory one, the counting of the examples in each subcorpus is based on the number of tagged adverbs. Note that more than one adverb may be attested in a given short example phrase (see coordination [section 2.2.3] and adverbs modifying other adverbs [section 2.2.1]). Inside the short example phrase, other tags besides the one for adverbs – possibly in combination with prepositions and articles or possessives (see section 2.1.5.) – can be used, such as the verb (see section 3) and the subject (see section 4) of the sentence.

Furthermore, the adverbs (and occasionally prepositions and verbs) of the examples are lemmatized in order to unify orthographic variation – especially regarding historical and dialectal data. This enables search queries via **lemma** and analysis of type-token-frequencies to be performed. The following description also explains the criteria for lemmatizing the tagged word-classes.

2. Annotations concerning the word-class *adverb*

2.1. Morphosyntactic structure (form) of the adverb

The morphosyntactic structure takes into account the formal structure of the adverbial as well as its possible morphological variation (inflection, suffixes). The annotation model suits different types of adverbials based on an adjectival root: simple (one-word) adverbs such as **adjective-adverbs** (2.1.1), **derived adverbs** (2.1.2) and **nouns** used as adverbs (2.1.3.), as well as adverbials formed by more than one word-form, such as **reduplicated adverbs** (2.1.4) and adverbial locutions (**prepositional phrases** (2.1.5)). These variants are lemmatized with the same **lemma** (the same underlying adjective): e.g. the Portuguese lemma *claro* delivers examples of *claro* (adjective-adverb), *às claras* (prepositional phrase) and *claramente* (derived adverb).

Since the database concerns the interface between adjectives and adverbs, it only includes adverbials that are lexically based on an adjectival root and, in exceptional cases, adverbials based

on nouns. Therefore, **lexical adverbs** like Spanish / French *bien*, *mal*, Portuguese *bem, mal*, Italian *bene, male* or Romanian *bine* are not tagged in this corpus.

2.1.1. Adjective-Adverbs

In the case of adjective-adverbs (also called *bare adverbs, adverbialized adjectives* or *short adverbs*), a word belonging to the word-class of the adjective takes adverbial functions (for the theoretical background, see Hummel 2017), such as the modification of the verb (manner adverb) as shown in the following examples: Spanish *claro* in *hablar claro* 'to speak clearly', French *haut* in *voler trop haut* 'to fly too high', Portuguese *rápido* in *aprender rápido* 'to learn fast', Italian *bello* in *scrivere bello* 'to write beautifully' or Romanian *drept* in *a merge drept* 'to go straight/right'. These adverbs are annotated with the category MORPHOSYNTACTIC STRUCTURE: adjectival. Only adjective-adverbs – that is, adjectives with adverbial functions (see section 2.2.1 for the syntactic functions of the adverbs) – are tagged and annotated, whereas adjectives with adjectival syntax are not tagged in the aaif-database (e.g. examples such as Spanish *alto* in *el arbol es alto* 'the tree is high' or French *droit* in *l'angle droit* 'the right angle').

For the **lemmatization** of the examples, the uninflected standard form in present-day language is used. It is identical to the masculine singular form of the base adjective, as in the case of the above-mentioned forms: Spanish *claro*, French *haut*, Portuguese *rápido*, Italian *bello*, Romanian *drept*. Regarding the rich diversity in varieties of southern Italy (*Napolitano, Siciliano, Salentino* and *Calabrese*, see subcorpus It_A_aaif), the corresponding standard Italian lemma is annotated in order to enable a simpler lemma search, e.g. dialectal *biello* in Calabrian texts is lemmatized as *bello*. In case there is no such corresponding standard Italian lemma, a reconstructed vernacular Latin form is marked with an asterisk and written in capital letters, for example, the lemma *MAIATICUM for the Neapolitan adjective-adverb *majàteco* 'solidly' in the following example:

(1)   *parlava Napoletano chiantuto e majàteco* (la Posilecheata, 1684, Napoli, It_A_aaif)
      'he was speaking Neapolitan robustly and solidly'.

The category INFLECTION refers to the number and gender of the adjectival form. Since most adjective-adverbs appear in their masculine singular form (as in the above-mentioned examples), they are considered an invariable and uninflected word-class. Despite the normative preference for invariable/uninflected adverbs, especially older examples in historical corpora show inflected uses of adverbs, as in this Spanish example:

(2)   *que se vayan luego derechos a Marbella* (Sp_AP_SH3, Epistolario del Conde de Tendilla, 1504-1506)
      'they should go straight to Marbella'

In this example, the adjective-adverb *derechos* (lemma: *derecho* 'straight') agrees with the subject; therefore, the inflectional category of this example is annotated as INFLECTION: masculine plural. Consequently, Spanish *derecha* is annotated as INFLECTION: feminine singular and *derechas* as INFLECTION: feminine plural. Since the masculine singular form of the adjective-adverb is identical to the unmarked, uninflected form, in examples such as (3) it is not possible to distinguish between an uninflected or inflected use of the adjective-adverb *derecho*:

(3)   *que él*[masc. sg.] *se vaya derecho*[masc. sg.] *a Marbella*

The only way to clearly identify uninflected/unmarked adverbs are cases such as (4). In this example, the subject would cause plural agreement (the form *derechos* would agree with the subject), whereas the adverb *derecho* is clearly uninflected:

(4)   *que se vayan*[plural] *derecho*[masc. sg.] *a Marbella*

However, in most cases the examples are simply ambiguous, as in (3). Therefore, the tag INFLECTION: uninflected is considered the default-tag which necessarily also covers the inflectional category masculine singular, whereas the other inflectional categories can be tagged separately (masculine plural / feminine singular / feminine plural and –for languages like Romanian– neuter singular and neuter plural).

The **lemmatization** enables searching for all **inflected variants** of one adjective-adverb. For example, the simple query for the Spanish lemma *derecho* (without specifying the category "inflection") delivers all inflected and uninflected examples, *derecho, derechos, derecha, derechas*, as well as diminutive forms such as *derechito, derechitas*, etc.

The specific inflectional categories (masculine/feminine/neuter and singular/plural) are annotated in all corpora except in the French ones. For French, the tags for the inflectional categories (besides the default category uninflected = masculine singular) are based on the distinction between INFLECTION: audible inflection (for example, *sain > saine* in *vivre saine*) and INFLECTION: inaudible inflection (for example, the inflection of *seul > seule* is only relevant for written French, as in the example *je m'en vais seule*, but not in spoken French).

2.1.2. Derived adverbs

Some subcorpora of the database include examples of **derived adverbs**, although they are not the main focus of the aaif-database. Derived adverbs are formed by a derivational suffix like *-mente* which is added to an adjectival base. Hence, contrarily to adjective-adverbs, where an adjective is directly converted into an adverb, derived adverbs show a specific adverbial suffix which clearly marks them as belonging to the word-class *adverb*. The tag MORPHOSYNTACTIC STRUCTURE: derived *-mente* covers all variants of this suffix in several Romance languages: French *-ment* and Spanish / Portuguese / Italian *-mente*, as well as historical variants such as Spanish *-mientre, -miente* or *-mentre* and regional varieties such as Calabrian and Sicilian *-menti*. Romanian only has a few borrowed *mente*-adverbs, which are not documented in the aaif-database.

The **lemma** of derived adverbs is the adjectival lemma of the derivational base: Spanish *ciertamente* 'certainly' is lemmatized as *cierto*, Portuguese *exatamente* 'exactly' as *exato*. Since the systematic derivational base of *mente*-adverbs is the feminine singular form of the adjective (Spanish *cierta > ciertamente* and Portuguese *exata > exatamente*), we consider that a fossilized form of the stem, but not an inflectional process on the whole adverb. Therefore, *mente*-adverbs are by default categorized as INFLECTION: uninflected. We would only consider *mente*-adverbs to be inflected if the whole adverb takes, for example, a plural morpheme: e.g. informal Spanish *los ríos meramentes flotables*, which is considered incorrect but still attested (not in the database).

Besides *-mente*, other language-specific derivational suffixes are the Romanian suffixes *-eşte* (e.g. *crestineşte* 'in a Christian manner'), *-iş* (e.g. *făţiş* 'openly, directly') and *-ul*; for each derivational suffix there is a specific tag: MORPHOSYNTACTIC STRUCTURE: derived *-eşte* / derived *-iş* / derived *-ul*.

The annotation tool also includes tags for the Italian suffix *-oni/-one* (used for actions that involve a body part touching the ground: *ginocchione/i* 'on one's knees', *bocconi* 'on one's mouth') and the English suffix *-ly*. Those two options are not shown in the search interface, since there have not (yet) been any annotated examples.

2.1.3. Nouns used as adverbials

This category applies to nouns that are used with adverbial functions. For example, in Spanish *pasarlo bomba* 'to have a very good time', the noun *bomba* 'bomb' is used as a manner adverb meaning (metaphorically) 'very well'. Nouns also appear in adverbial locutions (see section 2.1.5), e.g. Portuguese *com certeza* 'certainly' (literally 'with certainty') and *com clareza* 'clearly' (literally 'with clarity'), as well as in reduplicated locutions, e.g. Neapolitan *ardere miccia miccia* (see section 2.1.4). These nouns are tagged as adverbs and further annotated with the tag MORPHOSYNTACTIC STRUCTURE: noun.

The **lemma** of these adverbials is the basic (singular) form of the noun. Therefore, contrarily to *mente*-adverbs, adjective-adverbs and prepositional phrases of the type "preposition + adjective" –which are all lemmatized with an adjectival lemma– nouns are lemmatized as nouns. The category INFLECTION (feminine singular / masculine plural, etc.) is also tagged; note that the inflectional categories in the annotation tool were established for adjectives, thus *gender* (masculine, feminine, neuter) appears as an inflectional category even though it is inherent and nouns are only inflected for *number* (singular, plural).

This category is not systematically annotated in the database, since nouns are not the main target of the aaif-project. For example, in the Portuguese database Pt_APM_DeG, only the prepositional phrases *com certeza* and *com clareza* are tagged because they form paradigms with the adverbs *certo, ao certo, certamente* (common adjectival lemma: *certo*) and *claro, claramente, às claras, em claro* (common adjectival lemma: *claro*). On the contrary, other prepositional adverbials such as *sem interrupções* 'without interruptions' are not tagged in this subcorpus, since they share no lexemic root with any adjective-adverb.

Only nouns that are used adverbially (via direct conversion or via reduplication) or that form part of a prepositional phrase are tagged as nouns. In cases where the derivational basis of a derived adverb (see previous section) is a noun (instead of an adjective), the adverb has been classified as derived. For example, the Romanian adverb *fățiş* 'openly, directly' is formed by the noun *față* 'face, visage' and the derivational suffix *-iş*, and has been classified as a derived adverb with the MORPHOSYNTACTIC STRUCTURE: derived *-iş*.

2.1.4. Reduplicated adverbs

The category REDUPLICATION refers to the lexicalized or spontaneous consecutive repetition (doubling) of the same adverb. In some dialects of Southern Italy, the reduplication of **adjective-adverbs** is especially productive. For example, Neapolitan *sano sano* means 'entirely':

(5)   *ma la diceno tutta sana sana* (It_A_aaif, Aminta, 1725)
        'but they tell all [these things] entirely'

Reduplicated adverbs are tagged only once: e.g. only one word-form *sana* has been tagged with the same categories as a single-word adverb, that is: lemmatized as *sano* and annotated as an adverb with the MORPHOSYNTACTIC STRUCTURE: adjectival and its specific inflectional features (INFLECTION: feminine singular). Additionally, *sana* has been tagged with the separate category REDUPLICATED: true.

Likewise, one instance of *paso* 'slowly' was annotated as an adjective-adverb and tagged as reduplicated in the following Spanish example:

(6) *¡Paso, paso, Pármeno, no saltes!* (Sp_AP_SH3, La Celestina, 1499)
'Slowly, slowly, Pármeno, do not jump'.

As this example shows, adjective-adverbs can also be productively reduplicated in order to boost their meaning, without having a specific lexicalized or idiomatic meaning. Any type of reduplication is tagged in the aaif-database, regardless of its meaning.

In the same vein, reduplications of a **noun** may be used adverbially, e.g. the Neapolitan noun *miccia* 'fuse':

(7) *ardere miccia miccia* (It_A_aaif, La vita e morte de lo sciatamone 'mpetrato, 1750)
'to burn weak'

In the case of *miccia miccia*, one word-form *miccia* is annotated as an adverb with the tag MORPHOSYNTACTIC STRUCTURE: noun and INFLECTION: feminine singular. Additionally, *miccia* has been tagged with the category REDUPLICATED: true. In this case, *miccia miccia* (literally 'fuse') has a specific lexicalized idiomatic meaning.

Hence, in the case of reduplicated adverbials, only one form is annotated with the same tags as a single-word adverb (be it an adjective-adverb or a noun) and the feature of reduplication is not found under the category "morphosyntactic structure", but as a separate category: the tag REDUPLICATED: true. Therefore, the default tag for all other examples is REDUPLICATED: false.

2.1.5. Prepositional phrases (adverbial locutions)

Some periphrastic adverbials are formed by multi-word sequences introduced by a preposition. The following combinations can be tagged with the annotation tool, although not all possibilities are attested in the database:

Preposition + Adjective	Portuguese *de novo* 'again'
Preposition + Preposition + Adjective	Spanish *por de pronto* 'provisionally'
Preposition + Article + Adjective	Spanish *por lo seguro* 'surely'
Preposition + Possessive + Adjective	Spanish *a mis solas* 'on my own'
Preposition + Noun	Portuguese *com certeza* 'certainly'
Preposition + Article + Noun	Portuguese *na verdade* 'in truth'
Preposition + Adjective + Article	Romanian *cu dereptul* 'rightly, correctly'
Preposition + Derived Adverb	Romanian *de fățiș* 'openly, directly'

For each word-class (part of speech) there is a separate tag; the tag for the word-class **adverb** is thus combined with a separate word-class tag for **prepositions** and any **articles** or **possessives**. Therefore, in these adverbial locutions (prepositional phrases), each part of the multi-word sequence is annotated with the corresponding categories.

For the **lemma** of these Romanian adverbials, the underlying adjectival base is annotated, i.e. *cu dereptul* 'rightly, correctly' is lemmatized as *drept* (hence, the same etymology as Spanish *derecho*, see examples above).

For example, the adjective *novo* 'new' in the Portuguese locution *de novo* 'again' is annotated with the tag for the word-class **adverb**. Hence, *novo* is lemmatized and categorized as MORPHOSYNTACTIC STRUCTURE: adjectival. Additionally, *novo* is tagged with the category PART OF PREPOSITIONAL PHRASE: true (hence, the default for all other examples is PART OF PREPOSITIONAL PHRASE: false). Likewise, the

noun *certeza* 'certainty' in the prepositional phrase *com certeza* 'certainly' is categorized as an adverb with the MORPHOSYNTACTIC STRUCTURE: noun.

The Romanian derived adverb *fățiş* (*față* 'face' + derivational suffix *-iş*) occurs both alone and in the prepositional phrase *de fățiş* 'openly, directly'; in both cases it is tagged with MORPHOSYNTACTIC STRUCTURE: derived *-iş*. Whereas examples of *de fățiş* are tagged as PART OF PREPOSITIONAL PHRASE: true, examples of the simple adverb *fățiş* are tagged as PART OF PREPOSITIONAL PHRASE: false.

Therefore, the adjectives, nouns or adverbs that form part of a prepositional phrase are tagged and lemmatized as a single-word adverb (be it an adjective-adverb, a noun or a derived adverb, see previous sections), also regarding their possible INFLECTION.

Consequently, prepositional phrases can be queried in the search interface by selecting the category ADVERB: PART OF PREPOSITIONAL PHRASE: true. Additionally, the **prepositions** that form part of a prepositional phrase are lemmatized and can be found via lemma search. Therefore, it is possible, for example, to search prepositions cross-linguistically in several Romance languages, e.g. the **lemma** search for the preposition *de* in various subcorpora simultaneously delivers examples such as Portuguese *de novo* 'again'*, de leve* 'lightly', Spanish *de cierto* 'certainly', *de seguro* 'surely', and Romanian *de fățiş* 'openly', *de curând* 'quickly'. In the case of the subcorpus It_A_aaif, the lemmatization of the prepositions is based on the standard Italian form, therefore the dialectal form *de veru* 'really' (attested in Salento) can be found via lemma search for *di*.

In prepositional phrases which are formed by more than one preposition, each preposition is tagged and lemmatized separately. Therefore, examples such as Spanish *por de pronto* 'provisionally' show three tagged word-classes: two prepositions and one adjective-adverb. Since both prepositions are lemmatized, examples like the following can be found both via lemma search for por and lemma search for de:

(8)   *Supongamos, por de pronto un sustantivo, por ejemplo: "herencia".* (Sp_AP_CORDIAM, Documentos comentativos: El Coronel, Uruguay, 1880)
      'Let's assume, provisionally, a noun, for example: "inheritance"'

In some cases, the preposition is **contracted** with the adjective: Romanian *deplin* 'completely, fully' is an agglutinated form: preposition *de* + adjective *plin*. In these cases, both the preposition and the adjective are tagged and lemmatized. The contracted preposition is additionally tagged with the category CONTRACTED: true. Hence, the default tag for all prepositions that are not fused is CONTRACTED: false.

Occasionally, prepositional adverbials are formed with two contracted prepositions. The Romanian adverbial *îndelung* 'during a long time' is formed as follows: contracted preposition *în* + contracted preposition *de* + adjective *lung* 'long'.

In some cases, the periphrastic adverb may also include an **article** which is tagged as a separate word-class and annotated with the categories GENDER (masculine / feminine / neuter / undefined) and NUMBER (singular / plural / undefined). Articles are not lemmatized. Therefore, prepositional phrases like Spanish *a la fija, a lo seguro* contain three tagged word-classes (preposition + article + adverb), which can be queried in the search mask independently or in different combinations.

In some Romance languages, **articles** may be **contracted** with the preposition, e.g. Portuguese *ao* (preposition *a* + masculine singular article *o*) and *às* (preposition *a* + feminine plural article *as*). Therefore, prepositional adverbials like *ao certo* 'certainly' and *às claras* 'clearly' are tagged as follows:

ao certo:

PREPOSITION	lemma *a*
	CONTRACTED: false
ARTICLE	GENDER: masculine
	NUMBER: singular
	CONTRACTED: true
ADVERB	lemma *certo*
	MORPHOSYNTACTIC STRUCTURE: adjectival
	INFLECTION: uninflected
	PART OF PREPOSITIONAL PHRASE: true

às claras:

PREPOSITION	lemma *a*
	CONTRACTED: false
ARTICLE	GENDER: feminine
	NUMBER: plural
	CONTRACTED: true
ADVERB	lemma *claro*
	MORPHOSYNTACTIC STRUCTURE: adjectival
	INFLECTION: feminine plural
	PART OF PREPOSITIONAL PHRASE: true

In these examples, the prepositions themselves are tagged as CONTRACTED: false, since the category CONTRACTED: true applies to prepositions contracted to the adjective-adverb (see above). Simply for practical reasons and in order to differentiate the patterns, the contraction of *a + o* and *a + às*, in turn, is tagged by annotating the article as contracted.

In Romanian, the article is postponed and contracted to the noun or nominalized adjective. For example, the morphemes of the Romanian adverb *de amănuntul* 'detailed' are: *de a-mănunt-ul*. Hence, this adverbial is based on the adjective *mănunt* 'small, detailed' (< lat. *minutus*) and further consists of two prepositions (*de* and contracted *a*) and of the postponed and contracted article *-ul*. It is tagged with four tags:

de amănuntul:

PREPOSITION	lemma *de*
	CONTRACTED: false
PREPOSITION	lemma *a*
	CONTRACTED: true
ADVERB	lemma *mănunt*
	MORPHOSYNTACTIC STRUCTURE: adjective
	PART OF PREPOSITIONAL PHRASE: true
ARTICLE	GENDER: masculine
	NUMBER: singular
	CONTRACTED: true

In those cases where the periphrastic adverb includes a **possessive** (instead of an article), it is also annotated with a tag for this part of speech. Possessives are not lemmatized. The categories for possessives are GENDER (masculine / feminine / neuter / undefined), NUMBER (singular / plural /

undefined), and PERSON (1st / 2nd / 3rd person). For example, the Spanish prepositional phrase *a mis solas* 'on my own' has been tagged as follows:

a mis solas:

PREPOSITION	lemma *a*
	CONTRACTED: false
POSSESSIVE	NUMBER: plural
	GENDER: undefined
	PERSON: 1st
ADVERB	lemma *solo*
	MORPHOSYNTACTIC STRUCTURE: adjectival
	INFLECTION: feminine plural
	PART OF PREPOSITIONAL PHRASE: true

In sum, the annotation tool serves to annotate each word of a prepositional phrase separately and therefore allows different (or the same) word-class tags to be freely combined.

2.2. Syntactic properties and functions of the adverb

The following categories are annotated for all types of adverbs regardless of their morphosyntactic form (adjective-adverbs, derived adverbs, prepositional phrases or reduplicated adverbs). The categories offer a broad systematization of the diverse syntactic functions (see section 2.2.1) of the word-class *adverb* and other syntactic properties such as modification (see 2.2.2.) and coordination of two or more adverbs (see 2.2.3.).

2.2.1. Attribution target

The main function of the word-class *adverb* has been defined as *modification*: adverbs modify another part of the sentence (a word, a phrase or even the sentence as a whole). The syntactic scope of the adverb –the segment it modifies– is tagged under the category ATTRIBUTION TARGET. Depending on which types of adverbs have been compiled for a specific subcorpus (e.g. only manner adverbs or also other adverbial functions), more or fewer categories have been annotated and can be found via the search interface. Note that some examples may be ambiguous and can be interpreted differently; therefore, some categories may be vague and may have been used slightly differently (i.e. a broader or more general classification) by the different linguists who annotated the various subcorpora.

The following classifications were established for the aaif-annotation tool:

➢ ATTRIBUTION TARGET: verb

According to literature, the main function of adverbs is manner, that is, verb modification: Spanish *proceder de ligero* 'to proceed quickly', French *dire tout bas* 'to say very quietly', Portuguese *participar ativamente* 'to participate actively', Italian *scrivere bello* 'to write beautifully', or Romanian *a zice chiar* 'to speak clearly'. In these examples, the syntactic scope of the adverb is annotated as ATTRIBUTION TARGET: verb. The verb form modified by the adverbial is tagged and lemmatized too (see the tag for the word-class **verb**, section 3).

➤ ATTRIBUTION TARGET: verb and subject

In certain sentences, the manner adverb modifies both the action described by the verb of the sentence and the subject involved, e.g. the following Spanish example:

(9) *decía el Marquesito muy serio* (Sp_AP_SH3, La Regenta, 1884-1885)
     'said the little marquis very serious(ly)'

The adjective-adverb *serio* 'serious(ly)' describes both the action expressed by the verb *decir* 'to say' (the serious manner of speaking) and the attitude of the subject, *el Marquesito*, while speaking. Whereas secondary predicates modify exclusively the subject, albeit limited to the scope of the verb (e.g. *Juan compró una casa desesperado.* 'Desperate, John bought a house'), the adverbs of the category ATTRIBUTION TARGET: verb and subject also modify the verb at the same time.

The **inflection** of the adjective-adverb (see section 2.1.1.1) occasionally agrees in number and gender with the subject: in the French example *elle vivait saine* 'she lived healthy', the adjective-adverb *saine* (fem. sg.) agrees with the subject *elle* (fem. sg.) and describes both a quality of the subject ('she is healthy') and a characteristic of the verb *vivre* ('she lives in a healthy manner'). In these kinds of examples, in addition to the verb, occasionally the **subject** the adverb refers to is also annotated in some corpora (see the tag for the word-class subject, section 4). It should be added however that the distinction between "subject only", "verb and subject" and "secondary predicate" (not included in the database) is all but clear-cut. Hence, the database provides a rough classification that should be subject to more detailed analyses in research using the data.

➤ ATTRIBUTION TARGET: verb and object

Similarly, this tag applies to those manner adverbials that modify both the (transitive) verb of the sentence and the object involved, as in the following Spanish example:

(10) *los martirios [...] los ha de pagar bien caros* (Sp_AP_SH3, Los bandidos del Río frío, México, 1889-1891)
      's/he has to pay very dearly for the torments'.

In this example, the adjective-adverb *caro* 'dear(ly), expensive' is inflected (masc. pl. form *caros*) and it agrees not with the (sg.) subject of the sentence, but with *los martirios* (masc. pl.), i.e. the direct object of the verb *pagar* 'to pay'.

Likewise, reduplicated adverbs can also be inflected and may agree with the object of the sentence, as in the following Neapolitan example:

(11) *ma la diceno tutta sana sana* (It_A_aaif, Aminta, 1725)
      'but they tell all [these things] entirely'

The reduplicated adjective-adverb *sana sana* 'entirely' (literal translation: 'healthy-healthy') concords with the direct object *la* (fem. sg.).

Likewise, in the French sentence

(12) *On peut la [=la soupe] manger chaude, froide* (2015, Fr_A_Web)
      'you can eat it [=the soup] hot or cold',

the adjective-adverbs *chaud* and *froid* are inflected (fem. sg.), agreeing with the direct object *la* [=*la soupe*]. Hence, the adjective-adverbs *chaud* 'hot' and *froid* 'cold' refer to a quality of the direct object (the soup) at the moment the action described by the verb *manger* 'to eat' takes place.

Occasionally, the adverb refers to an implicit direct object of verbs like Portuguese *votar* 'to vote', as in the sentence

(13) *O povo brasileiro [...] continua votando errado, sempre* (Pt_APM_DeG, 1993, Rio de Janeiro, parte escrita)
'the Brazilian people always continue voting wrong(ly)'.

The adjective-adverb *errado* 'wrong(ly)' does not describe solely the manner of voting (the action of voting itself is not invalid or done in an incorrect manner), but refers to the implicit direct object (e.g. they voted the wrong party), hence it is also annotated as ATTRIBUTION TARGET: verb and object.

Besides, adverbs also modify other parts of the sentence than the verb. The following "attribution target"-categories do not consider the verb of the sentence, which is, consequently, not tagged:

➢ ATTRIBUTION TARGET: adjective

Adverbs may modify adjectives, e.g. the Romanian adjective-adverb *foarte* 'strong(ly)' modifies the adjective *mare* 'big' in this example:

(14) *Țara Ungurească mai denainte vreme era foarte mare* (1725, Grigore Ureche: Letopisețul Țărîi Moldovei, Ro_ADP_aaif)
'The country of Hungary was really large before'.

In this example, *foarte* is annotated with the category ATTRIBUTION TARGET: adjective; in turn, the adjective *mare* itself is not tagged (since adjectives with adjectival syntax are not included in the annotation model).

Adjective-adverbs may also be inflected and show agreement with the inflectional categories of the adjective they modify (which itself concords with the subject): the inflected (fem. sg.) Neapolitan adjective-adverb *meza* (standard Italian lemma: *mezzo*) modifies the adjective *vregognosa* (modern standard Italian: *vergognosa*) in the following sentence:

(15) *steva meza vregognosa* (It_AP_aaif, 1684, Napoli, Sarnelli, La Posilecheata)
'she was half shy/ashamed'.

As noted above, the adjective modified by the adverb (that is, the adjective as an attribution target) is not tagged itself. Nevertheless, deverbal adjectives (e.g. participles) are classified and tagged as verb forms (see section 4 for the tagging of the word-class **verb**), since this allows for annotating and lemmatizing of the participle. For example, in the Spanish sentence

(16) *como un cordero recién esquilado* (2007, El corazón helado, Sp_AP_SH3)
'like a recently sheared sheep'

the adjective-adverb *recién* (ATTRIBUTION TARGET: adjective) modifies the deverbal adjective *esquilado*, which itself has been annotated with the tag for the word-class **verb** and lemmatized as *esquilar*.

➢ ATTRIBUTION TARGET: adverb

Adverbs may also modify other adverbs: the syntactic scope of the Spanish adjective-adverb *harto* in *interpretar harto duramente* ('to interpret/perform extremely hard') is the *mente*-adverb *duramente*, which, in turn, is a manner adverb that modifies the verb *interpretar*. Therefore, *harto* is tagged with the category ATTRIBUTION TARGET: adverb. Depending on the types of adverbs tagged in each dataset (e.g. additionally to adjective-adverbs also *mente*-adverbs), the adverb modified

by an adverb has also been tagged itself. In the following Portuguese example, the *mente*-adverb *praticamente* modifies the adjective-adverb *sozinho*:

(17) *a pessoa trabalha praticamente sozinho … na sua … na sua criação* (1993, Natal, Pt_APM_DeG, oral corpus)
'the person works practically/nearly alone on its creation'.

In this example, the adverb *praticamente* has been tagged with the category ATTRIBUTION TARGET: adverb and the adjective-adverb *sozinho* with the category ATTRIBUTION TARGET: verb and subject.

In Romanian, an adjective-adverb (e.g. *foarte* 'strong(ly)') can also modify another adjective-adverb (e.g. *greu* 'hard(ly), difficult(ly)'):

(18) *foarte greu au stătut strâmbătate* (1774, Bertoldo, Ro_APD_aaif)
'with real difficulty they stood crooked'

In this sentence, both the adjective-adverbs *foarte* (ATTRIBUTION TARGET: adverb) and *greu* (ATTRIBUTION TARGET: verb) have been tagged.

➢ ATTRIBUTION TARGET: noun or syntagma without verb-reference

This category includes all types of nouns or noun phrases as well as prepositional phrases or any type of phrase which does not include the verb of the sentence (in the annotation model, we use the term *syntagma* for *phrase*). For example, the Spanish adjective-adverb *sola* (inflected: fem. sg., lemma *solo*) modifies the syntagma *una palabra* (which is a noun phrase) in the following example:

(19) *con sola una palabra* (1499, La Celestina, Sp_AP_SH3)
'with just/only one word'.

Likewise, in the Portuguese example

(20) *principalmente no verão* (1993, Rio de Janeiro, parte oral, Pt_APD_DeG)
'particularly during the summer'

the adverb *principalmente* modifies the prepositional phrase *no verão*.

Finally, in the Neapolitan example

(21) *simmo iusto quatto* (1934, Napoli, Salvatore Di Giacomo: Tutto il teatro, It_A_aaif)
'we are exactly four'

the adjective-adverb *iusto* (standard Italian lemma *giusto*) refers to the numeral. Hence, this heterogeneous category includes different types of phrases or words (nouns, numerals, etc. – all word-classes except adjective and adverb for which there is a proper category, see above). The common criterion is that the modified phrase or word does not include the verb, hence, the adverb does not characterize the action described by the verb.

➢ ATTRIBUTION TARGET: sentence

Adverbs of this category modify the sentence as a whole. For example, in the Neapolitan phrase

(22) *sì lo conosco, lo conosco sicuro* (1925, Eduardo Scarpetta, Tutto il teatro, It_A_aaif)
'yes, I know him, I know him surely/ for sure'

the adjective-adverb *sicuro* does not refer as a manner adverb to the verb *conoscere*, but its scope is the whole sentence; *sicuro* is a sentence-adverb that characterizes the whole proposition – the fact of knowing him – as 'sure'.

Similarly, the Spanish adjective-adverb *cierto* 'certainly' modifies the whole sentence (note that the adverb is also separated by commas, hence there is a prosodic pause):

(23) *Verdad sea que, cierto, duele más la pérdida* (1499, La Celestina, Sp_AP_SH3)
      'Certainly, it is true that the loss hurts more'.

Likewise, the Portuguese adverb *infelizmente* 'unfortunately, unhappily' refers to the whole sentence (expressing the speaker's point of view) in the following example:

(24) *Meu grupo infelizmente não foi o vencedor* (1993, Natal, parte escrita, Pt_APM_DeG)
      'Unfortunately, my group was not the winner'.

The tag ATTRIBUTION TARGET: sentence also applies for discourse markers, for example, affirmation markers or phatic markers (Spanish *bueno, claro*, Portuguese *bom, exatamente*, Neapolitan *certo*).

2.2.2. Modification of the adverb

The category MODIFIED: true applies if the adverbial itself is modified (mostly intensified) by another adverbial (hence, the default category for all non-modified adverbs is MODIFIED: false). For example, the Romanian adjective-adverb *scurt* 'short' is modified by the adverb *mai* 'more' in the following example:

(25) *alţii au scris mai scurt* (1620, Cronica universală, Ro_APD_aaif)
      'others wrote more briefly'

Similarly, in the French example

(26) *car ilz parloient bien bas* (1489, Philippe de Commynes, Fr_A_DHAA)
      'because they spoke very quietly'

the adjective-adverb *bas* is modified by the adverb *bien* 'very'. The feature MODIFIED: true is annotated for the adjective-adverbs *scurt* and *bas*, whereas the modifiers themselves, *mai* 'more' and *bien* 'very', are not tagged. As explained above, the aaif-database only annotates and lemmatizes adverbials which are based on an adjectival root. Therefore, modifiers such as Spanish *muy, tan, bien*, French *très, trop, bien*, Portuguese *tão, bem, demais*, Romanian *mai, aşa* or Neapolitan *tanto, cchiú* (standard Italian *più*), *così* are not tagged and annotated themselves.

In the case that the modifier itself is actually based on an adjectival root, both adverbials are tagged and annotated (see section 2.2.1 ATTRIBUTION TARGET: adverb), as in the case of the following Portuguese example:

(27) *a pessoa trabalha praticamente sozinho … na sua … na sua criação* (1993, Natal, Pt_APM_DeG, oral corpus)
      'the person works practically/nearly alone on its creation'

Both adverbs are annotated: *praticamente* (ATTRIBUTION TARGET: adverb; MODIFIED: false) and *sozinho* (ATTRIBUTION TARGET: verb and subject; MODIFIED: true).

Likewise, both adjective-adverbs, *foarte* and *greu*, in the Romanian example *foarte greu au stătut strâmbătate* 'with real difficulty they stood crooked' (see example 18 above) were tagged, and *greu* was annotated with the category MODIFIED: true.

2.2.3. Coordination of adverbs

If two or more different adverbs are coordinated, they share the same syntactic function and syntactic scope. For example, two manner adverbs can be coordinated, as in the following French example:

(28) *des accents qui sonnent beaux et étranges* (2016, Fr_A_Web)
'accents which sound beautiful and strange'

In this example, the two inflected adjective-adverbs *beau* 'beautiful(ly)' and *étrange* 'strange(ly)' are both manner adverbs which both modify the verb form *sonnent* 'to sound'. The adverbs *beau* 'beautiful(ly)' and *étrange* 'strange(ly)' are annotated separately and both appear separately in the table of results in the search interface. Additionally, they are tagged with the category COORDINATED: true (hence, the default tag for all non-coordinated examples is COORDINATED: false).

In the following Romanian example, two prepositional phrases are coordinated:

(29) *mai pe largu şi de agiunsu scriu* (1725, Letopisețul Țărîi Moldovei, Ro_ADP_aaif)
'I write longer/more in detail and sufficient'.

Both adverbials, *pe largu* 'long' (which is, by the way, modified by *mai*) and *de agiunsu* 'detailed' (modern standard Romanian lemma: *de-ajuns*), are annotated and both are tagged as COORDINATED: true.

In line with this, more than two adverbs can be coordinated, as in this Neapolitan example:

(30) *abbiannose chiano chiano adaso adaso e palillo palillo* (1634-1636, Lo cunto de li cunti, Napoli, It_A_aaif)
'they set off [standard Italian verb *avviarsi*] slowly, quietly and gently'

The three coordinated adverbs are each reduplicated (see section 2.1.4). Hence, one of the instances of the adjective-adverb *chiano* [standard Italian lemma *piano*] is lemmatized and tagged as REDUPLICATED: true and COORDINATED: true, and the same holds for the reduplicated and coordinated adjective-adverb *adaso* [standard Italian adjective-adverb lemma *adagio*] and for the reduplicated and coordinated noun *palillo*.

Coordination is also possible between different adverb types, i.e. with different morphosyntactic structures. Depending on which types of adverbs are tagged in a given subcorpus (adjective-adverbs, derived adverbs, prepositional phrases, etc.), the coordinated adverb may be tagged or not. For example, in the following Portuguese sentence, a *mente*-adverb is coordinated with a prepositional phrase (preposition + noun: *com clareza*) and both are annotated because this subcorpus includes both types of adverbials:

(31) *onde mostra objetivamente e com clareza como pratica-se o esporte* (1994, Relato de interação, Pt_APM_DeG)
'where he/she shows objectively and clearly how this sport is practised'

On the contrary, adverbs like Spanish *bien* 'well' are not tagged, as in the following example:

(32) *Francisco manejaba rápido y bien* (1996, Los últimos días de "La Prensa", 1996, Sp_A_CDH)
'Francisco drove quickly and well'

Therefore, only *rápido* is tagged and categorized as COORDINATED: true.

2.3. Semantic classification of the adverb

Given the diversity of possible (contextual) meanings and polysemy as well as diachronic semantic changes, the specific meaning of each adverb in a given example is not annotated, but depending on the annotated lemma, its meaning can be searched in dictionaries elsewhere. The French database Fr_A_DHAA exclusively contains a register of the specific meaning of each combination of the type "verb + adjective-adverb" (e.g. *tomber sec* 'tomber, arriver immédiatement, tout d'un coup').

In order to offer a rough **semantic classification** of the adverbials, six general categories were established and annotated for each example. Note that the meaning of some examples may be ambiguous and can be interpreted differently; therefore, some categories may be vague and may have been used slightly differently by the different linguists who annotated the various subcorpora.

➤ SEMANTIC CLASSIFICATION: manner

Manner-adverbs semantically modify the event expressed by the verb of the sentence, as in the following examples: Spanish *proceder de ligero* 'to proceed quickly, thoughtlessly', French *dire tout bas* 'to say very quietly', Portuguese *participar ativamente* 'to participate actively', Italian *scrivere bello* 'to write beautifully', or Romanian *a zice chiar* 'to speak clearly'. Therefore, most examples of manner adverbs take the verb as their syntactic scope (ATTRIBUTION TARGET: verb / verb and subject / verb and object, see section 2.2.1 above) and have also been tagged for the word-class **verb** (see section 3). In addition, adverbs that modify deverbal adjectives (ATTRIBUTION TARGET: adjective) also are often manner adverbs from a semantic point of view: e.g. the Portuguese example

(33)   *siluetas bem definidas e pinceladas esmeradamente alisadas* (1993, parte escrita, Pt_APM_DeG)
        'well defined silhouettes and carefully straight brushstrokes'.

The adverb *esmeradamente* 'neatly, carefully' characterizes a manner-property of the adjective *alisado* 'straight, smooth' (*alisado de modo esmerado* 'straightened in a careful manner').

Likewise, the Spanish adverbial prepositional phrase *a tuerto* 'one-eyed' typically describes a characteristic of verbs like *acusar* 'to accuse', *condenar* 'to condemn', etc. In the following example, *a tuerto* modifies the deverbal adjective *acusado* 'accused' and is classified as a manner adverb:

(34)   *si matarme quisieres acusado a tuerto, Dios me aya merçed* (1251, Calila e Dimna, Sp_AP_SH3)
        'if you wish to kill me when I am only accused in a one-eyed manner, I am at God's mercy'

➤ SEMANTIC CLASSIFICATION: time

Adverbs expressing temporal circumstances are tagged with this category, e.g. the examples in the following list.

Spanish adjective-adverb *recién* 'recently':

(35)   *como un cordero recién esquilado* (2007, El corazón helado, Sp_AP_SH3)
        'like a recently sheared sheep',

French adjective-adverb *bref* 'briefly':

(36)   *pour répondre bref* (2015, Fr_A_Web)
        'to answer quickly/briefly',

Calabrese derived adverb *novamenti* 'again' (standard Italian: *nuovamente*):

(37)   *si maritau novamenti* (1979, Calabria, Racconti populari calabresi, It_A_aaif)
        's/he got married again',

Portuguese prepositional phrase *de novo* 'again':

    (38)   *lavo a minha mão de novo* (1993, Rio de Janeiro, parte oral, Pt_APM_DeG)
          'I wash my hands again',

Romanian contracted prepositional phrase *îndelung* (preposition *în* + preposition *de* + adjective *lung*) 'during a long time':

    (39)   *nu îndelung au ținut Mihai vodă* (1675, Letopisețul Țărâi Moldovei, Ro_APD_aaif)
          'not for long they held Mihai Voda'.

Since the adverbials express a temporal circumstance of the action described by the verb, the **verb** in these examples is occasionally tagged too (see section 4).

> ➢   SEMANTIC CLASSIFICATION: location (place and direction)

Adverbs expressing a local circumstance of the event expressed by or referred to by the verb are tagged with this semantic classification, e.g. Spanish *alto* in *volar alto* and French *haut* in *voler haut* (both meaning 'to fly high'), as well as Portuguese *direto* in the following example:

    (40)   *vai direto lá* (1993, parte oral, Pt_APM_DeG)
          'go directly there'

Likewise, in Salentino dialect, the reduplicated adjective-adverb *rittu rittu* (standard Italian lemma *diritto* 'direct') has a directional meaning and therefore is tagged with this category:

    (41)   *cu pozzi rrivare rittu rittu a ncielu* (1995, Il Salento in vernacolo, It_A_aaif)
          'that you could arrive directly to heaven'.

Since the adverbials refer to a spatial circumstance of the action described by the verb, the **verb** in these examples is also tagged occasionally (see section 4).

> ➢   SEMANTIC CLASSIFICATION: quantity

Certain adverbs quantify on a scale or a semantic characteristic or the degree of a semantic property expressed by the segment they modify (the attribution target, be it a verb, an adjective, an adverb, a noun or a syntagma). In the following French example, the adjective-adverb *double* 'double' denotes the quantity the verb 'to pay' refers to (hence, ATTRIBUTION TARGET: verb and object):

    (42)   *dimanche serait payé double* (2015, Fr_A_Web)
          'Sunday would be paid double',

In the Spanish example *hombres harto sabios* 'very wise men' the adjective-adverb *harto* 'very' quantifies (more precisely: intensifies) the denoted quality of the adjective *sabios* 'wise'. Likewise, in the following Portuguese example the adverb *completamente* 'completely' intensifies the adjective *distorcido* 'distorted':

    (43)   *O som estava completamente distorcido* (1993, Rio de Janeiro, parte escrita, Pt_APM_DeG)
          'the sound was completely distorted',

Similarly, in the Romanian example *foarte greu au stătut strâmbătate* (see example 18) 'with real difficulty they stood crooked', the adjective-adverb *foarte* 'strongly, very' quantifies and intensifies the quality denoted by the adverb *greu* 'hard'. On the contrary, quantifiers may also downgrade: the Neapolitan adjective-adverb *meza* 'half' (standard Italian lemma: *mezzo*) quantifies the adjective *vregognosa* (modern standard Italian: *vergognosa* 'ashamed') in the following sentence:

    (44)   *steva meza vregognosa* (1684, Napoli, Sarnelli, La Posilecheata)
          'she was half shy/ashamed'

This adjective-adverb may also quantify the inherent property of a noun, e.g. the Spanish adverb *medio* quantifies the noun *mujeres* in this example:

(45) *éstos son medio mujeres* (1671, El rey gallo, Sp_AP_SH3)
      'these men are half women'

➢ SEMANTIC CLASSIFICATION: specification

Focus adverbs express a semantic specification of the segment they modify; their syntactic scope (attribution target) is mostly a noun or any syntagma without verb-reference. For example, French *juste* focuses on the temporal subordinate clause in this example:

(46) *J'arrive aux marches juste quand le maton donne l'ordre de couper la flotte* (2016, Fr_A_Web)
      'I arrive at the stairs just as the prison warden gives the order to cut the float'

Hence, *juste* 'exactly, just' specifies the temporal complement of the sentence. Likewise, in the Neapolitan example

(47) *simmo iusto quatto* (1934, Napoli, Salvatore Di Giacomo: Tutto il teatro, It_A_aaif)
      'we are exactly four'

the adjective-adverb *iusto* (standard Italian lemma *giusto*) specifies and focuses on the numeral.

The Spanish adjective-adverb *solo* specifies a noun phrase in the example

(48) *con sola una palabra de tu noble boca* (1499, La Celestina, Sp_AP_SH3)
      'with only one word from your noble mouth'.

Likewise, the prepositional phrase *en especial* 'especially' focuses on a noun phrase in the example

(49) *pasaron muchas cosas de notar, en especial para mí* (1492, Cárcel de Amor, Sp_AP_SH3)
      'a lot of noteworthy things happened, especially for me'

The Portuguese *mente*-adverbs *especialmente* 'especially' and *basicamente* 'basically' specify a noun phrase and an adjectival phrase in the following examples:

(50) *como todo filme especialmente filme de terror num tem fim* (1993, parte oral, Pt_APM_DeG)
      'like all films, especially horror-films, it has no end'

(51) *ela é basicamente azul e branco* (1993, parte oral, Pt_APM_DeG)
      'she is basically blue and white'.

➢ SEMANTIC CLASSIFICATION: discourse

Adverbs that express metalinguistic, pragmatic and/or discursive extra-propositional characterizations or speaker-attitudes (e.g. discourse markers and sentence adverbs) are subsumed under the general classification of *discourse*. This tag coincides mostly with the category ATTRIBUTION TARGET: sentence; as in the above cited examples:

Neapolitan *sicuro* 'sure(ly)':

(52) *sì lo conosco, lo conosco sicuro* (1925, Eduardo Scarpetta, Tutto il teatro, It_A_aaif)
      'yes, I know him, I know him surely/ for sure'

Spanish *cierto* 'certainly':

(53) *Verdad sea que, cierto, duele más la pérdida* (1499, La Celestina, Sp_AP_SH3)
      'Certainly it is true that the loss hurts more'

Portuguese *infelizmente* 'unfortunately, unhappily':

(54) *Meu grupo infelizmente não foi o vencedor* (1993, Natal, parte escrita, Pt_APM_DeG)
      'Unfortunately, my group was not the winner'

3. Annotations concerning the word-class *verb*

The tag for the word-class **verb** is optional depending on the characteristics of each subcorpus. Some of the corpora specifically document adjective-adverbs as manner adverbs (Fr_A_DHAA, Fr_A_Web, Sp_A_CDH), therefore, the annotations in these corpora target the combination "verb + adjective-adverb" and the verb in each example is tagged. In other corpora, the verbs are only tagged if the annotated adverb modifies the verb (e.g. SEMANTIC CLASSIFICATION: manner, and/or ATTRIBUTION TARGET: verb, verb and subject or verb and object).

The infinite form of the verb was annotated as the **lemma**: e.g. Spanish / Portuguese *cantar*, French *chanter*, Italian *cantare* and Romanian *cânta* (without preposition *a*) 'to sing'. Regarding the formal variation in varieties of southern Italy (*Napolitano, Siciliano, Salentino* and *Calabrese*, see subcorpus It_A_aaif), the corresponding standard Italian lemma was annotated in order to enable a simpler lemma search. In case there is no corresponding standard Italian lemma, a reconstructed vernacular Latin form was marked with an asterisk and written in capital letters (for example *COMMOGLIARE for the Neapolitan verb forms *accumigliateva* and *commoglio*). In the case of periphrastic verb forms, the verb carrying the main semantic information has been tagged and lemmatized: for example, Spanish *vestir* in the sentence

(55) *con vna manta blanca como andan vestidos de ordinario* (1602, Documentos cronísticos, México, Sp_AP_CORDIAM)
      'with a white blanket as they usually are dressed'

The tagging and lemmatization of the verbs enables queries for recurrent combinations of verbs and adverbs, such as French *parler fort* 'to speak loud', *parler bas* 'to speak quietly', *voir grand* 'to think big', etc. Furthermore, the tagged verbs have been annotated regarding three formal categories (SYNTACTIC CONSTRUCTION, COORDINATION, PART OF THE TEXT), which are explained in the following paragraphs.

3.1. Syntactic construction of the verb

The tag SYNTACTIC CONSTRUCTION offers three basic features: transitive, intransitive and reflexive.

As for the first category, the category SYNTACTIC CONSTRUCTION: transitive applies to those sentences in which the verb requires a direct or indirect object, for example the Spanish verb *pagar* 'to pay' in the sentence

(56) *los martirios [...] los ha de pagar bien caros* (1889-1891, Los bandidos del Río frío, México, Sp_AP_SH3)
      'he/she has to pay very dearly for the torments'

The category SYNTACTIC CONSTRUCTION: intransitive applies to those examples in which the verb does not select an object, such as the French verb *sentir* 'to smell' in the following example:

(57) *ça sent bizarre* (2015, Fr_A_Web)
      'that smells bizarre'

The category SYNTACTIC CONSTRUCTION: reflexive is annotated for verbs that are used with a reflexive pronoun, e.g. the Italian and Neapolitan verb *addormentarsi* 'to fall asleep' in this example:

(58) *t'adduorme doce doce* (1853-1925, Tutto il teatro, Napoli, It_A_aaif)
      'you fell asleep sweetly/gently' [standard Italian adjective-adverb lemma dolce]

In order to join reflexive and non-reflexive occurrences of the same verb under one common lemma, the reflexive pronoun was not annotated as part of the verb lemma, e.g. *t'adduorme* is lemmatized as *addormentare* (not as *addormentarsi*). In these cases, the tag SYNTACTIC CONSTRUCTION allows for differentiation between reflexive and non-reflexive (hence, transitive or intransitive) occurrences.

3.2. Coordination of verbs

Coordinated verbs are modified by the same adverb, e.g. the French verbs *manger* 'to eat' and *boire* 'to drink' are coordinated and both are modified by the manner-adjective-adverb *chaud* 'hot':

(59) *Pourquoi faut-il manger et boire bien chaud* (2015, Fr_A_Web)
      'Why is it necessary to eat and drink quite hot'.

Both verbs are annotated and lemmatized separately, and both are tagged with the category COORDINATED: true (hence, the default tag for all other verbs is COORDINATED: false).

3.3. Explicit and implicit verbs ("part of the text")

This third category of the verb-annotation allows the annotation of verbs which are not explicitly mentioned in the annotated sentence, but which can be reconstructed and inferred from the larger context. For example, the verb may be omitted (ellipsis) to avoid repetitions in sentences like

(60) *les basses sonnent "basses" et les aigues "aigues"* (2016, Fr-A_Web)
      'the basses [e.g. frequencies or notes] sound "low" and the trebles/high notes "sharp"'

In this sentence, the verb *sonner* 'to sound' can be easily reconstructed for the second adjective-adverb, *aigues* (*les aigues sonnent "aigues"*). In order to be able to search for the lemma combination "verb *sonner* + adjective-adverb *aigu*" (in addition to the combination "*sonner* + *bas*"), a placeholder (blank space) is inserted at the position the verb would be expected and annotated with the lemma *sonner* (*les aigues* [verb: lemma *sonner*] *"aigues"*). This placeholder is tagged with the category PART OF THE TEXT: false. Hence, the default category for all explicit verbs is PART OF THE TEXT: true, as in the case of *sonnent* in the first example: *les basses sonnent "basses"*.

4. Annotations concerning the *subject*

The **subject** is an optional category. Contrary to the other tags, which are word-class tags (adverb, preposition, verb [although limited to the verb of the predicate]), this tag refers to a syntactic function in a given sentence. It is usually annotated if the adverb refers to the subject of the sentence (ATTRIBUTION TARGET: verb and subject). The following grammatical categories are annotated for the subject:

GENDER: masculine, feminine, neuter, undefined

NUMBER: Singular, Plural, undefined

For example, in the Spanish sentence *decía el Marquesito muy serio* (see example 9 cited above) 'said the little marquis very serious(ly)', the sequence *el Marquesito* is tagged as the subject and annotated as GENDER: masculine, NUMBER: singular.

Furthermore, the category OVERT SUBJECT / NULL SUBJECT allows for differentiation between explicit and implicit subjects. The latter is the case of null-subjects in pro-drop languages (i.e. all Romance languages except French). In case the referent of the null-subject can be inferred or reconstructed from the wider context of the example (e.g. the verb inflection), a placeholder (blank space) has been annotated with the corresponding specifications of gender and number and the tag NULL SUBJECT (hence, the default tag for all other subjects is OVERT SUBJECT).

The annotation of the subject is helpful, for example, for contrasting inflected and invariable occurrences of adjective-adverbs, e.g. in this Portuguese example:

(61)   *vocês tenham cuidado … não andem sozinho* (1993, parte oral, Pt_APM_DeG)
       'be[pl.] careful, don't go[pl.] around alone[sg.]'

The plural-subject *vocês* appears as an overt subject in the first phrase but remains implicit in the second phrase, and can be inferred via the plural verbform *andem*. In this example, a subject-placeholder (blank space) is tagged: *vocês tenham cuidado …* [null-subject] *não andem sozinho*. This subject-placeholder is annotated with the categories NUMBER: plural and NULL SUBJECT. Since the implicit null-subject of the verb form *andem* is clearly a plural-subject, the adjective-adverb *sozinho* is invariable (masc. sg.).

On the contrary, inflected adverbs may show agreement with the subject (see section 2.1.1), e.g. in the Portuguese sentence

(62)   *emtão* [sic] *os dos Jão* [sic] *e Maria viveram felizes para sempre* (1994, parte escrita, Pt_APM_DeG) [errors are original]
       'so the two, João and Maria, lived happily forever'

In this example, in addition to the verb *viver* 'to live' and to the (inflected) adjective-adverb *feliz(es)* 'happy', the subject *os dos Jão e Maria* is annotated as an OVERT SUBJECT and with the specifications GENDER: masculine, NUMBER: plural.

5. Overview

Word-class	main properties	options
Adverb (lemmatized)	morphosyntactic structure	adjectival
		noun
		derived *-mente*
		derived *-eşte*
		derived *-ly*
		derived *-iş*
		derived *-ul*
		derived *-one/-oni*
		other
	inflection	uninflected / m.sg.
		audible inflection
		inaudible inflection
		feminine singular
		neuter singular
		masculine plural
		feminine plural
		neuter plural
	attribution target	verb
		verb and subject
		verb and object
		sentence
		adverb
		adjective
		noun or syntagma without verb-reference
		other
	semantic classification	manner
		quantity / intensity
		time
		location
		discourse
		specification
		undefined / other
	modified	(true / false)
	reduplicated	(true / false)
	coordinated	(true / false)
	part of prepositional phrase	(true / false)

Word-class	main properties	options
Preposition (lemmatized)	contracted	(true / false)

Word-class	main properties	options
Article (not lemmatized)	gender	masculine
		feminine
		neuter
		undefined
	number	singular
		plural
		undefined
	contracted	(true / false)

Word-class	main properties	options
Possessive (not lemmatized)	gender	masculine
		feminine
		neuter
		undefined
	number	singular
		plural
		undefined
	person	first
		second
		third
	part of text	(true / false)
	coordinated	(true / false)

Word-class	main properties	options
Verb (lemmatized)	syntactic construction	transitive
		intransitive
		reflexive

Word-class	main properties	options
Subject (not lemmatized)	gender	masculine
		feminine
		neuter
		undefined
	number	singular
		plural
		undefined
	overt subject	true (overt subject) / false (null subject)

6. References

Corpus

Schneider, Gerlinde / Pollin, Christopher / Gerhalter, Katharina / Hummel, Martin (2020): *Adjective-Adverb Interfaces in Romance. Open-Access Database* (=AAIF-Database). https://gams.uni-graz.at/context:aaif

Subcorpora

Fr_A_DHAA = Hummel, Martin / Gerhalter, Katharina / Schneider, Gerlinde / Pollin, Christopher (2020). *Corpus French Dictionnaire historique de l'adjectif-adverbe*. 2nd version. AAIF Database. https://gams.uni-graz.at/o:aaif.fradhaa

Fr_A_Web = Hummel, Martin / Telsnig, Helena / Korper, Greta / Gazdik, Anna / Höfferer, Julia / Gerhalter, Katharina / Schneider, Gerlinde / Pollin, Christopher (2020). *Corpus of French Adjective-Adverbs in Informal Texts from the Web*. AAIF-Database. https://gams.unigraz.at/o:aaif.fraweb

It_A_aaif = Ledgeway, Adam / Gerhalter, Katharina (2020): *Corpus of Adjective-Adverbs in Southern Varieties of Italy*. In Schneider, Gerlinde / Pollin, Christopher / Gerhalter, Katharina / Hummel, Martin (2020): *Adjective-Adverb Interfaces in Romance. Open-Access Database* (=AAIF-Database). https://gams.uni-graz.at/o:aaif.itaaaif

Pt_APM_DeG = Votre, Sebastião / Oliveira, Mariangela Rios de / Furtado da Cunha, Maria Angélica / Gerhalter, Katharina (2020): *Corpus of Brazilian Portuguese Adjective-Adverbs, mente-Adverbs and Prepositional Phrases in Corpus Discurso & Gramática.* In Schneider, Gerlinde / Pollin, Christopher / Gerhalter, Katharina / Hummel, Martin (2020): *Adjective-Adverb Interfaces in Romance. Open-Access Database* (=AAIF-Database). https://gams.uni-graz.at/o:aaif.ptapmdeg

Ro_ADP_aaif = Chircu, Adrian (2020): *Corpus of Old Romanian Adjective-Adverbs, Derived Adverbs and Prepositional Phrases*. In Schneider, Gerlinde / Pollin, Christopher / Gerhalter, Katharina / Hummel, Martin (2020): *Adjective-Adverb Interfaces in Romance. Open-Access Database* (=AAIF-Database). https://gams.unigraz.at/o:aaif.roadpdaaif

Sp_AP_CORDIAM = Striedner, Pascal / Gerhalter, Katharina / Hummel, Martin / Schneider, Gerlinde / Pollin, Christopher (2019): *Corpus of Spanish Adjective-Adverbs and Prepositional Phrases in CORDIAM*. AAIF-Database. https://gams.uni-graz.ato:aaif.spapcordiam

Sp_AP_SH3 = Gerhalter, Katharina / Hummel, Martin / Schneider, Gerlinde / Pollin, Christopher (2020). *Corpus of Spanish Adjective-Adverbs in Diachrony (for Sintaxis Histórica III)*. AAIF-Database. https://gams.uni-graz.at/o:aaif.spapsh3

Sp_A_CDH = Gerhalter, Katharina / Schneider, Gerlinde / Pollin, Christopher / Hummel, Martin (2020): *Corpus of Spanish "Verb + Adjective-Adverbs" in CDH. 2nd version*. AAIF-Database. https://gams.uni-graz.at/o:aaif.spacdh

Other references

Hummel, Martin (2017): *Adjectives with adverbial functions in Romance*, in: Hummel, Martin / Valera, Salvador (eds.), *Adjective adverb interfaces in Romance*, Amsterdam / Philadelphia: John Benjamins Publishing Company, 13–46.