Codebook

This codebook gives an overview of the different variables in the dataset: subject-related variables, verb-related variables and SC-related variables (SC = subject complement). Additionally, the variable LANGUAGE distinguishes between French and Dutch cases.

1 Variables related to the subject

1.1 SUBJECT_HEAD

This variable captures the distinct lexemes that constitute the head of the subject, = "types" according to the productivity literature. SUBJECT_HEAD is derived from the variable SUBJECT_SEQ, which spells out the complete literal sequence that corresponds to the subject.

1.2 SUBJECT_CAT

The morphosyntactic category of the subject. Note that this variable deals with the "semantic subject" (as it is formally present in the sentence containing the copular verb of interest), which does not necessarily function as the "grammatical subject".

adjective	Including the superlative (*het moeilijkst*, *le plus important*). Cases in which a noun is omitted but recoverable from the context ("ellipsis") are classified as "nominal".
other	- Bare shell nouns in specificational constructions, for example [*Probleem is* X]. - Autonymic use, for example *Looistoffen is een verzamelnaam*. - Subjects for which the coordination of two or more nouns entails the absence of the determiner, for example: *Smaak en voedingswaarde* < *blijven* > *ongewijzigd*. (**smaak blijft ongewijzigd*). - Other particular cases where the determiner seems omitted, including newspaper headlines.
infinitive	
nominalized_infinitive	Infinitive preceded by an indefinite article.

proper_noun	
phrasal	The group of words that forms the subject of the copular verb contains a subject and a conjugated verb.
impersonal_pronoun	- ce (French) - il (French) - het (Dutch) - er (Dutch) Only if *er* can be replaced by *het*, cf.: *Er < raakte > bekend dat Willem Frederik Hermans zich in 1942 als twintigjarige aanmeldde voor de Kulturkammer* […].
indefinite_pronoun	numbers (e.g. *800 d'entre elles*) and the following items: - certain, certain(_de_X), chacun, d'autres, on, personne, quelqu'un, quiconque, rien, tout, tout_le_monde, un(_de_X). - alles, andere, één, één(_van_X), geen(_van_X), eentje, heel_wat, ieder, iedereen, iemand, iemand(_van_X), iets, men, menigeen, niemand, niets, sommige, veel, veel(_van_X).
interrogative_pronoun	wie, qui
personal/demonstrative_ pronoun	- ça, ce, ceci, cela, celle, celle-ci, celui-ci, celui-là, ceux-ci, je, tu, nous, on, vous, il(s), elle(s). - dat, deze, die, dit, beide, het, ik, je/jij, we/wij, jullie, u, ie, ze/zij (singular), ze/zij (plural).
relative_pronoun	The relative pronouns, and not their antecedents, are encoded for this variable.
definite_NP	NP of which the noun is accompanied by: a definite article, a demonstrative determiner, a possessive determiner, or: - tout le/toute la/tous les/toutes les/tout ce/toute cette/tous ces/toutes ces. - al die/deze, beide, possessive constructions such as *Gods belofte*.
indefinite_NP	NP of which the noun is accompanied by: a covert indefinite article in Dutch, a number or: - un, aucun, beaucoup de, certain, chaque, d'autres, de(s), de tels, différent, la plupart de, l'un de X, moult, plusieurs, tout. - alle, andere, allerlei, iedere, de meeste, diverse, een X-tal, een aantal, een, één, elk, enkele, veel, (heel/nogal)wat, ieder, minder/meer dan, minder/meer, meerdere, sommige, tal van, verschillende, weinig, zo'n, geen.

interrogative_NP	NP of which the noun is accompanied by: combien de / quel / welke etc.
PP	one example: *au cœur du monde*
NA	- copular verb without subject - copular participle construction accompanying a noun, e.g. *de bewaard gebleven sleutelstukken, diverses circonstances demeurées mystérieuses* - imperative - nominalized construction, e.g. *het verloren gaan van* X - passive construction, e.g. *er werd op zoek gegaan naar* X - tout_en[gerund], e.g. *tout en demeurant* X

1.3 SUBJECT_CAT_2

Simplified version of the variable SUBJECT_CAT.

nominal	definite_NP, indefinite_NP, interrogative_NP, proper_noun, nominalized_infinitive
pronominal	impersonal_pronoun, indefinite_pronoun, interrogative_pronoun, personal/demonstrative_pronoun, relative_pronoun
adjectival	adjective
infinitival	infinitive
phrasal	phrasal
other	other, PP
NA	NA

1.4 SUBJECT_ANIM

The animacy of the subject.

animate	The subject is animate, including groups of humans (e.g. *population, équipe, humanité, team, groep, peloton*).
inanimate	The subject is inanimate, including: - thoughts / memories - emotions - appearance - countries / towns etc. used in an inanimate way

	- infinitives / phrasal subjects - etc.
intermediate	Heterogeneous category in-between animate and inanimate, which comprises: - body parts (*métabolisme, hoofd, voetje, lichaam, bekken, slokdarm, mond* ...) - companies / organizations / institutions referring to a group of humans (*gouvernement, la Haute autorité, vereniging, stichting, bank, brandweer, redactie* ...) - names of football teams (e.g. *Chievo*) - names of music groups - countries / towns etc. referring to a group of humans - political parties - media - army - other (*convoi*, *planète*)

1.5 SUBJECT_PERSONAL

This variable distinguishes between the subjects *ik/je* (first person singular) and all other subjects. Since *ik/je* corresponds automatically to an animate subject, SUBJECT_PERSONAL partly overlaps with SUBJECT_ANIM.

other_person	The subject is not *ik/je*. Note that adjectives, infinitives etc. are also part of this category.
first_person_sing	The subject is *ik/je*.

1.6 SUBJECT_DEFINITENESS

Definiteness of the subject. This variable is based on the variable SUBJECT_CAT. Note that, in the case of NPs, definiteness was not evaluated semantically, but was determined on the basis of the grammatical type of article.

definite	definite_NP, personal/demonstrative_pronoun, proper_noun, nominalized_infinitive
indefinite	indefinite_NP, interrogative_NP, interrogative_pronoun, indefinite_pronoun
NA	impersonal_pronoun, relative_pronoun, adjective, infinitive, phrasal, PP, other, NA

1.7 SUBJECT_IMPERSONAL

This variable identifies the instantiations of the impersonal construction.

yes	impersonal construction, including "presentative *er*" (for example, *er raakten vier mensen gewond*)
no	no impersonal construction

2 Variables related to the verb

The different copular verbs are available via the variable VERB.

2.1 COPULARITY

Binary distinction between prototypical copulas and semi-copulas, based on the criterion of a possible pronominalization of the subject complement by *le* (French) or *het/dat* (Dutch).

prototypical_copula	*être*/*zijn* ['be'], *rester*/*demeurer*/*blijven* ['remain'] and *devenir*/*worden* ['become']
semi_copula	*tomber* ['fall], *passer* ['pass'], *tourner* ['turn'], *virer* [+/- 'turn'], *(ge)raken* [+/- 'attain', 'touch'], *vallen* ['fall'], *gaan* ['go'], *komen* ['come']

2.2 COMPOUND_TENSE

For the subset of examples containing a conjugated copular verb, this variables encodes the opposition between simple tenses and compound tenses.

compound	compound tenses
no_compound	simple tenses
NA	

3 Variables related to the subject complement (SC)

3.1 SC_HEAD

This variable captures the distinct lexemes that constitute the head of the subject complement, = "types" according to the productivity literature. Note that the lexemes are neutralized for certain oppositions that are not relevant, for example: every feminine form is replaced by the masculine form. For prepositional subject complements, both the preposition and the head noun of the PP form together the type, for example: "in_[…]_duigen".

SC_HEAD is derived from the variable SC_SEQ, which spells out the complete literal sequence of the SC.

3.2 SC_CAT

This variable pertains to the morphosyntactic category of the SC. Moreover, a slightly different version of this variable, SC_CAT_2, categorizes particle SCs as a separate category.

adjectival	Note that the adjectival category also encompasses: - "particles", e.g. *los* in *loskomen* (lit. 'come free') - the SC *d'accord* - superlatives, e.g. *het ergst*
other	numbers + *nul* ; the adverbials *assez, trop, devant, derrière, debout, ensemble, boven, achter, overeind, overstag* ; *het eens*
infinitival	
nom-adj	SC that has a categorical status between nominal and adjectival (this category only applies to the examples with an animate subject).
nominal	
phrasal	SC that contains a subject and a conjugated verb.
prepositional	Note that infinitives preceded by a prepositional form are not considered to be prepositional SCs.
proform	*le, ainsi, comme, celle,* un(_de_X), één(_van_X), etc.

	This category also includes relative pronouns in the position of subject complement for which the replacement of the relative pronoun by its antecedent does not easily reconstitute the canonical phrase, for example: *du fauve qu'était < devenu > Dumbledore → *Dumbledore était devenu le fauve*.

3.3 SC_MODIFICATION

The modification of the SC: binary opposition between "modified SC" and "unmodified SC", for all morphosyntactic categories. In case of prepositional SCs, this pertains mostly to the modification of the head noun that complements the preposition. This variable examines the degree of conventionalization of the SC.

not_modified	
modified	
NA	SC that are not adjectival, nominal or prepositional.

3.4 ADJ_SC_MODIFICATION

The modification of the adjectival SC. Note that the same type of variable was not included for nominal and prepositional SCs.

mod_adv	adverb/adverbial phrase that modifies an adjectival head, e.g.: *ongemeen spannend*
mod_comparative	comparative construction, e.g.: *plus X que* (only if the comparant is also present, otherwise: mod_adv)
mod_superlative	superlative construction, e.g.: *het meest belangrijk*
mod_SC_inf	infinitival subordinate clause that modifies an adjectival head, e.g.: *erg leuk om te horen*
mod_SC_completive	completive clause that modifies an adjectival head, e.g.: *bewuster dat er meer is tussen hemel en aarde*
mod_SC_other	other subordinate clause that modifies an adjectival head, e.g.: *d'accord pour que la gestion soit confiée pour les cinquante prochaines années à un établissement public*
mod_nom-adv	adverbially used N(P) that modifies an adjectival head, e.g.: *ruim 0,4 °C warmer*

mod_nom-other	other type of N(P) that modifies an adjectival head; one example: *fourrée lapin*
mod_prep-adv	adverbially used PP, with a preposition that is not imposed by the adjectival head, e.g.: *ouvert pendant les travaux, caché derrière les arbres, onder je spullen en bezigheden verstopt, door het huis heen verspreid*
mod_prep-voor/pour	PP with a preposition *voor/pour* that is not imposed by the adjectival head, e.g.: *lourds pour le patient*
mod_prep-fixed	PP of which the preposition is (syntactically or semantically) imposed by the adjectival head, e.g.: *fidèle à ses promesses, incompatibles avec l'histoire d'Arte*
mod_other	other type of modification, e.g.: *si* X *que, zo* X *mogelijk,* X *oud,* X *geleden.*
not_modified	no modification
NA	non-adjectival subject complement

3.5 SC_UNIQUENESS

This variable distinguishes the SCs that contain only a single word (= non-prepositional, non-modified SC) and the SCs that contain multiple words (non-prepositional, modified SC and prepositional SC).

single_word	
multiple_words	

3.6 SC_COMPARATIVE

The variable SC_COMPARATIVE identifies the comparative SCs: synthetic comparatives (*mieux/pire* and X-*er*) and analytic comparatives ([*plus/moins/aussi*/*meer/extra/minder/even* X]), possibly followed by [*que* X / *dan* X]. “Comparative” is here defined as a morphological notion. Certain comparatives are also covered by the variable SC_SCALAR (e.g. *de plus en plus* X, ‘more and more X’), but not all of them.

comparative	Comparative structures, such as: - (bien/beaucoup/encore) plus X (e.g. *plus élevé*), plus X encore (e.g. *plus nécessaire encore*), un peu plus X (e.g. *un peu plus encombrant*), plus que X (e.g. *plus que solide*), de plus en plus X (e.g. *de plus en plus réduit*), de encore plus en encore plus X (e.g. *de encore plus en encore plus pressant*), d'autant plus X, de moins en moins X (e.g. *de moins en moins dupes*), moins X, mieux/pire, aussi X, plus X plus Y - X-er (e.g. *koeler*), wat X-er (e.g. *wat warmer*), iets X-er (e.g. *iets duurder*), nog X-er (e.g. *nog leuker*), aanzienlijk X-er (e.g. *aanzienlijk kleiner*), steeds X-er (e.g. *steeds mooier*), steeds meer X (e.g. *steeds meer het centrum*), langzaam maar zeker X-er (e.g. *langzaam maar zeker sterker*), alleen nog maar X-er (e.g. *alleen nog maar sterker*), (nog) veel X-er (e.g. *veel sterker)*, beter X (e.g. *beter beschermd*), (nog/wat) meer X (e.g. *meer verweven*), (nog/wat) minder X (e.g. *minder ondergedompeld*), meer dan X, minder dan *X*, maar weinig X-er (e.g. *maar weinig groter*), des te X-er (e.g. *des te natter*), hoe X-er hoe Y-er (e.g. *hoe langer, hoe natter*), almaar X-er, extra X, even X
no_comparative	
NA	The following morphosyntactic categories of the SC are not applicable for this variable: phrasal, proform, infinitival.

3.7 SC_SCALAR

This variable encodes the presence or absence of a scalar modifier.

neutral	no scalar modifier
scalar	scalar modifier (e.g. *de plus en plus* X)
NA	The following morphosyntactic categories of the SC are not applicable for this variable: phrasal, proform, infinitival.

3.8 SC_GRADATION_NORM

This variable verifies whether the SC is modified by a degree adverb that indicates a gradation with respect to a certain norm (e.g. *très* [‘very’]). The adverb has to occur at the beginning of the SC.

gradation_norm	Examples: - trop_X, assez_X, très_X, plus_ou_moins_X, essentiellement_X, absolument_X, plutôt_X, extrêmement_X, complètement_X, carrément_X, parfaitement_X, totalement_X, X_total (e.g. *virer pluriculturel total*), terriblement_X, quasi(ment)_X, pratiquement_X, folle(ment)/fou_X, raide_dingue_X, éperdument_X, gravement_X, relativement_X, un_peu_X, peu_X, vraiment_X, suffisamment_X, rudement_X, franchement_X (e.g. *franchement dommage*), hardiment_X, résolument_X, tout_X, bien_X, à_peu_près_X, un_minimum_X (e.g. *un minimum compréhensible*), le_moindrement_X (e.g. *le moindrement familier*), intimement_X (e.g. *intimement persuadé que*), inséparablement_X (e.g. *inséparablement associé à*), largement_X (e.g. *largement incompris*), presque_X, profondément_X (e.g. *profondément imprégné de*), limite_X (e.g. *limite nationaliste*), mal_X (e.g. *mal à l'aise*), si_X, tant_X, tellement_X, le_peu_de_X, relativement_X, vivement_X - te_X, erg_X, zeer_X, min_of_meer_X, aardig_X (e.g. *aardig gevuld*), ongemeen_X, helemaal_X, wat_X, wat_X_2 (e.g. *wat een fijn product* !), vrij_X (e.g. *vrij eenvoudig*), licht_X, lichtelijk_X, compleet_X, volstrekt_X, volledig_X, geheel_X, echt_X, ernstig_X, flink_X (e.g. *flink in de problemen*), fors_X (e.g. *fors onder de indruk*), relatief_X, bijna_X, hopeloos_X (e.g. *hopeloos verliefd*), voldoende_X, nagenoeg_X, vrijwel_X (e.g. *vrijwel gelijk*), lekker_X (e.g. *lekker warm*), heerlijk_X (e.g. *heerlijk warm*),

	onaanvaardbaar_X (e.g. *onaanvaardbaar laag*), reddeloos_X (e.g. *reddeloos verloren*), onherstelbaar_X (e.g. *onherstelbaar beschadigd*), gepassioneerd_X (e.g. *gepassioneerd betrokken bij*), onverminderd_X (e.g. *onverminderd van toepassing*), wereldwijd_X (e.g. *wereldwijd bekend*), uitgebreid (e.g. *uitgebreid aan bod*), intensief_X (e.g. *intensief in aanraking*), totaal_X, vol_X (e.g. *vol in de aanval*), stevig_X (e.g. *stevig bij jezelf*), enorm_X, X_genoeg (e.g. *concreet genoeg*), goed_X, optimaal_X (e.g. *optimaal behouden*), uitstekend_X, best_wel_X, hoe_X (e.g. *hoe dankbaar*), hard_X (e.g. *hard nodig*), druk_X (e.g. *druk op zoek*), een_beetje_X, een_stukje_X, enigszins X, deels_X (e.g. *deels in gebruik*), voor_een_deel_X (e.g. *voor een deel te wijten aan*), teveel_X (e.g. *teveel uit elkaar*), dodelijk_X (e.g. *dodelijk gewond*), zo_X, zodanig_X, perfect_X, vreselijk_X (e.g. *vreselijk jaloers*), ongelofelijk_X (e.g. *ongelofelijk vrolijk*), iets_X (e.g. *iets boven mijn marketingpet*). Note that comparative structures such as *wat* X-*er* are not identified as “gradation_norm”.
no_gradation_norm	
NA	The following morphosyntactic categories of the SC are not applicable for this variable: phrasal, proform, infinitival.

3.9 ADJ_SC_GRADABILITY

This variable pertains to the gradability of the adjectival subject complements.

gradable	gradable adjectives Gradability is determined on the basis of whether or not the adjective in question can combine with degree adverbs such as *très* and *un peu*. The context of the corpus example is also taken into account: if the degree adverb accompanies the adjectival SC in the corpus example, this case is encoded as gradable, even if the adjective is not prototypically gradable. Note that adverbs expressing approximation such as *plus ou moins* and comparative constructions such as *plus* X *que* are not used to test gradability.
not_gradable	adjectives that are not gradable
NA	non-adjectival subject complements

3.10 ADJ_SC_PARTICIPLE

This variable evaluates the participial nature of the adjectival subject complement.

adjective	Adjectives that are not derived from a verb and without the formal characteristics of a participle.
participle	Past participles used as adjective and derived from a verb. SCs such as *inaperçu, incompris, impayé, onopgemerkt* etc. are also considered to be participial, despite the prefix *in-*/*on-* that suggests an adjectival use.
participle_2	Present participles used as adjective and derived from a verb.

pseudo-participle	Adjectives that are not derived from a verb, at least in synchrony, but have the formal characteristics of a participle. French forms that end on *-ant, -é, -i, -it, -is* (except for the colour *gris*), *-u, -us, -ait* are considered as pseudo-participles.
NA	non-adjectival subject complements